Development and Evaluation of a Teaching Unit in Particle Physics to Promote Students' Critical Thinking

DISSERTATION

to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

Submitted to
School of Science of Technische Universität Dresden

by
Farahnaz Sadidi

This project (100284305) work was supported as a scholarship from the European Social Fund (ESF) and Free State of Saxony.

Submitted on:	12/12/2022
First referee:	Prof. Dr. Gesche Pospiech
	Technische Universität Dresden
Second referee:	Prof. Dr. Renee´ Schwartz
	Georgia State University

Abstract

Critical thinking (CT) is one of the desirable skills to be taught in school. It is not only considered an important 21st century skill for living in a democratic society, but also important for a deep understanding of domain-specific content. Despite its importance, studies show that students often lack the ability to think critically. Moreover, there is a lack of clear theory, supported by empirical findings, for developing domain-specific teaching-learning sequences to promote students' CT. This makes teaching CT challenging for teachers.

To address this gap, the presented study has two goals: to identify design principles for instruction that promotes critical thinking and to develop an exemplary instructional unit in particle physics on this basis. Particle physics is chosen because of its abstractness and complexity, as well as student interest in the subject. Another basis is a definition of CT that can be readily applied in the context of teaching physics. For this purpose, Halpern's classification of CT strategies and their measurable outcomes is used. Furthermore, a distinction is made between general CT skills that provide a framework for CT, such as understanding the need to define terms precisely, and domain-specific CT skills that represent the application of general CT skills in a specific domain and require domain-specific expertise, such as distinguishing between the concepts of mass and matter in the context of particle physics. This study examines the development of both general and domain-specific CT.

The teaching-learning sequences about antimatter (10 to 12 lessons) are developed for students in grades 10, 11, and 12 using the Design-Based Research (DBR) approach. Analysis of the data from pilot studies provides guidance for further development of the antimatter course and the creation of a teacher package that supports teachers both methodologically and in terms of content when implementing the antimatter course. In the main study, the course is implemented in 3 classes in different federal states of Germany. To evaluate the effectiveness of the course in promoting students' CT, the perspectives of students as well as of teachers are examined. To evaluate the effectiveness of the course from the students' perspective, the video and audio data, the students' works, students' interviews or questionnaires are inductively analyzed using the constant comparative method to identify the students' learning processes. The results show that students apply content knowledge, apply CT skills, and demonstrate a disposition toward CT. This corresponds to a developed CT. Further analysis is conducted to relate the design skeleton facets of the course (materials, activity structure, and participant structure) to the learning processes, using the conjecture map framework to support the results from the constant comparative method. A Particle Physics Critical Thinking (PPCT) test is also developed to triangulate the results. The results of administering the PPCT test as a posttest are consistent with the qualitative findings on the effectiveness of the course. A questionnaire is developed for teachers to elicit their perceptions of the relevance, practicality, and effectiveness of the course in promoting students' CT. The results show a positive perception.

Combining all the results shows that the antimatter course is an effective course in promoting CT. The design principles applied contribute to the theory of designing effective CT instruction. Furthermore, data analysis reveals the challenges students face in critical thinking and provides teachers with heuristics for designing a domain-specific course. Based on the findings, a model for teaching CT is developed.

This work leads to implications for teaching, in addition to other research questions. These include, for example, developing domain-specific CT instruction using 6 principles empirically tested in this study, considering heuristics for designing domain-specific CT instruction, and using the course materials for the purpose of developing CT. In addition, the PPCT can guide the development of other domain-specific CT tests.

Kurzfassung

Kritisches Denken (KD) ist eine der wünschenswerten Fähigkeiten, die in der Schule vermittelt werden sollten. Es gilt nicht nur als wichtige Kompetenz des 21. Jahrhunderts für das Leben in einer demokratischen Gesellschaft, sondern auch als wichtig für ein tiefes Verständnis von fachspezifischen Inhalten. Trotz dieser Bedeutung zeigen Studien, dass es den Lernenden oft an der Fähigkeit fehlt, kritisch zu denken. Zudem fehlt es an einer klaren, durch empirische Befunde gestützten Theorie für die Entwicklung von fachspezifischen Lehr-Lern-Sequenzen zur Förderung der KD-Fähigkeiten von SchülerInnen. Dies macht den KD-Unterricht zu einer Herausforderung für Lehrkräfte.

Um diese Lücke zu schließen, verfolgt die vorgelegte Studie zwei Ziele: Die Identifikation von Gestaltungsprinzipien für einen Unterricht, der die Fähigkeit zum kritischen Denken fördert, und die Entwicklung einer exemplarischen Unterrichtseinheit in Teilchenphysik auf dieser Grundlage. Die Teilchenphysik wurde aufgrund ihrer Abstraktheit und Komplexität sowie des Interesses der Schüler ausgewählt. Eine weitere Grundlage ist eine Definition von KD, die sich gut im Rahmen des Physikunterrichts anwenden lässt. Hierzu wurde Halperns Klassifizierung von KD-Strategien und ihre messbaren Ergebnisse verwendet. Darüber hinaus wird unterschieden zwischen allgemeinen KD-Fähigkeiten, die einen Rahmen für KD bilden, wie z. B. das Verständnis für die Notwendigkeit, Begriffe genau zu definieren, und domänenspezifischen KD-Fähigkeiten, die die Anwendung allgemeiner KD-Fähigkeiten in einer bestimmten Domäne darstellen und domänenspezifisches Fachwissen erfordern, wie z. B. die Unterscheidung zwischen den Konzepten von Masse und Materie im Kontext der Teilchenphysik. Diese Studie untersucht die Entwicklung sowohl der allgemeinen als auch der domänenspezifischen KD.

Die Lehr-Lern-Sequenzen über Antimaterie (10 bis 12 Unterrichtsstunde) werden für SchülerInnen der Klassenstufen 10, 11 und 12 mit Hilfe des Design-Based Research (DBR) Ansatzes entwickelt. Die Analyse der Daten aus den Pilotstudien liefert Anhaltspunkte für die Weiterentwicklung des Antimateriekurses und die Entwicklung eines Lehrerpakets, das Lehrkräfte methodisch und inhaltlich bei der Umsetzung des Antimateriekurses unterstützt. In der Hauptstudie wird der Kurs in 3 Klassen in verschiedenen Bundesländern Deutschlands durchgeführt. Um die Wirksamkeit des Antimateriekurses bei der Förderung des KD der SchülerInnen zu evaluieren, werden sowohl die Perspektiven der SchülerInnen als auch die der LehrerInnen untersucht. Um die Wirksamkeit des Kurses aus der Perspektive der SchülerInnen zu evaluieren, werden

die Video- und Audiodaten, die Schülerarbeiten, das Schülerinterview und der Fragebogen induktiv mit der Constant Comparative Methode analysiert, um die Lernprozesse der SchülerInnen zu identifizieren. Die Ergebnisse zeigen, dass die SchülerInnen inhaltliches Wissen und KD-Fähigkeiten anwenden und eine Disposition zeigen, die gemeinsam einer entwickelten KD entsprechen. Zusätzlich werden mit Hilfe von sog. „Conjecture Maps" die Gestaltungsfacetten des Kurses (Materialien, Aktivitätsstruktur und Teilnehmerstruktur) mit den Lernprozessen in Beziehung gesetzt, um die Ergebnisse aus der Constant Comparative Methode zu stützen. Ein Particle Physics Critical Thinking (PPCT) Test wurde ebenfalls entwickelt, um die Ergebnisse zu triangulieren. Die Ergebnisse der Durchführung des PPCT-Tests als Posttest stimmen mit den qualitativen Erkenntnissen über die Wirksamkeit des Kurses überein. Ferner wurde ein Fragebogen für Lehrkräfte entwickelt, um ihre Einschätzung der Relevanz, Praktikabilität und Wirksamkeit des Kurses bei der Förderung des KD der SchülerInnen zu erheben. Dieser zeigte eine positive Wahrnehmung.

Die Kombination aller Ergebnisse zeigt, dass der Antimateriekurs ein effektiver Kurs zur Förderung des KD ist. Die angewandten Gestaltungsprinzipien tragen zur Theorie der Gestaltung eines wirksamen KD-Unterrichts bei. Darüber hinaus zeigt die Datenanalyse die Herausforderungen auf, denen sich die SchülerInnen beim kritischen Denken gegenübersehen, und liefert den Lehrkräften Heuristiken für die Gestaltung eines domänenspezifischen Kurses. Auf der Grundlage der Ergebnisse wird ein Modell für den KD-Unterricht entwickelt.

Diese Arbeit führt neben weiteren Forschungsfragen auch zu Implikationen für den Unterricht. Dazu gehören z. B. die Entwicklung eines domainspezifischen KD-Unterrichts unter Verwendung von 6 Prinzipien, die in dieser Studie empirisch getestet wurden, die Berücksichtigung von Heuristiken für die Gestaltung eines domainspezifischen KD-Unterrichts, und die Verwendung der Kursmaterialien zum Zweck der Entwicklung von KD. Darüber hinaus kann der PPCT Test die Entwicklung anderer domainspezifischer KD-Tests anleiten.

Table of Contents

1 Introduction

This study focuses on the development and evaluation of a teaching unit in particle physics to promote students' critical thinking. Since critical thinking (CT) is the focus of this work, the importance of CT and its teaching are first discussed. Then, the specific research goals and the structure of the thesis are presented.

1.1 Importance of critical thinking and its teaching

Human beings are part of society, from small societies like family, friends, neighbors, to large ones like society and the world. Living in a society entails responsibilities for people. The decisions they make as individuals, e.g. waste separation, and as a society, e.g. elections, can affect the lives of future generations of all people around the world. The decision people make is influenced by the quality of their thinking. Critical thinking (CT) as thinking that is evaluative and reflective in nature (*cf.* Halpern, 2009; Lipman, 2003; Norris & Ennis, 1989), is promised to create a world of peace with adequate resources (National Education Goals Panel, 1991, as cited in Halpern, 2009, p. 2). It is considered as one of the 21st century skills (Halpern, 2009; Saleh, 2019; Wilson et al., 2000), necessary for living in a democratic society (Andrews, 2015; Lipman, 2003) and for an increasingly complex living and working environment (P21 Partnership for 21st Century Learning, 2019).

CT is defined as the ability to comprehend and defend against the persuasive techniques embedded in everyday language (Ennis, 1989, 2015; Halpern, 2009), which Lipman (2003) refers to as brainwashing (p. 209), the ability to devise hypotheses about the relationship among events and collect observation to test its validity (Ennis, 1989, 2015; Halpern, 2009; Lipman, 2003), and the ability to make decision about uncertain events and select the best alternative (Aristotle, as cited in Lipman, 2003, p. 208; Halpern, 2009).

The movement toward teaching critical thinking began in the late 19th century and peaked in the late 1980s and early 1990s through conferences such as Critical Thinking and Education in the United States (for a historical overview, see Lipman, 2003, pp. 28-46). It was discussed that students have the right to learn the skills that help them recognize propaganda and not fall prey to it, realize when there is intentional deception, examine the credibility of information sources, and think a problem or a decision through in the best way possible to interact with everyday life (Andrews, 2015; Barnett, 2015, pp. 64-65; Halpern, 2009; Hamby, 2015; Viennot, 2001). Moreover, the role of CT in understanding content knowledge was discussed (McPeck, 1990, p. 21; Viennot & Decamp, 2020, p. 101). Thus, CT was considered important as a daily life skill and also for deepening the understanding of a particular discipline.

Each of these functionalities of CT leads to a discussion of the nature of CT, whether general or domain-specific. General CT refers to the skills that are applicable and transferable to any subject in any discipline, including daily life (Norris & Ennis, 1989, p. 21; Wendland et al., 2015, p. 158), whereas Domain-specific CT consider the skills applicable to a specific subject or discipline, e.g. physics, (Norris & Ennis, 1989, p. 21), which require content-specific expertise (McPeck, 1990).

In physics education, it is discussed that both general and domain-specific CT skills need to be addressed (*cf.* Etkina & Planinšič, 2015; Redish & Hammer, 2009; Tiruneh et al., 2016; Viennot & Decamp, 2018). To achieve both goals of deepening students' understanding of school subjects and also enabling them to transfer CT skills to real-world contexts in everyday life, embedding CT skills in domain-specific instruction, i.e., designing domain-specific CT lessons, is promising (*cf.* Tiruneh et al., 2016).

1.2 Research goals

Despite the long-standing emphasis on teaching CT, studies show that students lack the ability to think critically (Davies, 2013; Garcia-Mila & Andersen, 2007; Larkin & Reif, 1979; Ogunleye, 2009; Reddy & Panacharoensawad, 2017; Ryu & Sandoval, 2012;

Shaheen, 2016; Temel, 2014; Viennot, 2001; Viennot & Decamp, 2020; Zohar, 2007). On the other hand, empirical research on the nature of cognitive development suggests that students of all intellectual ability levels can benefit from the teaching of CT (Idol & Jones, 1991; Kennedy et al., 1991, p. 18; Lai, 2011, p. 23; Lewis & Smith, 1993). Therefore, the reason for the lack of CT among students might be that there is no well-designed instruction that promotes CT (McMillan, 1987; Paul, 1993; Resnick, 1987).

While there are some proposals for designing effective CT instruction, the lack of empirical studies comparing the effectiveness of the proposed approaches (Abrami et al., 2008; Tiruneh et al., 2014) makes designing CT instruction challenging. In addition, the design of a CT instruction in a specific domain suffers from the lack of empirical studies showing that embedding CT skills in a domain-specific instruction improves both general and domain-specific CT skills (*cf.* Tiruneh et al., 2014). To this end, Tiruneh et al. (2016) and Sermeus et al. (2021) have recently made efforts to design effective CT instruction in the context of electricity and magnetism at the university and high school levels, respectively. They showed that embedding CT skills in electricity and magnetism courses can promote students' domain-specific CT, not general CT (Sermeus et al., 2021; Tiruneh et al., 2016).

