University of Helsinki, Institute of Behavioural Sciences,
Studies in Educational Sciences 260

Heidi Hyytinen

LOOKING BEYOND THE OBVIOUS

Theoretical, Empirical and Methodological Insights into Critical Thinking

Academic dissertation to be publicly discussed, by due permission of the Faculty of Behavioural Sciences at the University of Helsinki in the seminar room 302 at Athena (Siltavuorenpenger 3 A) on the 22nd of May, 2015 at 12 o'clock.

Helsinki 2015

Custos
Professor Sari Lindblom-Ylänne, University of Helsinki

Supervisors
Professor Sari Lindblom-Ylänne, University of Helsinki
Docent Katariina Holma, University of Helsinki and University of Eastern Finland
Docent Auli Toom, University of Helsinki
Emeritus Professor Richard Shavelson, Stanford University, USA

Pre-examiners
Lecturer, Dr Ben Kotzee, University of Birmingham, UK
Associate Professor Jeffrey Greene, University of Maryland, USA

Opponent
Associate Professor Krista Muis, McGill, Canada

Cover photo
Teemu Ylikoski

Unigrafia, Helsinki

ISBN 978-951-51-0307-9 (pbk)
ISBN 978-951-51-0308-6 (pdf)

ISSN-L 1798-8322
ISSN 1798-8322

University of Helsinki, Institute of Behavioural Sciences,
Studies in Educational Sciences 260

Heidi Hyytinen

Looking Beyond the Obvious

Theoretical, Empirical and Methodological Insights into Critical Thinking

Abstract

The central purpose of this doctoral thesis has been to deepen our understanding of the nature of critical thinking by combining theoretical, empirical and methodological perspectives. The concept of critical thinking has a central role both in research on the philosophy of education and in empirical research on learning and teaching in higher education. Although it is true that the philosophical and empirical analyses of critical thinking and knowledge differ fundamentally, the present thesis argues that there are shared concerns between these two scholarly traditions. The thesis consists of four studies, each of which approached this aim from different viewpoints. The methods involved both a philosophical approach and an empirical multi-method approach. The dialogue between the empirical and theoretical analyses offers new insights into conceptualising critical thinking and its prerequisites and extends our understanding of variations in critical thinking.

Based on the theoretical findings of these studies, I argue that normative elements (how things ought to be) and descriptive elements (how things are) are fundamentally intertwined in the educational research on critical thinking. Therefore, educational research on critical thinking requires both philosophical and empirical approaches and also dialogue between these two approaches. The theoretical part of the doctoral thesis further demonstrates how philosophical research can contribute to the normative elements of the prevailing empirically-based theorisation of critical thinking, particularly by revealing some conceptual inconsistencies within this framework. The research further introduces the notion of fallibilism (human knowledge is uncertain) as a way out. Epistemological fallibilism fits the presumption of critical thinking better than relativism from the theoretical and pedagogical points of view.

The empirical results revealed variation in (a) students' skills and dispositions to think critically, (b) students' ability to adapt their thinking and performance flexibility, (c) the nature of knowledge students consider to be relevant, (d) the knowledge that students use in problem-solving situations, as well as (e) the way

students process that knowledge. Based on these variations two profiles were identified: (1) superficial processing or (2) thorough processing. Superficial processing students reproduced information in the problem-solving situation. These students did not analyse, interpret, evaluate or synthesise knowledge, and their reasoning was very poor. They palmed off justification for knowing on authorities and others' testimonies. In contrast to previous research, the results show that these students did not share the belief that knowledge is absolutely certain or unquestionable. Nor did these students share the view that beliefs accurately represent or correspond to reality. These students emphasised the uncertain nature of knowledge. The thorough processing students, by contrast, evaluated the quality of the information and considered its premises, as well as the implications of different conclusions. They weighed different options, analysed connections between claims, connected related ideas and gave mainly well-reasoned explanations and convincing arguments. The findings also show that the thorough processing students believed knowledge to be tentative and fallible. However, these students did not argue that all knowledge is constructed by human beings nor did they believe that all interpretations, theories and beliefs are equally right. They thus avoided slipping into relativism. The results revealed that both deeply superficial and thorough processing thinking entails problems if it is connected to an inability to adjust thinking or actions to the demands of the task.

This thesis also identified several methodological challenges in assessing critical thinking. The results show that different performance-based critical thinking tests could give completely opposite pictures of a student's abilities. The results further indicated that the group-level analyses could overrun the rich variations in test performance that occurred among individual students. Additionally, the thesis reminds us that the theoretical framework has a great influence on how data are analysed and interpreted. Finally, the thesis argues that one assessment or analysis method is not enough to evaluate the complex cognitive processes of critical thinking. Instead of a sole focus on empirical or theoretical elements, more communication between the theoretical, empirical and methodological perspectives is required to deepen the understanding of critical thinking among students of higher education.

Keywords: critical thinking, performance assessment, epistemology, fallibilism, relativism

Helsingin yliopiston käyttäytymistieteiden laitos
Kasvatustieteellisiä tutkimuksia 260

Heidi Hyytinen

Nähdä ilmeiseltä näyttävien asioiden ja selitysten taakse

Teoreettisia, empiirisiä sekä metodologisia näkökulmia kriittisen ajattelun tutkimukseen

Tiivistelmä

Tämä väitöskirjatutkimus tarkastelee korkeakouluopiskelijoiden kriittisen ajattelun haasteita ja edellytyksiä, sekä kriittisen ajattelun ja tietokäsitysten välisiä yhteyksiä. Tutkimuksessani yhdistyy kasvatusfilosofinen ja kasvatuspsykologinen näkökulma. Tutkimukseni tavoitteena on selkiyttää korkeakoulututkimuksessa paljon hyödynnettyä teoreettista viitekehystä yksilöiden tiedosta ja tietämisen luonteesta. Aikaisemmissa, tähän viitekehykseen tukeutuvissa tutkimuksissa on todettu, että yksilöiden käsitys tiedosta ja tietämisestä on yhteydessä yksilön kriittisen ajattelun taitoihin. Väitöstutkimukseni muodostuu neljästä osatutkimuksesta, jotka tarkastelevat tutkimuksen tavoitteita eri näkökulmista.

Filosofisen analyysin perusteella argumentoin, että kasvatuspsykologisessa kriittisen ajattelun tutkimuksessa normatiiviset (esim. mitkä arvot, päämäärät ja tavoitteet suuntaavat kriittisen ajattelun oppimista tai millainen kriittinen ajattelu on hyvää) ja deskriptiiviset (esim. miten opiskelijat analysoivat tietoa) tutkimusintressit ovat kietoutuneet yhteen. Väitöskirjani osoittaa, että kriittisen ajattelun ja sen lähikäsitteiden kasvatuspsykologiset teoretisoinnit sisältävät selkeitä normatiivisia ulottuvuuksia. Tämän perusteella esitän, että kasvatuspsykologinen kriittisen ajattelun tutkimus tarvitsee sekä filosofista että empiiristä lähestymistapaa ja ennen kaikkea näiden lähestymistapojen välistä yhteistyötä. Tutkimuksessa kuvataan, miten filosofinen lähestymistapa voi edistää ja selkeyttää kasvatuspsykologisia kriittisen ajattelun teoretisointeja. Tutkimuksessa argumentoidaan lisäksi, että nykyisin vallalla olevat, lähinnä empiirisen tutkimuksen pohjalta rakennetut teoretisoinnit sisältävät useita käsitteellisiä jännitteitä, joiden yksi keskeinen syy on tietoteoreettisen fallibilismin puuttuminen. Fallibilismin mukaan kaikki inhimillinen tieto on epävarmaa, mutta ihmisellä on mahdollisuus korjata ja kehittää näkemyksiään, käsityksiään ja teorioitaan. Väitöskirjani teoreettisessa osuudessa argumentoidaan, että fallibilistinen tietokäsitys soveltuu kriittisen ajattelun taustateoriaksi.

Väitöskirjani empiirisen osuuden mukaan opiskelijoiden kriittisen ajattelun taidoissa löytyi isoja eroja. Tutkimusaineistoista tunnistettiin kaksi kriittisen ajattelun profiilia: 1) pinnallisesti tietoa prosessoivat ja (2) syvällisesti tietoa prosessoivat opiskelijat. Profiilit kuvaavat sitä, miten opiskelijat analysoivat, hankkivat, oikeuttivat ja käyttivät tietoa ongelmaratkaisutilanteessa. Pinnallisesti tietoa prosessoivat opiskelijat luottivat auktoriteetteihin ja empiirisesti todistettuun tietoon. Nämä opiskelijat tukeutuivat ongelmanratkaisussa auktoriteettien antamiin valmiisiin vastauksiin ja toistivat tietoa sellaisenaan. He eivät myöskään esittäneet perusteluja johtopäätöksilleen. Lisäksi heidän argumentaationsa oli heikkoa. Syvällisesti tietoa prosessoivat opiskelijat sen sijaan korostivat omaa vastuutaan tiedon analysoinnissa. Nämä opiskelijat suhteuttivat, analysoivat ja vertailivat eri tietoja. Lisäksi opiskelijat pohtivat tiedon sovellettavuutta ja luotettavuutta, sekä esittivät selkeän ja perustellun ratkaisun ongelmatilanteeseen omaan analyysiinsa pohjautuen. Väitöskirjatutkimuksen tulokset osoittavat, että molempiin profiileihin kuuluvat opiskelijat voivat kohdata ongelmaratkaisussa ylitsepääsemättömiä ongelmia, jos he eivät pysty joustavasti suuntaamaan omaa ajatteluaan ja toimintaansa ongelmatilanteen vaatimusten mukaisesti.

Aikaisemmissa tutkimuksissa heikot kriittisen ajattelun taidot liitetään absoluuttiseen ja naiviin realistiseen käsitykseen tiedosta, kun taas hyvät kriittisen ajattelun taidot yhdistetään relativistiseen näkemykseen tiedosta. Tässä väitöstutkimuksessa opiskelijat, joilla oli ilmeisiä ongelmia tiedon analysoinnissa ja arvioinnissa, ja sen seurauksena tukeutuivat auktoriteetteihin, eivät kuitenkaan olettaneet tiedon olevan erehtymätöntä ja lopullista. Nämä opiskelijat eivät myöskään kuvailleet, että yksilön käsitykset olisivat suoria kopioita todellisuudesta. Vastaavasti opiskelijat, jotka itsenäisesti arvioivat ja prosessoivat tiedon alkuperää ja esittivät perusteltuja johtopäätöksiä, eivät olettaneet, että kaikki tieto olisi heidän omaa konstruktiotaan. Näiden opiskelijoiden mukaan tietoa voidaan arvioida, suhteuttaa ja vertailla. Tämä tarkoittaa, että näiden opiskelijoiden tietoteoreettinen käsitys ei edusta relativismia. Molempien profiiliryhmien opiskelijat korostivat tutkitun tiedon epävarmaa, fallibilistista luonnetta.

Väitöstutkimuksen metodologisen osuuden tulokset osoittavat, että samoihin kriittisen ajattelun taitoihin keskittyvät testit voivat antaa hyvin poikkeavan kuvan opiskelijoiden taidoista. Testeihin liittyvien erojen lisäksi on erittäin tärkeää kiinnittää huomiota käytettyyn analyysimenetelmään sekä teoreettiseen viitekehykseen. Myös näillä on suuri merkitys siihen, millaisia näkökulmia ja tuloksia kriittisen ajattelun arvioinneissa esitetään tutkimustuloksina. Tutkijoiden on myös hyvä tiedostaa, että ryhmätason analyysit voivat peittää alleen yksilötasolla ilmenevän variaation opiskelijoiden kriittisessä ajattelussa. Myös käytetyllä teoriakehikolla on suuri vaikutus tutkijan tekemiin aineistotulkintoihin. Väitöstutkimukseni tulosten perusteella argumentoin, että yksittäinen analyysimenetelmä

tai testi ei ole sellaisenaan riittävä kuvaamaan kriittisen ajattelun kompleksisuutta. Tarvitaan enemmän vuoropuhelua teoreettisen, empiirisen ja metodologisten näkökulmien välillä.

Asiasanat: kriittinen ajattelu, epistemologia, tietokäsitys, fallibilismi, relativismi, osaamisen arviointi

ACKNOWLEDGEMENTS

Writing a doctoral thesis has been an amazing learning journey. It was made possible by the support and help of many people, even though it is not possible to mention all of them here.

I owe my deepest gratitude to my talented supervisors, Professor Sari Lindblom-Ylänne, Docent Katariina Holma and Docent Auli Toom. It has been an honour to know and work with all three of you. You have always been willing to share your expertise with me and offer advice. All of you have an amazing ability to ask the right questions at the perfect moment. You have encouraged me to grow not only as a researcher, but also as an independent thinker. When I started this process, my self-efficacy beliefs were rather low. You saw more in me than I saw in myself. Thank you for your support, encouragement, warmth and friendliness throughout this journey. I have learnt so much from all three of you. I am grateful to Sari Lindblom-Ylänne for the inspiring discussions and the invaluable feedback on my work during these years. Thank you providing me wonderful opportunities. I cannot thank you enough for everything you have done for me! I am deeply grateful to Katariina Holma for teaching and inspiring me to conduct theoretical research. Thank you for challenging and developing my thinking, supporting and motivating me. I would like to express my warmest thank to Auli Toom for encouraging me to follow my passion for research. Thank you for your warmest support, creative ideas and invaluable comments throughout this doctoral journey.

I am also very grateful to Emeritus Professor Richard J. Shavelson, who participated in supervising my work. Prof. Shavelson's proficiency in measuring performance assessment and critical thinking made a valuable contribution to the thesis. He was also my co-author in Study II. I would like to express my warmest thanks to my other co-authors, senior researcher Jani Ursin, Kari Nissinen, PhD, and Erika Löfström, PhD, for their valuable contributions to the third and fourth articles. The background for Study IV is the OECD's Assessment of Higher Education Learning Outcomes project (AHELO). I have had the privilege of working as part of the Finnish AHELO team and to co-operate with Jani and Kari. Thank you, Jani; your expertise in quality assurance in higher education has greatly enriched my understanding. I am deeply grateful to Kari for helping me with statistical questions. Our discussion of statistical methods has taught me very much. It was also a great pleasure to work with Erika in Study III. Her positive and open-minded attitude as well as her expertise in research integrity have greatly inspired me.

This doctoral thesis was carried out in the Centre for Research and Development of Higher Education (YTY) at the University of Helsinki. I owe my deepest thanks to all my colleagues at the Centre. I especially would like to express my warmest thanks to Docent Aino-Maija Lahtinen for the inspiring discussions during these years, to my fellow doctoral students Henna Asikainen, Milla Räisänen and Tarja Tuononen, as well as to the rest of the fantastic YTY´s research team. Henna, Milla and Tarja, we have made the doctoral journey together. I am grateful for your endless support, encouragement and sympathy. It has been a pleasure to work with such bright women. You are not only talented, but you also made this journey enjoyable. We have celebrated our successes together on so many occasions. Once we realised that the doctoral journey is not always pleasant, we decided to turn our sorrows to advantage and celebrate them too. You are the best!

I would like to thank the pre-examiners of my thesis for their valuable comments. I owe a special debt of gratitude to Dr. Ben Kotzee for his insightful and encouraging evaluations of the thesis. I am deeply grateful to Associate Professor Krista Muis for accepting the role of opponent at the public defence of the thesis. I appreciate her open-minded approach to the research on personal epistemology. In addition, I gratefully acknowledge the financial support of the Alfred Kordelin Foundation. I am also grateful to Dr. Roger Benjamin at the Council for Aid to Education for permitting me to use a performance task from the Collegiate Learning Assessment. And I want to thank the students in my empirical studies for their participation.

It has been extremely important to be able to share the ups and downs of the doctoral journey and my personal life with my dearest friends. I express my deepest thanks to Annukka, Carita, Jarmo, Seija, Arto, Liisa, Juha, Jenni and Tuomas. Your readiness to join in discussion and to encourage me to find my own path has made this journey meaningful and inspiring. I also want to express my gratitude to Helsingin taitoluisteluklubi's figure skating moms, especially Klara-Katariina. We have spent a lot of time watching our precious daughters' skating training sessions. Thank you for the great moments at the skating rink. Sometimes it feels as though the ice rink is our ice princesses' second home. Figure skaters and their parents know, perhaps better than anyone else, that willingness is not enough; you must also take action. Sometimes you need to try a new jump a thousand times before it succeeds. Fall down, get up, and try it again.

This journey would not have been possible without the dedicated support of my family. I would like to express a huge thank you to my parents for their love and support, to my lovely sisters Erja and Maiju for being both fantastic sisters and best friends —better than anyone could ask for— and to my brothers Tuomas and Markus for enriching my life. I also want to thank all of the relatives, both mine and Matti's, for being part of our lives. I am blessed to have such a caring family.

Äitiäni ja isääni haluan erityisesti kiittää rakkaudesta. Lähimmäisenrakkaus on lapsuuden kotini kulmakivi. Äidilläni oli, ja on edelleen, tapana haastaa ajatteluani käyttämällä vanhoja suomalaisia sanontoja ja runoja. Yksi suosituimmista runoista on Eino Leinon Hymyilevä Apollo, erityisesti seuraava kohta: ”Ei paha ole kenkään ihminen, vaan toinen on heikompi toista. Paljon hyvää on rinnassa jokaisen, vaikk' ei aina esille loista”. Tätä lapsuuden kodin perintöä toivon pystyväni vaalimaan elämäni tärkeimmässä tehtävässäni Mikaelin ja Matildan äitinä.

Above all, a thank you to my dear husband, Matti, who has always believed in me and encouraged me to do my best. Who has seen me succeed and seen me fail. Who knows when I need a hug. Who wipes my tears and hold tight my hand. Who celebrated our tenth wedding anniversary by running a marathon beside with me. Who knows how to enjoy the little things in life. Thank you for your endless love and care, and for all those thousand intellectually inspiring discussions over the years. This doctoral thesis is just as much yours as mine. I would have never come so far without you by my side. Thank you for these wonderful years, and for many, many more to come. There are not enough words to describe my love, appreciation and gratitude to you. I love you with all of my heart. Last but not least, thank you to my precious son Mikael and to my dazzling daughter Matilda. You are my life’s miracles. You have taught me what is really important in life. This thesis is dedicated to you, my beloved Mikael and Matilda. Never stop thinking and questioning. Rakastan teitä! Iikka-papan sanoin ”Ihiminen ossaa tehä ihan mitä vaan, jos se ite haluaa, mutta toisilla siihen mennee vähän pitemmän aikaa ja jotku ei sitte oppi millään.”

At home in Viikinmäki, April 2015

Heidi Hyytinen

CONTENTS

List of figures

List of tables

List of original publications

I Holma, K., & Hyytinen H. (equal contribution) (in press). The philosophy of personal epistemology. *Theory and Research in Education*.

II Hyytinen, H., Holma, K., Toom, A., Shavelson, R. J., & Lindblom-Ylänne, S. (2014). The complex relationship between students' critical thinking and epistemological beliefs in the context of problem solving. *Frontline Learning Research* 6: 1—15.

III Hyytinen, H., Löfström, E., & Lindblom-Ylänne, S. (2015). Challenges in argumentation and paraphrasing among beginning students in educational science. Manuscript submitted for publication.

IV Hyytinen, H., Nissinen, K., Ursin, J., Toom, A., & Lindblom-Ylänne, S. (2015). Problematising the equivalence of the test results of performance-based critical thinking tests for undergraduate students. *Studies in Educational Evaluation* 44: 1–8. doi:10.1016/j.stueduc.2014.11.001

1 INTRODUCTION

> "The value of an education in a liberal arts college is not the learning of many facts, but the training of the mind to think something that cannot be learned from textbooks."
>
> - Albert Einstein; quoted in Philipp Frank, *Einstein: His Life and Times* (1948, 185)

Critical thinking has been promoted as an educational ideal for centuries. The concept has its roots as far back as the history of Western philosophy—since the time of Socrates. It has been associated with such values as freedom and autonomy (Holma, in press; Winch, 2006). Critical thinking and reasoning are also fostered as individual capacities, which are seen to be the foundations for democratic citizenship as well as economic productivity (Arum & Roksa, 2011a). Critical thinking has thus been singled out as one of the most important skills for citizens of the twenty-first century (Halpern, 2014). Therefore, it is not surprising that mastering critical thinking is a goal that can be found in almost every higher education curriculum today.

Although critical thinking is considered a vital skill for learning and for coping with an uncertain future (Halpern, 2014), there is mounting evidence that all higher education students do not improve their critical thinking skills, including reasoning, argumentation and problem solving, during their higher education studies (e.g. Arum & Roksa, 2011a, 2011b; Bok, 2006; Pascarella, Blaich, Martin, & Hanson, 2011). As Arum and Roksa (2011a, 121) have argued based on their longitudinal studies, for many higher education students "the gains in critical thinking, complex reasoning and written communication are either exceedingly small or empirically nonexistent". Nevertheless, higher education has been identified as a suitable context in which to facilitate the learning of these skills (Rapanta, Garcia-Mila, & Gilabert, 2013). Yet graduate students have reported that their critical thinking skills are inadequate for their future working environments (Tynjälä et al., 2007), and employers have expressed similar concerns regarding students' preparedness for working life (Tynjälä, 2008).

1.1 An overview of the empirical research on critical thinking in higher education

Research on critical thinking is being pursued by scholar in various fields using two very different approaches, namely theoretical and empirical approach. In the field of higher education, empirical research on critical thinking has focused on

the development of critical thinking skills (e.g. Arum & Roksa, 2011a; Heijltjes, van Gog, Leppink, & Paas, 2014; King & Kitchener, 2002, 2004; Kuhn, 1999, 2005). Researchers have also highlighted the importance of understanding critical thinking as a social activity (e.g. Kuhn, 2005; Moore, 2004, 2013). Previous research on critical thinking has frequently applied quantitative multiple-choice tests, questionnaires (e.g. Australian Council of Education Research [henceforth ACER], 2001; Heijltjes et al., 2014; Phan, 2008; Tremblay, Lalancette, & Roseveare, 2012; Glaser, 1942) or qualitative interviews (e.g. King & Kitchener, 1994; Kaddoura, 2010; Kember, 2001). However, recently, many researchers have begun to question the reliability and adequacy of self-report questionnaires (e.g. Greene & Yu, 2014; Elby & Hammer, 2001). As a result, researchers have stated that there is a need to assess student performance directly (e.g. Elby & Hammer, 2001; Hofer, 2004; Stes, Min-Leliveld, Gijbels, & Van Petegem, 2010), for example, students' ability to construct arguments for and against a position using relevant and reliable information. Therefore, the focus on empirical research on critical thinking has begun to move to authentic performance assessment (i.e. Andiliou & Murphy, 2014; Klein, Benjamin, Shavelson, & Bolus, 2007; Shavelson, 2010). At the same time researchers have also assumed that one assessment method is not enough to evaluate complex cognitive processes such as reasoning (e.g. Baartman, Bastiaens, Kirschner, & Vleuten, 2007; Dierick & Dochy, 2001; Maclellan, 2004). This doctoral thesis addresses these concerns and responds to them by exploring students' critical thinking process in a problem-solving situation in which a qualitative multi-method approach was applied. The thesis also explores whether or not the measures of two performance-based critical thinking tests are equivalent. A strong variance in the test results would have profound implications, as this would mean that the form of assessment substantially affects the findings in the students' outcomes (Bowman, 2010).

1.2 Scarcity of theoretical analyses in the literature on educational psychology

Critical thinking is "regarded as a fundamental aim, and overriding ideal, of education" (Bailin & Siegel, 2003, 188; cf. Dewey, 1910; Winch, 2006) and it has an essential role in scientific practices. Considering that critical thinking has also been one of the most popular concepts in higher education for the last century, it is rather surprising that theoretical analyses in the context of *learning and teaching in higher education*[1] have been scarce. Nevertheless, the theoretical backgrounds for the concept of critical thinking have been discussed by many

[1] The research on learning and teaching in higher education is a multidisciplinary research field that primarily utilises psychological and educational knowledge, yet also integrates knowledge from other behavioural sciences as well as from the social sciences.

scholars in the philosophically-orientated literature over the decades (e.g. Dewey, 1910; Holma, in press; Siegel, 1988; Winch, 2006).

In the literature on educational psychology, students' creative and critical thinking is connected with the development of epistemological understanding, namely personal epistemology (e.g. Hofer & Bendixen, 2012; King & Kitchener, 2002, 2004; Kuhn, 2005; Nieminen, Lindblom-Ylänne, & Lonka, 2004). Many scholars of personal epistemology (Hofer & Pintrich, 1997; Southerland, Sinatra, & Matthews, 2001; Kitchener, 2011) have emphasised that their approach is based solely on empirical research. That is to say, their research focuses on empirical evidence, and thus has little if anything to do with philosophical analyses of knowledge. As Southerland et al. (2001, 333) explain: "[t]he key to understanding the psychological approach to the study of knowledge is that the scientists of human behaviour view knowledge and beliefs as *psychological constructs* (clearly within Popper's (1972) second world)." And, as elsewhere (2001, 331) they describe their interpretation of "Popper's second world": "the *second* is the world of subjective, individual, mental operations (the life of the mind or private consciousness)". Kitchener (2011, 89) also makes a distinction between philosophical epistemology and personal epistemology (which he recently has called PE) when he states that philosophical "[e]pistemology is concerned with providing an account of the justification condition—of when a belief or action is justified (warranted, appropriate)—whereas PE is concerned with determining the actual beliefs held by subjects along (perhaps) with causal or genealogical conditions. Justification and related concepts are, at their core, normative concepts". According to him (2011, 84), "[i]nsofar as PE research aims to be scientific—something most PE researchers would seem to want—it should be committed to the empirical testability of its claims. This motif of positivism (or better empiricism) is a legacy we should retain as an essential part of an adequate epistemology for PE research". Kitchener (2011, 85) goes on to state that "it [philosophical epistemology] could construct a completely philosophical theory of knowledge with no (or little) dependence on any scientific fact. It would follow therefore that a study of PE and the correlative concept of a PE would be different from this philosophical epistemology".

This thesis argues that it is true that the philosophical and empirical analyses of knowledge and critical thinking fundamentally differ from each other. Expressly put, a philosophical analysis of knowledge focuses on the adequate definition and justification of knowledge as a theoretical concept, whereas an empirical analysis of knowledge is interested in how human beings actually understand the nature of knowledge and how humans constitute and acquire knowledge. However, the purpose of this thesis is to argue that there are also shared concerns between these two scholarly approaches. To be exact, the descriptive assumptions concerning students' conceptions of knowledge or critical

thinking become normative by nature when these assumptions are regarded as a goal of higher education. This is not a problem as such; rather it is a necessary characteristic of a model or theory constructed in a research approach that strives both to obtain knowledge from students' actual conceptions and to make pedagogical recommendations for developing these conceptions. Therefore, the theoretical model constructed within this kind of research requires understanding students' actual conceptions as well as the normative standards of knowledge. In other words, both empirical and philosophical research must be consulted. [2] Descriptive and normative elements cannot be adequately determined without co-operation between the two types of research. Unfortunately, these two strong research approaches have remained relatively separate in today's research on higher education. This thesis argues that instead of focusing solely on empirical or theoretical elements, more communication between the theoretical, empirical and methodological perspectives is required to deepen understanding of the complex phenomenon of critical thinking. The thesis provides a multidimensional framework for analysing critical thinking by combining theoretical aspects from philosophical, educational and psychological traditions.

[2] This does not, of course, imply that either of these approaches can provide any certain or final answers, but simply points out that they are devoted to approaching these particular questions and are thus likely to have the best available answers to their core questions.