In the framework of this PhD project, an instruction in the context of particle physics was developed to promote students' general and domain-specific CT. The design principles were inspired by the work of Tiruneh et al. (2016), but also included a model for teaching CT (Halpern, 1998). The lessons focused on teaching antimatter and included 10 to 12 lessons for students in grades 10, 11, and 12. The Design-Based Research approach was used to optimize the antimatter course in an iterative cycle of design-enactment-analysis-redesign (Design Based Research Collective, 2003; McKenney & Reeves, 2019). Its effectiveness in promoting students' general and domain-specific CT was evaluated through a qualitative analysis of the data as well as an investigation of students' and teachers' perspectives about the course. Furthermore, a model for developing students' CT was developed that can serve as a guide for teaching CT. In addition, a Particle Physics Critical thinking (PPCT) test was developed to evaluate students' acquisition of domain-specific CT.

Thus, this work provides an empirical basis for designing domain-specific CT instruction in physics education by presenting design principles and a framework for teaching CT.

1.3 Structure of the work

This thesis is divided into a theoretical and an empirical part. Its structure is summarized in figure 1.1.

To provide a solid foundation for the study, critical thinking (CT) is first defined. Different definitions of CT are analyzed (see section 2.1 and 2.2) to identify the commonalities between them in terms of CT skills applicable in physics education. Then, the nature of CT, whether general or domain-specific, is discussed in relation to the teaching of CT in physics (see section 2.3). In addition, the state of students' critical thinking skills and teachers' perspectives on teaching CT are discussed (see section 2.4 and 2.5).

Chapter 3 first discusses the challenges for teaching CT, then discusses the design principles of a CT instruction based on a review of the literature on teaching CT in physics (see section 3.2). Then, the evaluation methods and criteria for assessing the quality of teaching are discussed at two levels: the design of teaching and its implementation (see section 3.3).

Chapter 4 discusses the methodological framework of the study, Design-based research (DBR). Here, the characteristics of DBR, the steps for conducting DBR in educational studies to develop and refine instruction, relevant research instruments, and evaluation methods are explained.

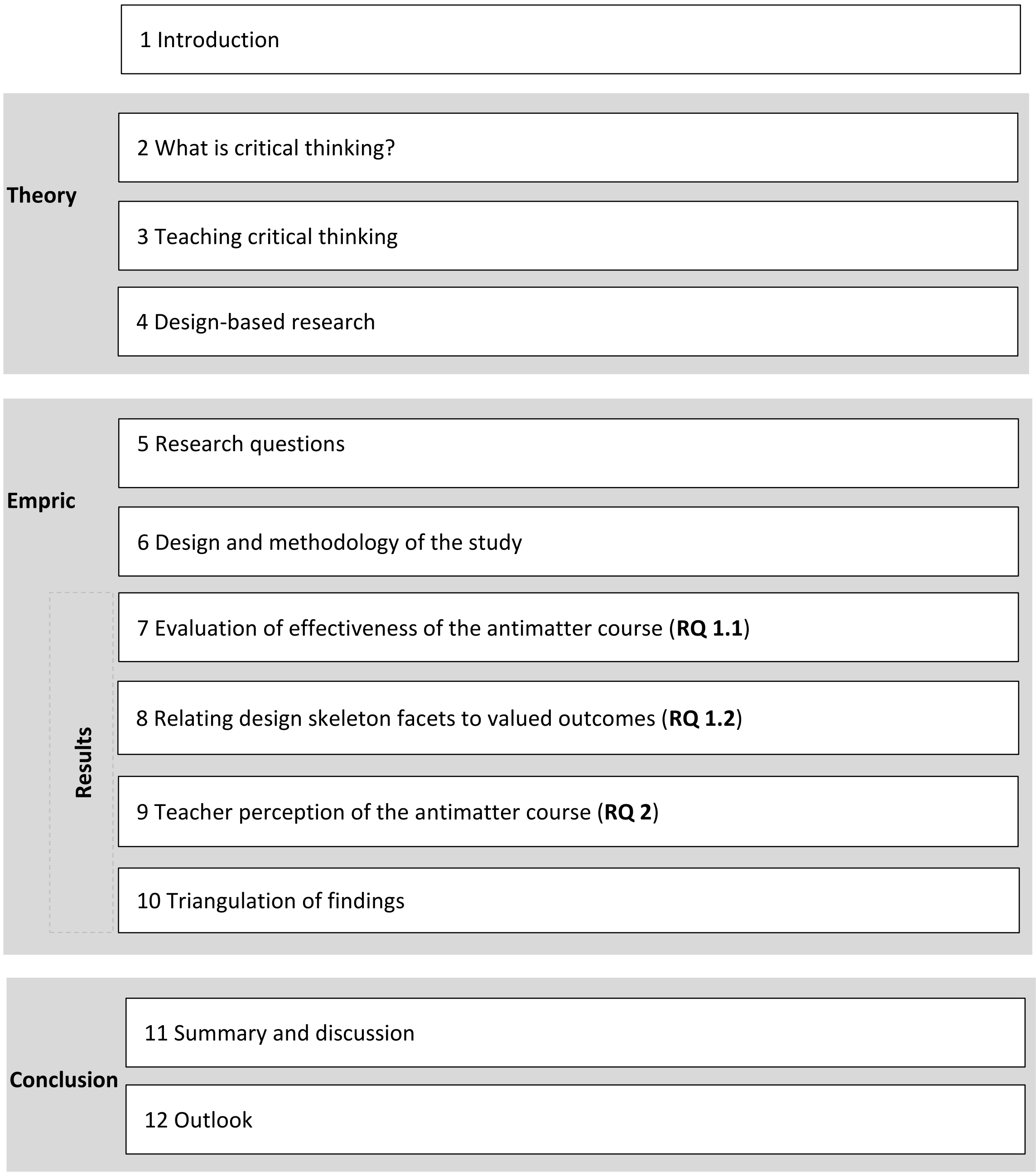


Figure 1.1: Structure of the work (RQ: Research Question).

The empirical part of the study begins with the design and methodology of the study. First, the design of the antimatter course using the DBR approach is discussed (see section 6.1). To this end, the target audience, the learning objectives, and the design principles are presented. Then, the research tools used or developed in the study to answer the research questions are presented (see section 6.2, 6.3, and 6.4).

The results of the study are presented in Chapters 7, 8, and 9. The effectiveness of the antimatter course in promoting students' general and domain-specific CT skills using qualitative analysis is demonstrated. In addition, a model of student CT development is provided (see section 7.5) that describes the processes students go through to achieve general and domain-specific CT. In addition, student and teacher views on the effectiveness of the antimatter course are discussed. Furthermore, Chapter 10 discusses the results of the quantitative analysis to show the extent to which the qualitative analysis can reflect the results of the qualitative analysis.

The conclusion part of the study summarizes and discusses the results (see section 11.1), discusses the contribution of the study to the theory (see section 11.2), and the limitations of the study (11.3). Finally, the implications of the study's findings for CT teaching and future researchers are discussed (see section 12.1 and 12.2).

Part I

Theory

2 What is critical thinking?

To improve students' critical thinking in the schools and colleges, a clear conception of critical thinking (CT) is needed. It is necessary to know "its defining features, its characteristic outcomes, and the underlying condition that make it possible" (Lipman, 2003, p. 209). The analysis of CT based on these three aspects helps to relate CT to teaching which is discussed in the next chapter.

In this chapter first an analysis of different definitions of CT is presented to give readers a clear picture of what CT is. Then, an attempt is made to find commonalities between the definitions in order to identify the CT skills applicable in this work. Next the nature of CT is discussed, whether it is a general or a domain-specific ability. Finally, the students' challenges to think critically on one hand and the teachers' perspective on CT and toward teaching CT on the other hand are addressed.

2.1 Definition of critical thinking

The way people think influences the way they live. To improve the quality of life, it is therefore necessary to improve the quality of thinking. Critical thinking is seen as an appropriate means to achieve this goal (Halpern, 2014, p. 3; Lipman, 2003, p. 208).

Thinking in general is the process of analysing data or information collected from different sources to take an action in service of the goal. This is a conscious mental process that includes making meaning out of information to build knowledge, comparing newly acquired information with existing knowledge and experience, and taking action toward achieving the desired state of affairs (Halpern, 2014, p. 7; Holyoak & Morrison, 2005, p. 2; Lipman, 2003, p. 74; Moseley et al., 2005, p. 12). This mental process can be transformed into action by raising questions, utilizing concepts, making inferences, making assumptions, and generating implications (Elder & Paul, 2012, p. 5; Moseley et al., 2005, p. 12).

There are various forms of thinking (Andrews, 2015, p. 60). Some categorize these various forms of thinking to critical, creative, and caring thinking (see Lipman, 2003), and others to reflective and productive thinking (see Moseley et al., 2005). Since the main goal of this PhD is developing students' CT skill, this chapter focuses on critical thinking.

The term critical thinking (CT) has developed over time and is defined from philosophical and psychological perspectives (see Moseley et al., 2005; see Pithers & Soden, 2000). Philosophers were engaged in the theoretical definition of CT and its nature, while psychologists were engaged in empirical research to investigate the nature of human mental abilities (Norris, 1986).

Following some definitions are discussed and are synthesized based on their "defining features, characteristic outcomes, and the underlying condition that make them possible" (Lipman, 2003, p. 209).

Robert Ennis

Ennis (1985) defines CT as "reasonable and reflective thinking that is focused on deciding what to believe or do" (p. 45). Based on this definition critical thinking is a consciously directed thinking that relies on criteria to use good reasons for evaluating statements (beliefs) and actions (Norris & Ennis, 1989, pp. 3-4). To give good reasons, it is necessary to consider the criterion "clarity" within the thinking process. To fulfil this criterion, all terms used for formulating a reason must be defined, and the implicit assumptions must be identified (Norris & Ennis, 1989, p. 11).

Furthermore, the reflective nature of critical thinking addresses disposition as an important factor throughout the whole critical Thinking process. That means a critical thinker has the attitude and tendency to think critically and consciously seeks and uses good reasons (Norris & Ennis, 1989, p. 4).

Summarizing his view, Ennis considers CT as a mental interplay between the cognitive (reasonable) and affective (reflective) levels of thinking. He sees CT as a combination of the ability to use knowledge to give reasons, and the disposition to seek clarity to give good reasons.

He sees CT as a process taking place within a problem-solving context (Norris & Ennis, 1989, p.8) to deal with some content, that might be part of a school subject or a situation or problem in everyday life (Norris & Ennis, 1989, p. 5). A broader view on Ennis' list of CT abilities involved in solving problems emphasizes on the abilities such as argument analysis, verbal reasoning (defining terms and judging definitions), and making and testing hypotheses (making and judging deductions and inductions) skills (Ennis, 2015, pp. 37-44; Moseley et al., 2005, pp. 154-155; Norris & Ennis, 1989, p.14, see also elaborated version of CT abilities pp. 183-187).

Table 2.1 shows a synthesis of Ennis' definition of CT based on "its defining features, its characteristic outcomes, and the underlying condition that make it possible" (Lipman, 2003, p. 209).

Table 2.1: Synthesis of Ennis' definition of critical thinking.

Ennis' central idea: Critical thinking is reasonable and reflective thinking that is focused on deciding what to believe or do.		
Defining features	**Characteristic outcomes**	**Underlying conditions**
It includes both disposition and abilities.	Evaluating statements and actions	Critical thinking • relies on criteria to evaluate outcomes. • takes place within a problem-solving context.

John E. McPeck

McPeck (1990) defines CT as a knowledge-based skill that enables one to reflect on the knowledge required to solve the problem at hand, to question that knowledge effectively, and to suspend judgment or belief if the knowledge is not complete (p. 28).

Based on this definition, CT is the ability of applying the relevant knowledge and skills to be engaged in a reflective activity (p. 21).

McPeck's (1990) definition includes a knowledge component when he defines CT as a knowledge-based skill and a critical component, which is the disposition to reflect on that knowledge (p. 21). Considering the knowledge component of CT, he argues that CT is always content-specific because thinking critically about something is thinking about that specific subject content (McPeck, 1990; Moseley et al., 2005, pp. 21-22). He values the importance of having complete knowledge of a particular content by having a higher degree of certainty and reliability of the data and being sensitive to how it is collected or generated (McPeck, 1990, p. 28).

To summarize his view, McPeck (1990) considers CT as a mental interplay between cognitive (content-knowledge and skills) and affective (disposition) levels of thinking. He emphasizes explicitly on the relationship between knowledge and critical thinking ability and says "critical thinking ability varies directly by the amount of knowledge required by the problem" (p. 28). Table 2.2 shows a synthesis of CT definition offered by McPeck.

Table 2.2: Synthesis of McPeck's definition of critical thinking.

McPeck's central idea: Critical thinking is the appropriate use of reflective scepticism within the problem area under consideration (1981).		
Defining features	**Characteristic outcomes**	**Underlying conditions**
It includes disposition, knowledge, and skills.	Reflective scepticism on the knowledge	Critical thinking • is content specific. • is sensitive to the data and data collection methods.

Mary A. Miller

Miller (1992) adopts Watson and Glaser's (1964) definition of CT in nursing education and defines critical thinking as the ability to define a problem, choose the best solution among alternative solutions, implement the solution, and finally reflect to evaluate the effectiveness of the solution (p. 1401). Miller sees CT as an active process that goes beyond thinking and implementing by reflecting on the results. In his view, reflection involves making a judgment about the decision by evaluating the results of its implementation. He considers CT as an interplay between the mental processes of thinking, acting and reflecting. Table 2.3 shows a synthesis of CT definition offered by Miller.

Table 2.3: Synthesis of Miller's definition of critical thinking.

Miller's central idea (adopted from Watson & Glaser, 1964)**:** Critical thinking is the composite of **attitudes** of inquiry to recognize existence of problems and the need for supportive evidence toward solutions, **knowledge** of the nature of valid inferences, abstractions, and generalisations to determine the weight and accuracy of evidence logically, and **skills** to apply the attitudes and knowledge.

Defining features	**Characteristic outcomes**	**Underlying conditions**
It includes disposition, knowledge, and skills.	Evaluating weight and accuracy of evidence toward solution	Critical thinking • is logical. • takes place within a problem-solving context. • includes reflection.

Facione and Facione

Facione and Facione (1996) analysed the 1990 American Philosophical Association (APA) consensus definition of critical Thinking proposed by 46 renowned experts and realized a clear reference to both cognitive skills and disposition. The CT cognitive skills and disposition mutually reinforce each other to make clear interpretations, to analyze the argument, to identify assumptions, and to test hypothesis (Facione & Facione, 1996, p. 4) in the complex context. Table 2.4 shows a synthesis of CT definition offered by Facione and Facione.

Table 2.4: Synthesis of Facione and Facione's definition of critical thinking.

Facione and Facione's central idea (adopted from the 1990 American Philosophical Association consensus definition of critical thinking)**:** Critical thinking is a purposeful and self-regulated judgment based on interpretation, analysis, evaluation, inference, and explanation of evidence. Moreover, it involves disposition toward thinking critically.		
Defining features	**Characteristic outcomes**	**Underlying conditions**
It includes cognitive skills and disposition.	Purposeful and self-regulated judgment	Critical thinking • is goal-directed. • takes place within a problem-solving context.

Matthew Lipman

Lipman (2003) defines critical thinking as "thinking that facilitates judgment because it relies on criteria, is self-correcting, and is sensitive to context" (p. 212). He considers judgment as the outcome of critical thinking. He emphasizes on good judgment which is made on the base of relevant information, is guided by appropriate data collection method and appropriate data analysis method, and is reflective on itself (on judgment). Moreover, he argues that CT is reliable thinking when it relies on valid, consistent, and relevant reasons which support the judgment (pp. 212-214). Furthermore, it is self-corrective which means critical thinker realizes its own weakness and also the faults in its own procedure (p. 218). Moreover, he emphasizes that CT is sensitive to the context which means based on the context the degree a person might present or perform the criteria could change, for example the test of truth is generally independent of the character of the speaker, but in a court trial, the character of a witness may become a relevant consideration (p. 219).

Lipman considers both cognitive and affective aspects of thinking. He accounts four varieties of cognitive skills: *inquiry, reasoning, concept formation, and translation*, and claims that these skills are the most relevant skills for educational purposes (see Lipman, 2003, pp. 178- 186). Moreover, he stresses the interplay between cognitive skills and disposition to develop CT (Lipman, 2003, pp. 166- 172). Table 2.5 shows a synthesis of CT definition offered by Lipman.

Table 2.5: Synthesis of Lipman's definition of critical thinking.

Lipman's central idea: Critical thinking is thinking that facilitates judgment because it relies on criteria, is self-correcting, and is sensitive to context.		
Defining features	**Characteristic outcomes**	**Underlying conditions**
It includes cognitive skills and disposition.	Relevant, valid and reflective judgment	Critical thinking • is sensitive to context. • relies on criteria.

Richard Andrews

Andrews (2015) argues that CT and argumentation are closely related and teaching students argumentation can foster students' CT (p. 60). He defines argumentation as the process of "expressions of ideas, joining them together logically, and positioning in relation to existing bodies of knowledge" (Andrews, 2015, p. 53). He considers having clear thought and self-discipline not only as necessary elements of critical thinking (Davies & Barnett, 2015, p. 28) but also essential for strong argumentations. He discusses that development of argumentation skills requires a disposition to be critical, a knowledge of argumentation skills, and an awareness of content-specific nature of arguments (Andrews, 2015, pp. 57-58). He acknowledges the role of argumentation in helping students to sharpen their focus and improving the quality of their engagement in speech, writing, and composing (p. 61) and prefers to take the argumentation perspective rather than critical thinking perspective (p. 60). Table 2.6 shows a synthesis of CT definition offered by Andrews.

Table 2.6: Synthesis of Andrews' definition of critical thinking.

Andrews's central idea: Critical thinking is closely related to argumentation. Argumentation is the process of expressions of ideas, joining them together logically, and positioning in relation to existing bodies of knowledge.		
Defining features	**Characteristic outcomes**	**Underlying conditions**
It includes disposition and general and domain-specific knowledge.	Strong argumentation	Argumentation • has domain-specificity nature. • needs general knowledge of argumentation theories and models.

Diane F. Halpern

Halpern (1998, 2014) considers CT as purposeful, reasoned, and effortful thinking that involves evaluating the thinking process, e.g. reasoning to reach a conclusion, and also evaluating the outcomes of thinking process, e.g. quality of decision made (2009, p. 4; 2014, p. 8). She defines CT as a "use of cognitive skills or strategies that increase the probability of a desirable outcome" (Halpern, 1998, 2014). Halpern classifies these thinking strategies as Verbal Reasoning, Thinking as Hypothesis Testing, Argument Analysis, Likelihood and Uncertainty Analysis, and Decision Making and Problem Solving (2009, 2014). She argues, however, the desirable outcomes may vary according to personal or society values, but the idea that CT increases the likelihood of desired outcomes offers a way to define CT (2014, p. 8).

Moreover, Halpern deals with the affective aspect of thinking and discusses that to improve students' CT, the thinking strategies and CT dispositions must be valued alike by teachers and learners (Moseley et al., 2005, p. 145).

Since thinking strategies are abstract if they remain at the level of thought, they must lead to an action to be measurable. Therefore, Halpern defines some measurable skills for these thinking strategies as outcomes of their application (2009, 2014). Table 2.7 gives an overview about some desirable outcomes of applying these thinking strategies.

Besides emphasizing on the cognitive and affective aspects of thinking, Halpern introduces a metacognitive framework to guide the thinking process (Moseley et al., 2005, p. 145) and stresses that applying this framework not only develops CT (Halpern, 2009, p. 12) but also helps transfer the thinking skills to the other contexts (Halpern, 2009, p. 15). The framework consists of four questions:

1. what is the goal?
2. What is known?
3. Which thinking skills will get you to your goal?
4. Have you reached your goal?

These questions direct the attention to the thinking process and the products of the thought (Halpern, 2009, p. 12).

Halpern does not claim that her categorisation of CT skills is a complete list (Moseley et al., 2005, p. 145) but her efforts on translating theories and researches from psychology to practices in everyday life, developing materials accompanies her ideas on CT, and developing tests to evaluate CT skills in the context of everyday life makes her work recognizable.

Table 2.7: Thinking strategies and their desirable outcomes addressed by Halpern (2009, 2014).

Thinking strategy	**Definition**	**Thinking skills (Examples of desirable outcomes)**

Verbal Reasoning	The skill needed to identify and avoid persuasive or misleading information contained in a variety of contexts.	• Detect misuse of definitions (Recognize vague or/and ambiguous terms)
		• Explain reason (give reason why the term is vague or/and ambiguous)
		• Use precise terms
		• Seek criterion “precision”
		  ○ Correctness
		    ▪ how **correct** are the terms based on content knowledge?
		    ▪ are the terms **acceptable** in relation to the context presented?
		  ○ Does the term need further specification?

Argument Analysis	The skill needed to judge how well reason and evidence support a conclusion including considering counterevidence, stated and unstated assumptions, and the overall strength of the argument.	• Identifying/giving premises
		• Identifying/giving reasons
		• Identifying/giving counterargument
		• Identifying/giving conclusion
		• Seeking criteria:
		  ○ Correctness: are the premises, reasons, and conclusion correct?
		  ○ Relevance: do the reasons support the conclusion?
		  ○ Sufficiency: which and how many components are missing?
		  ○ Validity: judging the credibility of an information source

Thinking as Hypothesis Testing	The skill needed when generating hypothesis about the relationships among events and then collecting observations to test its validity. The skill has much in common with the scientific research methods in the study of phenomena.	**Generating hypotheses** • Identifying variables • Being able to describe the hypothetical relationship between variables **Designing & executing experiments** • Understanding the need to control variables in order to make a strong causal claim • Recognizing the need for and using operational definition of the terminologies used in a claim **Evaluating evidence** • Checking for adequate sample size • Testing the validity of hypotheses
Likelihood and Uncertainty Analysis	The skill needed when making decision about uncertain events.	• Using probability judgment to improve decision making • Understanding the limits of extrapolation* • Avoiding overconfidence** in uncertain situations
Decision Making and Problem Solving	The skills needed when facing with generating and evaluating alternative solutions to solve a problem.	• Restating the problem and considering different alternative solutions • Identifying the best option from a number of alternatives in solving everyday problems • Seeking for more relevant information to make better decisions

* Using trends in data to make estimation of future events.

** Tendency to be more confident than necessary when making decisions about probabilistic event

To summarize her view on CT, Halpern sees CT as an interplay between cognitive and affective thinking skills in the context of everyday life. To make the CT skills transferable from one context to the other, she introduces a cognitive framework. Applying this cognitive framework and asking the relevant questions it includes when solving the problems in everyday life develop CT skills (Halpern, 2009, pp. 15-17). Table 2.8 shows a synthesis of CT definition offered by Halpern.

Table 2.8: Synthesis of Halpern's definition of critical thinking.

Halpern's central idea: Critical thinking is a use of cognitive skills or strategies that increase the probability of a desirable outcome.		
Defining features	**Characteristic outcomes**	**Underlying conditions**
It includes cognitive, affective, and metacognitive aspects of thinking.	Measurable skills of applying 5 thinking strategies: Verbal Reasoning, Argument Analysis, Thinking as Hypothesis Testing, Likelihood and Uncertainty Analysis, and Decision Making and Problem Solving	Critical thinking • classifies by skills. • takes place in everyday problem-solving context.

2.2 Commonalities between different definitions

As alluded before, CT is a reasonable and reflective thinking (Norris & Ennis, 1989, p. 1), is goal directed (Halpern, 2014, p. 8), and is relying on criteria (Ennis, 1985; Lipman, 2003, p. 212). It includes also the cognitive and affective aspects of thinking.

The cognitive aspect of CT includes the knowledge related to the problem at hand and also the CT skills to apply the knowledge in a way that achieves or explores the goal. In this context, CT skills are the goal-directed skills that contribute to the process of critical thinking and the achievement of the goal (Hamby, 2015, p. 78).

The affective aspect of CT considers the attitude and willingness to apply necessary skills to reach that goal. Since "the enactment of CT may run against dominant ideologies and power structures" (Davies & Barnett, 2015, p. 29), thinking critically requires effort, willingness (Hamby, 2015), and courage (Davies & Barnett, 2015, p. 29).

Given the goal of this dissertation to develop students' CT skills, it is necessary to reach a consensus on CT skills. The following is an overview of CT skills in some of the above CT definitions in which the author explicitly lists CT skills for educational purposes.

Ennis believes that CT occurs within the problem-solving context and often in the context of interacting with other people (Norris & Ennis, 1989, p. 8). Therefore, he defines CT skills (his term "abilities") as those strategies are needed to "keep the problem-solving process on task, communicate coherently with people, and deal intelligibly with information received from others" (Norris & Ennis, 1989, p. 8). His categorisation of CT skills includes four general groups (Norris & Ennis, 1989, p. 10):

- Skills involved in thinking clearly
- Skills involved in making and evaluating inferences
- Skills needed to establish a sound basis for inference
- Skills involved in carrying out the CT process in an orderly manner

A broader view on Ennis' list of CT skills involved in solving problems (Ennis, 2015, pp. 37-44; Moseley et al., 2005, pp. 154-155; Norris & Ennis, 1989, p.14, see also elaborated version of CT abilities pp. 183-187), emphasizes on the abilities addressed by Halpern (2009, 2014) such as argument analysis, verbal reasoning (defining terms and judging definitions), making and testing hypotheses (making and judging deductions and inductions), and decision making skills.

Lipman designed *Philosophy for Children* programme to improve students' CT skills in the community of inquiry by using key philosophical ideas. He considers CT skills as those strategies are required to make child's mind more precis to play an active role in an inquiry community (interview: https://p4c.com/julie-winyard-interviews-matthew-lipman/, 2008). Lipman (2003) considers the relevant thinking skills for educational purposes as the following:

- Inquiry skills: the skills involved in self-correcting practices such as identifying errors in their own thinking or each other's thinking, clarifying vague expressions in texts, and identifying inconsistences in discussion (p. 224).
- Reasoning skills: the skills involved in the process of making valid conclusion based on true premises such as looking for strong and relevant reasons, providing examples and counterexamples, and clarifying assumptions (pp. 170- 171).
- Concept formation skills: the skills involved in relating concepts to form principles, criteria, arguments, and explanations, such as organising information and then analysing and clarifying them to accelerate their application in understanding and judging (p. 184).
- Translation skills: the skills involved in transferring meanings from one language or symbolic scheme to another while keeping them intact. Here, Interpretation is necessary to give the translated meaning an appropriate sense in the new context (p. 185).

A review of the elaborated list of thinking skills offered by Lipman (2003, pp. 166-171) shows a consensus on verbal reasoning, argument analysis, thinking as hypotheses testing, and problem solving skills addressed by Halpern (2009, 2014) as CT strategies.

Apart from analysing Ennis and Lipman's overview on CT skills, since this PhD project was conducted in Germany, it is therefore necessary to review the skills considered by the National Standard of Physics Education (KMK, 2020) in Germany.