2 THE THEORETICAL FRAMEWORK FOR EXPLORING CRITICAL THINKING

2.1 Conceptualising critical thinking and its prerequisites

The concept of critical thinking derives from philosophical discussion. The definition and content of this concept have been debated in different philosophical traditions in the west (Holma, in press). In the early twentieth century, the American philosopher and psychologist John Dewey (1910, 9), the father of the modern critical-thinking tradition, referred to critical-thinking skills as reflective thinking and defined them as follows: "[a]ctive, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusions to which it tends". A number of scholars have built on Dewey's work (e.g. Ennis, 1991; Fisher & Scriven, 1997; Glaser, 1942; Paul, Elder, & Bartell, 1991; see also the American Philosophical Association, 1990). Nowadays, there are several definitions of critical thinking. One that is widely used comes from the philosopher Robert Ennis (1991, 6), who defines critical thinking as "reasonable, reflective thinking that is focused on deciding what to believe or do". This definition emphasises reflection and reasonableness as well as decision-making about belief and action. Ennis (1991) has further emphasised that critical thinking is an essential part of the problem-solving process. Many philosophical scholars have agreed that the concept of critical thinking is normative in nature (e.g. Bailin & Sigel, 2003; Ennis, 1991; Siegel & Biro, 1997). As such, critical thinking refers to the idea of *reasonable* and *good* thinking.

Critical thinking is often seen as a skilful activity in which a person may be more or less proficient (Fisher, 2011; Scheffler, 1965). Various definitions of critical thinking typically include a list of the thinking skills that characterise an ideal critical thinker. For example, Fisher (2011) lists the following: the ability to identify the elements in a reasoned case, especially reasons and conclusions; the abilities to identify and evaluate assumptions; the abilities to clarify and interpret expressions and ideas; to be able to judge the acceptability, especially the credibility, of claims; to evaluate arguments, analyse, evaluate and produce explanations; to be able to analyse, evaluate and make decisions; to draw inferences and produce arguments (see also Halpern, 2014). In the current empirically-orientated research on critical thinking (e.g. ACER, 2001; Tremblay, Lalancette & Roseveare, 2012) the notion of critical thinking as a skill has been pervasive (see also Holma, in press; Nicholas & Labig, 2013; Papastephanou & Angeli, 2007). However, many philosophers have argued that critical thinking

cannot be conceptualised solely by referring to a prescribed set of skills (Bailin & Siegel, 2003; Holma, in press; Fisher, 2011; Siegel, 1988, Scheffler, 1965; see also Halpern, 2014). It may happen that a person acquires the skills, but does not use them (Fisher, 2011). Therefore, it is not enough for students to have critical thinking skills; they also need to use these skills effectively (Holma, in press). Critical thinking always involves both the essential skills or abilities and the disposition to use them (Bailin & Siegel, 2003; Siegel, 1988). Siegel (1988) uses the term *critical spirit*, which combines some of the key elements of critical thinking. As he (1988, 39) puts it, "the critical thinker must have certain attitudes, dispositions, habits of mind and character traits which together may be labelled the critical attitude or the critical spirit". In the philosophical literature, the critical spirit is seen as an inseparable part of critical thinking (Holma, in press). According to Ritchhart (2002, 27) the main dispositions are open-minded, curious, metacognitive, seeking truth and understanding, strategic thinking and scepticism. He also defines motivation, awareness and inclination as intellectual dispositions (Ritchhart, 2002, 37).

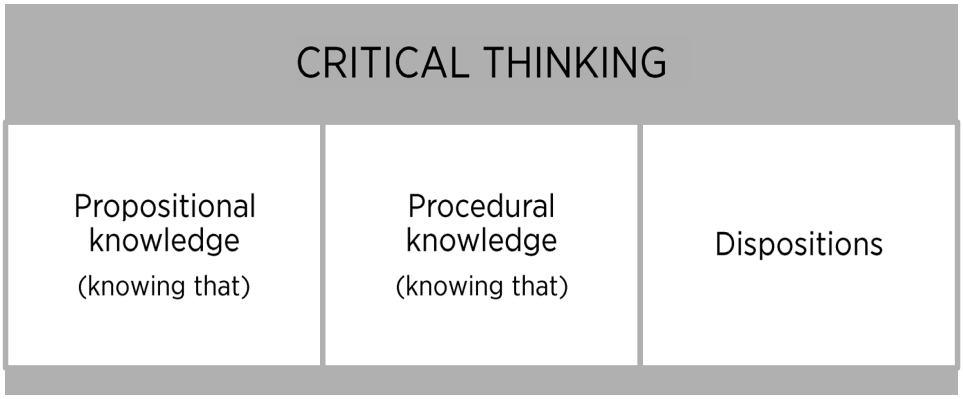
Critical thinking also demands a comprehensive use of different types of knowledge (Bok, 2006; Ennis, 1991). There is a reciprocal relationship among critical thinking, *knowledge* and *knowing*: on the one hand, students need knowledge about a phenomenon before they can think about it critically (Halpern, 2014); on the other hand, students must have the necessary skills to evaluate that knowledge. The concepts of knowledge and knowing are thus substantial aspects of conceptualising critical thinking. There are several definitions and classifications of the concept of knowledge. For example, philosophical epistemologists usually differentiate among three types of knowledge: propositional knowledge, procedural knowledge and knowledge by acquaintance (Everitt & Fisher, 1995; Ichikawa & Steup, 2012), although there is no consensus on the interpretation of knowledge or on the number of types of knowledge (Fenstermacher, 1994). For purposes of this thesis, the distinction between propositional and procedural knowledge is theoretically important.

Propositional knowledge is defined as *knowing that such-and-such is the case*. This is sometimes referred to as factual or declarative knowledge. Propositional knowledge (i.e. knowing that) is usually distinguished from procedural knowledge (i.e. knowing how) (Ryle, 1949). In philosophical discussions propositional knowledge is related to such epistemological concepts as truth, justification, reason and evidence (Ryle, 1949; Scheffler, 1965; Niiniluoto, 1999; Shope, 2004). Scheffler (1965) has argued that the *knowing that* attributes of a person may reveal his epistemological orientations, such as the criteria for justifying knowing.

Procedural knowledge, meaning *knowing how to do something* (i.e. knowing how to analyse, knowing how to solve problems; see Everitt & Fisher, 1995; Shope,

2004), is related to possessing a skill (Scheffler, 1965). In this sense critical thinking represents procedural knowledge. However, several researchers have assumed that procedural knowledge always involves some propositional knowledge as well (Everitt & Fisher, 1995; Smith, 2002; Markowitsch & Messerer, 2007). For example, if a person knows how to play chess, he will probably know certain facts (e.g. rules) about playing chess. Smith (2002) has emphasised that an individual has a certain skill only when his performance reflects both procedural and propositional knowledge. It follows that critical thinking entails both procedural and propositional knowledge. A lack of propositional or procedural knowledge limits the possibility of critical thinking (Winch, 2006).

In sum, an ideal critical thinker knows how to assess the strength of the evidence and the reasons that are relevant to the particular context or type of task. He or she also shows the disposition to draw on these skills (Bailin & Siegel, 2003; Halpern, 2014; Holma, in press; Scheffler, 1965). As Figure 1 illustrates, a critical thinker needs to have the propositional knowledge of what is reasonable, the procedural knowledge to evaluate and utilise that knowledge, as well as the disposition to do so. From this it follows that a critical thinker also needs to be autonomous: "free to act and judge independently of external constraint, on the basis of her own reasoned appraisal of the matter at hand" (Siegel, 1988, 89; see also Winch, 2002, 2006). These three elements are inseparable in the research and practice of critical thinking. However, it is important to notice that they are not necessary evenly distributed. It is thus possible, for instance, that a person has knowledge of what is reasonable and shows the disposition to use and assess that knowledge, but struggles with analysing and interpreting the knowledge.



Note: The chart is based on Bailin & Siegel (2003), Halpern (2014), Holma (in press), Scheffler (1965).

Figure 1. The prerequisites of critical thinking.

In the philosophical and educational literature (e.g. Nicholas & Labig, 2013; Bailin & Siegel, 2003), there has been a great deal of discussion about whether critical thinking is general or discipline-specific in nature. As a background assumption, I assume that there are both general and domain-specific elements in critical thinking; either alone can capture this complex phenomenon. Although the conventions of critical thinking are commonly embodied in social practices (e.g. Arum & Roksa, 2011a; Elby & Hammer, 2001; Kuhn, 2005), there are myriad subjective elements (such as students' prior knowledge, expectations, dispositions) related to and influencing critical thinking. Moreover, many dimensions of critical thinking (i.e. evaluating the reliability and relevance of evidence, identifying arguments, analysing information, addressing opposing viewpoints, reasoning) are relevant in each discipline. Moreover, it is possible that students' reasoning and critical thinking may vary within a discipline. Concerning this issue, Bailin and Siegel (2003, 184) have asserted that, although critical thinking is always bound to a particular context and often involves subject-specific knowledge, "it simply does not follow that nothing general can be said about the activity of thinking". Furthermore, they argue that the criteria of acceptability are not domain-specific. As far as I can see, the content-specific elements of critical thinking do not imply that critical thinking (such as reasoning or argumentation) can be evaluated only in relation to specific context, while outside this context nothing can be said of the credibility and justifiability of that. We can—and we need to—establish necessary assessment criteria in order to determine what kind of thinking represents valid or fallacious forms of reasoning. In this thesis, I mainly focus on general aspects of critical thinking.

2.2 Critical thinking is embedded in a progression of epistemological understanding

Research on personal epistemology today is one significant area in the investigation of learning and teaching in higher education. The research on personal epistemology explores individuals' conceptions of knowledge from the perspectives of educational and developmental psychology (Hofer & Pintrich, 1997, 2002). The literature on personal epistemology assumes that the development of creative and critical thinking is fundamentally linked with students' conceptions of knowledge (Hofer, 2001, 2005; Kember, 2001; King & Kitchener, 1994, 2002, 2004; Lucas & Tan, 2013; Phan, 2008). Previous studies on personal epistemology have concentrated especially on: (a) developing the conceptions and dimensions of epistemological beliefs (i.e. King & Kitchener, 1994, 2004; Kuhn, 1999; Perry, 1970; Schommer, 1990), (b) exploring how epistemological awareness is part of the thinking process (i.e. King & Kitchener, 1994, 2004; Kuhn 2005), (c) assessing domain-specific aspects of personal

epistemology (i.e. Hofer 2000; Hammer & Elby, 2003; Kaartinen & Lindblom-Ylänne, 2008; Muis, Bendixen, & Haerle, 2006) and (d) how personal epistemology is related to learning and student achievement (i.e. Hofer, 2001; 2004a, 2004b; Barzilai & Zohar, 2012).

According to Pintrich's widely used definition, the term *personal epistemology*[3] refers to "an individual's cognitions about the nature of knowledge and the nature of knowing" (Pintrich, 2002, 390). The term also includes a view of one's personal beliefs as a knower (Pintrich, 2002; Hofer, 2004). The term "personal epistemology can be described along a continuum from less sophisticated to more sophisticated ways of knowing" (Kaartinen-Koutaniemi & Lindblom-Ylänne, 2012, 2) or a progress "from a state of simple, absolute certainty into a multifaceted, evaluative system" (West, 2004, 61). This tradition has drawn heavily on William Perry's pioneering work of the late 1950s, when Perry began to explore Harvard freshmen's intellectual and ethical development during their college years (see Perry, 1970; Moore, 2002). On the basis of his research, Perry constructed a model demonstrating the variety and progression of students' epistemological and ethical thinking in which students' views of knowledge evolve from naive egocentrism, absolutism and dualism towards a relativist view. According to Perry (1970), the development from one position to a higher position does not follow a linear progression. Rather the progression is more likely to be wavelike and occur in surges. Perry (1970) also noted that there might be pauses between these surges, nor do all students reach the highest position on the model.

Today in the research on personal epistemology, there are several variants on its hierarchical models (e.g. Baxter Magolda, 1992; King & Kitchener, 1994, 2002, 2004; Kaartinen-Koutaniemi & Lindblom-ylänne, 2012; Kuhn, 1999, 2005; Perry, 1970; West, 2004). All these models entail claims of superiority of some conceptions of knowledge as compared with others (Hofer, 2002; Pintrich, 2002). In other words, the literature of personal epistemology makes a distinction between a lower level of epistemological beliefs (i.e. in which knowledge is perceived as consisting of unchanging facts and is acquired directly from external authorities) and a higher level of epistemological beliefs (i.e. in which knowledge is seen as uncertain and constructed by the individual) (Hofer & Pintrich, 1997, 2002; Hofer, 2005; King & Kitchener, 2002, 2004; Kuhn 2005; Kuhn &

[3] In the literature there are numerous synonyms for the term of personal epistemology, such as epistemological beliefs, epistemological assumptions, epistemic thinking, epistemic cognition, epistemological understanding and reflective judgement. According to Barbara Hofer (2005, 95), "[p]ersonal epistemology, an umbrella term that encompasses research referred to under a variety of names such as epistemological or epistemic beliefs". She further says that all "these terms have their foundation in the pioneering work of Perry (1970)". Recently, Barzilai and Zohar (2012, 41) have explained the use of the term personal epistemology in the following way: "To refer to the field as a whole, we use the term personal epistemology, the term most often used in the literature".

Weinstock, 2002; Valanides & Angeli, 2005). As Hofer (2002, 7) puts it: "[r]egardless of the number of stages, positions, or perspectives, the sequence invariable suggests movement from a dualistic, objectivist view of knowledge to a more subjective, relativistic stance and ultimately to a contextual constructivist perspective of knowing".

The idea of relativism as a sophisticated epistemological view is shared among different theoretical variants of personal epistemology (Briell, Elen & Clarebout, 2013; Hofer & Pintrich, 1997; Hofer, 2000, 2002, 2005, 2006b; Kember, 2001; Perry, 1970). The pioneering theory endorsing relativism is from Perry (1970, 109–111), who describes a revolutionary step in students' thinking in which their worldview becomes entirely relativistic. Perry (1970, 111) describes this level of personal epistemology as follows:

> Relativism is perceived as the common characteristic of *all* thought, *all* knowing, all of man's relation to his world. Against this ground, dualistic right-or-wrong thinking, and even "ideas of absolutes" becomes special cases *in the new relativistically structured context*.

Recently, some scholars have denied that relativism is the most sophisticated conception of knowledge (e.g. Elby & Hammer, 2001; Kuhn, 2005; King & Kitchener, 2004; Kitchener, 2011; Muis et al., 2006; Schommer-Aikins, 2002). It has also been argued that Perry's relativism has been misinterpreted as an ultimate form of relativism (Moore, 2002). However, it seems that today's literature on personal epistemology still endorses relativism, for example, in many of its pedagogical recommendations (see Brownlee, 2004; Kember, 2001; Lahtinen & Pehkonen, 2013). In Hofer's words (2006b, 74; see also 2001), "based on Perry's developmental scheme, faculty members have been advised for several decades to help students move from their black-and-white thinking toward a more relativist stance". In addition, the variants of personal epistemology, which have questioned the very idea of relativism as the highest level of knowing, assume that relativism is the necessary level through which one can develop towards the most sophisticated view of knowledge (Kuhn, 2005; King & Kitchener, 2004; see Table 1). According to Kuhn, Cheney and Weinstock (2000, 310):

> In what we take to be *a key event in the development of epistemological thought*, the multiplist [the relativist] relocates the source of knowledge from the known object to the knowing subject, hence becoming aware of the uncertain, subjective nature of knowing. This awareness comes to assume such proportions, however, that it overpowers and obliterates any objective standard that could serve as a basis for comparison or evaluation of conflicting claims. Because claims are subjective opinions freely chosen by their holders and everyone has a right to their opinion, all opinions are equally right.[4]

[4] Emphasis added.

Table 1. Kuhn's model of epistemological understanding.

Level	Assertions	Knowledge	Critical thinking
Realist	Copies of an external reality	Knowledge is received directly from the external world; Knowledge is certain	Unnecessary
Absolutist	Facts that are correct or incorrect in their representation of reality	Knowledge is received directly from the external world; knowledge is certain but not directly accessible, producing false beliefs	A vehicle for comparing assertions to reality and determining their truth or falsehood
Relativist (sometimes called multiplist)	Opinions freely chosen by and accountable only to their owners	Knowledge is generated by human minds; Knowledge is uncertain	Irrelevant
Evaluativist (in Perry's model this level is labelled 'commitment within relativism')	Judgements that can be evaluated and compared according to criteria of argument and evidence	Knowledge is generated by human minds; uncertain but susceptible to evaluation	Valuable as a vehicle that promotes sound assertions and enhances understanding

Note: Modified from "A Developmental Model of Critical thinking", by D. Kuhn, 1999, *Educational Researcher, 28*, p. 23.

In recent years, research on personal epistemology has shown that the ability to think critically is embedded in a progression of epistemological understandings (i.e. King & Kitchener, 2002, 2004; Kuhn & Weinstock, 2002; Kuhn, 1999, 2005; see also Table 1). It has also been demonstrated that students' epistemological beliefs play an important role in their ability to evaluate the credibility of competing claims (Barzilai & Zohar, 2012). Several researchers have stated that students with higher-level epistemological beliefs have more sophisticated critical thinking skills than students with lower-level epistemological beliefs (King & Kitchener, 2002, 2004; Kuhn & Weinstock, 2002; Kuhn, 1999, 2005). When students move on to the most developed epistemological level, their critical thinking tends to improve as well (Bok, 2006; Kuhn, 1999; Kuhn & Weinstock, 2002). Many scholars within the tradition have also emphasised that students with higher-level epistemological beliefs have the ability to make context-specific judgements (e.g. Barzilai & Zohar, 2012; Kuhn, 2005; King & Kitchener, 2004).

In the context of personal epistemology, there are contradictory assumptions among scholars related to the theoretical groundings of the concept of critical thinking. Some hold that "critical and creative thinking is only possible if relativism is recognised" (Kember, 2001, 217), whereas others see critical thinking as being irrelevant within a relativist framework (e.g. Kuhn, 2005). In a similar vein, philosophical literature holds that the idea that critical thinking presupposes the relativist view of knowledge is untenable (e.g. Bleazby, 2011; Holma & Hyytinen, in press-a, in press-b).

2.3 The manifold philosophical critiques of relativism

Relativism has been criticised as an extremely problematic epistemological position within philosophical epistemology (e.g. Boghossian, 2006; Niiniluoto, 1999; Siegel, 1987). According to philosophical tradition, the relativist position implies that all knowledge is relative to the person who believes or that all interpretations, theories and beliefs are equally right. To put it simply, if all beliefs are equally right or relative to the person, then there is no reason to compare and evaluate different beliefs; all are equally justified (Holma, 2012). Hence, a crucial philosophical question that relativists have difficulty in answering is *the problem of disagreement*, given that, within the relativist framework, no belief can be seen as incorrect. In Boghossian's words (2006, 39):

> If a given fact really does owe its existence to our intentional activities, it is hard to see how there could fail to be possible circumstances in which we might have chosen to construct a different fact incompatible with it.

However, the most fundamental problem of relativism is that it is *a self-refuting position*. According to Harvey Siegel (1987, 8–9):

> the relativist must appeal to non-relativist criteria, and assert relativism non-relativistically, in order to make the case for relativism. This is self-defeating for the relativist. But to fail to assert and defend relativism in this (non-relativistic) way is to fail to join the issue with the non-relativist who asserts that relativism is false (or incoherent). So the relativist can defend relativism only by rendering it incoherent. Conversely, to defend relativism relativistically is to fail to defend it at all. For if relativism is right, the very notion of rightness, and indeed that of rational defense, is given up, and so it cannot coherently be claimed that relativism is right or rationally defensible. In short: to defend relativism is to defend it non-relativistically, which is to give it up; to 'defend' it relativistically is not to defend it at all.

Boghossian (2006, 56) has argued that "either the formulation [the relativist] offers us does not succeed in expressing the view that there are only relative facts; or it consists in the claim that we should so reinterpret our utterances that they

express infinitary propositions that we can neither express nor understand". The problem of relativism becomes evident when it is related to the concept of critical thinking: there is no need to evaluate ideas or search for alternatives because all ideas are equally trustworthy and justifiable (Bleazby, 2011; Holma & Hyytinen, in press-a, in press-b; see also Kuhn, 2005). Therefore, the idea that critical thinking presupposes the relativist view of knowledge is untenable.

From the relativist point of view, the idea that all knowledge is generated by human minds is also considered as highly problematic (Boghossian 2006; Scheffler, 2000, 2001). It follows that knowledge can exist only if there are human beings who make and conceptualise it. For example, the world as constituted by the natural sciences can exist only when there are humans who perceive and conceptualise it. Israel Scheffler (1997, 199–200) has explained this problem in the following way:

> Now, whether a world answers to a version of our making is, in general, not up to us. Thus, if an "actual world" answers to a version of our making, we can hardly be supposed to have made it do so. Moreover, if a version of our making turns out to be true, it hardly follows that we have made its objects. Neither Pasteur nor his version of the germ theory made the bacteria he postulated, nor was Neptune created by Adams and Leverrier or by their prescient computations.

2.4 Various research methods on critical thinking

In examining the critical thinking of higher education students, researchers have used a variety of tests and assessment protocols. The assessments can be roughly divided into two main measurement protocols: *self-reports* and *performance-based assessments*. In the area of critical thinking, previous studies have compared the results of self-reports with performance-based assessments (e.g. Bowman, 2010; Bowman & Seifert, 2011). These studies have shown that self-reports and performance assessments measure different aspects of students' abilities and therefore yield different pictures of those abilities. In the present study, both of these methods are used to explore students' critical thinking.

Self-reports, such as surveys, questionnaires and qualitative interviews, focus on students' perceptions of their current attributes or how these attributes have developed over time (Bowman, 2010). The validity of self-report assessment has been discussed extensively (e.g. Bowman, 2010; Bowman & Seifert, 2011; Halpern, 1993; Pike, 1995, 1996, 1999). Halpern (1993, 279) summarises the problems of self-report instruments in the following way: "students may report that they have learned to think better when, in fact, they have not or, conversely, that they have not improved when they really have". Self-report assessments can be described as indirect measures, because they do not directly measure students'

performance or actual changes in their learning. Although such measures do not capture concrete evidence of changes in learning, they may be important indicators for improving educational processes (Shavelson, 2010).

Performance assessment is sometimes presented as a new assessment approach (Andrews & Wulfeck, 2014; Dierick & Dochy, 2001). However, there is a long history of using performance-based assessment as an indicator of students' learning and development in higher education in order to make educational decisions (see Douglass, Thomson, & Zhao, 2012; Ennis, 1991). The roots of today's performance-based assessment can be traced to the first third of the twentieth century with the beginning of standardised testing (Shavelson, 2010). What the various performance assessments have in common is the goal of eliciting what students know and can do (Andrews & Wulfeck, 2014). Performance-based assessments can be further grouped into two approaches, namely (1) multiple-choice tests or questionnaires and (2) constructed-response tasks. Below I differentiate between these two main forms of performance-based assessments.

Multiple-choice tests have been the dominant testing regime in the field of research on critical thinking (cf. Ennis, 1991; Shavelson, 2010). In the test situation the student must analyse a question and then identify and select the correct answers from a list of given options. In contrast to the constructed-response task, multiple-choice tests are often promoted as cost effective and objective (Brown, 2001), as there is no need for human evaluation in scoring them. The cognitive demands of multiple-choice tests have been debated for several decades (e.g. Bennet & Ward, 1993; Jensen, McDaniel, & Woodard, 2014; Lindblom-Ylänne, Lonka, & Leskinen, 1996; Nicol, 2007; Popham, 2003). Many researchers have argued that a multiple-choice test does not encourage students to use higher-order thinking processes (Nicol, 2007; Scouller, 1998). The reasoning behind that claim is that multiple-choice tests may be answered merely by low-level processing, such as factual recognition and selection (Lindblom-Ylänne et al., 1996; Nicol, 2007). The general view also suggests that it is more difficult to construct an answer than to recognise it. It is also possible to obtain a correct multiple-choice answer without really understanding a problem or knowing the various aspects related to it. For example, students can choose one item among the possible choices that best suits the question asked, and, of course, it is possible to guess the right answer from the alternatives given. Examinees can be assured that the correct answer is among the response options. Another weakness is that students "may be able to recognize a correct answer that they would never have been able to generate on their own. In that sense, multiple-choice items can present an exaggerated picture of a students' understanding or competence, which might lead teachers to invalid inferences" (Popham, 2003, 81–82). Although there is evidence that by applying a well-designed multiple-choice questionnaire it is possible to measure higher-order thinking (Jensen et al., 2014),

on the basis of a student's answer it is not possible to determine how the student processed the test questions (Lindblom-Ylänne et al., 1996). Nor can multiple-choice tests ever assess a student's skill in synthesising or generating an answer (Popham, 2003).

To address the limitation of multiple-choice tests, researchers have developed alternative assessment methods, specifically the *constructed-response tasks* (Bennet & Ward, 1993). There are several advantages to assessing students' critical thinking by applying constructed-response tasks. As Coates and Richardson (2011, 63) put it:

> while well-designed multiple choice questions can measure complex real-world thinking, many advanced forms of reasoning are best measured using tasks that require students to construct their response in the form of writing or drawing rather than selecting from a set of pre-defined alternatives.

In the constructed-response tasks examinees create their own answers to the questions (Coates & Richardson, 2011; Shavelson, 2010; Rodriguez, 2003). These types of measures are often open-ended tasks in which students need to analyse, evaluate and synthesise complex information, as well as provide reasoned explanation (cf. Popham, 2003; Shavelson, 2010). Therefore, the constructed-response tasks are said to promote higher-order thinking and encourage extended problem solving more than the multiple-choice tasks. Another advantage is that the constructed-response tasks can reveal the level of understanding (Popham, 2003). They also allow students to demonstrate their writing skills (VanTassel-Baska, 2014). The constructed-response task is sometimes referred to as *authentic assessment* because these tasks demonstrate the same thinking processes that individuals use when they solve complex problems in their daily lives (Andrew & Wulfeck, 2014; Baartman et al., 2006). In addition, Andiliou and Murphy (2014) have recently found that the constructed-response tasks support students' self-evaluation skills.

However, several disadvantages of the constructed-response task have been reported. The most important is the difficulty of scoring (Attali, 2014). The constructed-response assessment is characterised as subjective and open to scoring bias, because examinees' responses are traditionally scored using human evaluation. The scoring of constructed-response tasks is also considered time consuming and expensive; a large amount of time and effort is needed to train scorers and to score the responses (Attali, 2014). Recently, automated scoring of tasks has been developed (Almond, 2014). Popham (2003, 87) has also argued that constructed-response tasks "are tougher for test-takers", because "a student really needs to understand something in order to construct a response on that understanding". An example of a constructed response task is illustrated in Figure 2.

Performance task: Crime Reduction

Introductions: You will have 90 minutes to complete this task. This task will ask you to analyze a collection of different types of information. You will then use your analysis to prepare answers to a series of questions. Although you may not be familiar with some of the topics covered, you should be able to prepare appropriate answers by carefully using and thoughtfully reflecting on the information given to you. Your answers should clearly state what you mean. Please do your best.

Role: You are a consultant to Mayor Stone. Pat Stone is running for reelection as mayor of Jefferson, a city in the state of Columbia.

Scenario: Mayor Stone's opponent in this contest is Dr. Jamie Eager. Dr. Eager is a member of the Jefferson City Council. Dr. Eager made the following three arguments during a recent TV interview: First, Mayor Stone's proposal for reducing crime by increasing the number of police officers is a bad idea. Dr. Eager said "it will only lead to more crime." Dr. Eager supported this argument with a chart that shows that counties with a relatively large number of police officers per resident tend to have more crime than those with fewer officers per resident. Second, Dr. Eager said "we should take the money that would have gone to hiring more police officers and spend it on the STRIVE drug treatment program." Dr. Eager supported this argument by referring to a news release by the Washington Institute for Social Research that describes the effectiveness of the STRIVE drug treatment program. Dr. Eager also said there were other scientific studies that showed the STRIVE program was effective. Third, Dr. Eager said that because of the strong correlation between drug use and crime in Jefferson, reducing the number of addicts would lower the city's crime rate. To support this argument, Dr. Eager showed a chart that compared the percentage of drug addicts in a Jefferson zip code area to the number of crimes committed in that area. Dr. Eager based this chart on crime and community data tables that were provided by the Jefferson Police Department.

Task: Mayor Stone has asked you to prepare a memo that analyzes the strengths and limitations of each of Dr. Eager's three main points, including any holes in those arguments. Your memo also should contain your conclusions about each of Dr. Eager's three points, explain the reasons for your conclusions, and justify those conclusions by referring to the specific documents, data, and statements on which your conclusions are based. To do so, please answer the four questions that follow, using the documents provided.

You are provided with the following documentation:	
A: Investigator's Memo	E. Crime Statistics
B: Newspaper Story	F: Dr. Eager's Chart
C: Police Tables	G: Research Abstracts
D: Research Report	

Figure 2. An example of a constructed-response task.

Whether the multiple-choice test and the constructed-response task with the same content measure precisely the same characteristics has been debated for over 20 years (e.g. Bennet & Ward, 1993; Fellenz, 2004; Rodriguez, 2003). According to Rodriguez (2003, 180), there is evidence that "the more carefully items are designed to measure the same aspect, the more appropriate it is to combine the

scores from each format without concern". Recent research has shown that there are several aspects related to and influencing students' test performance, such as dispositions (including motivation, open-mindedness, curiosity, metacognition), skills, test anxiety, cognitive processing during the test, the knowledge structure assessed, institutional or cultural barriers and adequacy of the tests (e.g. Arum & Roksa, 2011a; Fellenz, 2004). It is worth mentioning that these differences are found not only between the multiple-choice and constructed-response formats, but also within each test format.

2.5 Summary of the key concepts

The above perspectives concerning the theory and practice of critical thinking offer a starting point for this doctoral thesis, which aims to deepen understanding of the emphases and gaps in prevailing research on critical thinking within the field of higher education. Table 2 provides a summary of the definitions of the key concepts in this study. Given this broader framework, it is possible to pin down different areas in critical thinking. Figure 3 illustrates the relationship between the key concepts of the present thesis, which is based on the view that the adequate conceptualising of critical thinking involves knowledge, skills and dispositions. The empirical part of the thesis concentrates on students' actions as well as on their thinking and decisions in the problem-solving situation. It focuses specifically on the following aspects of critical thinking: (1) identifying, interpreting and synthesising information from multiple sources to reach a conclusion; (2) evaluating the acceptability and reasonability of information; and (3) using information in producing explanations, decisions and arguments. Personal epistemology is seen here as students' conceptions about the nature of knowledge and knowing (i.e. the definition, content, justification and sources of knowledge), including personal beliefs about themselves as knowers. Recently, Harvey Siegel criticised the use of the term epistemology in science education; he (2014, 372) emphasised that "'epistemology' is the name of that branch of philosophy dedicated to the theory of Knowledge."[5] Therefore, I prefer the term *conceptions of knowledge* rather than epistemological beliefs or personal epistemology. I use term personal epistemology to refer the research tradition of personal epistemology.

[5] Similarly, in the contexts of personal epistemology some researches have stated that they prefer the term *epistemic* rather than epistemological because "*epistemological* relates to the study of epistemology"(Barzilai & Zohar, 2012, 41; see also Kitchener, 2002).

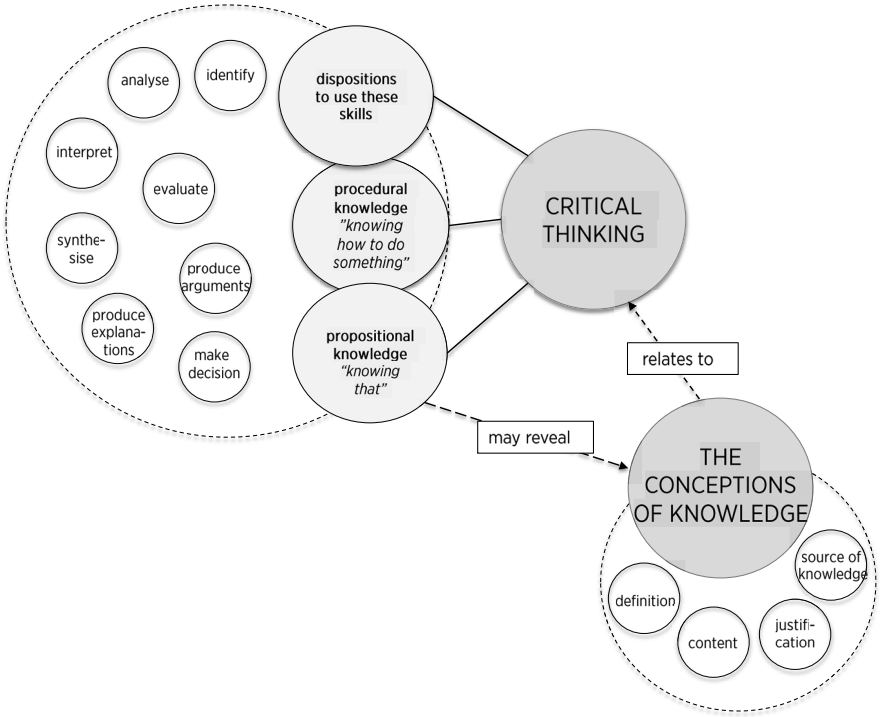


Figure 3. Relationships among the key concepts of this thesis.

Table 2. Summary of the key concepts of this study.

Concept	Description
Critical thinking	Process that enables an individual to make an informed decision about what to believe or do. It involves skills and dispositions (e.g. attitude, motivation, open-mindedness, curiosity, metacognition) to identify, analyse, interpret, synthesise information from a variety of sources to reach a conclusion, to evaluate the acceptability and reasonability of information, to use the information in producing explanations, decisions and arguments and to draw valid conclusions and address opposing viewpoints.[1] Critical thinking involves *knowing how to do something* (procedural knowledge) and *knowing that* (propositional knowledge).[2]
Conceptions of knowledge	Students' thoughts/beliefs about the nature of knowledge (the definition, content, justification and sources of knowledge) and the nature of knowing, including personal beliefs about themselves as knowers.[3]

[1] Based on Bailin & Siegel (2003), Ennis (1991), Fisher (2011), Fisher & Scriven (1997), Siegel (1988).
[2] Based on Scheffler (1965); cf. also Ryle (1949).
[3] Modified from Pintrich (2002).

3 THE AIMS OF THE DOCTORAL THESIS

The present thesis is based on four original publications, which are referred to in the text by Roman numerals (Studies I–IV). The general aim of the thesis is to deepen understanding of the nature of critical thinking by combining theoretical, empirical and methodological perspectives. More specifically, the focus of this research is threefold:

Firstly, the *theoretical* aim is to clarify the empirically-based theorisation[6] of critical thinking (Studies I & II). The purpose is to offer new, sustainable, insights into conceptualising critical thinking and its prerequisites, especially knowledge and knowing. The nature of knowledge and knowing has been discussed by educational psychologists exploring personal epistemology as well as by philosophers. As noted above, these two strong research traditions have remained relatively separate. Educational psychologists focus on empirical evidence concerning students' conception of knowledge; philosophers focus on theories of the conditions of knowledge. The present study sets out to identify conceptual differences and similarities between the research traditions of educational psychology and philosophy as well as to integrate these viewpoints where appropriate. I assume that although empirical research on personal epistemology and philosophical research on the conditions of knowledge are two different fields of study, these traditions have shared concerns. In particular, normative assumptions concerning individuals' conceptions of knowledge and critical thinking would benefit from the theoretical and conceptual analysis used in philosophical epistemology (i.e. what kind of critical thinking or conception of knowledge is worth pursuing in the first place).

Secondly, the *empirical* focus of this study is to examine and describe qualitative differences in students' critical thinking skills in a specific problem-solving situation (Studies II, III & IV), and to analyse the interconnections between students' conceptions of knowledge and their critical thinking skills (Study II). The aim is to bring the theoretical insights into dialogue with empirical elaboration. The empirical part of this doctoral thesis combines theoretical and methodological perspectives and thereby endeavours to bring forth new empirical data on critical thinking and conceptions of knowledge and extend understanding of the variations in critical thinking.

Thirdly, the *methodological* aim is to enhance understanding of the important role of the analysis method (Studies II & IV) and explore the critical role of the type of assessment instrument selected to measure students' critical thinking (Studies II & IV). In the area of critical thinking, previous studies have more often

[6] Theoretical formulations are based on empirical research findings.

compared the results of self-reports and performance-based assessments (e.g. Bowman, 2010; Bowman & Seifert, 2011). These studies have shown that self-reports and performance assessments measure different aspects of students' abilities and therefore yield different pictures of those abilities. Study IV addresses this issue by comparing the test results of two different performance-based critical thinking tests. The aim is to explore how closely the measures of the two critical thinking tests are aligned. A strong variance in the test results would have profound implications, as this would mean that the form of assessment substantially affects the findings about student outcomes (Bowman, 2010). Information about the differences between various performance assessment instruments is valuable for interpreting the results of critical thinking tests.

The present thesis utilises a multi-method approach (Johnson, Onwuegbuzie & Turner, 2007), including think-aloud protocol, interviews, performance tasks and a multiple-choice questionnaire, to explore the complex connections between critical thinking and conceptions of knowledge in a specific problem-solving situation.

These aims were approached through the following research questions

Theoretical questions:
A1 To what extent does empirical research on personal epistemology and philosophical analysis of the conditions of knowledge relate to each other?
A2 What new perspectives does the philosophical understanding of the conditions of knowledge offer empirical research on the conception of knowledge?

Empirical questions:
B1 How is critical thinking presented in a specific problem-solving situation? How does it vary from one individual to another?

- How do students process and approach problems and solve them?
- What kinds of knowledge do students use to solve problems?
- How do students analyse, evaluate, synthesise and interpret that knowledge?
- How do students utilise that knowledge in producing arguments and counter-arguments?
- How do students come to a conclusion and validate their conclusions?

B2 How do the conceptions of knowledge vary from one individual to another?
B3 How are students' conceptions of knowledge linked with critical thinking?

Methodological question:
C1 How do the analysis method and the type of assessment instruments influence the findings of students' critical thinking?

4 METHODS AND PROCEDURES

This section presents an overview of the context and methodological choices of the four studies. The content of each study is explained in detail in the original publications.

Study I: Analysing the normative elements of personal epistemology

> Holma, K., & Hyytinen H. (equal contribution) (in press). The philosophy of personal epistemology. *Theory and Research in Education*.

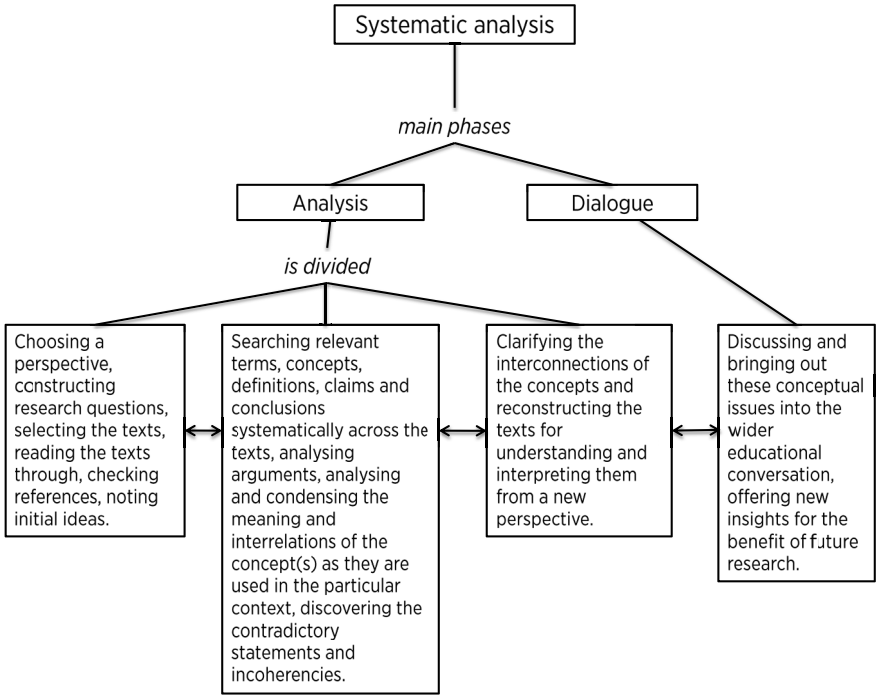
Study I analysed the theoretical background assumptions of personal epistemology (henceforth PE). Previous research on PE has provided evidence that an individual's epistemological positions influence learning, critical thinking and academic success (Kuhn, 2005; Schommer-Aikins & Easter, 2006; Hofer & Pintrich, 2002; King & Kitchener, 2004). Several scholars in the tradition of PE have endeavoured to analyse and describe the relationship between psychological and philosophical epistemology (e.g. Greene et al., 2008, 2010; Muis, 2004; Muis et al., 2006; Murphy, 2003; Southerland et al., 2001). However, it has been noted that the connection between these two research fields is still not clear (Buckland, 2010; Chinn et al., 2011). Murphy, Alexander and Muis (2012) have recently argued that a deeper understanding of the nature and conditions of knowledge is needed in research on PE.

The aim of Study I was twofold. Firstly, the study set out to identify the conceptual differences between the research traditions of educational psychology and those of philosophy, as well as to integrate appropriate viewpoints. Secondly, the aim was to explain the conceptual inconsistencies of personal epistemology from a philosophical point of view and suggest a remedy for solving these problems within the prevailing theoretical framework of PE. The study utilised a systematic analysis (Holma, 2009; Hyytinen, 2011), which is a common analytical method in the philosophy of education. This kind of theoretical analysis endeavours to build a bridge between theory and practice by "combining philosophical ways of analysing and arguing with the dialogical and pluralist way of thinking needed in educational research" (Holma, 2009, 325). I conducted this study with Dr. Katariina Holma, and we made equal contributions. Next I will explain the methodological choices of Study I. It is important to note that these aspects are not traditionally explicated in philosophically-orientated texts. That is the reason why a method section is missing from the original publication.

The theoretical analysis can be described as "a thoroughgoing process of analysis and synthesis" (Holma, 2009, 325). The phases of analysis are described

in Figure 4. The analytical process focused on the normative elements of PE and it was divided into two main phases—analysis and dialogue. During the first phase the analytical perspectives were determined and the texts selected. We chose pioneering and classic studies and hierarchical models of PE, as well as other articles that make pedagogical recommendations for developing students' conceptions of knowledge. To ensure that the selected texts generated an overview of the personal epistemology literature, the relevant texts were also sought in the electronic databases ISI, EBSCO, Proquest and Scopus. The following keywords and phrases were combined in searching for texts in these databases: *"epistemological belie*", "personal epistemolog*", "nature of knowledge", epistemic*, "epistemological assumption*", "epistemic thinking*", "epistemological belie*", knowledge, knowing, "higher education", college*, university, "university student*"*. The search through the databases was limited to articles published in international, academic, peer-reviewed journals. Furthermore, the snowball method was used to identify the relevant texts. The final database consisted of the following texts: 55 journal articles (including reviews, theoretical and empirical articles) and 4 books.

All the texts were read through, and the relevant terms, concepts, definitions, claims and conclusions among them were identified. We focused on the following aspects: a definition of PE, theoretical variants of hierarchical models of PE, relativism, evaluativism, realism, absolutism, truth, objectivity and certainty. The aim was to understand these concepts as they were used in this particular context. At the same time the interconnections of the concepts as well as contradictory statements and inconsistencies were observed. We compared, among other things, the theoretical variants in PE and definitions of the above-mentioned concepts. Our aim was to condense the meaning and interrelations of the key concepts that are crucial for understanding the PE debate. The last phase of the analysis was reconstruction. Here we considered solutions to several theoretical inconsistencies identified in the analysis. The aim of the second main phase—the last phase of the systematic analysis—was to introduce the findings into a broader educational conversation and endorse a constructive dialogue between the research traditions of educational psychology and those of philosophy.



Note: Based on Holma (2009), Hyytinen (2010) and Jussila et al. (1989).

Figure 4. The phases of systematic analysis.

Study II: Qualitative analyses of the relationship between students' critical thinking and conceptions of knowledge

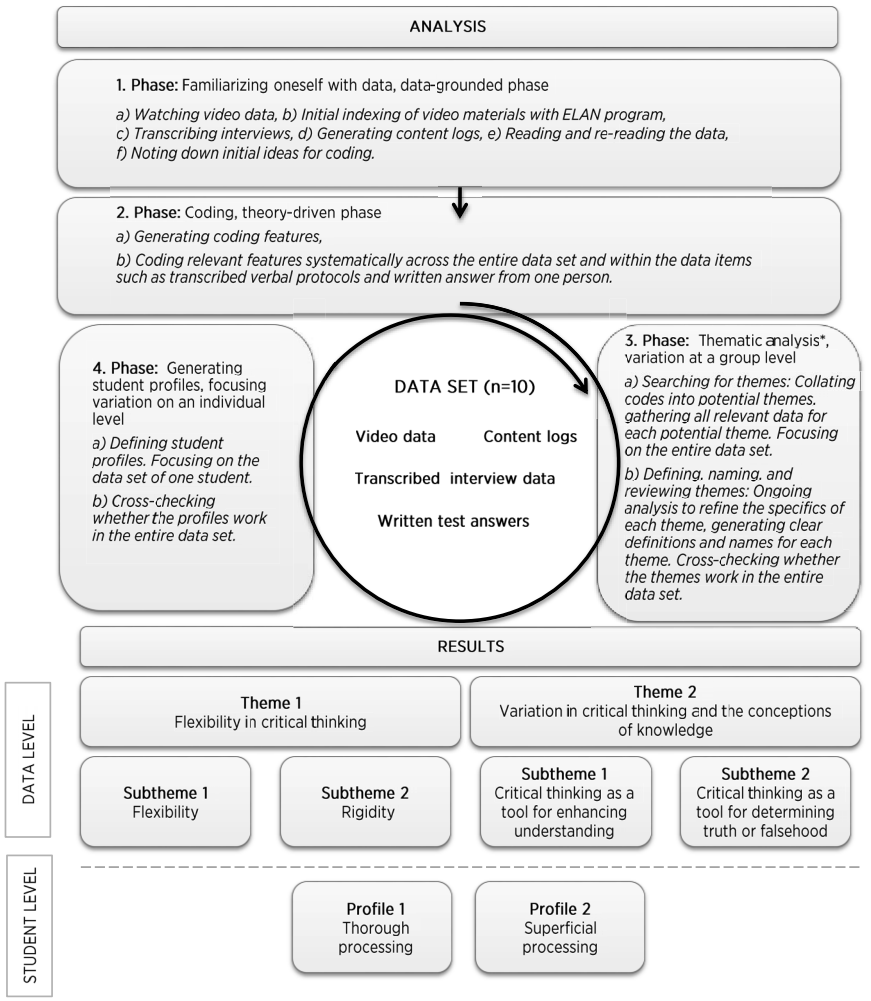
> Hyytinen, H., Holma, K., Toom, A., Shavelson, R. J., & Lindblom-Ylänne, S. (2014). The complex relationship between students' critical thinking and epistemological beliefs in the context of problem solving. *Frontline Learning Research* 6: 1–15.

Study II utilised a multi-method approach in order to identify and describe qualitative differences in students' critical thinking skills and to explore the connection between critical thinking and the conception of knowledge in a specific problem-solving situation. The target population for this study consisted of the entire class of third-year bioscience students in a certain Finnish university. Firstly, 40 students were selected at random (approximately one-half of the target population). All the students selected were then invited to participate in the study. Ten of the 40 volunteered. Seven of the participants were female and three were male. The students' ages varied from 22 to 29, the mean age being 24. All came from a homogeneous cultural background, and all shared the same first language

(Finnish). In addition, the students had equivalent national high school certificate and had enrolled in the same bachelor's study programme. The participants were at the same phase of their studies, that is, near the end of their bachelor's degree, with the exception of one student whose study pace had been slower. During their university careers, the students had participated in lectures, practical laboratories, seminars, field courses and web-based teaching. The purpose of Study II was to deepen understanding of the critical thinking and conceptions of knowledge among university students in order to describe how these phenomena varied among individuals in this specific group.

For Study II a large body of data for each participant was collected using a multi-method approach (Johnson, Onwuegbuzie, & Turner, 2007), including think-aloud protocol, interviews and a Collegiate Learning Assessment (CLA) performance task (the latter is described in more detail below). The CLA performance task was the same as in Study IV. The data collection was carried out in the spring of 2010 and consisted of ten cognitive labs. The students came to a classroom and were given the details of the study. They then spent two to three hours reading and responding to the performance task. In responding to the task, the students were asked to verbalise their thoughts (to think aloud). In the course of carrying out the task while thinking aloud, the students were also asked to write a memorandum addressing critical issues in the task and recommending—and justifying—a course of action. Following the task, the students were interviewed about their processes in carrying out the task. Students were also asked questions about the learning of critical thinking and how they understood the nature of knowledge and knowing.

The cognitive-lab data were analysed using thematic analysis (Braun & Clarke, 2006) with an abductive approach (Timmermans & Tavory, 2012; Haig, 2005). An abductive strategy means that the themes identified in the data were linked to theoretical understanding based on previous studies. Abduction is a process that combines things which were not previously associated by creating a new interpretation, that is, the relationship of a new combination of study features (Timmermans & Tavory, 2012). Hence, the analytical process was nonlinear, moving back and forth among all the data, data items, analysed qualities and understanding of the phenomenon based on prior studies. The analysis included four phases (Figure 5) representing the unique combination of data-grounded and theory-driven phases, as well as group-level and individual-level analyses. In the analysis investigator triangulation (Denzin, 1970, 2012; Creswell & Miller, 2000) was utilised to confirm the reliability of the findings.



*Based on Braun & Clarke (2006, 87).

Figure 5. A visualisation of the analysis process in Study II.

During the first phase, video recordings were initially indexed with the ELAN programme, which allows the addition of as many tiers and annotations in the video stream as needed (see Lausberg & Sloetjes, 2009; Max Planck Institute for Psycholinguistics, 2012). The purpose of indexing was to make the large video data set easier to handle. In this study the indexing tiers corresponded to the parts of cognitive labs including training, think-aloud methods and interviews. In addition, students' interviews on the videos were transcribed. After the indexing, content logs were created for each video in which accurate descriptions and

summaries of events were systematically recorded. Transcriptions of relevant sections of verbalisations of students' critical thinking and conceptions of knowledge (e.g. whenever a student evaluated the quality and reliability of the information in a document or reached a conclusion based on her or his analysis) and nonverbal acts (e.g. a student did not read in detail or skipped over the document) were also included in the log.

The second phase of the analysis was data coding (see Table 3 for definitions). This phase was theory-driven, meaning that the features guiding the coding were based on prior studies. The coding focused on the following qualities: *the process by which the student approached the task and solved the problem, the knowledge that the student used to carry out the task, the critical thinking exhibited, and the conceptions of knowledge*. These different qualities were coded systematically across the entire data set and within the data items, such as the transcribed interviews and the think-aloud videos of each person. By this means, all the data items from one student, including the video data, content log, written test answers and transcribed interviews, were coded and analysed separately, after which the data from all the students were combined and compared (see Table 4 for an example of the codes). All extracts were labelled with a student code (S1-S10) and a method code (I= interview, T=think aloud, W= written test answer).

Table 3. Data sources and focal points for coding in Study II.

Data Sources	Coding Features
Video data, content logs, transcribed interviews	1. The process: how does the student approach the task and solve the problem?
Video data, students' written answers, content logs, transcribed interviews	2. What *knowledge/information* does the student use to solve the task? 2.1 What kind of knowledge/information did the student use? 2.2 Why? 2.3 How does the student use that knowledge/information?
Video data, students' written answers, content logs, transcribed interviews	3. *Critical thinking* 3.1 How does the student identify, analyse and evaluate information, ideas and arguments? 3.2 How does the student judge the acceptability (especially the credibility) of documents? 3.3 How does the student interpret data/ graphs/ maps? 3.4 How does the student recognise the relationship between assumptions? 3.5 How does the student evaluate background information? 3.6 How does the student make a decision? 3.7 How does the student identify reasons and come to a conclusion? 3.8 How does the student produce explanations and arguments?
Video data, students' written answers, content logs, transcribed interviews	4. *Epistemological beliefs* 4.1 What does the student think about knowledge, knowing and the credibility of knowledge? 4.2 How does the student determine the trustworthiness, acceptability and justification of different types of information? 4.3 How does the student describe herself or himself as a knower?

Table 4. An example of codes.

Data Extract	Coded for
You could consider this a good argument; the expert has gone [to the place where events took place] to see for himself. (S9T)	4.1 What does the student think about knowledge, knowing and the credibility of knowledge? 4.2 How does the student determine the trustworthiness, acceptability and justification of the different types of information?
- - yeah, I don't believe [the chair of the stakeholder group] is completely off the mark either. [Reliability] is just always case-specific. (S8I)	4.1 What does the student think about knowledge, knowing and the credibility of knowledge?

In the third phase the codes and coded extracts were grouped under potential themes, and all the relevant data were gathered under each theme (Braun & Clarke, 2006). The aim was to identify themes which capture a holistic picture of a phenomenon. A variety of preliminary themes were identified on the basis of the codes. During the analysis, the preliminary themes were defined and combined several times. In the end two main themes and two subthemes remained (see Figure 5). The final themes were refined, labelled and cross-checked to see if they worked in relation to the coded extracts and the entire data set. The focus of the thematic analysis was the variation of study features on the phenomenon level. After completing the thematic analysis, I found that the students could be placed in one of several profiles based on our themes as well as on the patterns of behaviour and cognition observed. This phase focused on the variation in study features at the individual level. Thereafter, final descriptions, interpretations and revisions of the results were conducted.

Study III: Qualitative analyses of challenges in argumentation and paraphrasing

Hyytinen, H., Löfström, E., & Lindblom-Ylänne, S. (2015). Challenges in argumentation and paraphrasing among beginning students in educational science. Manuscript submitted for publication.

Study III aimed to identify difficulties in writing at the beginning of educational science programmes in the Finnish Open University by analysing the students' written argumentation and use of sources at the textual level. Study III focused on common problems identified in the review of the literature (Darab, 2006; Jurowska & Thompson, 2012; Keinonen & Kärkkäinen, 2010; Sandoval, 2009), such as using sources and constructing solid arguments, and it attempted to understand how paraphrasing and argumentation are problematic for higher education students. Even though there are many field-specific conventions in academic writing, it is also possible to identify similarities in texts at the micro level: we can analyse the construction of a solid argument (Walton, 1990, 1995) and the nature of source citation if we have the original source at our disposal.

The participants were 138 new students in educational science (117 women and 21 men) attending a Finnish Open University (see below). These students had diverse educational backgrounds: 23 per cent of the participants had university degrees (a bachelor's, master's, or doctoral degree); 36 per cent had a polytechnic degree; 16 percent had completed vocational upper-secondary education; and 25 per cent had a general upper-secondary education before undertaking educational science studies. Thus, not all students were beginners in the sense that they were unfamiliar with university study. However, all students were beginning students in educational science. Their ages varied from 20 to 64, the mean age being 35.

Finnish Open Universities, which provided the context for Study III, are part of the Finnish university system. Almost all Finnish universities provide Open University education in co-operation with their departments and faculties. Open Universities are thus not autonomous institutions. They offer university-level study for everyone, regardless of age or educational background. Course content and criteria for assessing learning outcomes are equivalent to the teaching at the affiliated university. However, Finnish Open Universities do not award academic degrees. In the year 2011, over 43,000 individuals attended Open University courses in Finland (Statistics Finland 2012). Many Open University students are adults who want to develop their potential in general or expand their knowledge in a specific subject. Another group of Open University students consists of graduates of upper-secondary school who seek to improve their chances of being admitted to the university as a regular student, either by taking courses in their subject of interest or by taking basic level courses that provide an alternative route to university admission. Educational sciences, for instance, offer such an alternative route.