Based on the **National Standard of Physics Education** (KMK, 2020), skill is defined as a combination of knowledge, attitude, and application (Wissen+ Wollen+ Handeln). The relevant skills for physics education are divided into the following categories:

- Content knowledge expertise: the ability of using theories, models, experiments to deal with a task or problem (p. 9).
- Generation knowledge skills: the skills to apply scientific thinking and scientific reasoning method, such as making and testing hypotheses, and interpreting and reflecting on the results (p. 9).
- Communication skills: the skills to use correct terminologies (verbal reasoning) and to argue rationally (p. 9)
- Evaluation skills: the skills to make a decision or judge based on criteria and reflect on the process of decision making (it is necessary to have knowledge of content, knowledge of interdisciplinary perspectives, and knowledge of assessment procedures) (pp. 9-10).

A broader look of the subskills addressed in each abovementioned category (see KMK, 2020, pp. 12-18) indicates a focus on the skills addressed by Halpern (2009, 2014) such as verbal reasoning, argument analysis, thinking as hypothesis testing, and decision making and problem solving.

Table 2.9 presents common aspects between Halpern's categorisation of CT skills and the CT skills defined by Ennis, Lipman, and the national standard of physics education in Germany (KMK).

As it is shown in the table 2.9, the "Likelihood and uncertainty analysis skill" is only addressed in Halpern's categorisation of CT skills. Since lots of decisions in the life are probabilistic like medical diagnoses and treatment decisions, business, college admission, and also research, therefore it is necessary to train students this skill.

On the other hand, the purpose of the comparison in Table 2.9 is not to show that the CT skills addressed by Halpern are very complete, since Halpern did not claim this either. For example, Lipman (2003) explicitly considers translation skill in the list of necessary skills for educational purpose which is defined as carrying meanings from one context in language, symbolic scheme, or sense modality to another (pp. 185-186), whereas one cannot find this in Halpern's list.

In addition, KMK (2020) includes the ability to apply skills learned in the physics context in other contexts as one of the evaluation skill outcomes, while one cannot find this skill explicitly in Halpern’s list in Table 2.9. However, Halpern discusses elsewhere that the application of CT skills in a variety of contexts enhances skill development and emphasizes the important role of transfer (Halpern, 2009, p. 8).

In summary, considering the fact that all abovementioned definitions come from different perspectives and motives (from philosophy like Ennis or education Like Andrews) can make it clear why the terminologies are used are a little different but they have the same core. All authors deploy a language of dispositions and skills.

In addition, one can find more commonalities between the competencies addressed in the National Standard of Physics Education in Germany (KMK, 2020) and Halpern’s definition of CT. The outcomes of applying CT skills based on Halpern’s definition fit well with the addressed skills in the KMK. Moreover, in the context of particle physics, the likelihood and uncertainty analysis skill seems a relevant skill to be improved, since probabilistic decisions play a role in particle physics when a new particle is discovered. Overall, Halpern’s definition of CT is considered an appropriate basis for designing relevant teaching-learning activities in the context of particle physics.

Table 2.9: Commonalities of Halpern's definition of critical thinking (CT) skills with defined CT skills by Ennis, Lipman, and the national standard of physics education in Germany (KMK).

Halpern Critical Thinking Skills			**Commonality with**											
			Ennis (1989, 2015)				**Lipman (2003)**				**KMK (2020)**			
	Skill	**Outcomes**	*Thinking clearly*	*Making and evaluating inferences*	*Carrying out the CT process in an orderly manner*	*Establishing a sound basis for inference*	*Inquiry skills*	*Reasoning skills*	*Concept formation skills*	*Translation skills*	*Content knowledge expertise*	*Generation knowledge skills*	*Communication skills*	*Evaluation skills*
	Verbal reasoning	Detecting misuses of definitions	*										*	
		Using precise terminologies	*										*	
		Using questioning and paraphrasing to comprehend text	*											
		Explaining reasons to make valid conclusion	*				*	*					*	

	Argument analysis	Identifying key parts of an argument like premises, reasons, and conclusion	*						*				*	
		Identifying counterargument		*									*	
		Making strong argument in which reasons support conclusion							*				*	*
		Judging the credibility of an information source				*							*	*
		Recognizing the missing components in an argument							*				*	
	Thinking as hypothesis testing	Recognizing the need for and using operational definitions (recognize the need to define a term and specify how to measure it)											*	
		Understanding the need to control variables in order to make strong causal claim		*			*					*		
		Being able to describe the relationship between any two variables as positive, negative, or unrelated							*			*		
		Checking for adequate sample size before drawing a conclusion		*								*		
	Likelihood and uncertainty analysis	Estimating the probability of occurrence of an event												
		Using probability judgment to improve decision making												
		Understanding the limits of extrapolation (make estimations of future events)												
		Understanding the need for relevant and valid information for being able to conclude about the most probable event												

	Decision making and problem solving	Considering different sorts of solutions or alternatives			*									*
		Selecting criteria to judge possible solutions		*	*			*	*					*
		Recognizing the need to seek counterevidence		*				*						
		Examining the relevance of the procedures in solving problems (monitoring the decision making process)			*							*		*
		Using analogies to solve problems						*						

2.3 Nature of critical thinking: general and domain-specific

As alluded before, critical thinking is a purposeful, reasoned, and effortful thinking which leads to an action to reach desirable outcomes. It is clear that depending on the context or type of thinking task, the skills a critical thinker uses to achieve desired outcomes can differ (Halpern, 2014, p. 8). Sometimes the context is everyday life and sometimes a specific subject or discipline. There is a discussion whether CT skills are generally applicable skills across different contexts or content specific skills (Davies, 2006; Ennis, 1990; McPeck, 1990; Moore, 2004; Norris & Ennis, 1989). Therefore, it is important to distinguish between general CT skills and domain-specific CT skills. Being clear about CT as general and/or domain-specific competence helps to design courses that promote CT skills and tests that evaluate those skills more realistically (McPeck, 1990, p. 22).

General CT skills refer to the skills applicable and transferable to any issue in any discipline including everyday life (Norris & Ennis, 1989, p. 21; Wendland et al., 2015, p. 158). Here the CT skills are seen independent of the disciplines and can be taught for example in a class about informal logic or argument techniques (Davies, 2013, p. 530). Here, the emphasis is more on the general abilities like considering different perspectives, checking the validity of information sources, making strong arguments, and making decision in everyday situations. The proponents of general CT skills believe that the CT skills are “generic in nature” but they also respect differences in applying CT skills from discipline to discipline (Davies, 2013, p. 530).

Domain-specific CT skills consider the skills applicable to a specific subject or discipline (Norris & Ennis, 1989, p. 21), e.g. physics. Here the CT skills are seen discipline-dependent and can be taught by using the language of the discipline (McPeck, 1990; Davies, 2013, p. 530). However, there are different views among the domain-specificity perspective (e.g. domain specificity, epistemological subject specificity, and conceptual subject specificity; see Ennis, 1990; McPeck, 1990) but all the proponents of domain-specific CT skills in general emphasize knowledge of context as the major ingredient of critical thinking. They discuss that since any task or problem requiring critical thought are combinations of different types of knowledge, the sufficiency of a set of general CT skills seems rather unlikely (McPeck, 1990, p. 26). They argue that understanding knowledge of discipline provides the necessary skills for dealing with everyday problems. For example, McPeck (1990) argues that, since “the disciplines had their origins in the human condition” (p. 30) and provide “knowledge and understanding which goes beyond the present and the particular” (p. 31), therefore the disciplines provide “the best set of knowledge and skills for dealing with problems affecting society” (p. 31).

Each of these perspectives about the nature of CT, whether general or domain-specific, includes differences in designing CT instruction. Taking domain-specific perspective into account requires teaching critical thinking as part of teaching the subject, as a means of developing subject expertise (McPeck, 1990; Moseley et al., 2005, p. 22). Whereas, the perspective on the general nature of CT emphasizes on teaching important general skills in a separate timetable where students learn and apply the skills within a special pedagogical framework, for example Lipman’s approach in designing the “community of enquiry” (Lipman, 2003; Moseley et al., 2005, p. 23). More discussion about the teaching approaches in promoting CT skills is given in chapter 3.

Since there is an overlap between CT skills learned in a specific domain and CT skills required in everyday life (Davies, 2006; Ennis, 1990, p. 16), it is necessary to consider a balance between promoting general and domain-specific CT skills. The general CT skills provides a framework for reasonable and effortful thinking and the domain-specific CT skills outlines how the general skills are used in the service of the disciplines (Andrews, 2015; Davies, 2006; Ennis, 1997).

To summarize, the proponents of both perspectives on the nature of CT skills agree that an in-depth content knowledge of a particular domain is required for CT competency (McPeck, 1990, pp. 22-27; Norris & Ennis, 1989, p. 21; Tirunel et al., 2017). Furthermore, since there is an overlap between general and domain-specific CT skills, it is necessary to use a definition for CT skills which has the potential to translate the skills applicable in everyday life to subject matter context (Ennis, 1990, p. 16). As alluded in section 2.2, Halpern's definition (2009, 2014) can serve this point.

2.3.1 Nature of critical thinking in physics

Generally, instruction in science is considered as a means to prepare learners "to make sound judgments about what to accept and what to question, to reconsider past assumptions and adapt to new discoveries, and to respond effectively and productively to new situations and new knowledge as it develops" (Redish & Hammer, 2009, p. 630). Furthermore, improving critical thinking is explicitly mentioned as a goal of science education by the European Commission (2015).

To discuss the extent to which CT in physics is domain-specific a study by Viennot and Decamp (2018) can be addressed. To find the link between critical attitude and conceptual understanding in physics context, they offer students incomplete and invalid explanations, misleading statements, restricted valid conclusions, and invalid arguments. Their findings show the important role of content knowledge for critical thinking and argue reaching a "threshold of comprehension" is necessary to activate students' abilities to formulate CT comments (Viennot & Decamp, 2018, p. 4). So far students reach this point, they ask crucial questions, criticize their own line of reasoning and the others, and search actively for a satisfactory level of comprehension with a valid and complete explanation (Viennot & Decamp, 2018, p. 5). Moreover, being master in the terminologies used in the specific domain (Viennot & Decamp, 2018, p.7) increases the probability of analysing critically an explanation.

Furthermore, this study argues the importance of development of general critical thinking abilities if students' content knowledge is weak or medium. Here the ability to identify and detect the missing link in the line of reasoning, understand the need to define terms used in a text, and rely on logic is considered important to evaluate the internal consistency of the explanation or argument (Viennot & Decamp, 2018, p. 12).

Besides emphasizing both aspects of CT skills in physics, general and domain-specific aspects, they also argue that conceptual understanding is not a sufficient condition for critical thinking (Viennot & Decamp, 2018, p.5). They warn of a syndrome of "expert anaesthesia of judgment" which is observed in the absence of critique regardless of the incompleteness of the offered explanations (Viennot & Decamp, 2018, p. 5). This group involves the students who knows the topic very well and express satisfaction with the incomplete explanation. Viennot and Decamp (2018) hypothesize that this group "unconsciously completed what they were reading" (p. 5). This indicates the importance of developing CT attitude in addition to content knowledge to improve CT skills (Viennot & Decamp, 2018, p. 12).

Other effort for improving students' CT skills in the specific domain of physics was done by Etkina and Planinšič (2015). Besides teaching students the skills of making hypotheses and designing experiments to test them, which is content-related, the general abilities to tolerate the ideas of others, understand that all hypotheses are accepted as equal prior to testing, and test hypothesis not their intention (Etkina & Planinšič, 2015, p. 436) are needed to be developed and considered along teaching physics.

Furthermore, another effort is reported by Redish and Hammer (2009) to redesign physics course for biology students to help them learn to "think scientifically—to build coherence, think in terms of mechanisms, and to follow the implications of assumptions." (p. 629). Using everyday examples from physics, students have the opportunity to improve their understanding of physics knowledge and beyond to learn the general skills such as seeking coherence, looking for relevant and valid sources of information or considering limits to generalization (Redish & Hammer, 2009).

To summarize, the results of practical efforts in teaching CT skills in the domain of physics show that improving students' CT skills needs a focus on both general and domain-specific nature of CT skills. Although based on the topic and issue some tightly content-related skills may differ but there are common CT skills that are applicable across wide variety of domains like understanding the need to define terms, seeking for relevant and valid sources of information, and neutrality in testing hypotheses (Viennot & Decamp, 2018; Etkina & Planinšič, 2015; Redish & Hammer, 2009).

2.4 Students' challenges in critical thinking

To develop students' CT skills, it is necessary to know students' challenges toward thinking critically. This helps to design appropriate teaching-learning activities to engage students in thinking.

Studies show that critical thinking must be learned (*cf.* Black, 2005; Duron et al., 2006; Ritchhart & Perkins, 2005) because the skills of drawing conclusions, making decisions, and solving problems through a systematized process of reasoning with evidence, weighting it, and drawing reasoned conclusions are complicated and need to be practiced (Ritchhart & Perkins, 2005, p. 776).

Considering "students as persons" emphasizes the role of students as "actors in the world, not just thinker" (Barnett, 2015, p. 64). A critical actor applies CT skills and disposition (critical self-reflection) to interact with everyday life (Barnett, 2015, pp. 64-65). However, studies show that most educated adults fail to think critically in many situations (Gelder, 2005; Halpern, 1998; Kennedy et al., 1991; Kuebli et al., 2008; Perkins et al., 1983) and they often consider their "personal experience" to be more convincing evidence than a carefully conducted scientific study (Lai, 2011, p. 22).

On the other hand, empirical research on the nature of cognitive development suggests that students of all intellectual ability levels can benefit from CT instruction (Idol & Jones, 1991; Kennedy et al., 1991, p. 18; Lai, 2011, p. 23; Lewis & Smith, 1993). Furthermore, a study about the factors that inspire first-year university students to study physics shows that students see physics as a special way of thinking which gives them the chance to think critically (Levrini et al., 2016, p. 11).