The students' writing was assessed using the CLA performance task. The task was translated into Finnish and adapted to a Finnish context. The adaptation and translation process followed the same detailed protocols as Studies II and IV (see Appendix A) to ensure consistency among Studies II, III and IV. The CLA performance task is designed to measure general academic skills, such as critical and analytical thinking, problem-solving and written communication (Klein et al., 2007; Shavelson, 2010). The CLA performance task requires students to apply critical thinking and written communication skills, and it endeavours to measure "general skill-based competencies, rather than academic aptitude, general intelligence, or subject- and content-specific skills" (Arum & Roksa, 2011a, 147; see also Shavelson, 2010). The questions and other materials in the CLA performance task were not specialised, thus ensuring that the work could be completed by students from a variety of disciplines. The task simulated several aspects of academic writing and included instructions, four open-ended questions and reading material about a fictitious, but realistic societal problem. In order to respond, the students needed to (a) organise, synthesise, assess and analyse information (which might be reliable/unreliable or relevant/irrelevant to the task); (b) identify judgemental errors, such as 'correlation proves causation' from several sources (email correspondence, memoranda, research abstracts, research reports, newspaper articles, statistics, etc.); (c) make a reasoned explanation for a problem and propose a solution; and (d) write arguments and counter-arguments for and against a particular solution using information from the reading material (Shavelson, 2010). Students had 90 minutes to complete the task. The instructions and each sub-question advised them to cite sources as well as justify their positions.

The data for Study III were analysed using qualitative content analysis. The analytical process was divided into five phases (see Figure 6). In the first phase, the first and second authors read the students' texts through several times independently to obtain a sense of the whole and make notes on relevant information. Thereafter, the definition of coding features was negotiated jointly. In the second phase, the first author coded the data using the ATLAS.ti programme. This phase was theory-driven, meaning that the features of coding were based on prior studies. In analysing the argumentation, the coding concentrated on the following aspects: (1) the conclusion and (2) the reasons and claim(s) that students used as evidence to support the conclusion. In a solid argument, claims warrant "conclusion by providing good reasons for them" (Siegel & Biro, 1997, 278). In order to identify problems in paraphrasing and citing references and fallacious arguments, we used the referencing problems identified by Walker (2010) and the argumentation schemes defined by Walton (1995). Furthermore, we utilised the findings of thematic analysis from Study II. In the third phase the codes were grouped into categories and sub-categories. The

purpose of grouping the data was to classify similar or related coded phenomena in the same category (see Elo & Kyngäs 2007; Mayring, 2000). Thereafter, categories were defined and reviewed, and a final description and interpretation were generated. The abstraction process and the final categories are described in Table 5. Altogether three categories of problems in argumentation and five categories of referencing problems were identified. In the last phase, the occurrence of the categories throughout the data set was examined and a cross-tabulation of the categories generated. The students' texts were then divided into three groups.

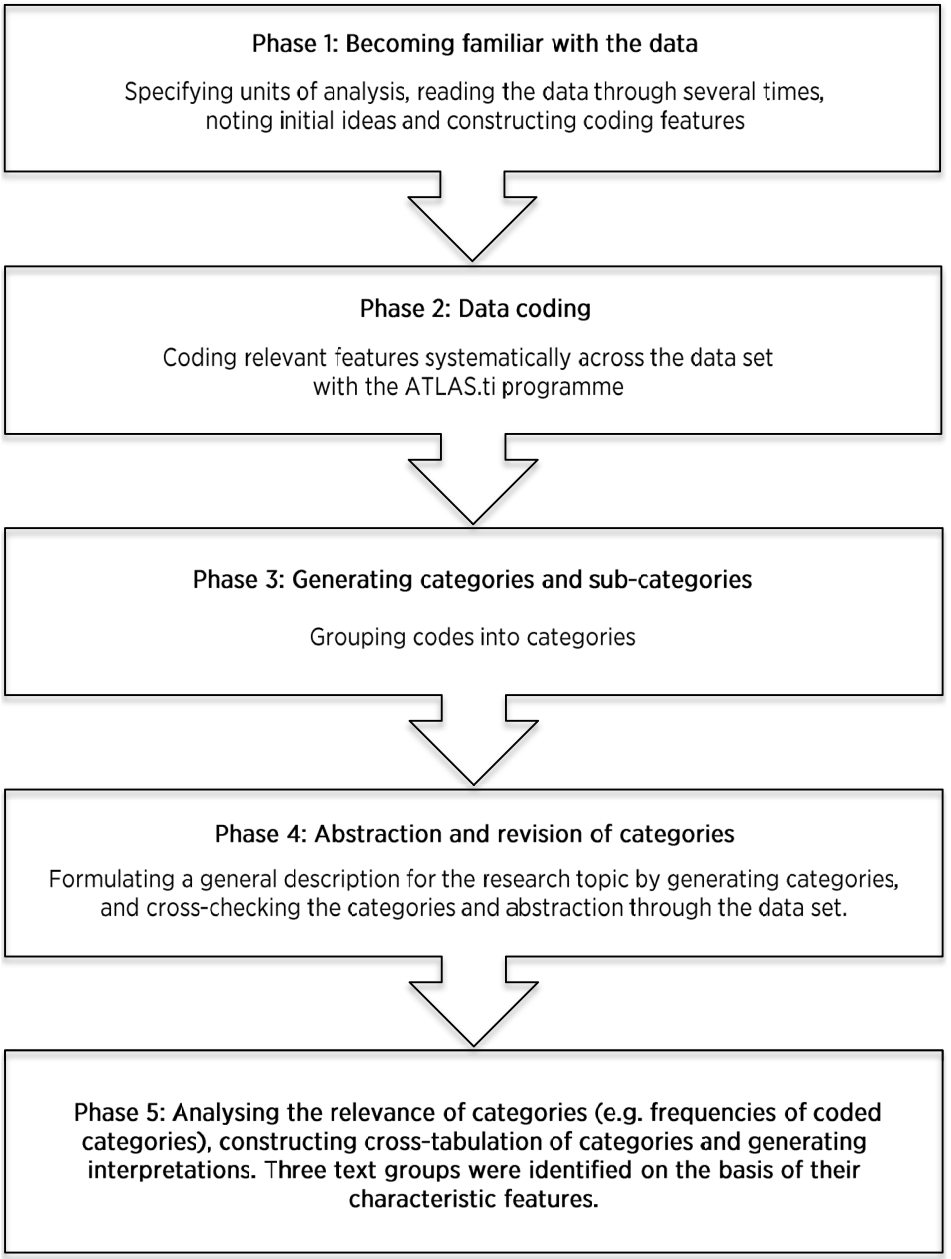


Figure 6. A visualisation of the analysis process in Study III.

Table 5. Categories and sub-categories of problems in argumentation and writing.

Code	Sub-category	Category
Problems in argumentation	Claim is missing Rationales/ reasons are missing	Unclear Argument
	Student provided a single reason Student provided unconnected list of facts as rationales (reasons are not explained or elaborated) Student provided unconnected list of sources as rationales	Presenting isolated fact(s) as rationale(s)
	Hasty generalisation Irrelevant conclusion Correlation proves causation Appealing to probability Appealing to authority Appealing to tradition	Incorrect argument/ fallacy (claim is inconsistent with reason)
Paraphrasing problems	The text is the same as the original	Copy without source
		Copy with source
	Sources are cited, but the text is almost the same as the original	Patchwork paraphrasing
		Poor paraphrasing
		Conclusions disguised as one's own

Study IV: The equivalence of test results on two performance-based assessments

Hyytinen, H., Nissinen, K., Ursin, J., Toom, A., & Lindblom-Ylänne, S. (2015). Problematising the equivalence of the test results of performance-based critical thinking tests for undergraduate students. *Studies in Educational Evaluation* 44: 1–8. doi:10.1016/j.stueduc.2014.11.001

The main goals of Study IV were to explore the critical role of the type of assessment instrument and enhance understanding of the importance of the analytical method. Study IV compared the test results of two different performance-based assessments of critical thinking used in OECD's AHELO (Assessment of Higher Education Learning Outcomes) Feasibility Study: a

constructed-response task from the Collegiate Learning Assessment (CLA)[7] developed by the Council for Aid to Education (CAE) and a multiple-choice questionnaire (MCQ) from the Australian Council for Educational Research (ACER).[8] These tests ostensibly measure the same critical thinking skills, such as analysing, interpreting and evaluating information and problem solving. The CLA task, intended specifically to evaluate reasoning, problem solving and written communication (Shavelson, 2010), presents a realistic problem and includes directions, open-ended questions and reading materials (such as letters, memos, summaries of research reports, articles, graphs, maps, interview notes). Some of the information is relevant, some is not. Students are expected to organise, analyse, synthesise and evaluate these multiple sources of information to arrive at a solution or an explanation of a problem (Klein et al., 2007; Shavelson, 2010). The MCQ also requires students to analyse, evaluate, interpret information and solve problems representing real-world perspectives. However, ACER's MCQ "does not attempt to assess real-world performance directly" (ACER, 2001, 1). MCQ includes several questions followed by four response options, and students arrive at a decision by selecting a correct answer from these alternatives.

The material for this study consisted of AHELO data collected in Finland in the spring of 2012. A total of 330 higher education students at the end of their bachelor's degree studies participated. Half the students were university students and the rest came from polytechnics. The Finnish higher education system consists of both, universities and polytechnics. Universities focus on scholarly research on which the education they provide is based. Polytechnics emphasise a more practical approach and have close links to the labour market. Their mission is to train professionals for working life as well as to participate in new product design and development. Finnish polytechnics and universities enjoy extensive autonomy. It follows that higher education institutions are free to decide on student selection and to design the contents of their degree programmes (Ministry of Education and Culture, 2015a). Higher education is publicly funded in Finland and free of charge for students. The general target time for completing a bachelor's degree is three years at the universities and from three and half to four years at the polytechnics. Student selection for higher education is based on previous academic achievement (i.e. success on matriculation examinations) and entrance examinations. Students come to higher education through various routes. Some continue their tertiary studies immediately after upper-secondary school. Others

[7] An example of a CLA task is available in the AHELO report, volume 1 (see Tremblay et al., 2012: 220–236). See also Figure 2. The CLA performance task is proprietary and consequently cannot be described in further detail here.

[8] The specific MCQ used in this study is proprietary and consequently cannot be described in further detail here. An example of the MCQ sample items is available on the website www.acer.edu.au/documents/CRT-OnlineSampleQuestions.pdf.

apply for higher education from working life. The Finnish educational system has no dead-ends, meaning that it is possible for students to progress from one level of education to the next or return to education later in life (Ministry of Education and Culture, 2015b).

Although the MCQ and the CLA focus on many of the same critical thinking skills, they entail different uses of these skills. The MCQ asks students to recognise a correct answer from a set of response options using critical thinking, whereas the CLA, with its open-ended tasks, asks students to synthesise information and produce an answer in their own words. In addition, each item on the MCQ is devised to measure a single dimension of critical thinking, whereas the CLA is based on the idea that critical thinking cannot be divided and measured in single dimensions (see further in Shavelson, 2010). However, according to the AHELO report, both assessments measure core cognitive skills (Tremblay et al., 2012). Study IV explored how the test results of the CLA and the MCQ align. It focused on the differences in the test scores both at the group level and at the individual level. A strong variance in the test results would call into question whether or not these assessments measure the same underlying dimensions of critical thinking (Bowman & Seifert, 2011; Rodriguez, 2003). Contradictory test results would also have profound implications, as this would indicate that the type of measure has a great influence on a student's outcome (Bowman, 2010).

The computer delivered and invigilated a test situation lasting a maximum of 150 minutes in which the students first had 90 minutes for the CLA and then 30 minutes for the MCQ. At the end of the session the students filled in a short background questionnaire. Students could not proceed to the background questionnaire in less than 120 minutes unless they had responded to all questions, first on the CLA and then on the MCQ. The students' CLA responses were evaluated for Analytical Reasoning and Evaluation (ARE), Problem Solving (PS) and Writing Effectiveness (WE) based on detailed criteria by two trained scorers using a scale of 1 to 6. An example of the scoring criteria is included in Appendix B. On the multi-choice questionnaire the administered items were randomly divided into four subtests with 23 or 24 items each. The translation, adaptation and validation process of the AHELO test instruments is explained in more detail in the AHELO report (Tremblay et al., 2012; see also Appendix A).

Study IV was based on the materials of the AHELO study including the raw data as well as the test scores. The data were analysed using a mixed-method strategy, integrating both quantitative and qualitative analyses (Johnson & Onwuegbuzie, 2004; Onwuegbuzie, Johnson, & Collins, 2009). The analysis of Study IV consisted of three phases (see Figure 7). In the first phase, *quantitative statistical analyses* focused on the relationship between the CLA and MCQ scores at the *group level*. Pearson correlation coefficients were employed in examining the association between the scores. Differences of score means were tested by the

paired t-test. In this phase students were classified into performance groups on the basis of the test scores. The first phase thus informed the second by providing a framework from which to select cases for further analysis.

The second phase of the analysis explored the differences between the MCQ and CLA scores at the *individual level*, meaning that a unit of analysis was an individual student. By analysing the entire data set, we endeavoured to determine whether individual students' test performance on the MCQ and the CLA differed (cf. Lindblom-Ylänne, Parpala, & Postareff, 2013). In this phase students were classified into performance groups on the basis of their test scores. The individual level analysis showed that some students had completely opposite test results. In the next phase we selected these opposite test responses (n= 32) for qualitative analysis. The second phase thus informed the third phase by providing a framework from which to select cases for further analysis. The students who had completely dissonant test scores could be divided into two groups. This phase focused on analysing the raw data. Before proceeding to the qualitative analyses of the CLA responses, we examined each student's MCQ answers carefully and identified the range of correct, wrong and unanswered items.

The qualitative analysis of CLA responses was divided into several phases. In the first phase students' answers were read through several times. The second phase was data coding. This phase was theory-driven, meaning that the features of coding were based on prior studies. In order to examine and identify argumentation in a more detailed way than was done in the original scoring, we used the argumentation schemes identified by Walton (1995). Furthermore, we utilised our findings of thematic analysis (Study II) and content analysis (Study III). The coding focused on the following qualities of critical thinking: (1) identifying, interpreting and synthesising information from multiple sources to reach a conclusion; (2) evaluating the acceptability of information; and (3) producing explanations and arguments. These various qualities were coded systematically within each test response. Then the similar or related coded phenomena were grouped together. Thereafter, a short description of each student's CLA response was written. After that, the differences between and within dissonant groups were identified. In the last phase a final description and interpretation were generated. All the authors collaborated on the analyses. In the analysis we utilised investigator triangulation (Denzin, 1970, 2012; Creswell & Miller, 2000) to confirm the reliability of the findings.

The aims, measures, and data analyses of Studies I–IV are summarised in Table 6.

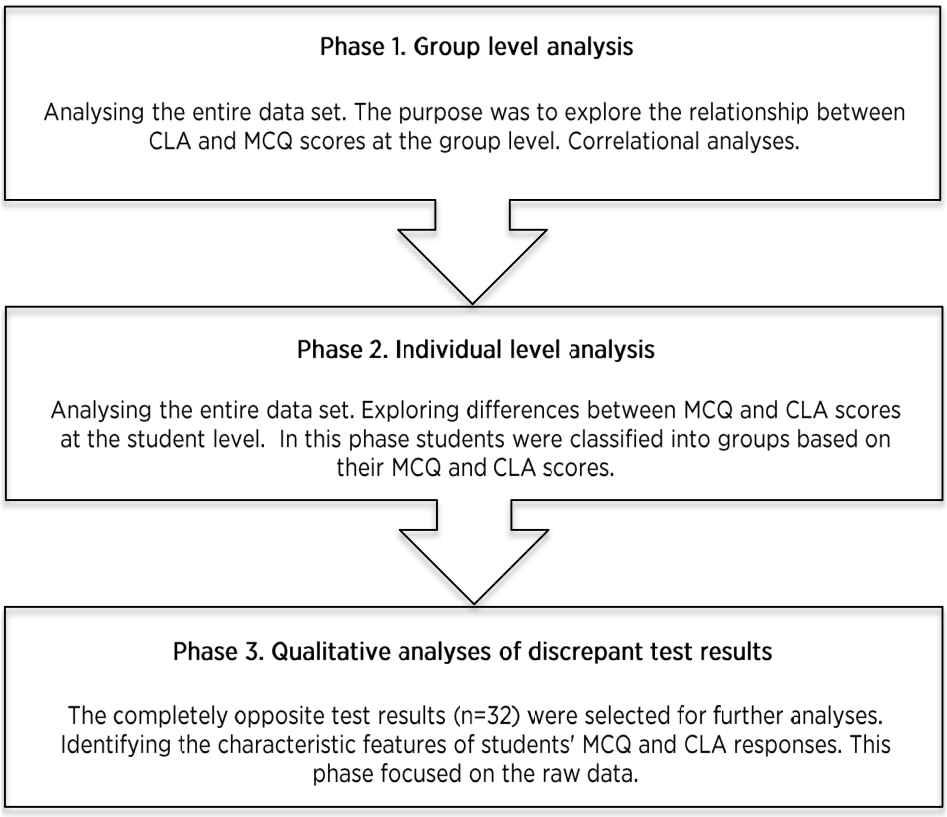


Figure 7. Three main phases of analysis: combining the individual level and group level analysis.

Table 6. Summary of the methodological choices of the four studies.

	Research question	Aims	Data	Analyses method	Unit of analysis
STUDY I	A1-A2	• To identify the conceptual differences between the research tradition of educational psychology and philosophy • To analyse the normative elements of personal epistemology • To explain the conceptual inconsistencies of personal epistemology from a philosophical point of view • To suggest a remedy	Literature review	Philosophical analysis, systematic analysis	Theoretical variants of PE
STUDY II	A1-A2, B1-B3, C1	• To identify and describe qualitative differences in students' critical thinking skills • To examine how students' critical thinking skills and their conceptions of knowledge intertwine and jointly contribute (n=10)	Multi-method approach: CLA, think-aloud protocol, interview (self-report)	Thematic analysis with an abductive approach	Both a group of students and an individual student
STUDY III	B1	• To identify challenges in students' written argumentation and their use of sources at the textual level • To identify and describe qualitative differences in students' argumentation, paraphrasing and use of sources (n=138)	CLA	Content analysis with combined data-driven and theory-driven approaches	A group of students

STUDY IV	C1, B1	• To enhance understanding of the importance of the analytical method • To explore the critical role of the type of assessment instrument by comparing the measures of two performance-based assessments • To identify and describe qualitative differences in students' critical thinking skills (n=330)	Multi-method approach: CLA, MCQ, background question-naire	Mixed-method approach Group level analyses: Pearson correlation, t-test Qualitative analyses: Content analysis with data-driven and theory-driven approaches	Both a group of students and an individual student

5 RESULTS AND DISCUSSION

In this chapter, the main findings of the doctoral thesis are presented and discussed. The first section introduces the theoretical findings. The second section presents information relating to the empirical findings, and the last section presents the methodological results.

5.1 Theoretical findings

5.1.1 The normative elements of PE

Study I analysed to what extent empirical research on PE and philosophical analysis of the conditions of knowledge relate to each other. Contrary to the claims of many advocates of PE (Hofer & Pintrich, 1997; Kitchener, 2011; Southerland et al., 2001), the results of the theoretical analyses suggested that PE cannot be understood merely in descriptive terms, but clearly involves normative elements.

Study I emphasised the importance of the distinction between normativity (how things ought to be) and descriptivity (how things are). In other words, on the basis of a philosophical analysis alone, one cannot say how things are in the real world (i.e. how human beings see and understand the nature of knowledge or construct their systems of belief), while on the basis of an empirical analysis one cannot say how things ought to be (i.e. what kind of personal epistemology is sophisticated or is worth pursuing in the first place). Thus, if we are interested in understanding the most adequate conception of knowledge, we need to consult philosophical research, and when we want to know what people actually think about knowledge and knowing, we should consult empirical research. Therefore, on the basis of empirical research alone, we cannot derive conclusions about the superiority of one conception of knowledge over another.

According to Study I, the following three examples demonstrate that PE's theoretical models include normative elements:

1. PE's theoretical models entail claims of superiority of some conceptions of knowledge over others. These models are clearly hierarchical, because they hold that some epistemological positions are more 'sophisticated' or 'higher' than others (Kaartinen-Koutaniemi & Lindblom-Ylänne, 2012, 2; King & Kitchener, 1994, 2002, 2004; Kuhn, 1999, 2005; Perry, 1970; West, 2004). This is to say that the PE theories maintain that there are differences in the adequacy and justifiability of students' personal conceptions of knowledge and that PE's hierarchical models represent this order of superiority.

2. Researchers have provided numerous educational recommendations based on empirical findings (i.e. Brownlee, 2004; Brownlee et al., 2009; Hofer, 2001, 2006b; King & Kitchener, 2002; Lahtinen & Pehkonen, 2013; Perry, 1970). As an example, Brownlee et al. (2009, 612) write: "[e]ssentially, we are advocating for pedagogy in higher education that is informed by personal epistemology rather than the implementation of particular teaching strategies per se. This new pedagogy requires a culture change in learning; one that engages us all in a more sophisticated way of knowing and learning in higher education".

3. The PE theories assume that a change in students' conceptions of knowledge can be interpreted as development from lower (less adequate) to higher (more adequate) conceptions (i.e. Kuhn, 2005; Hofer & Pintrich, 1997). As Pintrich (2002, 400) puts it: "there is fairly high agreement on the nature of developmental change. Again, at some level, all the models represented in this volume are in line with the proposition that an individual's thinking about epistemological issues not only changes over time, but that it develops towards a more sophisticated perspective or stance toward knowledge and knowing". However, there is always the possibility of some kind of adaptation instead of genuine development, at least from a theoretical point of view. It is possible, for example, that students have learned from their teachers what would be the most adequate way to think about knowledge and have adopted this view regardless of its real epistemological benefits.

As noted before, normative elements belong to the field of philosophy. It is thus justifiable to analyse the normative elements of PE in the light of philosophical understanding. It is worth mentioning that, in the language of philosophy, if a researcher does not recognise the normative nature of some parts of the theory—and interprets these as descriptive—she or he commits the naturalist fallacy.[9] To be exact, on the basis of an empirical analysis one cannot say how things ought to be. For example, if the researcher finds relativist features in the empirical data, it does not imply that relativism is an epistemologically justifiable position or that it should be enhanced by teaching.

5.1.2 Different interpretations of the concept of relativism

Study I suggested that there are different interpretations of the concept of relativism among PE scholars, and it appears that the interpretations are not all

[9] Fallacies are "failures of rationality" (Siegel & Biro, 1997, 278). The naturalistic fallacy by which one derives '*ought*' from *'is'* refers to a situation in which a person makes claims about what ought to be on the basis of statements about what is or vice versa.

necessarily relativistic in its ultimate form. Some researchers state that relativism is a view that understands knowledge as "contingent and contextual" (Hofer, 2000, 379; see also Briell et al., 2013, 482), whereas others interpret relativism as referring to the idea that "everyone has a right to his or her opinion, all opinions are equally right"[10] (Kuhn & Weinstock, 2002, 123; see also Hofer, 2005). These interpretations differ so fundamentally from each other that some of them do not appear to be epistemologically relativistic at all, whereas others cannot be interpreted in any other way.

It is therefore important to distinguish three positions, all labelled relativistic in the context of PE: 1) knowledge is context-dependent in a weak sense, 2) knowledge is context-dependent in a strong sense, 3) all beliefs are equally right. The first position states that particular human beliefs, statements and theories can be understood only in relation to some context, and they cannot be directly compared, say, with the facts of independent reality. This position does not necessitate relativism, but actually is a better fit with *epistemological fallibilism*. In the second case beliefs, statements and theories can be evaluated only in relation to some context, but outside this context, nothing can be said of their credibility and justifiability. This position is a version of (modest) relativism. Finally, the position that all opinions are equally right is, of course, a version of relativism in its ultimate form.

As noted before, the essential problem of a modest and ultimate form of relativism is that we cannot be incorrect, since relativism admits no criteria for evaluating our beliefs and theories. To put it simply, in the context of relativism there is no need to evaluate ideas or search for alternatives, because all ideas are equally trustworthy and justifiable. This is worth noting also in relation to Kember's (2001, 217) aforementioned connection of relativism with the possibility of critical thinking. Along the lines of the previous argument, within the framework of epistemological relativism, critical thinking is altogether pointless. If all beliefs and opinions are equally right, then why should one think critically? Study I argued that the learning aim of critical thinking is conceptually connected with the epistemological ideal of rationality (cf. Scheffler, 1973). Rationality as well as its development through teaching requires having some criteria for evaluating beliefs, conceptions and theories. Furthermore, relativism undermines the possibility of teaching at the general level as well. If all knowledge is equally valid, then teachers have no criteria with which to assess students' understanding; students are free to hold whatever beliefs or positions they want. Nor is it justified, within the relativist framework, to assume that a teacher has some sort of epistemic authority (Kotzee, 2010). We actually cannot justifiably reject any competing presumptions as false.

[10] Emphasis added.

5.1.3 Introducing the concept of epistemological fallibilism as a solution to theoretical inconsistencies

Study I argued that there are several other conceptual inconsistencies in the PE literature. For instance, PE does not sufficiently distinguish between the epistemological notions of objectivity, certainty and truth. Study I suggested that these conceptual inconsistencies are connected with a deeper problem: the theoretical framework of PE does not identify the concept of epistemological *fallibilism*. Fallibilism implies that human knowledge is uncertain (Reed, 2002; Niiniluoto, 1999; Holma, 2012). However, it does not imply that all beliefs, conceptions or theories are equally right. In contrast to relativism, it thus presumes the possibility of improving our current conceptions, theories or beliefs, seeking criteria for evaluating, comparing and justifying these beliefs or theories (Holma, 2012; Holma & Hyytinen, in press-a, in press-b; Peirce, 1934a. 1934b). In philosophical discussion epistemological fallibilism is often committed to the idea of realism, which states that reality exists independent of human beings.

In previous studies of PE, epistemological notions of objectivity and certainty are often tied and connected with a *naïve version* of realism and *absolutism*. Naïve realism is an epistemological position that assumes that "our knowledge and symbol systems [i.e. theories] directly reflect the structure of reality" (Holma, 2004, 421; Putnam, 1981). The PE literature seems to understand realism as naïve realism (see e.g. Kuhn 2005; Kuhn & Weinstock, 2002), and furthermore appears to connect with naïve realism the assumption of there being a possibility of the certainty of human knowledge (i.e. absolutism). Hilary Putman describes this kind of realism as *metaphysical realism* (Putnam 1978, 1981). As King and Kitchener (2004, 7) put it, knowledge is "obtained with certainty by direct observation".[11] Also within this framework, critical thinking turns out to be pointless, because human knowledge directly copies reality.

However, the possibility of objectivity implies neither the possibility of certainty nor the idea that our beliefs are direct copies of external reality. Contrary to the absolute nature of certainty, objectivity is best understood as a matter of degree. It follows that individuals' descriptions of reality can be more or less objective, depending on the adequate use of evidence and reasons. At one end of the continuum are the merely subjective beliefs, while at the other end are the beliefs based on all the relevant evidence and reasons (Holma, 2011, 536). Theories of science, for example, strive to approach the highest level of objectivity. But even a theory that has reached the highest level of objectivity could never be assumed to be certain.

[11] King and Kitchener (2004) do not call the lowest level of reflective thinking realism. However, in their model they maintain that, at the most limited level of thinking, knowledge is certain and is obtained from direct observation (King & Kitchener, 2004, 7). This position fits naïve realism.

The theoretical analyses in Study I pointed out that without fallibilism, several PE theorists have considered the only alternative to relativism as being some kind of naïve version of realism (i.e. human beliefs are direct copies of reality) and epistemological absolutism (i.e. that human knowledge can achieve certainty) (e.g. Kuhn, 2005, 30; Kuhn & Weinstock, 2002, 124; Kuhn et al., 2000; see also King & Kitchener, 2004). One can agree that in comparison with these views of realism and absolutism, relativism may indeed seem a sophisticated epistemological position. However, although many contemporary philosophers defend realism (Niiniluoto, 1999), none of them accepts the view that human beliefs are direct copies of reality, nor that human beliefs could be certain. Realism means only that reality exists independent of human beings. Different variants of realism may disagree both in relation to what part of reality exists independent of human beings and to what extent the reliable knowledge can be acquired about reality. In the philosophical literature realism implies neither the certainty of human knowledge nor the idea that human beliefs are direct copies of reality.