One can recognize here a discrepancy between what students desire to get from education and what they show in the future as educated people. One reason for this gap may be deficient teaching-learning activities (domain-specific instruction) to encourage the development of critical thinking (Paul, 1993). The design of these types of activities is discussed in 3.2. Here, the focus is on the students' challenges toward CT which should be considered when designing effective instruction.

Following is a discussion of the challenges related to the 5 CT skills as defined by Halpern (2009). It is noticeable that systematically separating the challenges in these 5 CT skills is not an easy task. As discussed earlier, all CT skills are interrelated, e.g., one cannot make a good argument without considering the skill of verbal reasoning.

2.4.1 Verbal reasoning skill

Paying attention to the meaning of the terms involved in an explanation is necessary for learning and teaching (Viennot & Decamp, 2020, p. 84). In a study about reasoning in physics, Viennot (2001) confronts students with an uncompleted explanation to study students' common trends of reasoning. She reports the lack of coherence in students' reasoning (p. 115) and also the lack of considering the importance of applying non-ambiguous terminologies, especially in thermodynamics (p. 109). She argues that verbal synthesis of calculations or problems in physics is necessary for understanding physics by identifying the essential meaning of a relationship, the need for coherence in a text, and the need for unambiguous terminologies (pp. 213-214).

2.4.2 Argument analysis skill

A very general phenomenon reported by cognitive psychologists is that "when people believe a conclusion is true, they are also very likely to believe an argument that appears to support it, even when these arguments are unsound" (Kahneman, 2012, cited in Viennot & Decamp, 2020, p. 27). Viennot and Decamp (2020) call this phenomenon “confirmation bias” (p. 27) and in their studies they observe this phenomenon in students’ argumentation in physics.

Moreover, studies show that students lack the ability to argue or analyse arguments (Davies, 2013; Garcia-Mila & Andersen, 2007; Ryu & Sandoval, 2012; Shaheen, 2016; Viennot, 2001; Zohar, 2007). Students’ challenges are reported as the lack of the ability to provide valid reasons for the conclusion, lack of coherence and clarity in making argument, and lack of reaching a sound conclusion (Ryu & Sandoval, 2012, p. 492; Shaheen, 2016, pp. 24-25).

In science (biology lesson) typical students’ problems are reported as the lack of providing enough support for their conclusion, and ignoring alternative points of view that would potentially generate counterarguments (see review in Garcia-Mila & Andersen, 2007, pp. 32-33; Zohar, 2007, p. 245). Furthermore, the results of a study on the role and value of counterargument in learning science show that students have difficulty in identifying a flaw in one’s argument and constructing a counterargument (Henderson et al., 2015, p. 1686).

Explicitly in physics Viennot (2001) reports that in making conclusion students overly restricted themselves to obvious reasons and deny some reasons when there is no obvious effect (p. 206).

2.4.3 Thinking as hypothesis testing skill

Lawson (2004) believes that teachers should teach science as a process of making hypotheses and testing them (his term “process of critical inquiry”, p. 334) to develop students’ scientific reasoning skill, which Halpern refers to as thinking as hypothesis testing skill. Lawson (2004) sees the scientific reasoning skill as an inquiry not only into decision making and problem solving, but also into understanding complex concepts (p. 333).

In physics Viennot (2001) reports students' difficulties in identifying variables, in explaining the relationship between variables to solve multi-variable problems, and in considering limitations in validity of an explanation (pp. 115-116). Furthermore, students' failure to control variables is reported (Boudreaux et al., 2008).

2.4.4 Likelihood and uncertainty analysis skill

It is discussed that the ability to analyse probability and uncertainty is important in constructing knowledge in science (Ford, 2008; Henderson et al., 2015; Viennot & Decamp, 2020). Considering the openness of an explanation in physics to improvement and consequently considering alternatives and evaluating their probabilities while making decisions about the most precise explanation emphasizes the importance of likelihood and uncertainty analysis skill in constructing knowledge (Viennot & Decamp, 2020, p. 101). The fact that the explanation consists of data and information (top-down view) proves how important it is to consider the probability of the correctness of data and, consequently, the interpretation of the data.

To evaluate the probability of the occurrence of an event, it is also important to define criteria (Ford, 2008; Viennot & Decamp, 2020). For example, to evaluate the likelihood of an explanation in physics to be the most accurate explanation, it is necessary to check its internal consistency, its compatibility with generally accepted laws, and its logical coherence (Papadouris et al., 2018; Viennot & Decamp, 2020, p. 101).

Considering challenges people face in making probabilistic decision, it is discussed that people tend to ignore the probabilistic nature of an event and look for a deterministic cause to explain the event (Halpern, 2009). Moreover, the results of an investigation of the overconfidence phenomenon show that people tend to be more confident in their decision about probabilistic events than they should be (Halpern, 2009; Kahneman & Tversky, 1973). Furthermore, people overestimate the probability of occurrence of two or more uncertain events by neglecting base-rate information (Halpern, 2009, pp. 171-172). In addition, studies show that people often ignore given data or information to calculate probability and mostly use their personal estimations of likelihood of events (Halpern, 2009; Kahneman & Tversky, 1973).

In education, likelihood and uncertainty analysis skill is mostly discussed in a mathematical statistical sense in calculating the probability of occurrence of events (Díaz & De La Fuente, 2007; Halpern, 2009; Tarr & Jones, 1997; Tarr & Lannin, 2005). The results of studies on students' difficulties in mathematical statistical context report that students ignore the nature of probabilistic events, believing the occurrence of events to be equally likely and believing in the control of the outcomes of an event, and also that students ignore given numerical information in making probabilistic decision (Díaz & De La Fuente, 2007; see also review in Tarr & Lannin, 2005).

To my knowledge, there is no study to address students' difficulties in making probabilistic decisions in the context of physics. Since one of the goals of this dissertation is to improve students' likelihood and uncertainty analysis skill, the analysis of the data gives the opportunity to discuss students' challenges in the specific context of particle physics (see section 7.3.1).

Furthermore, it is necessary to address that the mathematical statistical aspect of likelihood and uncertainty analysis is out of scope of this PhD. The main focus on this PhD is to improve students' likelihood and uncertainty analysis skill by making several interpretations and evaluating them using content knowledge and considering probability (see section 7.3.1).

2.4.5 Decision making and problem solving skill

The results of a study on students' perception of problem solving skills show a lack of confidence to engage in problem-solving activities, a lack of the ability to consider advantages and disadvantages of a solution, and a lack of the ability to make a decision impartially (Temel, 2014).

In physics, studies address students' challenges in problem-solving ability from two perspectives: quantitative problem-solving and cognitive development of problem-solving ability (Larkin & Reif, 1979; Ogunleye, 2009; Reddy & Panacharoensawad, 2017). From the quantitative perspective, students' challenges are addressed, such as lack of equation recall, insufficient mathematical knowledge, and insufficient practice of physics problems (Ogunleye, 2009; Reddy& Panacharoensawad, 2017). In terms of cognitive development, challenges for students include constructing meanings of the problem statement, internally linking the meaning of the statements in the problem statement, and moving from more general to more detailed aspects of a problem (Larkin & Reif, 1979; Ogunleye, 2009; Reddy & Panacharoensawad, 2017).

2.5 Teachers' perspective on critical thinking and teaching critical thinking

It is consistently argued that teachers play an important role in the implementation of CT skills, as their perspective on CT can influence whether and how these CT skills are integrated into their classroom (Evagorou & Dillon, 2011; Hashweh, 1996; Kowalczyk et al. 2012; McNeill, 2009; McNeill et al., 2016; Williams, 2005). Teachers who believe that engaging students in CT confuses them and that CT instruction is only appropriate for high-achieving students are those who avoid involving students in problem-solving and critical thinking experiences, while teachers who see essential such opportunities to student learning are willing to provide these learning experiences (McNeill et al., 2016; Zohar et al. 2001; Zohar, 2008). Furthermore, studies show a high correlation between teachers' perception of CT skills and their confidence to integrate CT skills into instruction as well as to assess students' CT skills (Black, 2005; Kowalczyk et al., 2012; McNeill et al., 2017, p. 440). Therefore, developing teachers' self-perceptions of CT skills are considered essential (McNeill et al., 2016).

In addition, studies show that while most college and university faculty recognize the importance of critical thinking, effective critical thinking instruction does not occur frequently (Paul et al., 1997; Reed & Kormrey, 2001, p. 9). For example, the results of a study by Paul et al. (1997) showed that while 89% of the faculty surveyed claimed that critical thinking was a major goal of their teaching, only 19% provided a clear explanation of what critical thinking is, and only 9% clearly taught critical thinking in their courses. This again highlights the importance of improving teachers' perception of CT.

To support developing teachers' perception of CT, it is necessary to review the literature regarding teachers' perspectives on CT skills and their perspective on teaching CT skills. This knowledge helps to identify aspects of CT where teachers need special help to be made aware of. Furthermore, it gives hints to develop a teacher package including guidelines and materials for integrating CT skills into instruction.

2.5.1 Teachers' perspective on critical thinking

Since this thesis has distinguished between general and domain-specific CT skills, teachers' perspectives on CT are discussed under these two categories: their perceptions of general CT skills and their perceptions of domain-specific CT skills. Teachers' perceptions of general CT skills are the result of reviewing studies that have examined teachers' perceptions using a general context or a general CT test, and teachers' perceptions of domain-specific CT skills are the result of reviewing studies that have examined teachers' perceptions using a specific context, such as biology and physics.

2.5.1.1 Teachers' perceptions of general critical thinking skills

Studies on teachers' perception of CT show that teachers of different disciplines teaching different ages from elementary school to university agree that CT is an important learning goal for students however, their self-perception of CT is different from what CT means in education (Choy & Cheah, 2009; Kowalczyk et al., 2012; McNeill et al., 2016; Noula, 2018; Paul et al., 1997; Rini & Prabawanto, 2021; Shaheen, 2016).

Studies report a lack of deep understanding of CT among teachers and also a lack of being familiar with the strategies to integrate CT in the classroom (Alazzi, 2008; Demir, 2015a; 2015b; McNeill et al., 2016; Zohar, 2007).

A study by Noula (2018) shows that teachers' perceptions of CT are associated with a number of skills that are instrumental in nature, such as those of writing and reading, rather than considering CT as a necessary skill for everyday life that can be developed in classroom practice (p. 879). Furthermore, some teachers consider CT as a tool for developing content knowledge, without considering the reflective and evaluative nature of CT (Choy & Cheah, 2009).

Besides reports of teachers' challenges in perceiving CT, however, there are also some studies showing a high level of teachers' perceptions of CT. For example, the results of a study of UK university teachers' perspectives on CT (Shaheen, 2016) show that teachers perceive CT in terms of skills and disposition and define it generally as the ability to analyse text and beliefs, discuss alternative perspectives with open-mindedness, and evaluate ideas, assumptions, and arguments to reach a reasonable conclusion (pp. 22-23).

2.5.1.2 Teachers' perceptions of domain-specific critical thinking skills

Studies on physics teachers' perspectives on CT shows that teachers consider CT skills as an important competency for students (Haviz & Metiza Maris, 2020; Viennot & Decamp, 2020, p. 93), but their understanding of CT in physics still has some deficits. It is reported that argument analysis skill is not considered a relevant CT skill in physics classroom (Haviz & Maris, 2020) and also in science (McNeill et al., 2016; Zohar, 2007). Although, science teachers are familiar with thinking as hypothesis thinking skills like "defining a research question, controlling variables, planning experiments, describing experimental results, and drawing conclusion" (Zohar, 2007, p. 247), there is a deficit in their understanding of argument analysis skills. For instance, some science teachers generally describe argument analysis skill as the skill of students' literacy, while failing to focus on the important elements of argument analysis skill in science, such as making sense of data or evidence, considering multiple views, and evaluating the validity of an argument (McNeill et al., 2016, p. 2036). Furthermore, science teachers' lack of identifying assumptions, and of the ability to construct arguments and counterarguments is reported (Zohar, 2007).

The results of another study with four pre-service science teachers about their argumentation skills in the context of evolution indicate that while teachers support their claims with evidence, their arguments do not include alternative causes, and suffer from inadequate sampling of evidence for making conclusion or generalizations (Zembal-Saul et al., 2002). Due to the small sample size, it is not possible to make a sound generalization, but these findings give hints to be realistic about the expectations from teacher in implementing argument analysis skills in the science class (Zembal-Saul et al., 2002).

Furthermore, in a study by Viennot and Decamp (2020) they faced future physics teachers with an incomplete and inconsistent text to experience for themselves the situation a potentially critical person faced with a text. Here a critical analysis was expected from students which includes extracting precise information, evaluating the logical consistency, and evaluating the generalization the text suggests (p. 98). The results show that students express a feeling of difficulty with the activity of critical analysis of the text (p.91). Furthermore, in their critical analysis of the text, the idea of coherence is less emphasized than that of precision (p. 93).

Moreover, the results of a study on the future physics teachers' problem solving skill in the specific domain of nuclear physics shows that pre-service physics teachers fail to formulate a problem, to make a precise conclusion, and to use the right strategy to solve the problem (Hartini et al., 2021).

2.5.2 Teachers' perspective on teaching critical thinking

In addition to teachers' self-perception of CT skills, teachers' perception of teaching CT influence their practices toward including CT in their classrooms.

Physics teachers consider CT skills as an important competency for students (Haviz & Maris, 2020; Viennot & Decamp, 2020), however concerning classroom implementation of CT, some future physics teachers express concern about confusing students or frustrating them when using the activity to improve students' critical analysis (Viennot & Decamp, 2020, p. 92).

Furthermore, some teachers only consider cross-thematic activities suggested by curriculum in which students work on different projects appropriate for developing students' CT skills and do not consider their regular teaching practice as a way of improving CT (Noula, 2018, pp. 879-880).