Study I argued that naïve realism and epistemological absolutism are not the only alternatives to relativism. Study I also introduced the concept of fallibilism as an epistemologically justifiable alternative to solve the above-mentioned conceptual inconsistencies, because it preserves the benefits that PE identifies with relativism without slipping into relativism (i.e. which claims that it is not possible to evaluate, compare and improve our conceptions) or into conceptual problems in terms of the notion of absolutism (i.e. which claims that some sources of certainty or fixed starting points exist on which to base our search for understanding). In other words, fallibilism differs from relativism by not taking all beliefs, conceptions or theories as being either equally right or right merely in relation to some context, but allows us to seek the criteria for evaluating, comparing and justifying our beliefs, theories and conceptions. Fallibilism can further be considered as a responsible epistemic position, because it admits to the uncertainty knowledge, but does not end up with the relativist conclusion that all beliefs are equal. Fallibilism, by accepting the uncertainty of all human knowledge, can also be seen as a psychologically mature view of knowledge. As PE scholars King and Kitchener (2004, 9) have themselves written, clearly in a fallibilist spirit:

> Reflective thinkers consistently and comfortably use evidence and reason in support of their judgements. They argue that knowledge claims must be understood in relation to the context in which they were generated, but that they can be evaluated for their coherence and consistency with available information. Because new data or new perspectives may emerge as knowledge is constructed and reconstructed, individuals using assumptions of reflective thinking remain open to reevaluating their conclusion and knowledge claims.

The same conceptual inconsistencies appeared during the analytical process in Study II.[12] I realised that Study II's empirical data could not be explicated and interpreted in depth using previous PE models; in other words, the commonly used dimensions ranging from absolutist to relativism or evaluativist did not capture the variations in the students' conceptions of knowledge. The philosophical assumptions of knowledge and knowing opened new insights into understanding the results of our analysis, especially the variations in students' epistemological notions. Among other things, the results demonstrated that all students who participated in Study II shared a fallibilist view of knowledge. The concept of fallibilism is missing from the theoretical framework of PE. The empirical findings are discussed in more detail in the next section.

5.2 Empirical findings

5.2.1 Critical thinking as a tool for determining the relevance of knowledge

The empirical results of Study II suggested that students' conceptions of knowledge were woven into their critical thinking: students used critical thinking as a tool for enhancing understanding or determining truth or falsehood. Based on this difference, students were classified in one of two profiles, either (1) thorough processing or (2) superficial processing. These profiles captured the diversity of the students' skills and dispositions to think critically. In addition, these two profiles characterised the variation in how students determined what is needed to evaluate knowledge and how they justified, acquired and used the knowledge in the problem-solving situation. The same profile groups were also identified in Studies III and IV. These two profiles were not distinguishable by gender or educational background.

Superficial processing students

The results showed that the students who palmed off justification for knowing on authoritative figures or scientific proof showed superficial processing: these students *did not make a serious effort to analyse, evaluate, interpret or synthesise the information in the materials* (Studies II & IV). This profile consisted of students who used critical thinking as an instrument for determining truth or falsehood. Their goal in the problem-solving situation was to find the right answer to the problem.

[12] The findings of Study II were the starting point for the theoretical analysis of PE (Study I).

In Study II the superficial processing students thought that there was one definite answer to the problem. These students acquired knowledge through testimonies. They trusted in scientific proof as well as testimonies from authoritative specialists or experts. Students in this profile believed that *knowledge is trustworthy only if it was produced through a reliable process*, for example, by using empirical methods or consulting suitable experts. They considered scientific and verified knowledge to be the most reliable (i.e. to hold the highest epistemic status), because that kind of knowledge is based on evidence. They further believed that scientific knowledge is objective and therefore most often unbiased. However, in the problem-solving situation, they *did not evaluate or interpret* the empirical knowledge at all, nor did they consider presuppositions or the contextual nature of knowledge. For that reason, they did not recognise that some material for the task, which included empirical or verified knowledge, was biased. However, in the interview these students believed that empirical knowledge could be fallible too. They expressed the idea that some theories based on empirical knowledge might be false and that it is possible to improve current conceptions and theories. In contrast to earlier findings (e.g. Kaartinen-Koutaniemi & Lindblom-Ylänne, 2012; King & Kitchener, 2004; Kuhn, 2005), the findings of Study II pointed out that the students who appealed to authorities did not consider knowledge as absolutely certain or irrefutable. Nor did these students share the view that beliefs accurately represent or correspond to reality. The students thus did not share a sense of absolutism, namely the idea that knowledge is certain. From the epistemological perspective, these students took the *fallibilist position*.

In Study II some of the superficial processing students had serious difficulties evaluating the quality of information, because they did not consider themselves capable of evaluating knowledge. They believed that if a person who is said to be an authority on something makes an argument about that something, then the argument should be trustworthy and therefore usable. It follows that the right answers can be found by consulting the right expert. These students repeated the arguments and conclusions as these were presented in the documents. The students mentioned a few external criteria for evaluating knowledge (such as an authority, expert opinion, publication, openness, journal citations). However, in practice these students did not know how to use these criteria independently. The following example illustrates these uncertain knowers:

> *I don't know what the right approach is in order to grasp those overall concepts from that huge mass of teensy-weensy details. Because a candidate [for a bachelor's thesis] has to read a huge number of articles to find the ones that are, like, related to one's own topic and all. So it's really hard when you're, like, reading an article to judge why this one might be better than that one. So. But I got a tip from my supervisor that I should pay attention to the*

reliability of the journal. To be honest, it's the research articles, the ones we have at the university, that are actually the only ones we're told we can cite. And then it's like… they're easy to evaluate based on which publications are more credible. And on the Web on [sic] Science, they have this one like… what is it, like an indicator that they have, just based on the number of citations and other factors, of the accuracy of the research data… It's hard! In a way, to make that distinction between what's true and what isn't. At least I don't have the know-how to say what's true and what isn't.

(Study II, Student 5, interview)

Further analysis demonstrated that the students in the superficial profile group *provided isolated reasons and reproduced ideas* (Studies II, III & IV). For example, their written responses were typically composed of reproduced or slightly modified portions of the source texts, meaning that the students had not explained the content of the materials in their own words. Superficial processing students identified only a few of the claims presented in the documents and disregarded many relevant aspects of those claims. They did not identify alternative solutions or conclusions or approaches to the problem. Nor did they provide any reasons or explanations for their own conclusions. As a result, they had problems reaching a conclusion. Furthermore, the analysis indicated several *weaknesses in producing arguments*. Common to all answers in this profile group was that the students provided vague arguments: they made claims or gave conclusions, but did not supply valid reasons or explanations to support their claims, thereby leaving their arguments weak. The following extract from Study IV describes a typical written response in the superficial processing profile (reproduced or slightly modified portions from the source texts are underlined):

Question 1a: What specific information, evidence, facts and/or theories discussed in the Document Library support the inbreeding explanation?

Document 6, transcription of the radio interview with Dr. Leusid

QUESTION 1b: What are the arguments against the inbreeding explanation?

In Document 2 Charles Stone writes that the catfish were found in the same pond that Bonaventure Mills poisoned to death 10 years ago. He thus blames the mills for what has happened.

In Document 6 Thomas Leusid states that duplications of limbs in frogs and catfish are caused by a parasite.

(Study IV, Student 28, written answers)

Some of these students had no disposition to use critical thinking skills in the test situation or in everyday life; as one student put it in the interview:

> *In everyday life it's rare that, if you're discussing something, it's rare that anything like this happens. Or I never, really rarely discuss anything argumentatively in any way. In real life I simply don't like it, discussing issues.*

(Study II, Student 9, interview)

Thorough processing students

The students who *deeply analysed* the content of the documents created their own understanding of the problem-solving situation (Studies II, III & IV). For them, critical-thinking skills were tools for deepening and enhancing understanding. These students believed that theories and beliefs could be understood in relation to some context, as the following extract from Study II shows:

> *- - yeah, I don't believe [the chair of the stakeholder group] is completely off the mark either. [The reliability of] knowledge is just always context-specific.*

(Study II, Student 8, interview)

Study II pointed out that these students considered it possible to improve current theories and beliefs. In the problem-solving situation, these students were open to new evidence that could disprove a previously-held position or belief. For them, scientific knowledge is probably reliable. They believed that both objective and subjective knowledge could attain the highest epistemic status, meaning that, for example, their own subjective perceptions or experiences could be a reliable source of knowledge. These students thought that the credibility of knowledge could be affected by vested interests or bias, for example. Although these students emphasised their own role in constructing their understanding, they did not believe that all knowledge is generated by human minds. In the problem-solving situation these students justified and enhanced their own understanding with reasoning. Besides reasoning, they considered intuition, testimonies, perceptions and experiences as possible sources of knowledge. These students saw that all human knowledge (i.e. theories, beliefs and conceptions) is fallible. However, the students did not argue that all knowledge is constructed by the knower nor did they believe that all interpretations, theories and beliefs are equally right. Therefore, they avoided slipping into relativism. From the epistemological perspective, these students also took the *fallibilist position*.

Studies III and IV indicated that the thorough processing students *evaluated the quality of the information and considered its premise*s, as well as the implications of different conclusions. They *weighed different options, connected*

related ideas and gave reasoned explanations. The following extract from Study IV describes a typical CLA written response in the thorough processing profile:

> Question 1a: What specific information, evidence, facts and/or theories discussed in the Document Library support the inbreeding explanation?
>
> Thomas Leusid says in his interview that many catfish are quite young—at most three years old, which strongly indicates that the catfish have come to this area only a short time ago, so it could not be about Db09-caused toxin concentration—especially since the Db09 toxin has not, according to Sandy Evans' study as referred to by Dr. Munt, caused any mutations. In addition, Leusid's research team has found some multiplied structures in the catfish populations of Miracle Lake, although the feature of three eyes had not been observed earlier. Leusid's theory—that only a number of fertilised eggs have come to Bush Creek—is supported in particular by the fact that Bush Creek and Miracle Lake greatly restrict the mobility of fish, and adult catfish live at the bottom of waterways [and] are not likely to get across the rapids.
>
> QUESTION 1b: What are the arguments against the inbreeding explanation?
>
> Researcher Thomas Leusid had not seen the fish physically, and the phenomenon of multiple eyes had not been previously observed in multiplied structures. Moreover, Thomas Leusid cannot exclude the possibility of parasitic organisms or some other as yet unknown factor. For confirmation of the inbreeding explanation, Leusid's research team should raise catfish from Millpond in laboratory conditions. The Milltown Clarion tells about Malcolm Reis's research, according to which people only 15 kilometers away had discovered parasite-inflicted mutations in fish as well as in frogs studied by Reis. The investigation of the environmental damage caused by the Db09 toxins of Bonaventure Mills can also be regarded as biased and partial, because the company provides financial backing for the study. In addition, Sandy Evans' study of mutagens was conducted on bacteria only, so this study cannot necessarily be directly compared with mutation effects in fish.
>
> (Study IV, student 5, written answer)

Study II showed that some of the thorough processing students had problems with decision-making. These students further acquired knowledge in a rather uncritical way. In the problem-solving situation they rarely evaluated the reliability of the documents. Indeed, these students did not have clear criteria for evaluating the reliability or the relevance of information. They simply trusted their intuition:

> *This just seems scientific somehow.*
>
> (Study II, student 6, think aloud)

I don't know how I should formulate this, but I'll start by saying that when I read, for instance… or when I'm taking classes, I don't spend a whole lot of time wondering if some piece of information is reliable or not.

(Study II, student 6, interview)

These students started to analyse and interpret thoroughly all information presented in the documents. They identified all major facts and ideas. They also considered various decisions or explanations, but could not explain what decision was the best or why. There were too many options available. These students endlessly weighed the different options. Because the students did not reach clear conclusions, they did not present any arguments for accepting a particular conclusion. They showed an inability to adjust their thinking to new evidence or make changes in their actions. They somehow analysed the problem too deeply.

Table 7 provides an overview of the student profiles presented in Studies II, III and IV. The aspects that distinguished the profiles were the differences in (1) the skills and dispositions to think critically, (2) the nature of knowledge that a student considered as relevant (3) the knowledge that a student acquired and used in the problem-solving situation and (4) the way the student processed that knowledge.

Table 7. Conceptions of knowledge in two qualitatively different student profiles of critical thinking.

Student profile	Critical thinking	Conceptions of knowledge	Characteristics of written responses
Superficial processing (Studies II, III & IV)	Critical thinking as a tool for determining truth or falsehood (Study II)	*Definition and content of knowledge*: - Knowledge is truthful only if it is produced by a reliable process, (e.g. using empirical methods). - All human knowledge is fallible. - Does not consider the contextual nature of knowledge. (Study II) *Justification of knowledge:* - Appealing to testimonies and proofs (Studies II, III & IV) *Source of knowledge*: - Obtaining knowledge through testimonies and proofs (Study II)	*Identifying, interpreting, synthesising information:* - Presents isolated facts - Reproduces information, no personal analysing - Disregards major part of information *Evaluating information:* - Does not evaluate the quality of information *Producing explanations and arguments:* - No reasoning - Vague argumentation (Studies II, III & IV)
Thorough processing (Studies I, III & IV)	Critical thinking as a tool for enhancing understanding (Study II)	*Definition and content of knowledge*: - Both objective and subjective knowledge can hold the highest epistemic status (i.e. can be truthful). -All human knowledge is fallible and contextual. (Study II) *Justification of knowledge:* - Reasoning (Studies II, III & IV) *Source of knowledge*: - Reasoning, experience, perception, testimony and intuition - All knowledge is not generated by human minds (Study II)	*Identifying, interpreting, synthesising information*: - Demonstrates thorough understanding of materials - Interprets and synthesises information - Identifies major ideas presented in the materials - Considers implications *Evaluating information:* - Evaluates quality of information* *Producing explanations and arguments*: - Provides mostly convincing arguments and rationales - Weighs different options - Gives reasoned explanations* (Studies II, III & IV)

*Note: In Study II, there were students who belonged in the thorough processing profile but they did not evaluate the quality of information, nor did they give reasoned explanations. These students endlessly weighed the different options (for more, see Figure 8).

It is important to notice here that Study III focused solely on students' written CLA responses, while Study IV focused on students' tests responses and self-reported questionnaire items. Therefore, we were not able to investigate more deeply the extent to which differences in the students' answers arose from conscious and intended strategies, as was the case in Study II. However, I argue that among these three studies there are shared perspectives. Firstly, in all three studies the students' critical thinking was measured by using the CLA performance task. Secondly, the results of Studies III–IV demonstrated that the students' CLA written answers reveal the level of processing, and thus extended the understanding of two qualitatively different profiles of critical thinking as identified in Study II. In addition, in Study IV all students reported on their background questionnaires the effort they had put into completing the tasks. Thorough processing students stated that they had done their best or close to their best, while students who belonged to the superficial processing profile reported that they had put some effort into completing the tests. These findings are also consistent with the findings of Study II.

5.2.2 Challenges in constructing convincing arguments and paraphrasing

The results of Studies II, III and IV suggested that a significant number of the higher education students had problems constructing arguments and paraphrasing the meaning of original passages in their own words. Statistical analysis of the CLA scores in Study IV showed that, with regard to the CLA sub-scores, the mean of written effectiveness was significantly lower than the means of the other two CLA sub-scores, indicating that students tended to have lower scores in writing than in analytical reasoning and problem solving. Therefore, Study III focused on deepening understanding of why argumentation and paraphrasing are problematic for higher education students.

Argumentation, especially the tradition of written argumentation, has had an essential role in academic discourse (Keinonen & Kärkkäinen, 2010). Even though there are many field-specific conventions in academic writing, it is possible to identify commonalities in texts at the micro level: we can analyse the construction of a solid argument (Walton, 1995), as well as the nature of paraphrasing if we have the original source at our disposal (Walker, 2010). Furthermore, argumentation and paraphrasing are core components in any discipline. By analysing qualitative differences in argumentation and referencing, it is possible to identify when and where higher education teachers need to intervene. Strategies that support writers can then be identified. There are myriad factors related to and influencing academic writing, such as institutional procedures, field-specific conventions and students' prior knowledge, expectations and social background, as well as the amount of time and effort the student spends on his or her studies (Arum & Roksa, 2011a; Lea & Street, 1998). Individual teachers have

little or no influence over many of these factors. Therefore, Study III focused specifically on the aspects of academic writing that every higher education teacher can address in teaching.

The term *argument* is traditionally defined as "an attempt to present evidence for a conclusion by providing claims that support the conclusion" (Groarke 2013, para. 9). In a good argument reasons and claims are consistent with the conclusion (Siegel & Biro, 1997). Study III identified three main problems in argumentation, which we labelled 'unclear argumentation', 'isolated facts' and 'incorrect arguments/fallacies' (see Table 5). In 'unclear argumentation', substantial claims or reasons that support the conclusion were missing, thereby leaving the argument obscure or lacking in logic. 'Isolated facts' refers to a situation in which students provided only a single reason or rationale in cases of parallel or competing reasons, offered disconnected or random facts and connected these with each other without proper justification. 'Fallacy/incorrect argument' refers to errors in reasoning: students drew a conclusion, but the rationales given did not support the conclusion. The fallacies were hasty generalisations, i.e. a student argued from a special case to a general rule based on insufficient evidence; an irrelevant conclusion, i.e. a student used irrelevant information as evidence; a correlation proved causation, i.e. a student argued that a correlation implied causation. Incorrect arguments also included appealing to authority, tradition or probability rather than a solid argument based on the careful analysis of the information provided. In 'appealing to authority', a student argued that something was true or correct because an authority on something had asserted it was so. According to the philosophical literature (Siegel & Biro, 1997; Walton, 1995), appealing to authority is not necessarily a fallacy. This kind of argument may sometimes be judged as fallacy and sometimes not, depending on the situation. The question is whether "the authority is a good one (for the conclusion in question)" (Siegel & Biro, 1997, 286). In this case the authority was nominally competent, but the problem was that the argument appealed to an authority who was biased and thus not relevant to the question. The same phenomenon was found in Studies II and IV. In 'appealing to probability', a student took something for granted because it would probably be true, whereas in 'appealing to tradition' a student argued that something was true or correct because it correlated with tradition.

Study III identified five types of paraphrasing problems, namely 'copy without source', 'copy with source', 'poor paraphrasing', 'patchwork paraphrasing' and 'conclusions disguised as one's own'. 'Copy without source' refers to a situation in which a student copied short portions of a text without a source citation. 'Copy with source' refers to situations in which a student cited a source, but presented the material without quotation marks as if it were paraphrased, when in fact the student had copied the material word-for-word. In 'poor paraphrasing', a student

properly cited the sources, but changed and modified a source text slightly, for example, by changing or adding words or using synonyms. Poor paraphrasing appeared to include minimal or no processing of the materials. In 'patchwork paraphrasing', a student pieced together small parts of poorly paraphrased or directly copied texts from several sources and possibly some formulations of their own. In patchwork paraphrasing some relevant processing of the materials may have occurred. 'Conclusion disguised as one's own' refers to a situation in which a student paraphrased conclusions from the materials without a citation, giving the impression that the conclusions were the result of the student's own analysis.

5.2.3 Problems in critical thinking seem to accumulate

In the empirical findings of this thesis, variations in the student' critical thinking were evident. The empirical results showed that students' critical thinking skills were unevenly developed (Studies II, III & IV). In other words, a student might have the ability to identify and evaluate information, for example, yet at the same time struggle with other abilities, such as arriving at a conclusion, adjudicating conflicting claims or producing arguments. Furthermore, Studies II, III and IV emphasised that the nature of the difficulties and their variance among students were interlinked with the students' ability to utilise and process the available materials. For example, in Study IV, the written responses of the superficial processing students revealed several problems in identifying, interpreting, synthesising and evaluating information, as well as in producing explanations and arguments in their own words. By contrast, the written responses of the thorough processing students demonstrated that the task materials were processed in great depth. The latter group showed that they knew how to define problems, analyse, interpret and evaluate the information on all sides of an issue and use relevant data to produce arguments and explanations. However, it is important to recognise that even the students who processed materials in depth had some trouble producing convincing arguments.

The same issue also emerged in Study III. The specific types of problems in academic writing and the ways in which the problems were related to each other differed among the participants, which allowed us to identify the nature of the problems encountered by the student writers. Three text groups were identified and, on the basis of their characteristic features, labelled 'superficially processed texts', 'patchy texts', or 'thoroughly processed texts'. These three categories characterise how students used and processed the available source materials; they also illustrate how coherent and organised the students' written answers were. In addition, the groups captured the diversity of problems in academic writing, including problems in argumentation and paraphrasing. The group labelled 'superficially processed texts' included answers in which the students did not present evidence of their own understanding or processing of the materials. In the

group labelled 'patchy texts' the answers were uneven in quality. In the group labelled 'thoroughly processed texts', the written answers demonstrated a deep processing of the materials. These answers were also organised in a cohesive way. The problems identified in argumentation and paraphrasing were evident in all groups, but were fewer among students who had thoroughly processed the materials as compared to the two other groups of writers.

Table 8. Frequency of categories in each text group in Study III.

	Text groups					
	Superficially processed texts n= 46 (33%)		**Patchy texts** n= 73 (53%)		**Thoroughly** processed texts n=19 (14%)	
	Number of students' answer*	Frequency of coded category **	Number of students' answer*	Frequency of coded category **	Number of students' answers*	Frequency of coded category **
Problems in argumentation						
Unclear argumentation	31 (67%)	58	50 (68%)	69	4 (21%)	4
Isolated fact	19 (41%)	31	17 (23%)	24	4 (21%)	4
Fallacy /incorrect argument	29 (63%)	70	54 (74%)	89	9 (47%)	12
Problems in paraphrasing						
Copy without source	6 (13%)	8	9 (12%)	9	0	0
Copy with source	26 (57%)	53	30 (41%)	65	3 (16%)	5
Patchwork paraphrasing	4 (9%)	4	2 (3%)	2	1 (5%)	2
Poor paraphrasing	39 (85%)	118	60 (82%)	140	11 (58%)	18
Conclusions disguised as one's own	4 (9%)	4	6 (8%)	9	0	0

* Note: The frequencies refer to the number of answers.
** Note: These frequencies refer to the number of times the particular category was identified in the data.

Table 8 illustrates how many students in each group struggled with each specific problem in argumentation and paraphrasing in Study III. There were substantial differences in how the categories and their frequency varied among the groups. In the groups 'superficially processed texts' and 'patchy texts' the problems tended to repeat within one and the same answer, whereas in the group 'thoroughly processed texts', the problems tended to be merely isolated occurrences.

5.2.4 Flexibility in critical thinking

The results of Study II indicated that students had various skills in their ability to adapt their thinking and their performance flexibility to the demands of the task. There was clear variation in the students' ability to change their actions or ways of critical thinking, and we identified both rigidity and flexibility. Flexibility meant that the students could modify their actions and processes and change their behaviours as needed, whereas rigidity refers to situations in which students could not change their processes or look at things from a new perspective or adjust to new evidence in a problem-solving situation. Students who were able to make changes in their actions showed open-mindedness and an inquiring attitude. The students who acted in a flexible manner in the problem-solving situation were also able to reflect on their thinking process and use their self-evaluation skills.

In the following extract, one student describes how he adjusted his performance and ended up analysing and interpreting the documents correctly instead of copying conclusions from the materials:

> *I approached [this assignment] maybe a little too much as if I had simply copied what they say here in these papers and put them down [in my answer]. But then when I started thinking, like about my own views on the topics, then right off in [question] number one, it took me a really long time to answer this question.*
>
> (Study II, student 8, interview)

On the other hand, there were students who could not adjust their thinking or performance. Some of these students said that they always act in the same way:

> *Well, I'm always like this time-management catastrophe. Like in exams and everything, especially exams, it always feels like I run out of time. And in general I notice that in all comprehension and analysis assignments and things like that, they always take me a really long time.*
>
> (Study II, student 5, interview)

Students who had several problems in critical thinking, yet had flexibility, somehow coped with the demands of the task. For example, in Study II, some students had problems evaluating documents and did not form a general picture of the situation presented in the documents. Because the students were struggling with the demands of the task, they selected documents and reproduced arguments and conclusions just as these were presented in the materials. Eventually, the students reached a limited conclusion. On the other hand, there were students who were skilled in specific critical-thinking skills, such as analysing and interpreting information, but lacked other abilities, such as evaluating conflicting claims or producing explanations. These students could neither reach a conclusion nor were they able to determine the weaknesses of alternative solutions. In addition, these students were unable to change their actions or thinking; for example, they were not flexible in time management. In addition, some of the thorough processing students somehow over-analysed the problem (i.e. endless weighing of different options) and, in the end, they failed in the problem-solving process. Also in Study IV, one group of thorough processing students failed the MCQ. The reason was that they apparently did not have enough time to analyse the different response options.

These results match those observed in earlier studies. For example, Elen et al. (2011) have argued that individuals need flexible problem-solving skills in today's rapidly changing society. According to Elen and his colleagues, flexibility is not solely a skill, though it implies that an individual is also disposed to act in a flexible way. Likewise, Winch (2006, xi) has highlighted the critical role of self-knowledge in critical thinking: "[c]ritical appraisal requires not merely the ability to frame and evaluate arguments for and against a position, but also knowledge; not just of what different ends in life entail, but also their personal suitability. In other words, a crucial component of knowledge required is personal self-knowledge, the achievement of which itself involves a critical perspective".

A summary of the empirical findings of this thesis is presented in Figure 8. The figure combines the two main aspects of empirical findings (i.e. student profiles and flexibility in critical thinking) to form a comprehensive picture of the nature of critical thinking.

Critical thinking as a tool for enhancing understanding
Generating personal understanding through thorough processing

Epistemological beliefs: Both objective and subjective knowledge can hold the highest epistemic status. Human knowledge is fallible. Human beliefs, statements and theories can be understood only in relation to the context.

Flexibility in critical thinking

Reaching a well-reasoned solution + Figuring out how to complete multidimensional tasks and planning action + Defining the problem, evaluating, analysing, interpreting information, identifying alternative reasons, considering relationships between assumptions and ultimately reaching a reasoned conclusion.	*Endless weighing of different options* + Defining the problem, identifying ideas, analysing, and interpreting information - Problems in time management, decision-making, reaching reasoned conclusions, evaluating knowledge and judging the acceptability of information. - Problems in producing explanations
Reaching a limited solution - Identifying only a few ideas - Problems in evaluating, analysing and interpreting information - Problems in decision-making and reaching conclusions + Searching for alternative ways to complete a task, changing one's own routines or seeking help from authoritative specialists	*Problems in reaching a conclusion* - Identifying only a few ideas - Problems in evaluating, analysing and interpreting information - Problems in decision-making, reaching conclusions and producing explanations - Expectation that a problem has a definite, right answer - Disposition to think critically may be low

Rigidity in critical thinking

Critical thinking as a tool for determining truth or falsehood
Seeking the right answer through superficial processing

Epistemological beliefs: only objective (i.e. empirical and scientific) knowledge can achieve the highest possible epistemic status. Human knowledge (such as theories and beliefs) is fallible.

Figure 8. Summary of empirical findings.

5.3 Methodological challenges in assessing critical thinking

Both theoretical and empirical analyses of critical thinking brought out several methodological challenges in assessing and analysing this concept.

Firstly, the *theoretical findings* of the present thesis demonstrated that the confusion between normative and descriptive elements can lead to methodological challenges in analysing critical thinking. As noted, determining an individual's conception of knowledge does not reveal what kind of conception represents a sophisticated way of thinking. Thus, without theoretical and philosophical analyses, the normative dimensions easily remain a researcher's personal interpretation of what kind of thinking is sophisticated. Furthermore, it is important to note that the theoretical framework always exerts great influence on the way in which data are interpreted and thus directs the analysis. Therefore, this thesis suggests that there is a real risk that conceptual problems in the framework disturb researchers' interpretation of empirical data.

Secondly, the *empirical results* of the thesis show that measuring critical thinking is challenging. Critical thinking tests attempt to capture an individual's thinking skills, that is, something that exists in the mind. As mentioned in the theoretical framework, there is no particular assessment method that sufficiently captures the whole phenomenon of critical thinking (see also Bailing & Siegel, 2003). The methodological aim here was to enhance understanding of the crucial role played by the assessment instrument as well as the method of analysis in measuring critical thinking skills. The results of Study II and Study IV supported the finding that the *type of assessment* instrument selected to measure students' critical thinking skills can yield different pictures of those abilities (cf. Bowman, 2010; Bowman & Seifert, 2011). In this thesis I argue that the different tests can give dissimilar results, and separate measurements can capture a one-sided view of students' performance. The following results support these claims:

1. Study II found that students' belief in themselves as critical thinkers and knowers (i.e. the self-report assessment) is not necessarily equivalent to how they actually perform in a problem-solving situation. Although some students described themselves in the interview as critical (i.e. able to analyse arguments and determine the truthfulness of claims), in the problem-solving situation they did not evaluate information in the sources critically. Because these students did not interpret the documents they selected nor did they consider presuppositions, they did not recognise that some of the documents were biased. They analysed and interpreted information superficially and focused only on isolated details. In order to draw conclusions these students mainly reproduced details from the documents. Similarly, some students could cite criteria for evaluating knowledge, but in practice they did not know how to use these criteria.

2. Study IV dealt with test performance differences as measured by two different critical thinking tests: a constructed-response task from the Collegiate Learning Assessment (CLA) and a multiple-choice questionnaire (MCQ). The results of Study IV indicated that 10 per cent of students had completely opposite test scores when measured by two different critical thinking tests, meaning that some students had low scores on the CLA and high scores on the MCQ and vice versa (i.e. these students had problems producing convincing arguments and explanations and integrating information from several sources in writing, but showed few if any problems identifying and selecting the right answers from a list of pre-formulated alternatives). This finding is significant because if tests are supposed to measure the same characteristics and dimensions of critical thinking, then the results should be similar. Altogether, one third of the students performed better on the MCQ than on the CLA, while one fourth performed better on the CLA. Study IV suggested that to succeed in both the open-ended CLA task and the MCQ, students needed to know how to apply different skills and thinking processes flexibly. The MCQ asked students to recognise a correct answer from a set of response options, whereas the CLA, with its open-ended tasks, asked students to synthesise information and produce an answer in their own words.

Thirdly, the results of Study IV showed that the *methods of analysis* have an important role in assessing critical thinking: different analytical approaches gave different perspectives on students' performance. The results indicated that the group-level analyses overran the rich variation in test performance that occurred among individual students. To determine whether the students scored differently on the two tests, the CLA scores were first converted to follow the same scale as the MCQ scores. In other words, we redefined the CLA scores as a proportion of the score obtained vis-à-vis the maximum possible score. The means of the converted test scores were compared at the group level using paired t-tests. The group-level analysis did not show statistically significant differences between the CLA and MCQ tests, illustrating that on average the means of students' test scores were similar on both tests (see Table 9).

Table 9. Means of test scores (n=330).

	Mean	Std Deviation
MCQ	0.537	0.170
CLA	0.546	0.153

The individual-level analysis, however, showed variation among the scores. In order to examine how many students had discrepancies in their scores on the two assessments, the students were classified into groups based on their MCQ and CLA scores. According to their success on the CLA test, the students were classified as *high performers*, corresponding to the best-performing third of the students; *average performers*, corresponding to the middle third; and *low performers*, corresponding to the worst-performing third. The respective groups were then created from the MCQ test scores. Table 10 illustrates the interrelation and rich variation between these groups. The individual-level analysis showed that the correspondence between the CLA and the MCQ was fully comparable in 45.5 per cent of the students' test performances. Less than one fifth were high performers (n= 50; 15%), average performers (n=42; 13%) or low performers (n= 58; 18%) on both tests. One fourth of the students (n=79; 24%) performed slightly better on the MCQ than on the CLA, while one fifth (n=67; 20%) performed slightly better on the CLA than on the MCQ. Ten per cent of the students (n=34) performed in completely opposite ways on the two tests. We selected these opposite test responses for further analysis because if tests are supposed to measure the same characteristics and dimensions of critical thinking, then the results should be similar.

Table 10. Cross tabulation of students' test performances (n=330) in Study IV.

Frequency	Low CLA	Average CLA	High CLA	Total
Low MCQ	n=58	32	17	107
Average MCQ	37	42	35	114
High MCQ	17	42	50	109
Total	112	116	102	330

The qualitative analysis was used to deepen the understanding of the characteristic features of the completely opposite test responses and the possible reasons behind these test scores. The qualitative analysis of students' written responses revealed that students who had low CLA and high MCQ[13] scores had problems in producing explanations and arguments in their own words. By contrast, the written responses of the students who had high CLA and low MCQ[14] demonstrated that the task materials were processed in great depth. These students showed that they knew how to define problems, analyse, interpret and evaluate the information on all sides of an issue and use relevant data to produce arguments and explanations. On the background questionnaire these students reported that they had put their best efforts into completing the tests. It is thus somewhat surprising that these thorough processing students performed so low on the MCQ. However, the further analyses in Study IV showed that the problem with the MCQ was not the wrong answers, but the high number of unanswered items (see Table 11). In the test situation the students could not leave any questions blank. However, once the allotted time for each section was called, the students were forced to move on to the next section. Each item of MCQ measures a separate dimension of critical thinking and therefore, requires students to start the analysis process from the very beginning over and over again. This could be very exhausting for thorough processing students.

Table 11. Range of correct, wrong and unanswered MCQ items within the profiles in Study IV.

Profile group	Range of correct items*	Range of wrong items	Range of unanswered items
(1) Superficial processing students High MCQ and low CLA (n= 17)	15–19 Mean 16.9	2–9 Mean 7.1	0–6 Mean 1.4
(2) Thorough processing students High CLA and low MCQ (n= 17)	4–10 Mean 7.8	5–14 Mean 8.2	0–11 Mean 7.2

*Note: Max.= 23 or 24.

[13] At least 63% of the MCQ items were correct, whereas an average CLA score by two scorers ranged from 5.5 to 8.5 out of 18.

[14] An average CLA score by two scorers ranged from 11.5 to 15 out of 18, but fewer than 43% of the MCQ items were correct.

The results of Study IV suggested that thorough processing strategies lead to analysing each multiple-choice alternative deeply, and therefore some of the thorough processing students failed on the MCQ. As a consequence, without taking into account the number of unanswered items and the number of right and wrong answers, the multiple-choice test seemed to provide a narrow view of students' critical thinking. By exploring the relation between right, wrong and unanswered items at the individual level, as well as by identifying the qualitative differences in the students' constructed-response answers, Study IV showed that both overly superficial and thorough processing strategies entail problems. Furthermore, Study II also revealed that both processing strategies entail problems if a student cannot adjust his/her thinking processes and actions to the demands of the task.

6 GENERAL DISCUSSION

6.1 Theoretical reflections

The theoretical aim of the studies that make up this thesis was to offer new insights into conceptualising critical thinking and its prerequisites, especially conceptions of knowledge. Study I showed that the philosophical and psychological research traditions partly use the same epistemological concepts, such as relativism, realism and absolutism. However, there are several examples showing that these two academic traditions give different meanings to these concepts. In psychological research on the conceptions of knowledge this difference is explained as deriving from the different nature of the research traditions (Kitchener, 2011; Southerland et al., 2001). In other words, empirically-orientated research on knowledge is traditionally considered descriptive in nature and therefore is distinct from the normative philosophical research on knowledge. However, Study I revealed shared elements between these two research areas: the theory of PE clearly involves normative assumptions. One normative element is the claim of superiority of some conceptions of knowledge over others, and the other involves pedagogical recommendations. It is important to point out that the normative elements are generally characteristic of all such educational research, which is intended to develop teaching and learning practices by offering pedagogical recommendations based on empirical findings. Therefore, PE as well as other educationally-orientated research needs both philosophical and empirical approaches as well as dialogue between these two academic fields. Problems occur if the empirical findings (descriptive elements) are confused with normative assumptions, that is, based on empirical findings a researcher interprets how thing should be. Such conclusions and interpretations commit the naturalistic fallacy. It follows that on the basis of descriptive features, we cannot formulate any conclusion concerning the superiority of one conception of knowledge over another. Study I illustrated how the naturalist fallacy may result in problems both in relation to empirical research and to pedagogical recommendations.

In previous studies of PE, several researchers have questioned the appropriateness of the terms *naïve* or *sophisticated* and have presented an argument for alternative terms considered to be less judgemental (Elby & Hammer, 2001; Muis, 2004). This thesis encourages us to rethink the theoretical bases and criteria for why certain beliefs are considered sophisticated or not. In my view, changing or replacing the terms is not a sustainable solution. Bailin and Siegel (2003) have also problematised the idea of constructing the hierarchical models of knowledge. They use Bloom's (1956) taxonomy as an example, in which

the recall of information is represented as the lowest category of thinking, whereas evaluation is considered the highest. They argue that this kind of hierarchical classification does not actually refer to the nature of knowledge nor does it provide epistemic criteria for why certain kinds of knowing are judged to be relevant or irrelevant. Rather such a classification describes the outcomes of thinking. For these reasons Bailin and Siegel (2003, 188) note that "rather than attempting to categorize thinking into different kinds to be placed on a hierarchy, it is much more fruitful to focus on what is involved in fulfilling the relevant critical criteria, no matter what the task or context—that is, to focus on critical thinking".

The findings here suggest that the scarcity of theoretical analyses in the empirically-orientated research may lead to a situation in which the meanings and prerequisites of the concepts are not adequately defined. As a result, researchers may use the same concept to describe different meanings or vice versa. As Study I showed, there are three different positions labelled relativist in the context of PE. Previous research on PE has not thoroughly distinguished the differences between these positions or the consequences that follow from them. This kind of conceptual diversity can easily cause confusion: one cannot be sure which form of relativism a researcher is committed to unless the meaning of the concept is carefully explicated. Furthermore, researchers are continuously producing new concepts along with the old. As Hofer and Bendixen (2012, 227) explain, while there are some advantages in conceptual diversity, there are also some serious problems: "[t]hese diverse explorations across fields and subfields have not only enriched the study of personal epistemology but also made it difficult for researchers to communicate effectively and hampered the coherence of this substantial research area". These examples indicate that theoretical analyses are needed to clarify prerequisites and make distinctions among concepts. Theoretical analysis provides a bridge between theory and practice. It is possible to elucidate the background assumptions as well as the contradictory statements and inconsistencies in the theoretical framework by analysing the interconnections among the concepts. For these reasons I argue that theoretical and conceptual analyses are not insignificant matters. Theoretical framework is the researcher's tool for analysing and interpreting data. If the tool is not adequate, then there is a real risk that the analyses will be distorted.

This thesis demonstrates how philosophical research could contribute to the normative dimensions of educational theory, particularly by revealing some conceptual inconsistencies and contradictory assumptions in the theory. The thesis further introduces the notion of fallibilism as a way out of conceptual inconsistency. Contemporary philosophical epistemology offers the concept of epistemological fallibilism to avoid relativism without committing one to naïve realism or epistemological absolutism.

As a conclusion, this thesis makes the claim that the theoretical framework of PE as well as its pedagogical recommendations would benefit by including fallibilism among its epistemological conceptions. The notion of epistemological fallibilism has two kinds of implications for the research on PE: some of these implications are related to the theoretical framework and others to the pedagogical recommendations. At the theoretical level, the concept of fallibilism can expand and refine the theoretical framework of PE, as was the case in Study II. At the pedagogical level, a fallibilist view of knowledge can be seen as a responsible epistemic position. It can further be seen as a psychologically mature view of knowledge, because this position admits to uncertainty of knowledge, but does not embrace the relativist conclusion that all beliefs are equally valid. Fallibilism is also consistent with the evolutionary understanding of knowledge: the bodies of knowledge we now have may be mistaken and thus possible subjects for revision, but they have nevertheless survived the process of evolution to this point; as such, they provide the best available starting point for further inquiry into the choices and actions of the present moment. Epistemological fallibilism fits better than relativism with the presumption of critical thinking from the theoretical and pedagogical points of view, and would be worth promoting in higher education studies. Furthermore, the idea of *research-based* teaching to which the research on learning and teaching in higher education strives to contribute relies on there being some criteria for defining teaching methods that are adequate in comparison with other methods. This idea is untenable in the relativist framework, but not in the fallibilist framework. However, simply replacing the problematic terms with fallibilism would be a risky alternative. In my view, more dialogue is needed between theoretical and empirical research to refine both the theory of epistemological development and how fallibilism is integrated into this theory, as well as its educational implications. For example, the idea of evaluativism as the most developed epistemological position has some similarities (i.e. the possibility of evaluating knowledge in the light of reasons and evidence) to the epistemological fallibilism defended in this thesis. Unfortunately, the concept of evaluativism seems to include some problematic assumptions. One of these concerns the idea of humans as constructing facts. Kuhn's model, which states that *all* knowledge is generated by human beings, leads to an absurd conclusion (see Table 1). It follows, for example, that the findings of natural science exist only if there are human beings who made and conceptualised these findings. However, if Kuhn (2005) means in her model that at the highest level of knowing human beings have an active role in constructing their own understanding, this view would not lead to a philosophically problematic conclusion.

It is important to emphasise that Study I as a philosophical analysis is not criticising the empirical findings—the descriptive elements—of previous PE

research. For example, on the basis of empirical research, it can be true that students' conceptions of knowledge develop towards relativism or that students share the absolutist view of knowledge. However, the point of Study I was that even if it is empirically true that students share the relativist view of knowledge, this would not place relativism in an epistemologically justifiable and sophisticated position, and thus relativism should not be fostered as a goal of higher education. Of course, it is also possible that the theoretical model used in interpreting empirical data has distorted the research findings, but that problem is an empirical matter. It is also important here to understand that interpretations of the superiority of knowledge are researchers' constructions and interpretations. In line with this argument, Perry (1970, 206) himself has noted:

> Our discovery of this particular sequence of challenges as an element common to all our students' experience was a product of our own relation to our data. It was salient to us, and our judges confirmed its generality throughout the reports of our students. This is not to say, however, that other observers with other concepts might not find other common elements or developments.[15]

6.2 Empirical reflections

The goal of the empirical part of the present thesis was to extend the understanding of variation in critical thinking and analyse the interconnections between students' conceptions of knowledge and critical thinking skills. This descriptive aim attempted to depict how different aspects of critical thinking and the conception of knowledge are intertwined. The research involved three empirical studies, each of which contributed to the understanding of the nature of critical thinking from different perspectives. The results demonstrated the complex and reciprocal relationship between students' conceptions of knowledge and critical thinking. Critical thinking emerged as a tool for understanding and determining the relevance and reliability of knowledge. Students' conceptions of knowledge reflected how the students acquired, evaluated, justified and utilised knowledge in a problem-solving situation (see Figure 9). For instance, those students who believed that knowledge could attain truth only when produced by a reliable process (i.e. using empirical methods or consulting experts) palmed off justification for knowing on authorities, and they did not analyse or evaluate that knowledge.

[15] Emphasis added.

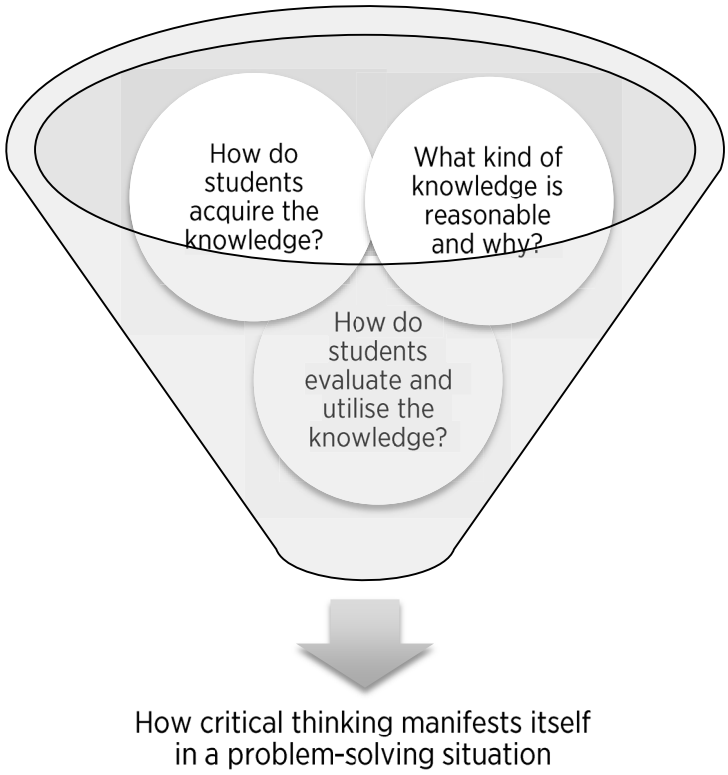


Figure 9. The conception of knowledge and critical thinking.

The results demonstrate that some undergraduate students had real problems in critical thinking. For example, Study IV demonstrated that almost one fifth of the students had low scores on both critical thinking assessments. While some problems were related to the lack of disposition or skill, such as an inability to evaluate the credibility of information, examine presuppositions, make interpretations, develop a personal perspective or generate arguments or conclusions, other problems were related to an inability to modify the whole critical-thinking process in a flexible manner or resulted from a combination of these aspects. The same phenomenon has been observed in previous research on critical thinking (e.g. Arum & Roksa, 2011a; Bok, 2006; Pascarella et al., 2011). The thesis suggests that the nature of the problems and their variance are related to the students' ability to utilise and process the available materials. As an example, incorrect arguments were prevalent in both profile groups, but these were fewer among students who thoroughly processed the materials as compared to the superficial processing students (Study II-IV).

The results of Study IV showed that 18 per cent of participants were low performers on both tests. All students who participated in this study were volunteers. It is thus possible that the students who agreed to participate were more academically inclined than those who did not and that they felt more confident in their test performance and writing skills. This suggests that the difficulties in critical thinking with which students struggle may be greater than Study IV indicates. We know from the background questionnaire that the majority of the students (82%) reported putting their best effort or almost best effort into taking these two tests; only four students reported that they had put little or no effort into the tests. This indicates that most of participants were committed to trying their best in using these assessment tools.

Although the aim of the present research was not to seek explanations for why students' critical thinking skills varied, the empirical results raise the question of what is behind the difference in student profiles. The results showed that some superficial processing students did not really know how to produce convincing arguments and explanations or how to integrate information from several sources, whereas some students in this profile did not share open-mindedness or an inquiring attitude to think critically and thus struggled with the task. Previous research has shown that there are several other aspects that influence students' performance, such as institutional procedures, field-specific conventions and students' prior knowledge, expectations and social backgrounds, as well as the amount of time and effort that the student spends on his or her studies (Arum & Roksa, 2011a; Lea & Street, 1998; Nortedge, 2003; Ylijoki, 2001). The problems may also indicate that the students were not familiar with the assignment, meaning that the learning patterns developed by students in their prior studies were not suited to the new and changed demands. Problems in reasoning may also be explained by the fact that students who are at an intermediate level of expertise, as was the case in Studies II and IV, commonly have problems integrating different bodies of knowledge in a problem-solving situation (Boshuizen, 2004). Based on the findings of previous studies and this thesis, I suggest that the best results in coping with different demands can be achieved by increasing students' self-evaluation skills and flexibility in critical thinking.

6.3 Methodological reflections

The empirical data for this research were analysed by combining data-driven and theory-driven approaches, which offered different perspectives on critical thinking. At the beginning of my doctoral study my goal was to explore the data utilising a theory-driven approach, but during the analytical process in Study II, I realised that the data would not fit the pre-existing theoretical framework used in most of the research in educational psychology. I found that by using a

multifaceted approach (combining the data- and theory-driven approaches as well as the empirical and theoretical analyses) it was possible to deepen my understanding of the emphases and gaps in the prevailing empirical research on critical thinking and personal epistemology. The dialogue between the empirical and theoretical analyses offered new insights into conceptualising critical thinking and its pre-requisites and an opportunity to deepen my understanding of the variations in critical thinking. By combining the theoretical, empirical and methodological perspectives, I was able to use these new perspectives, such as the concept of fallibilism, to analyse the empirical data on critical thinking. The findings here thus contribute to our understanding of the important role of the theoretical framework.

The strength of this thesis is its use of a multifaceted approach to assess critical thinking. The approach here included (a) combining different methods in data collection (including students interviews/self-reports, two different performance-based assessments, the think-aloud-method and video observation), and (b) combining different analytical methods. The focus of the analyses varied from theoretical to empirical, and ranged as well as from a phenomenon-level analysis and a group-level analysis to an individual-level analysis. These approaches provide a way of considering the overlapping aspects of critical thinking examined and increase the reliability and interpretability of the findings (Creene, Carecelli, & Graham, 1989; Abowitz & Toole, 2010). The advantage of combining the individual-level and group-level analyses is that this offers multiple insights into understanding of critical thinking from different standpoints—from a macro-level to a micro-level. The thesis findings support the view that the group-level analyses need to be complemented by individual-level analyses in order to capture a more nuanced picture of performance differences (Lindblom-Ylänne et al., 2013). Through a multifaceted approach it was also possible to introduce new perceptions and gain a more complete understanding of the phenomenon (Johnson et al., 2007). And while the sample size was small, the multi-method approach produced a large amount of data on each participant.

The thesis results are consistent with other studies and suggest that one assessment or analysis method is not enough to evaluate complex cognitive processes (e.g. Baartman et al., 2007; Bowman, 2010; Dierick & Dochy, 2001; Maclellan, 2004). The key challenge in assessing critical thinking, which involves mental processes, is that this kind of phenomenon is not directly observable. Therefore, interpretations of an individual's critical thinking are more or less indirect. The research literature provides ample arguments both for and against the different assessment formats. However, researchers have yet to agree on the best way to measure critical thinking. Each assessment instrument and each format for critical thinking has strengths and weaknesses. Students' self-reports do not necessarily correspond to their abilities to solve real problems and to utilise

critical thinking skills in different situations. The most important drawback to the multiple-choice tests is not being able to know how students processed the test items: the results show whether or not the student has selected the right answer, but reveals nothing about how the student ended up with the solution. Nor do the multiple-choice tests measure students' abilities to produce arguments or give reasoned explanations, which are considered the essential elements of critical thinking (e.g. Ennis, 1991; Halpern, 2014). Although the scoring of the constructed-response task might be challenging (i.e. it is time-consuming, risks scoring bias and so on), the students' written answers reveal the level of processing as well as the level of understanding (cf. Popham, 2003). Additionally, there is evidence that the constructed-response task can facilitate reflection and support the students' self-evaluation skills (Andiliou & Murphy, 2014). However, because of the scoring challenges, the constructed-response tasks are not necessarily the best option for research that focuses on a large dataset.

There has been widespread concern about the effects of testing. For example, testing schemes are assumed to exert considerable influence on what universities emphasise in teaching and the qualities of learning they promote (Bennet & Ward, 1993; Brooks, 2012). Critical thinking tests are also used in educational selection for predicting students' future performance; in other words, test results are used to select the students believed to be the best suited to graduate degree programmes (Kuhn, 2005). Testing results may have other significant implications for policy-making, especially for educational policy within the area of higher education (Morgan & Shahjahan, 2014; Douglass et al., 2012). Another criticism concerns the focus of the tests. Banta and Pike (2012) have argued that the skills and outcomes measured in testing critical thinking represent only a small part of what is important in higher education and in working life (see also Sackett, Borneman, & Connelly, 2008). This thesis demonstrated that different tests can give completely opposite pictures of a student's abilities. For these reasons it is extremely important that researchers, educators and policy-makers know the limitations that are associated with different assessment types and analytical methods and know, first and foremost, the conclusions and interpretations that can be drawn from individual students' test performance. As Arum and Roksa (2011a, 141) have put it: "while the CLA instrument as a measure of learning tracks remarkably well with sociological factors at the aggregate group or at the institutional level, there are still limitations to its precision at the individual level that should caution policy makers from imposing high-stakes accountability schemes based on it or similar assessment indicators". Finally, it is worth mentioning that performance-based assessments as well as other testing formats do not directly assess the quality of teaching, nor do they reflect a student's educability. Thus, over-interpreting the results of performance

assessment as an indicator of the quality of teaching or a student's capacity to be educated is problematic.

6.4 Limitations and ethical reflections

Through the use of a multifaceted approach, it has been possible to enhance our understanding of variations in critical thinking. However, the present findings should not be interpreted as an accurate prediction of the level of undergraduate students' critical thinking. Rather the findings of this study illustrate the nature of the phenomenon of critical thinking and how different aspects combine to contribute to this crucial skill. The small sample size was one common limitation in all of the empirical studies. Another limitation is that none of the studies focused on the domain-specific aspects of critical thinking. The domain-specificity differences were not reported here or in the original articles because these aspects were not a focus of this thesis. However, the domain-specific elaborations could facilitate and enrich the insights into critical thinking even further. There are also several other limitations. For example, Study II involved a small homogeneous sample of students in only one discipline. In Study III only one, ninety-minute writing task was used to measure the students' critical thinking. It is thus possible that a certain aspect of the test situation (such as a life situation, activity level and mood) may have affected the written responses. In Study IV a total of 2,400 students were invited to take the test; however, only 330 Finnish students participated, leaving the response rate at 13.8 per cent. For the purpose of Study IV the low response rate was not a problem because the aim was to compare the test instruments (the CLA and MCQ) rather than to make generalisations from a small sample to a target population. The small response rate may indicate the risk of potential bias in the data. However, this does not necessarily mean that real bias exists in the data. Another limitation of Study IV concerned the experimental design in the original AHELO study. All students started with the constructed-response task and thereafter proceeded to the multiple-choice questionnaire. Furthermore, Study I focused on a limited number of texts. Given these limitations, more communication between the theoretical, empirical and methodological perspectives is needed to increase understanding of this complex phenomenon in the different spheres and in the different phases of higher education.

It is important to point out that the quality and credibility of small-scale qualitative research has been widely discussed (e.g. Flyvbjerg, 2006; Gobo, 2007; Larsson, 2009). The conclusion has been that while case studies and small-set qualitative studies are more limited than large sample studies in breadth and generalisability, they nevertheless provide depth and are especially relevant in exploring phenomena from new perspectives, as is the case in the present thesis.

In the event of a large sample study, the situation is the reverse. That being said, both small and large sample studies are seen as being essential to the development of science (Flyvberg, 2006). Abowitz and Toole (2010) have argued that while no single method or approach is ideal, integrating different methods and approaches in a research design can enhance the reliability of the results and strengthen inferences by providing the opportunity to combine different viewpoints. These aspects also support the use of a multifaceted approach.

All the sub-studies here followed the ethical guidelines of the Finnish Advisory Board on Research Integrity (2012) for the responsible conduct of research, which defines honesty and accuracy as the key aspects (see also Steneck, 2007). The different strategies were used to enhance the accuracy and reliability of the results. Firstly, the multi-method approach was used both in data collection and analysis. Secondly, all test instruments (MCQ and CLA) were translated into Finnish and adapted to a Finnish context using the detailed procedure (see Appendix B). Three certified translators participated in this process. The content validity of the CLA constructed-response task as well as the validity of the translation were confirmed by using the students' interviews and think-aloud method (i.e. a pilot study before the large-scale research was undertaken). Furthermore, in Study IV the students' CLA responses were double-scored by two independent, trained scorers.

The analytical process for each sub-study is explained in the original publications and in the Methods and Procedures section of the current document. The analysis processes of sub-studies were nonlinear; they involved moving back and forth among all the data, data items, analysed qualities and understanding of the phenomenon based on prior studies. The final results of each sub-study were revised and cross-checked several times to see if they worked in relation to the coded extracts and the entire data set. In Study III the first author was responsible for coding the data using the ATLAS.ti programme. The credibility of the coding was checked by the second author, who coded 25 per cent of the data. The consistency was 95 per cent. Thereafter, the first author considered the suggested changes and changed the coding where appropriate. In Studies I, II and IV the analyses were conducted in collaboration with all the authors. In the analysis investigator triangulation (Denzin, 1970, 2012; Creswell & Miller, 2000) was utilised to confirm the reliability of the findings, meaning that the final results were obtained by means of a thorough discussion with all the authors.