Looking at the barriers from teachers' perspective that prevent them from implementing CT skills alongside teaching content knowledge, lack of time in the classroom, lack of appropriate instructional materials, and lack of knowledge about how to implement CT instructional methods are reported (Fox, 1962; Kowalczyk et al., 2012).

In addition, in a study of teachers' views on the most difficult skills to teach in schools, Fox (1962) reported that teachers believe that teaching the ability to analyse, interpret, and evaluate information, the ability to evaluate sources of information, and the ability to determine the most reasonable and logical conclusion are the most difficult skills (p. 336).

2.5.3 Rationale for developing supportive materials for teachers

To expect physics teachers to integrate CT skills into the instruction, first they need sound understanding of these skills (Choy & Cheah, 2009; Kennedy et al., 1991; Kowalczyk et al., 2012; Zohar, 2007). Furthermore, they need knowledge of teaching strategies to embed CT skills into physics class to support students' CT

(Driver et al., 2000; Kowalczyk et al., 2012; Lauer, 2005; Zembal-Saul et al., 2002; Zohar, 2007). Therefore, it is necessary to develop supportive materials for teachers to help them to teach CT (Fox, 1962; Kowalczyk et al., 2012; Zohar, 2007).

Furthermore, it is reported that teachers' perception of CT influences their use of teaching-learning materials and consequently results in achieving learning gain in terms of students' ability to think critically (McNeill, 2009). In addition, teachers' familiarity with the instructional theories upon which these supportive materials are designed may increase opportunities for students to be engaged in critical thinking (Zohar, 2007, p. 250). Therefore, it is necessary to prepare a supportive material for teachers in which the theories behind the instruction and behind the designed materials are explained, the goals of instruction are clearly discussed, and the steps toward implementing are clearly defined.

Furthermore, some studies discuss the importance of teacher training programs for improving teacher' perception of CT. In these training programs teachers themselves engage in high-level critical thinking activities and also practice and reflect teaching strategies toward implementing critical thinking strategies in a domain-specific context. Since developing science teacher education and professional development programs for critical thinking is not the goal of this thesis, for those interested, see the reports by Zohar (2007), and Idol and Jones (1991, pp. 447-448).

3 Teaching critical thinking

It is important to teach students the skills they need for everyday life such as critical thinking (CT). To provide students with opportunities to learn and practice CT in the context of school subjects, it is important to develop appropriate teaching-learning activities in which students actively engage in the process of learning and applying the skills. Here the three questions may arise: what CT skills should be taught? What instructional design is appropriate to teach these skills? Are CT skills taught separately or in association with a subject (discussion about the nature of CT whether general or domain-specific)?

To answer the first question, synthesis of different definitions of CT showed that the 5 CT skills defined by Halpern (2009, 2014) are appropriate for teaching general and domain-specific CT skills (see section 2.2). This is also supported by a review of studies in science and mathematics education on teaching the necessary CT skills (Aizikovitsh-Udi & Amit, 2011; Etkina & Planinšič, 2015; Henderson et al, 2015). It is discussed that teaching argument analysis skill (Andrews, 2015, p. 60; Davies, 2013, pp. 542-553) particularly teaching counterargument in science education (Henderson et al., 2015), teaching likelihood (probability) analysis skill using real examples from everyday life in mathematics education (Aizikovitsh-Udi & Amit, 2011), and teaching making and testing hypothesis skill in physics (Etkina & Planinšič, 2015) can develop students' CT. Halpern's (2009) definition of CT encompasses all these skills and includes also measurable outcomes of applying these skills which makes evaluating the skills possible.

The answer to the questions of appropriate instructional design and teaching of CT skills, whether separately or in association with a subject, is presented in this chapter. This chapter first discusses the challenges of teaching CT. Then, the design of a CT instruction is discussed. In addition, the ways to evaluate the effectiveness of CT instruction are presented.

At this point, it should be mentioned that the term "instruction" encompasses a wide range of applications, e.g. dance. In this thesis, however, the term "instruction" is used in the sense of teaching instruction, classroom instruction, and teaching-learning activities. In addition, the term "CT instruction" is used to capture the concept of instruction that intends to improve students' CT skills.

3.1 Challenges of teaching critical thinking

There is a recurring discussion of the importance of teaching CT (McMillan, 1987; McPeck, 1990; Resnick, 1987; Viennot & Decamp, 2020). In general, CT is seen as a driving force for improving the quality of understanding (McPeck, 1990, p. 21), and in the specific context of science, CT is considered an essential prerequisite for developing scientific explanations and learning (Viennot & Decamp, 2020, p. 101). However, teaching CT requires efforts in designing appropriate teaching-learning activities, courses, and programs and needs research and empirical evidence (McMillan, 1987, p. 16; Resnick, 1987, p. 37) to identify effective approaches toward teaching CT.

The results of a review study by McMillan (1987) on the effects of teaching methods on enhancing college students' CT (from 1950 to 1985) showed that despite the claim of these studies in improving students' CT skills, there is a lack of clarity regarding three essential aspects for designing teaching-learning activities to improve CT: a clear definition of CT, a reliable and valid instrument for specific measurement, and a clear theoretical description of the nature of the learning experiences that should enhance CT (supported also by McPeck, 1990). Therefore, to improve CT, McMillan (1987) suggests to design programs that are characterized by a clear definition and observable outcomes to permit strong causal inference about their effectiveness (p. 16). Furthermore, he emphasizes on the development of an appropriate assessment tool that is sensitive to the specific changes to improve understanding of the conditions that enhance CT (McMillan, 1978). Finally, he concludes that despite the importance of developing these programs, it is not easy to design and implement these programs (McMillan, 1987, p. 16).

In addition to the challenges of teaching CT skills discussed by McMillan (1987), there are also different views of the nature of CT, whether general or domain-specific (see section 2.3), which consequently direct the discussion toward teaching CT as a general or domain-specific skill. In addition, there are different perspectives on approaches to teaching CT, whether implicit or explicit. Moreover, the limitation of empirical studies on teaching approaches that promote CT skills is a challenge (Abrami et al., 2008; Tiruneh et al., 2014). As a result, there is no consensus on effective instructional approaches to improve CT.

3.1.1 Teaching critical thinking as a general or domain-specific skill

McPeck (1990) as one of the advocators of the domain-specific nature of CT emphasizes on teaching CT trough the disciplines arguing that everyday problem is multidimensional and therefore solving everyday problem needs to employ different knowledge from different disciplines (p. 46). He discusses that to improve CT these CT skills should be embedded in the normal curriculum and not be offered in another specific course (McPeck, 1990, p. 32). Furthermore, Resnick (1987) argues that embedding CT skills in disciplines can give students more learning opportunities to practice and develop CT skills that are characteristic of a specific domain (pp. 35-36) and consequently, it is more likely that students learn the wide range of skills that are applicable in different contexts.

On the other side, the advocators of the general nature of CT emphasize on teaching general CT skills in a separate course where students learn and apply the skills within a special pedagogical framework (Lipman, 2003, Moseley et al., 2005, p. 23). For example, Lipman's Philosophy for Children program is designed for fifth- and sixth-grade students to improve their thinking skills (Lipman, 2003, pp. 156-161). In this program students discuss intensively the traditional philosophical problems that arise in a story-like context.

Considering the importance of both general and domain-specific CT skills introduces another perspective on teaching CT skills. Barnett (2015) emphasizes that general CT skills can be enhanced by the domain-specific CT skills (p. 67) and self-reflection (p. 68). Furthermore, Resnick (1987) argues that while each discipline has its characteristic way of CT (pp. 35-36), at the same time a wide range of skills are shared across disciplines and can be thought of as “enabling skills” for learning and teaching (p. 46), e.g., identifying the importance of using correct terminologies, being sensitive to the definition of the terms, and constructing and evaluating an argument.

Furthermore, as discussed in section 2.3.1, to improve students’ CT skills in the specific domain of physics, it is important to focus on both general and domain-specific CT skills. Therefore, in teaching physics both general and domain-specific skills should be taught. However, whether implicitly or explicitly will be discussed in the next section.

3.1.2 Teaching critical thinking implicitly or explicitly

Teaching CT skills implicitly or explicitly introduces new terminologies in the area of teaching approaches toward CT; infusion and immersion approach. In an infusion approach the CT skills are taught explicitly in a subject-matter instruction while in an immersion approach students get deeply immersed in the subject without getting explicit CT instruction (Ennis, 1989, p. 5). However, in both approaches it is necessary to design a deep, thoughtful, and well understood subject-matter instruction (Ennis, 1989, p. 5).

In addition to infusion and immersion approaches, there is also a mixed approach. In the mixed approach students are taught general CT skills separately and explicitly (for example at the beginning of the subject-matter course) and also involved in a subject-matter CT instruction (Ennis, 1989, p. 5). The mixed approach makes general CT skills explicit (Ennis, 1989, p.5) and encourages students to apply the skills in a subject-matter instruction by making them domain-specific.

Teaching students the argument analysis skill, Andrews (2015) discusses the importance of the explicit teaching of general and also domain-specific argument analysis skills to promote higher education students' CT. He considers general argument analysis skill as a framework for making strong arguments and evaluating them which is also applicable in the specific domain (Andrews, 2015). He argues that focusing on both general and domain-specific argument analysis skills can help students "to think clearly, argue well, and take their place in a democratic society where difference is tolerated, understood, and, where possible, resolved to allow consensus and action" (Andrews, 2015, p. 61).

Moreover, in the specific context of physics, Viennot and Decamp (2018) emphasize the explicit teaching of general CT skills as an approach to conceptual understanding. They discuss "a topic can be considered as mastered by an individual or a group when appropriate explanations can be produced, their conditions of validity analysed, their informative value possibly extended, and when contestable arguments can be appropriately discussed and possibly rejected" (Viennot & Decamp, 2018, p. 14).

Furthermore, the results of review studies examining the effects of instructional interventions on CT skills (Abrami et al., 2008; Tiruneh et al., 2014) suggest that the mixed approach, where CT is taught as a stand-alone subject within a specific content course, can improve students' CT skills. Moreover, an explicit and detailed explanation of CT principles is needed to improve the effectiveness of a subject-matter CT instruction (Abrami et al., 2008; Tiruneh et al., 2014, p. 8). Therefore, "embedding CT instruction within specific subject matter domains rather than teaching in separate courses is being considered as a more promising route to help students become critical thinkers." (Tiruneh et al., 2014, p. 8).

Therefore, to design an effective instruction, it is important to be clear about which CT skills to focus on (Tiruneh et al., 2014, p. 9) and then design teaching-learning activities to teach and develop theses CT skills in both everyday life and domain-specific contexts (Abrami et al., 2008, p. 1121).

3.1.3 Valuing disposition alongside teaching critical thinking

Since CT includes both disposition and skill (Ennis, 1985; Facione & Facione, 1996; Halpern, 1998; Hamby, 2015; Lipman, 2003; McPeck, 2016, p. 17), therefore for training critical thinkers it is necessary to consider both the cognitive and affective domains of students' learning in an area (McPeck, 2016, p. 17). Regarding this, McPeck (2016) discusses that "to the extent that CT is a skill, it is teachable" (p. 18), but teaching particular skills "is not sufficient to produce a critical thinker". Therefore, it is necessary to "develop the disposition to use those skills" (p. 19). Furthermore, Resnick (1987) discusses that "shaping disposition to critical thought is central to developing higher order cognitive abilities in students." (p. 41).

Besides the importance of developing the disposition to engage students in critical thinking, the term disposition is defined with a great variability (see Hamby, 2015; see Nieto & Valenzuela, 2012). A current study by Nieto and Valenzuela (2012) on investigation the internal structure of critical thinking disposition, defines disposition as an involvement of motivation and mental habits (attitude). They show that "motivation to think critically" has a great weight in the initial stage of practicing CT while developing mental habits is a process and will achieved by exercising CT skills (Nieto & Valenzuela, 2012, p. 36). When thinking critically becomes a mental habit, students apply CT skills outside the classroom in real life, when it is optional and even when there is no external force (e.g. getting a good mark in the class) (Hamby, 2015, p. 80).

Furthermore, Hamby (2015) introduces "willingness" as an additional essential fundamental for being a critical thinker. Hamby (2015) defines "willingness" as the internal motivation to employ CT skills which drives the critical thinker to "evaluate arguments, clarify the meanings of terms and statements, evaluate authorities and sources, and consider different alternatives" in order to reach a reasoned judgment when examining an issue (Hamby, 2015, p. 78).

On the other hand, despite the importance of disposition, the results of a study of the effect of science instruction on the disposition of eleventh-grade students toward CT show that students' dispositions do not improve significantly during one-year participation in regular science classes (biology, Physics, and chemistry) (Ben-Chaim et al., 2000). This demonstrates the importance of special attention to disposition when constructing an instruction for teaching CT skills.

Therefore, to teach students CT skills, it is important to teach them to value efforts to think critically (Hamby, 2015, p. 86; Resnick, 1987, pp. 40-44). This might help move students from external motivation to internal motivation, and eventually foster their willingness to engage in critical thinking and apply CT skills. Besides that, the other CT virtues such as open-mindedness and valuing fallacious-free reasoning should be taught and modelled in the classrooms (Hamby, 2015, p. 11).

To develop a disposition toward CT, the learning environment must be changed from a lecturing environment to a community where efforts to think critically are valued, where opportunities are provided to support the development of intrinsic motivation (working to accomplish a task) rather than external motivation (achieving grades), and where more opportunities are provided to practice CT skills to develop a mental habit of thinking critically.

In addition, the results of a review study by Resnick (1987) show that the quality feature of the task correlates with the kind of motivation that keeps people on task. He argues that continuing to work on a complex task is an indicator of students' intrinsic motivation (Resnick, 1987, p. 43).