From the ethical point of view, it is also important to make sure that all sub-studies were carried out without causing harm to the individuals involved. Voluntary participation, informed consent, anonymity and confidentiality are seen as key ethical concerns in behavioural science research (Finnish Advisory Board, 2009). In the present thesis each database for the three original empirical studies was processed in such a way that the participants could not be identified. Additionally, all participating students were volunteers. All students gave their

written consent for participation, and all were informed that consenting or refraining from consenting would not affect their status or subsequent grades in any way. In Studies III and IV students were also informed that they would not receive individual feedback on their performance. The purpose was to collect baseline data about incoming students' critical thinking skills. Based on this data, necessary interventions can be used in teaching.

6.5 Practical implications

Higher education studies are intended to provide students with the necessary knowledge and skills required in different positions of expertise. Many scholars have pointed out that if undergraduate studies are assumed to contribute to the development of critical thinking skills, then teaching and learning activities need to be purposefully designed to that end (Arum & Roksa, 2011a; Badcock, Pattison, & Harris, 2010; Halpern, 2014). According to Biggs and Tang's (2007) widely used recommendation, students learn what they do. In a similar vein Arum and Roksa (2011a) explain in their recent study that expectations play a key role in facilitating students' learning outcomes. In their (2011a, 93) words: "when faculty have high expectations, students learn more". In other words, critical thinking should be an integral part of higher education and should be explicitly acknowledged in curricula and course outlines. If the topic is not addressed, then there is a risk that critical thinking will remain coincidental in most courses and will depend on the views that the teacher holds about the effectiveness and methods of teaching critical thinking. Furthermore, there is the real risk that the concept of critical thinking will remain only "part of the vocabulary of higher education audit-speak" (Furedi, 2004, 2nd paragraph).

Students are expected to obtain both subject-specific knowledge and critical thinking skills as part of their higher education. However, the present thesis argues that the idea of developing knowledge and skills without the notion of individual disposition is superficial. It follows that students should also be guided to adopt an outlook of thinking critically (Siegel, 1988). As mentioned above in the theoretical framework, from a theoretical point of view the core elements of critical thinking are knowledge, skills and disposition. Students need to have knowledge of what is reasonable for the task or problem, the relevant skills to evaluate and utilise that knowledge and the disposition to do so. The empirical part of this doctoral thesis has identified differences between students in all these aspects of critical thinking. According to Halpern (2014, 18), the key aspects of teaching critical thinking are (1) focusing explicitly on critical thinking skills (e.g. teaching explicit ways to analyse and evaluate arguments, recognise biases and weak arguments, provide arguments and make decisions), (2) developing a

disposition for effortful thinking and learning, (3) focusing on learning activities that increase the probability of transfer and (4) focusing on metacognitive skills.

From the practical point of view, one important implication of the present thesis relates to acknowledging the variations in critical thinking. Although students enter higher education with widely varying skills, they are all expected to begin developing critical thinking and learning the proper conventions from the very beginning of their study. However, previous studies have called attention to the fact that students' critical thinking skills do not always develop during their years in higher education (Arum & Roksa, 2011a; Bok, 2006; Pascarella et al., 2011). The findings here showed that after three years of higher education, students' critical-thinking skills differed greatly. Because there was substantial variation in the students' critical thinking, it seems that a *one-size-fits-all* model of teaching is inadequate. In the future it would be fruitful to investigate how different pedagogical solutions may facilitate and improve the development of critical thinking for different kinds of students. Such a study would help us support students in developing better critical thinking skills by creating learning settings and tasks that more appropriately meet students' different needs.

However, before we can develop the necessary settings and give more detailed pedagogical recommendations, we should first determine what kind of critical thinking can reasonably be developed in higher education—what are the desirable aims of critical thinking, and why? If we bypass the theoretical analysis of the normative elements of teaching critical thinking, there is a real risk that we will offer recommendations based only on researchers' implicit interpretations of what is worth pursuing rather than offering recommendations based on adequate, critical discussion and argumentation. Owing to the normative nature of pedagogy, pedagogical recommendations cannot follow directly from empirical research. That is to say, we require both empirical evidence about the effectiveness of different pedagogical models and also philosophical understanding about the normative elements of teaching critical thinking (see Bailin & Siegel, 2003). As noted, this kind of dialogue is very important, because the relationship between normative and descriptive elements of teaching cannot be adequately formulated without co-operation between these two research approaches. Therefore, if we want to make pedagogical recommendations regarding the teaching of critical thinking, we need to understand both areas of research.

6.6 Conclusions and directions for future research

In the present work a variety of perspectives has been emphasised in order to understand better the various aspects and dimensions of critical thinking. Based on the findings of this thesis, the following seven issues concerning critical thinking are seen as important for further elaboration. The first two concern the

empirical findings, the next two deal methodological issues, and the last three concern theoretical issues:

1. This doctoral thesis has identified two student profiles: superficial processing and thorough processing. The results indicated that thorough processing students mainly solved problems in a critical manner, whereas superficial processing students more often addressed the problems in an uncritical manner. In other words, the students in the superficial profile did not demonstrate a serious attempt to analyse, interpret, evaluate or synthesise information from the materials provided for the task. In the problem-solving situation these students disregarded much of the information and used only one or two sources. Nor did the students consider presuppositions or connections between different points of view. Their reasoning was poor, and they provided vague arguments. This means that these students provided claims or conclusions, but did not provide valid reasons or explanations to support their claims, thereby leaving their arguments weak. Several types of fallacies were also identified.

   By contrast students in the thorough processing profile identified the major ideas in the problem-solving situation. These students demonstrated thorough and accurate understanding of the materials: they evaluated the quality of the information and considered its premises as well as the implications of different conclusions. Most of them weighed different options, analysed connections between claims, connected related ideas and, moreover, gave reasoned explanations. These students further identified and used several criteria in evaluating the reliability of the information: corroborating claims from different sources, evaluating the context in which the claim was made, exploring who interpreted the data and evaluating the presuppositions. Their approach to analysing, evaluating and utilising knowledge can be seen as a psychologically mature view of critical thinking.

   Testing critical thinking can be considered as an intervention. The instructions of the CLA constructed-response task used in this thesis advised students to analyse the materials as well as to justify their positions. Moreover, after the tests many students reported that the constructed-response task guided their analytical process and pushed them to think independently. Therefore, it was rather surprising to find students who seemed to be satisfied with superficial processing. Based on the thesis findings, I suggest that reflection and self-evaluated skills are connected with flexibility and the intention to use critical thinking skills. Future studies would thus benefit from focusing more deeply on students'

self-evaluation skills, that is to say, why some students were unable to adjust their thinking process and why some students decided on superficial analyses.

It is also worth of mentioning that sometimes thinking tasks can demand a superficial, or at least quicker, thinking processing as was the case in Study IV. Therefore, the threshold question is that how well students are able to adapt their thinking process to the demands of the situation.

2. In this thesis I argued that students' performance in critical thinking assessments is associated with the level of processing and the effort made during the test, their dispositions (such as motivational aspects and open-mindedness), their conceptions of knowledge, as well as the students' ability to adjust their actions and thinking processes in a flexible manner. Previous studies have shown that students' expectations, prior knowledge, social backgrounds, institutional or cultural aspects also have an influence on their test performance (e.g. Arum & Roksa, 2011a; Fellenz, 2004; Snow, 1993). In future studies on critical thinking, it would thus be fruitful to investigate students' performance in relation to their expectations. Furthermore, the communication between general and discipline-specific perspectives of critical thinking can be used to examine contextual differences. This kind of research would deepen our understanding of how contextual aspects affect students' performance. Additionally, longitudinal measurements and course work effort (i.e. how much time the students spend studying) would further extend our understanding of the development of critical thinking. A recent longitudinal study has shown that students who study by themselves and show a high level of engagement with reading and writing (i.e. read more than 40 pages a week and write more than 20 pages per semester) develop more reasoning and written communication skills than students who do less (Arum & Roksa, 2011a).

3. The empirical and methodological findings of the doctoral thesis enhance our understanding of the important role of the analysis method and the assessment format used. Combining various assessment and analytical approaches made it possible to obtain a more comprehensive picture of students' critical thinking (see also Johnson & Onwuegbuzie, 2004). Additionally, the thesis demonstrates that by using a multifaceted approach, it was possible to bring out new perspectives that are not currently found in theorisation. As an example, Study II showed that students shared the fallibilist view of knowledge. Previous research on PE has not identified students who are committed to epistemological fallibilism. As noted in Study I, this would have been difficult in any case,

as the concept of fallibilism has not previously been included in the PE theoretical framework.

This thesis has drawn its assumptions and viewpoints from philosophical, educational and psychological theories and procedures. During the analytical process of Study II, I discovered several conceptual inconsistencies in terms of the relationship of the empirically-based theory of knowledge and its justifications for the conceptions of knowledge. The theoretical analysis in Study I provided a tool with which to analyse and clarify these inconsistencies and integrate different viewpoints. In my view, theoretical analysis has an essential role in evaluating the background assumptions of theories. Through theoretical analysis, it was also possible to bring out interconnections between the key concepts as well as contradictory statements and inconsistencies in the theoretical framework (see also Holma, 2009). Theoretical analyses can be used in dialogue with empirical analyses in order to refine the prevailing theorisation.

4. The thesis argues that normative and descriptive elements are fundamentally intertwined in the research on critical thinking. Therefore, research on critical thinking requires co-operation between both philosophical and empirical approaches. Here I demonstrate some methodological problems that follow from the confusion of these two elements and that influence the interpretations of the data. These problems also have broad and serious implications, because they have led to problematic educational guidelines; for example, researchers have promoted relativism as a learning aim and as a background assumption of critical thinking.

   In this thesis I have argued that more dialogue between theoretical and empirical research is needed. Although the theoretical analysis in Study I focused on one particular theory, my broader aim was to demonstrate the importance of philosophical analysis to educational theories more generally. As with PE, educational theories contain both empirical and philosophical research elements. The relationship between these two groups of elements cannot be adequately determined without co-operation between the two types of research. Researchers should thus be aware of the difference between normative and descriptive elements. Both educational/pedagogical and philosophical theories include a strong normative dimension, whereas (purely descriptive) empirical theories do not share such assumptions. However, it is important to understand that when a researcher makes interpretations based on empirical theory or on the empirical findings of how teaching and learning should be developed

or makes claims of superiority of some conceptions as compared with others, she or he simultaneously moves into the normative dimension. Therefore, the arguments presented in this thesis in relation to PE have wide-ranging relevance in educational research.

5. According to the present thesis, relativism, naïve realism and fallibilism have theoretical importance for conceptualising critical thinking. The thesis explained why the idea that critical thinking presupposes a relativist view of knowledge is untenable. Also, in the context of naïve realism and absolutism, critical thinking turns out to be pointless, because certain knowledge is received directly from the external world. As mentioned several times, fallibilism admits of uncertain knowledge, but does not arrive at the relativist conclusion that all beliefs are equal. Fallibilism further differs from the variants of absolutism, which claim that there are some sources of certainty on which to base our search for understanding. This position presumes the possibility of improving our current conceptions, theories or beliefs. As Holma (2012, p. 399) aptly states of fallibilism, "this position, like the belief that all human knowledge is uncertain, coheres with the evolutionary understanding of knowledge: the bodies of knowledge we now have may be mistaken and thus [are] possible subjects for revision, but they have, nevertheless, survived the process of evolution to this point; as such, they provide the best available starting point for choices and action of the present moment concerning further inquiry" (see also Peirce, 1934ab). From this point of view, epistemological fallibilism fits the presumption of critical thinking.

6. This doctoral thesis has further argued that relativism is an epistemological position, which is very difficult to defend not only from a philosophical perspective, but also from an educational point of view (see also Phillips, 1995, 2007). As Study I explained, relativism is problematic in the context of research on learning and teaching in higher education. Many of its aims and assumptions contradict relativism. As an example, the endorsement of critical thinking as a learning aim (Hofer & Bendixen, 2012; Kember, 2001; King & Kitchener, 2002, 2004; Kuhn, 2005; Lucas & Tan, 2013; Phan, 2008) is pointless in a relativist framework. Furthermore, the idea of *research-based* teaching to which the research on learning and teaching in higher education endeavours to contribute, both on theoretical and practical levels, relies on the idea that there are criteria for defining the kind of teaching methods that are adequate in comparison with other methods. If, in contrast, researchers' beliefs about teaching methods were only relative to their own subjective constructions, culture

or some other frame of reference, universities would not be justified in promoting or advancing this kind of teaching.

7. In contrast to previous research on PE (King & Kitchener, 2002, 2004; Kuhn, 1999, 2005; Kuhn & Weinstock, 2002), this doctoral thesis maintains that trusting in objective knowledge or relying on authorities does not imply either the possibility of certainty or the idea that our beliefs are copies of external reality. Nor is such trust a sign of lower-level thinking as such. Actually, to some extent it is sensible to trust in objective knowledge and in authorities. If this were not so, why would we need education or teachers? Objectivity, truth and certainty refer to different points of view. For example, although scientific theories try to achieve the highest level of objectivity, it does not follow that any scientific theory is certain.

   However, in this thesis, from the point of view of critical thinking, the challenge for those students who trusted solely in objective knowledge was that they solved the given problem in an uncritical manner. In other words, they did not analyse, evaluate or interpret the information contained in the documents they were given. They acquired knowledge by appealing to authoritative opinion, trusting in verified empirical evidence and listening to others' testimonies. These students palmed off justification for knowing on testimonies or on authoritative experts; either they were simply not disposed to use their critical thinking skills or they had problems applying those skills in a problem-solving situation. Furthermore, some of these students tried to justify biased knowledge by appealing to an authority. And yet the results showed that these same students did not share the belief that knowledge is absolutely certain or unquestionable. Nor did these students share the view that beliefs accurately represent or correspond to reality. In effect, the students did not share a sense of naïve realism. In the future it would be significant to investigate in more depth how critical thinking is related to the students' attitude of testimonies.

   The present thesis introduced the concept of epistemological fallibilism as an epistemologically justifiable solution to this theoretical inconsistence. The recognition of epistemological fallibilism would also be beneficial both to the PE theoretical framework and to PE-based university pedagogy. However, the thesis does not argue that fallibilism is the most sophisticated position that a person could hold. Actually, the results of the present research show that the fallibilist view of knowledge can be connected with different profiles in critical thinking. Therefore, more research combining theoretical and empirical approaches is needed to clarify the theory of epistemic development. It has been suggested that

students who share the most sophisticated view of knowledge have the ability to make context-sensitive judgements (e.g. Barzilai & Zohar, 2012). However, the findings of this thesis suggest that students who use their critical thinking skills at remarkably high levels have the ability to make context-sensitive judgements. In the future, it would be interesting to examine students' conceptions of knowledge within a larger dataset and in different contexts. Besides utilising communication between empirical and theoretical research, another unexplored aspect of students' conceptions of knowledge might be identified. Such a study would help to continue refining the theories of PE.

REFERENCES

Abowitz, D., & Toole, M. (2010). Mixed method research: Fundamental issues of design, validity, and reliability in construction research. *Journal of Construction Engineering and Management* 136: 108–116.

Almond, R. G. (2014). Using Automated Essay Scores as an Anchor When Equating Constructed Response Writing Tests. *International Journal of Testing* 14: 73–91.

The American Philosophical Association. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction "the Delphi Report". Committee on Pre-College Philosophy. Millbrae, CA: The California Academic Press.

Andiliou, A., & Murphy, P. K. (2014). Creative solutions and their evaluation: Comparing the effects of explanations and argumentation tasks on students reflections. *Frontline Learning Research* 4: 92–114.

Andrews, D. H., & Wulfeck H. (2014). Performance Assessment: Something Old, Something New. In J. M. Spector, M. D. Merrill, J. Elen, M. J. Bishop (eds.), *Handbook of Research on Educational Communications and Technology*. New York: Springer, pp. 303–310.

Arum, R., & Roksa, J. (2011a). *Academically Adrift. Limited Learning on College Campuses*. Chicago: The University of Chicago Press.

Arum, R., & Roksa, J. (2011b). Limited learning on college campuses. *Society*: 48, 203-207. DOI 10.007/s12115-011-9417-8.

Attali, Y. (2014). A Ranking Method for Evaluating Constructed Responses. *Educational and Psychological Measurement online*, DOI: 10.1177/0013164414527450.

The Australian Council for Educational Research (ACER). (2001). Graduate Skills Assessment. Summary Report. 01/E Occasional Paper Series. Higher Education Division. Retrieved from http://www.acer.edu.au/documents/GSA_SummaryReport.pdf

Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & van der Vleuten, C. P. M. (2007). Evaluating assessment quality in a competence-based education: A qualitative comparison of two frameworks. *Educational Research Review* 2: 114–129.

Badcock, P.B.T., Pattison, P. E., & Harris, K-L. (2010) Developing generic skills through university study: a study of arts, science and engineering in Australia. *Higher Education* 60(4): 441–458.

Bailin, S., & Siegel, H. (2003). Critical thinking. In N. Blake, P. Smeyers, R. Smith, & P. Standish (eds.), *The Blackwell guide to the philosophy of education*. Oxford: Blackwell Publishing, pp. 181–193.

Banta, T., & Pike G. (2012). "Making the Case Against – One More time", Commentary on R. Benjamin. (2012), The Seven Red Herrings About Standardized Assessments in Higher Education, September 2012, National Institute for Learning Outcomes Assessment. Retrieved 24.2.2015 from http://learningoutcomesassessment.org/documents/HerringPaperFINAL.pdf.

Barzilai, S., & Zohar, A. (2012). Epistemic thinking in action: evaluating and integrating online sources. *Cognition and Instruction* 30: 39–85.

Baxter Magolda, M. (1992). Students' epistemologies and academic experiences: implications for pedagogy. *Review of Higher Education* 15: 265–287.

Bennet, R. R., & Ward, W. C. (Eds.). (1993). *Construction versus choice in cognitive measurement: issues in constructed response, performance testing, and portfolio assessment*. Hillsdale: Lawrence Erlbaum Associates.

Biggs, J., & Tang, C. (2007). *Teaching for Quality Learning: What student does*. Maidenhead, UK: SRHE and Open University Press Imprint.

Bleazby, J. (2011). Overcoming relativism and absolutism: Dewey's ideals of truth and meaning in philosophy for children. *Educational Philosophy and Theory* 43, 453–466. doi: 10.1111/j.1469-5812.2009.00567.x

Boghossian, P. (2006). *Fear of Knowledge: Against Relativism and Constructivism*. Oxford: Clarendon Press.

Bok, D. (2006). *Our underachieving colleges. A candid look at how much students learn and why they should be learning more*. Princeton, NJ: Princeton University Press.

Boshuizen, H. P. A. (2004). Does practice make perfect? A slow and discontinuous process. In H. P. A. Boshuizen, R. Bromme, & H. Gruber (eds.), *Professional learning: Gaps and transitions on the way from novice to expert*. Netherlands: KluwerAcademic Publishers, pp. 73–95.

Bowman, N. A. (2010). Can 1st-year college students accurately report their learning and development? *American Educational Research Journal* 47: 466–496.

Bowman, N. A., & Seifert, T. A. (2011). Can College Students Accurately Assess What Affects Their Learning and Development? *Journal of College Student Development* 52: 270–290.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology* 3: 77–101.

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.

Brew, A., & Jewell, E. (2012). Enhancing quality learning through experiences of research-based learning: implications for academic development. *International Journal for Academic Development* 17: 47–58.

Briell, J.E., Elen, J., & Clarebout, G. (2013). Seeking convergent evidence of epistemological beliefs: a novel survey. *Electronic Journal of Research in Educational Psychology* 11(2): 473–500.

Brooks, P. (2012). Outcomes, testing, learning; what's at stake? *Social Research* 79: 601–611.

Brown, R. W. (2001). Multi-choice versus descriptive examinations. Frontiers in Education Conference 31st Annual.

Brownlee, J. (2004). Teacher education students' epistemological beliefs. Developing a relational model of teaching. *Research in Education* 72: 1–17.

Brownlee, J., Walker, S., Lennox, S., Exley, B., & Pearce, S. (2009). The first year university experience: Using personal epistemology to understand effective learning and teaching in higher education. *Higher Education* 58: 599–618.

Buckland. L. A. (2010). Implications of philosophy for assessing epistemic cognition. ICLS '10 Proceeding of the 9th International conference of the Learning Sciences – Volume 2, pp. 40–42.

Chinn, C. A., Buckland, L. A. & Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: arguments from philosophy and psychology. *Educational Psychologist* 46(3): 141–167.

Coates, H., & Richardson, S. (2011). An international assessment of bachelor degree graduates' learning outcomes". *Higher Education Management and Policy* 23: 1–19.

Creswell, J.W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory and Practice* 39: 124–131.

Darab, S. (2006). A preventive approach to plagiarism: an empirical study of a first-year unit for undergraduates. *International Journal for Educational Integrity* 2: 3–15.

Denzin, N. K. (1970). *The research act*. Chicago, IL: Aldine.

Denzin, N. K. (2012). Triangulation 2.0. *Journal of Mixed Methods Research* 6: 80–88.

Dewey, J. (1910). *How we think*. Boston: D.C. Heath & Co.

Dierick, S., & Dochy, F. (2001). New lines in edumetrics: new form of assessment lead to new assessment criteria. *Studies in Educational Evaluation* 27: 307–329.

Douglass, J. A., Thomson, G., & Zhao, C-M. (2012). Learning outcomes race: the value of self-reported gains in large research universities. *Higher Education* 64: 317–335.

Elby, A., & Hammer, D. (2001). On the substance of a sophisticated epistemology. *Science Education*: 554–567.

Elo, S., & Kyngäs H. (2007). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1): 107–115.

Ennis, R. (1991). Critical thinking: a streamlined conception. *Teaching Philosophy* 14: 5–24.

Everitt, N., & Fisher, A. (1995). *Modern epistemology: a new introduction*. New York, McGraw-Hill.

Fellenz, M. R. (2004). Using assessment to support higher level learning: the multiple choice item development assignment. *Assessment & Evaluation in Higher Education* 29: 703–719.

Fenstermacher, G. D. (1994). The Knower and the Known: The Nature of Knowledge in Research on Teaching. *Review of Research in Education* 20: 3–56.

Finnish Advisory Broad on Research Integrity. (2009). Ethical principles of research in the humanities and social and behavioural sciences and proposals for ethical review. Retrieved 24.2.2015 from http://www.tenk.fi/sites/tenk.fi/files/ethicalprinciples.pdf.

Finnish Advisory Broad on Research Integrity. (2012). Responsible conduct of research and procedures for handling allegations of misconduct in Finland. Retrieved 24.2.2015 from http://www.tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf.

Fisher, A. (2011). *Critical Thinking: An Introduction*. Cambridge: Cambridge University press.

Fisher, A., & Scriven, M. (1997). *Critical thinking: its definition and assessment*. Edgepress and Centre for Research in Critical Thinking. University of East Anglia.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative inquiry* 12: 219–245.

Frank, P. (1948*). Einstein: His Life and Times*. London.

Furedi, F. (2004). It's now no longer critical and nor is it thinking. *Times Higher Education*. Retrieved 24.2.2015 from http://www.timeshighereducation.co.uk/news/its-now-no-longer-critical-and-nor-is-it-thinking/191406.article.

Glaser, E. M. (1942). An experiment in the development of critical thinking. *Teachers College Record*: 409–410.

Gobo, G. (2007). Sampling, representativeness and generalizability. In C. Seale, G. Gobo, J. F. Gubrium, & D. Silverman (eds.), *Qualitative research practice*. London: SAGE Publications, pp. 405–426.

Green, J. A., & Yu, S. B. (2014). Modelling and measuring epistemic cognition: A qualitative re-investigation. *Contemporary Educational Psychology* 39: 12–28.

Greene, J. A., Azevedo, R., & Torney-Purta, J. (2008). Modeling epistemic and ontological cognition: Philosophical perspectives and methodological directions. *Educational Psychologist* 43(3): 142–160.

Greene, J. A., Torney-Purta, J. & Azevedo, R. (2010). Empirical evidence regarding relations among a model of epistemic and ontological cognition, academic performance, and educational level. *Journal of Educational Psychology* 102(1): 234–255.

Greene J. A. & Yu S. B. (2014). Modeling and measuring epistemic cognition: A qualitative re-investigation. *Contemporary Educational Psychology* 39: 12–28.

Groarke, L. (2013). Informal Logic. In E. N. Zalta (ed.), The Stanford Encyclopedia of Philosophy (Spring 2013 Edition). Stanford University. Retrieved 24.2.2015 from http://plato.stanford.edu/archives/spr2013/entries/logic-informal/.

Haig, B. D. (2005). Abductive theory of scientific method. *Psychological Methods* 10: 371–388.

Halpern, D. F. (1993). Assessing the effectiveness of critical thinking instruction. *The Journal of General Education* 42: 238–254.

Halpern, D. F. (2014) *Thought and Knowledge*. Fifth edition. NY: Psychology Press.

Hammer, D., & Elby, A. (2003). Tapping epistemological resources for learning physics. *The Journal of the Learning Sciences* 12 (1): 53–90.

Healey, M. (2005). Linking research and teaching to benefit student learning. *Journal of Geography in Higher Education* 29(2): 183–201.

Heijltjes, A., van Gog, T., Leppink, J., & Paas, F. (2014). Improving critical thinking: Effects of dispositions and instructions on economics students' reasoning skills. *Learning and Instruction* 29: 31–42.

Hofer, B. K. (2000). Dimensionality and disciplinary differences in personal epistemology. *Contemporary Educational Psychology* 25: 378–405.

Hofer, B. K. (2001). Personal epistemology research: implications for learning and teaching. *Journal of Educational Psychology Review* 13(4): 353–383.

Hofer, B. K. (2002). Personal epistemology as a psychological and educational construct: an introduction. In B. K. Hofer & P. R. Pintrich (eds.*), Personal epistemology: The Psychology of Beliefs about Knowledge and Knowing*. New Jersey: Lawrence Erlbaum Associates, pp. 1–14.

Hofer, B. K. (2004). Epistemological understanding as a metacognitive process: Thinking aloud during online searching. *Educational Psychologist* 39: 43–55.

Hofer, B. K. (2005). The Legacy and the challenges: Paul Pintrich's contributions to personal epistemology research. *Educational Psychologist* 40: 95–105.

Hofer, B. K. (2006a). Domain specificity of personal epistemology: resolved questions, persistent issues, new models. *International Journal of Educational Research* 45: 85–95.

Hofer, B. K. (2006b). Beliefs about knowledge and knowing: integrating domain specificity and domain generality: A response to Muis, Bendixen and Haerle (2006). *Educational Psychology Review* 18: 67–76.

Hofer, B. K. & Bendixen, L. D. (2012). Personal Epistemology: Theory, research, and future directions. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra & J. Sweller (eds.), *APA Educational Psychology Handbook, Vol 1: Theories, Constructs, and Critical Issues*. Washington, DC, US: American Psychological Association, pp. 227–256.

Hofer, B. K. & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research* 67: 88–140.

Hofer, B. K. & Pintrich, P. R. (2002). *Personal epistemology: The psychology of beliefs about knowledge and knowing*. New Jersey: Lawrence Erlbaum Associates.

Holma, K. (2004). Plurealism and education: Israel Scheffler's synthesis and its presumable educational implications. *Educational Theory* 54(4): 419–430.

Holma, K. (2009). The strict analysis and the open discussion. *Journal of Philosophy of Education* 43(3): 325–338.

Holma, K. (2011). The epistemological conditions of moral education: the notions of rationality and objectivity revisited. *Educational Theory* 61(5): 533–548.

Holma, K. (2012). Fallibilist pluralism and education for shared citizenship. *Educational theory* 62(4): 397–409.

Holma, K. (in press). The Critical Spirit: Emotional and moral dimensions critical thinking. Studier i Peadgogisk filosofi.

Holma, K., & Hyytinen, H. (equal contribution) (in press-a). The philosophy of personal epistemology. *Theory and Research in Education*.