3.2 Design of critical thinking (CT) instruction

Designing teaching-learning activities to improve CT skills requires efforts (McMillan, 1987, p. 16; Resnick, 1987, p. 37). Clear principles are needed on which to base a coherent framework for classifying thinking and learning outcomes (Moseley et al.,

2005) and also to allow instructors to plan carefully sequenced teaching-learning steps to engage students in the complex process of critical thinking. However, a review study on CT instructions demonstrating improvement in students' CT skills (Tiruneh et al., 2014) reports a lack of clear design principles.

Moseley et al. (2005) review the instructional designs dealing with building understanding and promoting thinking strategies such as analysis, evaluation, application, and problem solving (pp. 44-118). In their work they analyze a variety of structured learning environments for promoting thinking skills and conclude that in order to design an instruction to help students to gain an understanding of content knowledge, to foster their critical thinking, and to engage them in the process of critical thinking, it is helpful to combine the frameworks which address the desired goals of instruction (Moseley et al., 2005, p. 318). This "integrated model" can be used to design effective instruction towards achieving its goals.

Since the goal of this PhD is to design effective instruction to develop students' CT skills, the integrated model could include a critical thinking model (Allegretti & Frederick, 1995; Reed & Kromrey, 2001; Williams, 2005) and an appropriate instructional design theory. The critical thinking model can provide guidelines for focusing on aspects of thinking skills and developing relevant activities and materials to provide students with the opportunity to practice and learn the skills. Furthermore, an appropriate instructional design theory can provide a framework, in which teaching-learning sequences are systematically established to reach the goal of instruction. These two components of instruction, critical thinking model and appropriate instructional design theory, should be aligned to achieve the desired outcomes of CT instruction.

3.2.1 First component of CT instruction: Critical thinking model

As mentioned earlier, Halpern's (2009, 2014) definition of CT, which focuses on measurable outcomes of applying CT skills, can answer the question of what aspects of

CT should be the focus of a CT instruction. Furthermore, based on empirical research in cognitive psychology Halpern (1998) presents a 4-part model for teaching CT:

(a) **A skills approach to CT:** Instruction should focus on skills that are clearly defined, based on observable outcomes and therefore can be taught, and that are also necessary in the global marketplace and for functioning as effective citizens in a complex democratic community. These skills are verbal reasoning, argument analysis, thinking as hypothesis testing, likelihood and uncertainty analysis, and decision making and problem solving.

(b) **A dispositional component:** instruction should value and encourage disposition for effortful thinking and willingness to apply CT skills.

(c) **Structure-training activities**: instruction should highlight the structural aspects of problems and arguments and practice them in different sorts of examples so they can act as retrievals queus in a novel context, regardless of a subject matter. For example, when teaching students the skill of verbal reasoning, one of the structural aspects is to pay attention to the use of precise terminologies. Regardless of the context, sensitivity to the use of terminologies can be applied in any field when a better understanding is sought.

(d) **Metacognitive monitoring:** instruction should ask well-structured questions to help students to reflect on their learning process before beginning the task, while working on the task, and after completing the task. These questions serve as metacognitive monitoring, improve students' self-awareness, and help them to direct and assess their thinking. These questions require students to think about the time and effort required to solve the problem, the reason for working on the problem, the estimated difficulty level of the problem, the likely useful CT skills for solving the problem (before beginning the task), the assessment of the progress toward the goal (while working on the task), and the evaluation of the quality of the solution (after completing the task).

Halpern (1998) argues that considering these 4 elements in designing a domain-specific instruction helps students to apply the CT skills they learned in a variety of contexts (p.

451). Focusing explicitly on the structural aspects of problems or arguments and embedding metacognitive monitoring by asking reflective questions might serve this goal (Brown & Kane, 1988; Halpern, 1998; McBride & Bonnette, 1995, p. 384; Williams, 2005).

To reach this goal, the teacher also plays an important role. The teacher acts as facilitator in the class who asks reflective questions, shows interest in students' ideas, prompts them to think about alternative solutions or strategies, and gives them corrective feedback (Brown & Kane, 1988; McBride & Bonnette, 1995, p. 385).

3.2.2 Second component of CT instruction: Appropriate instructional design theory

Based on the "integrated model", designing instruction to promote students' CT skills also requires appropriate theory for instructional design. The literature review of studies that explicitly address teaching CT and provide guidelines for designing teaching-learning experiences to promote CT reveals two directions. Some discuss the strategies involved in designing instruction to enhance CT skills (e.g., problem-solving strategies), while others focus more on the interactional aspects of instruction (the ways in which teacher-student and student-student interact).

Each aspect is discussed in the appropriate section below. However, it must be mentioned that one cannot ignore the interactional aspects of using a strategy when discussing effective strategies. For example, a problem-solving strategy cannot work alone, and teachers and students need to interact appropriately for the problem-solving strategy to work effectively and ultimately promote students' CT skills. The separation was made based on the emphasis on these particular aspects, which are also supported by experimental results, while the other aspects are sometimes taken for granted.

3.2.2.1 Instructional strategies

It is discussed that well-designed problem-based courses, in which a problem stands in the center of the course, are likely to encourage student to think critically about the content (Pithers & Soden, 2000; Rini & Prabawanto, 2021). Furthermore, it is discussed that providing an authentic instruction where students are exposed to authentic or situated problem improve students' scientific reasoning skills (Chinn & Malhotra, 2002). Moreover, providing students with the opportunity to apply their knowledge and critical skills to take a critical action in the real world to solve an authentic problem can improve students' disposition toward CT (Barnett, 2015).

Moreover, demonstrating CT skills and consistently practicing the skills are considered as a way to improve students' CT skills (Barak & Dori, 2009; Shaheen, 2016, p. 29; Viennot & Decamp, 2020). For example, in the specific domain of physics, Viennot and Decamp (2020) offer students an explanatory text and give them chance to practice critical analysis. While working on critical analysis of the explanatory text, students learn to extract precise information, to evaluate the logical consistency and to evaluate the generalization the text suggests (Viennot & Decamp, 2020, p. 98). Another example is from in-service science teachers who are involved in critical evaluation of science education articles during one semester (Barak & Dori, 2009). They analyze, evaluate and compare articles on the same topic, and present their results in a classroom discussion. The results of study show an improvement in science teachers' CT skills and disposition (Barak & Dori, 2009).

3.2.2.2 Instructional interactions

The way a teacher interacts with students and also the way students interact with themselves (self-reflection) and with others can influence promoting CT skills. When students' thoughts and content knowledge are challenged by asking critical questions, students learn to develop and evaluate their thoughts, to be open-minded, and to respect the other point of view (Barnett, 2015, pp. 70-71; Henderson et al., 2015, p. 1690; Williams, 2005). Critical questions are questions that "force an individual to defend their own reasons and construct rebuttals" (Henderson et al., 2015, p. 1677). They constitute a framework to guide the thinking process and to force students to justify their believes. They can be asked in general terms to challenge students' knowledge: *How do you know? What is your evidence? What is wrong with this explanation? What is flawed about this model? How can this argument be made more convincing?* Or in the specific domain, e.g., science: *what is wrong with Bohr atomic model or what are the critical challenges which help to establish that day and night is not caused by a moving sun*? (Henderson et al., 2015, p. 1690).

Moreover, it is discussed that providing discussion opportunities in the class and mentoring (giving feedback/guidance to students individually) can have a great positive effect on improving students' CT skills (Chinn & Malhotra, 2002; Garside, 1996; Henderson et al., 2015, p. 1690). Furthermore, changing the role of teacher as instructor to a "model of thoughtfulness" who acknowledges the difficulty in solving many of the critical thinking tasks, and encourages students to think about alternative solutions or strategies (McBride & Bonnette, 1995, p. 385) is discussed as an approach to improve CT skills. For example, McBride and Bonnette (1995) suggest a framework for at-risk students, in which teacher modeling and group collaboration play an important role. They implement the instruction with 43 boys in an outdoor summer camp and engage them actively in group collaboration and CT to solve the problems in games class. They conclude that teacher's role as facilitator who demonstrates applying thinking strategies, engages students in thought-process and self-reflection by asking critical questions, and helps them to provide a rational for decision making can improve students' CT skills (p. 387).

In addition, peer interaction and group discussions are discussed to improve students' argument analysis skills (see Anderson & Soden, 2001) by engaging students in argument analysis tasks and asking them to draw and evaluate conclusions from available evidence (Williams et al., 2004, p. 39).

3.2.2.3 First Principles of Instruction Model

Summarizing the main features of instruction to improve CT skills from the literature in search of an appropriate theory for instructional design shows that instruction should:

- Focus on solving an authentic problem.
- Demonstrate CT skills and provide practice opportunities for students.
- Provide opportunities for discussion, ask critical questions, and provide feedback.

In light of these features, Merrill's First Principles of Instruction (2013) is considered appropriate for CT instructional design. This model is the result of a synthesis of instructional design theories and provides practical guidelines (*cf.* Tiruneh et al., 2018) by providing checklists for instructional design activities and a variety of examples (first aid course, art, Excel course, history, etc.). In addition, this model is used in designing a domain-specific instruction in physics (specifically electricity and magnetism) to improve students' general and domain-specific CT skills (Sermeus et al., 2021; Tiruneh et al., 2017). However, the results showed the effectiveness of instruction in improving domain-specific CT skills, but not general CT skills (Sermeus et al., 2021; Tiruneh et al., 2018; Tiruneh et al., 2017).

The Model of First Principles of Instruction (Merrill, 2002, 2013) is based on five principles: problem-centered, activation, demonstration, integration, and application (Figure 3.1). Merrill (2013) argues that an effective instruction is one where

- Learners acquire skill in the context of real-world problems (**problem-centered**).
- Learners' prior knowledge is considered as a foundation for the new knowledge (**activation**).
- Learners observe a demonstration of the skills they need to solve the problem (**demonstration**).
- Learners apply their newly acquired knowledge and skills to solve the problem (**application**).
- Learners reflect on, discuss, and defend their new knowledge and skills (**integration**).

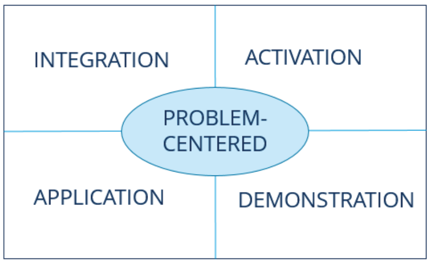


Figure 3.1: First Principles of Instruction (Merrill, 2013, p. 22).

Furthermore, Merrill (2013) emphasizes on the importance of giving feedback, asking reflection questions, and peer-interaction while students work to solve problem centered in the center of a domain-specific content.

3.2.3 A proposal for CT instruction

In the light of the integrated model proposed by Moseley et al. (2005), the CT instruction should include Halpern's (1998) 4-part model for teaching CT and Merrill's (2013) First Principles of Instruction model. A clear analysis of both models shows that there is overlap between the two models, which is helpful when creating CT instruction, as this overlap brings more consistency between the components included in the instruction, strengthening its structure as a whole.

For instance, "structure training activities" from Halpern's model involves practicing the structural aspects of problems and arguments in different types of examples, which overlaps with Merrill's "demonstration" principle. Furthermore, incorporating "structure training activities" and "metacognitive monitoring" from Halpern's model into an instructional design enables the transfer of skills across domains. Similarly, the principles pf "problem-centered" and "application" in Merrill's model serve the goal of skill transfer in instruction, since students are expected to apply their skills to solve an authentic problem that is the focus of the course.

In addition, "cognitive monitoring" emphasizes embedding reflection questions into instruction, which is aligned with the principle of "integration", which states that learning is enhanced when learners reflect on, discuss, and defend their skills. The importance of teacher-student and peer-interaction is also recognized.

Moreover, this overlap reduces the number of components that should be inserted into CT instruction and consequently reduces the complexity of designing CT instruction. Adhering to the clarity and simplicity of this proposal (incorporating Merrill's model with Halpern's model) improves the chance of its application by teachers and instructors. Therefore, incorporating the First Principles of Instruction model (Merrill, 2013) and the 4-part model (Halpern, 1998) in designing a domain-specific CT instruction seems to be a good proposal, and appropriate implementation of this proposal can improve students' CT skills. In addition, explicit teaching of general CT skills at the beginning of the course is considered important for improving CT skills. The practical application of these ideas is discussed in chapter 6, where a CT instruction in particle physics is developed.

3.3 Evaluation of CT instruction

There is theoretical and empirical evidence in the literature supporting the effectiveness of instruction based on the First Principles of Instruction model (Merrill, 2013, pp. 393-435). Halpern's (1998) model is also based on solid theoretical and empirical findings. Therefore, goal of this study is not to evaluate these theories or models upon which domain-specific CT instruction is based. Rather, it is to evaluate the effectiveness of an application and implementation of these models, which are consistent with the theories but can be influenced by many variables, such as the instructor's perception of these models, the subject matter, the characteristics of the students, and also the characteristics of the teacher.

Theory suggests that domain-specific CT instruction designed based on an authentic problem progression (Merrill's framework) and an explicit focus on CT skills (Halpern's framework), along with planning teacher-student and student-student interactions, can improve students' CT skills. To evaluate the effectiveness of the developed instruction in promoting students CT skills, clear criteria, appropriate evaluation methods, and accordingly appropriate data collection methods are needed. In the following sections, these aspects are discussed.

3.3.1 Evaluation criteria

The term evaluation in the educational context is defined as a systematic assessment of the quality of instruction in terms of its intended goals (Merrill, 2013, p. 370; Nieveen & Folmer, 2013, p. 158). With regard to a systematic evaluation, the goals of instruction should be specified, and the evaluation criteria should be defined (Lesh et al., 2008, p. 145).

Generally speaking, to evaluate whether an instruction has the potential to achieve its goals or in the other words whether an instruction is a high quality instruction, it is necessary to consider whether it is relevant to the educational problem or to the need at hand, whether it is logically designed, and whether it is practically applicable (McKenney & Reeves, 2019, p. 167; Nieveen & Folmer, 2013, p. 155). This introduces the criteria of high quality instruction such as **effectiveness, relevance, consistency, and practicality** (Doyle & Ponder, 1977; Nieveen, 1999, 2010).