Holma, K. & Hyytinen, H. (equal contribution) (in-press-b). Filosofian ja empirian dialogi: normatiivisten ja deskriptiivisten ulottuvuuksien yhteenkietoutuminen kasvatustutkimuksen metodologisena haasteena. *Kasvatus* 3/2015.

Hyytinen, H. (2011). Biggsin linjakkaan opetuksen perusteista. In K. Holma & K. Mälkki (eds.) *Tutkimusmatkalla: Metodologia, teoria ja filosofia kasvatustutkimuksessa*. Helsinki: Gaudeamus, pp. 94–107.

Hyytinen, H., Holma, K., Toom, A., Shavelson, R., & Lindblom-Ylänne, S. (2014). The complex relationships between critical thinking and epistemological beliefs in the context of problem solving. *Frontline Learning Research* 6: 1–25.

Hyytinen, H., Nissinen, K., Ursin, J., Toom, A., & Lindblom-Ylänne, S. (2015). Problematising the equivalence of the test results of performance-based critical thinking tests for undergraduate students. *Studies in Educational Evaluation 44*: 1–8. doi:10.1016/j.stueduc.2014.11.001

Hyytinen, H., Löfström, E., & Lindblom-Ylänne, S. (2015). Challenges in argumentation and paraphrasing among beginning students in educational sciences. Manuscript submitted for publication.

Ichikawa, J. J., & Steup, M. (2012). The Analysis of Knowledge. In E. N. Zalta (ed.), The Stanford Encyclopedia of Philosophy (Winter 2012 Edition). Retrieved 24.2.2015 from http://plato.stanford.edu/archives/win2012/entries/knowledge-analysis/>.

Jensen, J., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test…or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology review* 26: 307–329.

Johnson, J. R., & Onwuegbuzie, A. J. (2004). Mixed methods research: a research paradigm whose time has come. *Educational researcher* 33: 14–26.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research* 1: 112–133.

Jurowska, J. E., & Thompson, J. P. (2010). "Opening doors to early academic integrity" – aiding the transition to and managing expectations of academic practice at university. *International Journal for Educational Integrity* 8: 4–20.

Kaartinen-Koutaniemi, M., & Lindblom-Ylänne, S. (2008). Personal epistemology of psychology, theology and pharmacy students: A comparative study. *Studies in Higher Education*, 33: 179–191.

Kaartinen-Koutaniemi, M., & Lindblom-Ylänne, S. (2012). Personal epistemology of university students: individual profiles. *Education Research International*: 1–8.

Kaddoura, M. A. (2010). New graduate nurses' perceptions of the effect of clinical simulation on their critical thinking, learning, and confidence. *Journal of Continuing Education in Nursing* 41: 506–516.

Keinonen, T., & Kärkkäinen, S. (2010). University students' argumentation in science and environmental education. *Problems of Education in the 21st Century* 22: 54–63.

Kember, D. (2001). Beliefs about knowledge and the process of teaching and learning as a factor in adjusting to study in higher education. *Studies in Higher Education* 26: 205–221.

King, P. M. & Kitchener, K. S. (1994). *Developing reflective Judgment: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults*. San Francisco: 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*. New Jersey: Lawrence Erlbaum Associates, 37–62.

King, P. M., & Kitchener, K. S. (2004). Reflective Judgment: Theory and Research on the Development of Epistemic Assumptions Trough Adulthood. *Educational Psychologist* 39: 5–18.

Kitchener, R. F. (2002). Folk epistemology: An introduction. New Ideas in Psychology 20: 89–105.

Kitchener, R. F. (2011). Personal epistemology and philosophical epistemology: The view of a philosopher. In J. Elen, E. Stahl, R. Bromme & G. Clarebout (eds.), *Links between Beliefs and Cognitive Flexibility: Lessons learned*. New York: Springer, pp. 79–104.

Klein, S., Benjamin R., Shavelson, R., & Bolus, R. (2007). The Collegiate Learning Assessment. Facts and Fantasies. *Evaluation Review*, 31 (5): 415–439.

Kotzee, B. (2010). Seven posers in the constructivist classroom. *London Review of Education* 8(2): 177–187.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher* 28: 16–25.

Kuhn, D. (2005). *Education for thinking*. Cambridge, MA: Harvard University Press.

Kuhn, D., & Weinstock, M. (2002). What is epistemological thinking and why does it matter? In B. K. Hofer & P. R. Pintrich (eds.) *Personal epistemology: The psychology of Beliefs about Knowledge and Knowing*. New Jersey: Lawrence Erlbaum Associates, pp.121–144.

Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development* 15: 309–328.

Lahtinen, A-M, & Pehkonen, L. (2013). 'Seeing things in a new light': Conditions for changes in the epistemological beliefs of university students. *Journal of Further and Higher Education* 37: 397–415.

Larsson, S. (2009). A pluralist view of generalization in qualitative research. *Journal of Research & Method in Education* 32: 25–38.

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments, & Computers* 41(3): 841–849. Retrieved 24.2.2015 from http://www.springerlink.com/content/d53722q3k3314374/.

Lea, M. R., &. Street, B. V. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education* 23: 157–173.

Lincoln Y. S., & Guba, E. G. (1990). Judging the quality of case study reports. International Journal of *Qualitative Studies in Education* 3: 53–39.

Lindblom-Ylänne, S., Lonka, K., & Leskinen E. (1996) Selecting students for medical school: What predicts success during basic science studies? A cognitive approach. *Higher Education* 31: 507–527.

Lindblom-Ylänne, S., Parpala, A., & Postareff, L. (2013). Challenges in analysing change in students' approaches to learning. In D. Gijbels, V. Doche, J. Richardson & J. Vermunt (eds.) *Learning patterns in higher education: Dimensions and research perspectives*. New York: Routledge, pp. 232–248.

Lucas, U., & Tan, P. L. (2013). Developing a capacity to engage in critical reflection: students' 'ways of knowing' within an undergraduate business and accounting programme. *Studies in Higher Education* 38: 104–123.

Maclellan, E. (2004). How convincing is alternative assessment for use in higher education. *Assessment & Evaluation in Higher Education* 29: 311–321.

Markowitsch, J., & Messerer, K. (2007). Practice-oriented methods in teaching and learning in higher education. In P. Tynjälä, J. Välimaa, & G. Boulton-Lewis (eds.), *Higher education and working life: collaborations, confrontations and challenges*. Oxford: Elsevier, pp. 177–194.

Max Planck Institute for Psycholinguistics. (2012). ELAN. [Software]. Retrieved 24.2.2015 from http://tla.mpi.nl/tools/tla-tools/elan/.

Mayring, P. (2000). Qualitative Content Analysis. *Forum: Qualitative Social Research* 1: 105–114.

Ministry of Education and Culture (2015a). University Education in Finland. Retrieved from 17.3.2015 http://www.minedu.fi/OPM/Koulutus/yliopistokoulutus/?lang=en.

Ministry of Education and Culture (2015b). Education System in Finland. Retrieved from 17.3.2015 from http://www.minedu.fi/OPM/Koulutus/koulutusjaerjestelmae/?lang=en.

Moore, W. S. (2002). Understanding learning in a postmodern world: reconsidering the Perry scheme of intellectual and ethical development. In B. K. Hofer & P. R. Pintrich (eds.), *Personal epistemology: The Psychology of Beliefs about Knowledge and Knowing*. New Jersey: Lawrence Erlbaum Associates, pp. 17–36.

Moore, T. (2004). The critical thinking debate: how general are general thinking skills. *Higher Education Research & Development* 23: 3–18.

Moore, T. (2013). Critical thinking: seven definitions in search of a concept. *Studies in Higher Education* 38: 506–522.

Morgan, C., & Shahjahan, R. A. (2014). The legitimation of OECD's global educational governance: examining PISA and AHELO test production. *Comparative Education*, DOI: 10.1080/03050068.2013.834559.

Muis, K. R. (2004). Personal epistemology and mathematics: a critical review and synthesis of research. *Review of Educational Research* 74(3): 317–377.

Muis, K. R., Bendixen, L. D., & Haerle, F. C. (2006). Domain-Generality and Domain-Specificity in Personal. Epistemology Research: Philosophical and Empirical Reflections in the Development of a Theoretical Framework. *Educational Psychology Review* 18: 3–54.

Murphy, P. K. (2003). The philosophy in thee: tracing philosophical influences in educational psychology. *Educational psychologist* 38(3): 137–145.

Murphy, P.K., Alexander, P. A., & Muis, K. R. (2012). Knowledge and knowing: the journey from philosophy and psychology to human learning. In K. R. Harris, S. Graham, T. Urdan , C. B. McCormick, G. M. Sinatra & J. Sweller (eds), *APA Educational Psychology Handbook, Vol 1: Theories, Constructs, and Critical Issues*. Washington, DC, US: American Psychological Association, pp. 189–226.

Nicholas, M. C., & Labig Jr., C. E. (2013) Faculty approaches to assessing critical thinking in the humanities and the natural and the human sciences: Implications for general education. *Journal of General Education* 62(4): 279–319.

Nicol, D. (2007). E-assessment by design: using multiple-choice test to good effect. *Journal for Further and Higher Education* 31: 53–64.

Nieminen, J., Lindblom-Ylänne, S. & Lonka, K. (2004). The development of study orientations and study success in students of pharmacy. *Instructional Science* 32: 387–417.

Niiniluoto, I. (1999). *Critical Scientific Realism*. Oxford: Oxford University Press.

Onwuegbuzie, A., Johnson, R. B., & Collins, K. (2009). Call for mixed analysis: A philosophical framework for combining qualitative and quantitative approaches. *International Journal of Multiple Research Approaches* 3: 114–139.

Paul, R.W., Elder, L., & Bartell, T. (1997). California Teacher Preparation for instruction in critical thinking: research finding and policy recommendations. California Commission on Teacher Credentialing, Sacramento. Research 143.

Papastephanou, M. & Angeli, C. (2007). Critical Thinking Beyond Skill. *Educational Philosophy & Theory* 39(6): 604–621.

Pascarella, E. T., Blaich, C., Martin, G. L., & Hanson, J. M. (2011) How robust are the findings of academically adrift? Change: *The Magazine of Higher Learning* 43: 20–24.

Peirce, C. S. (1934a). Some consequences of four incapacities. In C. Hartshorne & P. Weiss (eds.), *Pragmatism and Pragmaticism, volume 5 of Collected Papers of Charles Sanders Peirce*, pp. 156–189.

Peirce, C. S. (1934b). Questions concerning certain faculties claimed for man. In C. Hartshorne & P. Weiss (eds.), *Pragmatism and Pragmaticism, volume 5 of Collected Papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press, pp. 135–155.

Perry, W. G. (1970). *Forms of Intellectual and Ethical Development in the College Years.* Harvard University.

Phan, H. P. (2008). Predicting change in epistemological beliefs, reflective thinking and learning styles: a longitudinal study. *British Journal of Educational Psychology* 78: 75–93.

Phillips, D. C. (1995). The good, the bad and the ugly: The many faces of constructivism. *Educational Researchers* 24: 5–12.

Phillips, P. J. J. (2007). *The Challenge of Relativism: Its Nature and its Limits*. London: Continuum.

Pike, G. R. (1995). The relationship between self-reports of college experiences and achievement test scores. *Research in Higher Education* 36: 1–21.

Pike, G. R. (1996). Limitations of using students' self-reports of academic development as proxies for traditional achievement measures. *Research in Higher Education* 37: 89–114.

Pike, G. R. (1999). The constant error of the halo in educational outcomes research. *Research in Higher Education* 40: 61–86.

Pintrich, P. R. (2002). Future challenges and directions for theory and research on personal epistemology. In B. K. Hofer & P. Pintrich (eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing*. New Jersey: Lawrence Erlbaum Associates, pp. 121–144.

Popham, W. J. (2003). *Test Better, Teach Better. The instructional Role of Assessment*. Alexandria, Virginia: ASCD.

Popper, K. R. (1972). *Objective Knowledge: An Evolutionary Approach*. London: Clarendon Press.

Putnam, H. (1978). *Meaning and the Moral Sciences.* London: Routledge and Kegan Paul.

Putnam, H. (1981). *Reason, Truth and History*. Cambridge: Cambridge Press.

Rapanta, C., Garcia-Mila, M., & Gilabert, S. (2013). What is meant by argumentative competence? An integrative review of methods of analysis and assessment in education. *Review of Educational Research* 83: 483–520.

Ritchhart, R. (2002). *Intellectual Character: What it is, why it matters and how to get it*. San Francisco: Jossey Bass.

Rodriguez, M. C. (2003). Construct equivalence of multiple-choice and constructed-response items: a random effects synthesis of correlations. *Journal of Educational Measurement* 40: 163–184.

Ryle, G. (1949). *The concept of mind*. London: Hutchinson's University Library.

Sackett, P. R., Borneman, M.J., & Connelly, B. S. (2008). High-Stakes Testing in higher Education and employment. Appraising the evidence for validity and fairness. *American Psychologist* 63: 215–227.

Sandoval, W. A. (2009). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences*, 12: 5–51.

Scheffler, I. (1973). *Reason and Teaching*. London: Routledge & Kegan Paul.

Scheffler, I. (1974). *Four Pragmatists: A Critical Introduction to Peirce, James, Mead, and Dewey*. London: Routledge.

Scheffler, I. (1997). *Symbolic Worlds: Art, Science, Language, and Ritual*. New York: Cambridge University Press.

Scheffler, I. (2000). A plea for plurealism. *Erkenntnis* 52(2): 161–173.

Scheffler, I. (2001). My quarrels with Nelson Goodman. *Philosophy and Phenomenological Research* 62(3): 665–677.

Scheffler, I. (1965). *Conditions of knowledge. An introduction to epistemology and education*. Glenview: Scott, Foresman and Company.

Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology* 82: 498–504

Schommer-Aikins, M. (2002). An evolving theoretical framework for an epistemological belief system. In B. K. Hofer & P. Pintrich (eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing*. New Jersey: Lawrence Erlbaum Associates, pp. 103–118.

Schommer-Aikins, M., & Easter, M. (2006). Ways of Knowing and Epistemological Beliefs: Combined effect on academic performance. *Educational Psychology* 26: 411–423.

Scouller, K. (1998). The influence of assessment method on students' learning approaches: multiple choice question examination versus assignment essay. *Higher Education* 35: 453–472.

Shavelson, R. J. (2010). *Measuring college learning responsibly: Accountability in a new era*. Stanford, CA: Stanford University Press.

Shope, R. K. (2004). The Analysis of Knowing. In I. Niiniluoto, M. Sintonen, & J. Woleński (eds.), *Handbook of Epistemology*. Dordrecht: Kluwer Academic, pp. 3–56.

Siegel, H. (1987). *Relativism Refuted: A Critique of Contemporary Epistemological Relativism*. Dordrecht: D. Reidel.

Siegel, H. (1988). *Educating Reason: Rationality, Critical Thinking, and Education*. NY: Routledge.

Siegel, H. (2014). What's in a name?: epistemology, "epistemology," and science education. *Science Education* 98(3): 372–374.

Siegel, H. & Biro, J. (1997). Epistemic Normativity, Argumentation, and Fallacies. *Argumentation* 11: 277–292.

Smith, G. (2002). Are There Domain-Specific Thinking Skills? *Journal of Philosophy of Education* 36: 207–227.

Snow, R. E. (1993) Construct validity and constructed-response tests. In R. R. Bennet & W. C. Ward (eds.), *Construction versus choice in cognitive measurement: issues in constructed response, performance testing, and portfolio assessment*. Hillsdale: Lawrence Erlbaum Associates, pp. 45–73.

Southerland, S. A., Sinatra, G. M., & Matthews, M. R. (2001). Belief, knowledge, and science education. *Educational Psychology Review* 13: 325–351.

Steneck, N. H. (2007). *The ORI Introduction to the Responsible Conduct of Research*. Washington, D.C.: Department of Health and Human Services, Office of the Secretary, Office of Public Health and Science, Office of Research Integrity.

Stes, A., Min-Leliveld, M., Gijbels D., & Van Petegem, P. (2010). The impact of instructional development in higher education: The state-of-the-art of the research. *Educational Research Review* 5: 25–49.

Timmermans, S., & Tavory, I. (2012). Theory construction in qualitative research: from grounded theory to abductive analysis. *Sociological Theory* 30: 167–186.

Tremblay, K., Lalancette, D., & Roseveare, D. (2012) AHELO Feasibility Study Report. Volume 1 –Design and Implementation. Organization for Economic Co-operation and development (OECD). Retrieved 24.2.2015 from http://www.oecd.org/edu/skills-beyond-school/AHELOFSReportVolume1.pdf.

Tynjälä, P. (2008). Perspectives into learning at the workplace. *Educational Research Review* 3: 130–154.

Tynjälä, P., Slotte, V., Nieminen, J., Lonka, K., & Olkinuora E. (2007). From University to working life: Graduates workplace skills in practice. In P. Tynjälä, J. Välimaa & G. Boulton-Lewis (eds.), *Higher Education and working Life. Collaborations, confrontations and challenges*. Oxford: Elsevier, pp. 73–88.

Valanides, N. & Angeli, C. (2005). Effects of instruction on changes in epistemological beliefs. *Contemporary Educational Psychology* 30: 314–330.

VanTassel-Baska, J. (2014). Performance-Based Assessment: The road to authentic learning for the gifted. *Gifted Child Today* 37: 41–47.

Walker, J. (2010). Measuring plagiarism: Researching what students do, not what they say they do. *Studies in Higher Education* 35(1): 41–59.

Walton, D. N. (1990). What is reasoning? What is an argument? *The Journal of Philosophy* 87(8): 388–419.

Walton, D. N. (1995). *A Pragmatic Theory of Fallacy*. Tuscaloosa, AL: University of Alabama Press.

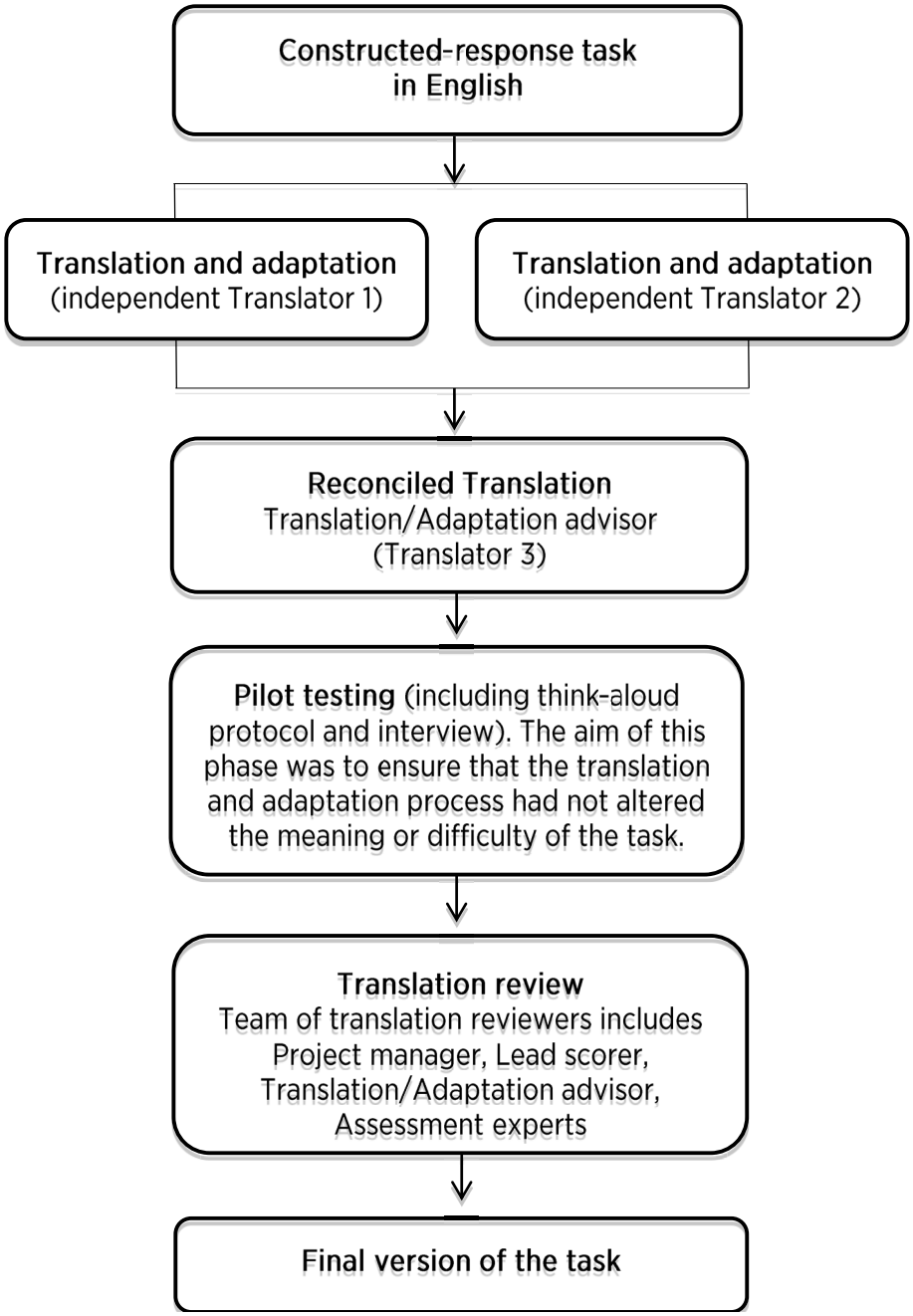
West, E. J. (2004). Perry's Legacy: Models of Epistemological Development. *Journal of Adult Development* 11: 61–70.

Winch, C. (2002). Strong autonomy and education. *Educational Theory* 52 (1): 27–41.

Winch, C. (2006). *Education, autonomy, and critical thinking*. London: Routledge.

APPENDICES

Appendix A. Translation and adaptation procedure



Appendix B. Scoring criteria for CLA performance task

		Scores						
Scale	**Description**	**0**	**1**	**2**	**3**	**4**	**5**	**6**
Analytic Reasoning and Evaluation (ARE) *How well does the student analyse, evaluate and synthesise data and information?*	Interpreting, analysing, and evaluating the quality of information. This entails identifying information that is relevant to a problem, highlighting connected and conflicting information, detecting flaws in logic and questionable assumptions, and explaining why information is credible, unreliable, or limited.	▪ Not scorable. Student made no attempt to answer the questions, so the response cannot be evaluated.	▪ Does not identify facts or ideas that support or refute arguments presented in the Document Library (salient features of objects to be classified) or provides no evidence of analysis. ▪ Disregards or severely misinterprets important information. ▪ Does not make claims about the quality of information and bases response on unreliable information.	▪ Identifies very few facts or ideas that support or refute arguments presented in the Document Library (salient features of objects to be classified). ▪ Disregards or misinterprets much of the Document Library. May restate information "as is." ▪ Does not make claims about the quality of information and presents some unreliable information as credible.	▪ Identifies a few facts or ideas that support or refute several arguments presented in the Document Library (salient features of several objects to be classified). ▪ Disregards important information or makes minor misinterpretations of information. May restate information "as is." ▪ Rarely, if ever, makes claims about the quality of information and may present some unreliable information as credible.	▪ Identifies a few facts or ideas that support or refute all major arguments presented in the Document Library (salient features of all objects to be classified). ▪ Briefly demonstrates accurate understanding of important Document Library content, but disregards some information. ▪ Makes very few accurate claims about the quality of information.	▪ Identifies several facts or ideas that support or refute all major arguments presented in the Document Library (salient features of all objects to be classified). ▪ Demonstrates accurate understanding of much of the Document Library content. ▪ Makes a few accurate claims about the quality of information.	▪ Identifies most facts or ideas that support or refute all major arguments presented in the Document Library (salient features of all objects to be classified). Provides analysis that goes beyond the obvious. ▪ Demonstrates accurate understanding of a large body of information from the Document Library. ▪ Makes several accurate claims about the quality of information.

Problem Solving (PS) *How well does the student form a conclusion from his or her analysis?* *How well does the student consider other options?*	Considering and weighing information from discrete sources to make decisions (draw a conclusion and/or propose a course of action) that logically follow from valid arguments, evidence, and examples. Considering the implications of decisions and suggesting additional research when appropriate.	▪ Also 0 if and only if Analytical Reasoning and Evaluation is 0.	▪ Provides no clear decision or valid rationale for the decision. <u>When applicable:</u> ▪ Does not propose a course of action that follows logically from the conclusion. ▪ Does not recognise the need for additional research or does not suggest research that would address unanswered questions.	▪ Provides or implies a decision, but very little rationale is provided or such rationale is heavily based on unreliable evidence. <u>When applicable:</u> ▪ Briefly proposes a course of action, but some aspects do not follow logically from the conclusion. ▪ May recognise the need for additional research. Any suggested research is vague or would not adequately address unanswered questions.	▪ Provides or implies a decision and some reason(s) to favour it, but the rationale may be contradicted by unaccounted for evidence. <u>When applicable:</u> ▪ Briefly proposes a course of action, but some aspects may not follow logically from the conclusion. ▪ May recognise the need for additional research. Any suggested research tends to be vague or would not adequately address the unanswered questions.	▪ Provides a decision and credible evidence to back it up. Possibly does not account for credible, contradictory evidence. May attempt to discount alternatives. <u>When applicable:</u> ▪ Proposes a course of action that follows logically from the conclusion. May briefly consider implications. ▪ Recognises the need for additional research. Suggests research that would address an unanswered question.	▪ Provides a decision and a solid rationale based largely on credible evidence from multiple sources and discounts alternatives. <u>When applicable:</u> ▪ Proposes a course of action that follows logically from the conclusion. May consider implications. ▪ Recognises the need for additional research. Suggests research that would address some unanswered questions.	▪ Provides a decision and a solid rationale based on credible evidence from a variety of sources. Weighs other options, but presents the decision as the best option given the available evidence. <u>When applicable:</u> ▪ Proposes a course of action that follows logically from the conclusion. Considers implications. ▪ Recognises the need for additional research. Recommends specific research that would address most of the unanswered questions.

		Scores						
Scale	**Description**	**0**	**1**	**2**	**3**	**4**	**5**	**6**
Writing Effectiveness *Does the student clearly articulate the argument?* *Does the student logically organise the argument?* *Does the student present evidence in an order that contributes to a persuasive and coherent argument?*	Constructing organised and logically cohesive arguments. Strengthening the writer's position by providing elaboration on facts or ideas (e.g., explaining how evidence bears on the problem, providing examples, and emphasising especially convincing evidence).	Also 0 if and only if Analytical Reasoning and Evaluation is 0.	▪ Does not develop convincing arguments. Writing may be disorganized and confusing. ▪ Does not elaborate on facts or ideas.	▪ Provides limited, invalid, overstated, or very unclear arguments. May present information in a disorganized fashion or undermine own points. ▪ Any elaboration on facts or ideas tends to be vague, irrelevant, inaccurate or unreliable (e.g., based entirely on the writer's opinion). Sources of information are often unclear.	▪ Provides limited or somewhat unclear arguments. Presents relevant information in each response, but that information is not woven into the arguments. ▪ Provides elaboration on facts or ideas a few times, some of which is valid. Sources of information are sometimes unclear.	▪ Organises response in a way that makes the writer's arguments and the logic of those arguments apparent, but not obvious. ▪ Provides valid elaboration on facts or ideas several times and cites sources of information.	▪ Organises response in a logically cohesive way that makes it fairly easy to follow the writer's arguments. ▪ Provides valid elaboration on facts or ideas related to each argument and cites sources of information.	▪ Organises response in a logically cohesive way that makes it very easy to follow the writer's arguments. ▪ Provides valid and comprehensive elaboration on facts or ideas related to each argument and clearly cites sources of information.

Sources:

Tremblay, K., Lalancette, D., & Roseveare, D. (2012) AHELO Feasibility Study Report. Volume 1.Design and Implementation. Organization for Economic Co-operation and Development (OECD). Retrieved 24.2.2015 from http://www.oecd.org/edu/skills-beyond-school/AHELOFSReportVolume1.pdf.

Shavelson, J. (2010) Measuring college learning responsibly. Accountability in a new era. Stanford, CA: Stanford University Press, p. 52–53.