Since designing a high quality instruction is desired, it is necessary to evaluate instruction at different phases of its development, e.g., its design (development) phase and its implementation. In order to gain insight into the quality of the instruction and, if necessary, get revision decision, it is necessary to conduct ongoing evaluation before implementing the instruction with a large group of students (Merrill, 2013, p. 376).

Therefore, it is important to conduct formative assessment in the design (development) phase. Here, professional review by subject-matter experts can ensure the accuracy of the content and the validity of the instruction, and furthermore, interaction between learners and instructor in small groups can identify the need for modification of the instruction (Merrill, 2013). In other words, in the design phase, the evaluation criteria like relevance, consistency, expected effectiveness, and expected practicality should be applied to improve the quality of the instruction (Nieveen & Folmer, 2013) while in the implementation phase it is important to determine the actual practicality and actual effectiveness of the instruction.

To evaluate the actual effectiveness of instruction in the implementation phase, students' skills in solving complex tasks concerning the goal of instruction should be assessed (Merrill, 2013 p. 370; Resnick, 1987). Furthermore, since the effectiveness of instruction cannot be evaluated in isolation from its implementation, teachers' perceptions and the way they implement the instruction should be evaluated. Therefore, **fidelity of implementation** must be also considered as a criterion to interpret the effectiveness of an instruction (Mowbray et al., 2003, p. 333; Songer & Gotwal, 2005, p. 4; Stokhof et al., 2019).

3.3.1.1 Relevance, consistency, practicality, and effectiveness

As alluded before, a high quality instruction should be relevant to the educational problem, be logically constructed, be usable in a real setting, and result in the desired outcomes. Furthermore, the quality of instruction should be evaluated at both its design and its implementation phase. Therefore, the following criteria can be considered (Nieveen, 1999; Nieveen & Folmer, 2013). Each criterion addresses a particular aspect of the instruction and its fulfilment requires answering a corresponding question:

- Relevance (content validity): Is the content of the course based on scientific knowledge?
- Consistency (construct validity): Are the tasks and teaching sequences logically designed?

- Practicality:
  - Expected: Are the tasks and teaching sequences expected to be usable in a real setting?
  - Actual: Are the tasks and teaching sequences usable in a real setting?
- Effectiveness:
  - Expected: Are the tasks and teaching sequences expected to result in the desired outcomes?
  - Actual: Do the tasks and teaching sequences result in the desired outcomes?

In addition, based on the evaluation criteria, the evaluation methods and accordingly the empirical data collection methods must be clarified. Building on earlier research on approaches to formative evaluation (Nieveen, 1997, 1999), four major evaluation methods are distinguished: developer screening, expert appraisal, micro-evaluation, and tryouts (McKenney & Reeves, 2019, pp. 173-7; Merrill, 2013, p. 390; Nieveen & Folmer, 2013, pp. 161-2). Depending on which phase the instruction is in (design phase or implementation phase), different evaluation methods can be conducted, which are often combined (McKenney & Reeves, 2019, p. 173; Merrill, 2013, p. 390). These evaluation methods and corresponding data collection methods are discussed as the following (McKenney & Reeves, 2019; Merrill, 2013; Nieveen & Folmer, 2013):

- Developer screening: the content, construct, expected practicality, and expected effectiveness of the instruction are critically assessed by members of the design group (e.g. instructor and supervisor). Data is usually collected through checklists that include the required features of the instruction.
- Expert appraisal: the content, construct, expected practicality, and expected effectiveness of instruction are discussed by a group of experts (e.g. external experts in the specific domain). The experts "walk through" the instruction and provide feedback. Data is usually collected through the experts' tinkering with the resource itself, interviews, or checklists.
- Micro-evaluation: the instruction is implemented with small target group outside its normal user setting (e.g. outside school and/or teaching lesson by researcher)

to evaluate its actual practicality and actual effectiveness. Data is usually collected through interviews, assessment, questionnaire, observing, discourse analysis, and participants' logbooks.

- Tryout: the instruction is implemented in its normal setting (e.g. teacher teaches lessons in the regular class in the school) to evaluate its actual practicality and actual effectiveness. To evaluate its practicality, data is usually collected through observation, interviewing, and administrating a questionnaire. To evaluate its effectiveness in achieving the desired outcomes, participant performance should be assessed. Data is usually collected through a test or participants' reports.

It is clear that evaluation methods like developer screening, expert appraisal, and micro-evaluation are appropriate for the development phase of instruction, while tryout is applicable in the implementation phase.

3.3.1.2 Fidelity of implementation

Instruction should be implemented by teachers and the way teachers implement instruction can influence the effectiveness of instruction. Therefore, examining the extent to which teachers implement instruction in accordance with how instruction should be delivered is important for both assessing instruction and assessing student learning (Mowbray et al., 2003, p. 333; Songer & Gotwals, 2005, p. 4). To this goal, fidelity is defined as "the extent to which delivery of an intervention adheres to the protocol or program model originally developed" (Mowbray et al., 2003). To assess fidelity of implementation, it is necessary to examine the extent to which the implementation of instruction is consistent with the underlying theories used to design the instruction (structural fidelity) and the extent to which the quality of the implementation of instruction is comparable to the intended implementation (process fidelity) (*cf.* Mowbray et al., 2003; Songer & Gotwals, 2005; Stokhof et al., 2019).

Structure fidelity

The structure fidelity of an instruction is operationalized as adherence to the structural components of the intervention and duration of implementation of the structural components of the intervention (Lynch, 2005; Songer & Gotwals, 2005; Stokhof et al., 2019). Teachers' adherence to the instruction can be evaluated by measuring whether the instruction was delivered consistently with how it was designed (Lynch, 2005). Furthermore, the information about the length of time given to each lesson can help understanding of structure fidelity of the instruction (Lynch, 2005).

Process fidelity

The process fidelity of an instruction is operationalized as the relevance, practicality, and effectiveness of the instruction for improving students' CT skills (Doyle & Ponder, 1977; Nieveen, 1999, 2010; Stokhof et al., 2019). Here teachers' perception of the relevance, practicality, and effectiveness of the instruction can be considered relevant (Lynch 2005). Relevance addresses teachers' perceptions of an instruction's ability to achieve desired goals (Nieveen, 1999). Practicability refers to teachers' perceptions that working with the instruction was possible within practical limits of time, resources, and knowledge (Nieveen, 2010). Effectiveness refers to the perceived support of instruction in relation to the goals of instruction (Doyle & Ponder, 1977).

To evaluate the structure and process fidelity of an implementation, it is necessary to develop appropriate instruments such as questionnaires or interview questions for data collection. Furthermore, analyzing video data of the implementation of instruction in a real setting can help to draw conclusions about structure and process fidelity (*cf.* Stokhof et al., 2019). In addition, pretest and posttest administration is considered relevant, as studies show that high fidelity of implementation can lead to higher scores on the posttest (Lynch, 2005), possibly showing that students are learning more (Songer & Gotwals, 2005).

3.3.2 Evaluation approaches

To assess the effectiveness of a CT instruction and therefore to assess changes in students' critical thinking, it is important to use more than one measure, both quantitative and qualitative (Behar-Horenstein & Niu, 2011; Resnick, 1987; Williams, 2005). The quantitative approach emphasizes administering pre- and posttests to assess the transferability of the skills, while the qualitative approach focuses on analyzing student discourse and tasks to verify that students learn and apply the intended skills.

McPeck (1990) argues that the statistical results of a pretest and posttest alone are not sufficient to evaluate the effectiveness of an instruction and that it is therefore important to focus on the qualitative evidence in order to interpret how the instruction actually works (pp. 60-63). Using multiple measures of CT (e.g. measurement of student and teacher perceptions of CT, analysis of student essays, interviews, observations) in addition to an appropriate standardized test helps triangulate results for interpreting CT development (Behar-Horenstein & Niu, 2011; Ennis, 1993, p. 186; McMillan, 1987, p. 15; Spicer & Hanks, 1995; Williams, 2005, p. 173). This increases the opportunity to identify factors affecting critical thinking development (Behar-Horenstein & Niu, 2011) and improve also the validity and reliability of evaluation (Merrill, 2013, p. 371).

3.3.2.1 Quantitative approach

In the field of education and human cognition, there are a large number of assessment instruments covering a wide range of formats, origins, psychometric properties, and areas of application (Abrami et al., 2008; Ennis, 1993, p. 183; Spicer & Hanks, 1995). Standardized CT tests (see below) are one such assessment tool used to evaluate the effectiveness of CT instruction. The decision about an appropriate test depends, of course, on the intended outcomes of CT instruction. Considering those outcomes, it is sometimes necessary to develop a test. In this case, it is important to consider how to systematically develop a valid and reliable test that reflects the intended CT outcomes of the instruction (Ennis, 1989; Henderson et al., 2015; McPeck, 1990; Resnick, 1987, p. 32).

General and domain-specific critical thinking test

As alluded before, in science both general and domain-specific CT skills should be taught. Regarding this, a domain-specific CT instruction in physics should include both general and domain-specific CT skills. To evaluate the development of general CT skills a general CT test is required that focuses on the application of critical thinking skills in everyday life (e.g. Halpern Critical Thinking Assessment). To evaluate the development of domain-specific CT skills a domain-specific CT test is required that focuses on the application of CT skills in a specific domain of physics that requires content expertise (e.g. CTEM in the context of electricity and magnetism).

There are some general CT tests which are mostly employed to evaluate the effectiveness of a CT instruction, e.g. the Cornell Critical Thinking Test– Level Z (CCTT), the California Critical Thinking Skills Test (CCTST), the Ennis-Weir Critical Thinking Essay test, the Watson-Glaser Critical Thinking Appraisal, and the Halpern Critical Thinking Assessment (HCTA) (Abrami et al., 2008; Abrami et al., 2015; Spicer & Hanks, 1995; Tiruneh et al. 2014). The most commonly cited reason for using these standardized tests is that they are valid, reliable, and also widely used (Behar-Horenstein & Niu, 2011).

It is obvious that the CT skills targeted by each test differ depending on the definition of CT skills on which the test is based. Tiruneh et al. (2017) review the five general CT tests mentioned above based on criteria such as clarity of definition and more commonalities between the targeted skills in the test across all tests. Among these five general CT tests, the Halpern Critical Thinking Assessment (HCTA) was selected as best meeting the criteria to assess the general CT skills.

Halpern Critical Thinking Assessment (HCTA)

Halpern Critical Thinking Assessment (HCTA) is a standardized general CT test that focuses on the five CT skills introduced in Halpern's (2009, 2014) definition of CT: Verbal Reasoning, Argument Analysis, Thinking as Hypothesis Testing, Likelihood and Uncertainty Analysis, and Decision Making and Problem Solving. The target group is respondents aged 15 years and older. Answering the test items requires the application of these five CT skills in various scenarios of daily life, e.g., shopping, exercising, looking for a job, and so on. Each item focuses on assessing a desirable outcome of applying a CT skill in order to interpret respondents' CT skills. However, it is important to keep in mind that CT skills cannot be considered as stand-alone skills, since solving many real-world tasks require the application of several of these skills (Halpern, 2016, p. 8). Therefore, one cannot draw a line between the different skills. They are interconnected, and focusing on one desired outcome means that the main focus of the task is on that particular outcome, while the other CT skills and outcomes can also be considered.

The HCTA has two forms: open-ended (S1 and S3) and multiple-choice (S2 and S4) forms. For each form, there are two versions of the HCTA that focus on the same CT skills but in different scenarios. This allows researchers to use the HCTA as a pre- and post-test. The open-ended form of the HCTA evaluates the active application of critical thinking in everyday situations and conversations, where respondents must consciously search for and select appropriate knowledge and skills from their own memory to construct an answer (Halpern, 2016, p. 11) while the multiple-choice form evaluates the passive application of critical thinking, where respondents must select the appropriate response from a predetermined list of alternatives (p. 11).

The main variable in the HCTA is critical thinking (CT) which is sum of the five CT sub-skills: Verbal Reasoning, Argument Analysis, Thinking as Hypothesis Testing, Likelihood and Uncertainty Analysis, and Decision Making and Problem Solving. The validity and reliability of the HCTA are measured by administrating the test to numerous diverse samples and the results showed a high validity and reliability of the test (see Halpern, 2016; see also Verburgh et al., 2013).

Domain-specific critical thinking test

To evaluate the acquisition of domain-specific CT skills, it is necessary to develop a valid and reliable domain-specific test (Behar-Horenstein & Niu, 2011; Ennis, 1989, p. 9; Resnick, 1987, p. 34). This test can help the instructor to interpret about how the instruction is actually working (Abrami et al., 2015; Behar-Horenstein & Niu, 2011; McMillan, 1987, p. 15; McPeck, 1990, p. 61; Spicer & Hanks, 1995, p. 9).

This domain-specific CT test should include open-ended items (Behar-Horenstein & Niu, 2011; Resnick, 1987, p. 34; Spicer & Hanks, 1995) which give students chance to apply the content knowledge they gained in a specific domain e.g. physics and also the CT skills. However, a review of several domain-specific CT tests (see Tiruneh et al., 2017) illustrates that some domain-specific CT skills such as Lawson's (1978, 2004) Classroom Test of Scientific Reasoning and the Critical Thinking Test in Biology developed by McMurray in 1991, assess general CT skills and do not require domain-specific expertise, which contradicts the definition of domain-specific CT (Davies, 2013 p. 530; McPeck, 1990). Content expertise is the focus of proponents of the domain-specific nature of CT and cannot be ignored in such a CT test.

A recent effort has been done by Tiruneh et al. (2017) to design a domain-specific CT test in the specific domain of physics, electricity and magnetism. As mentioned above, by comparing some mostly used general CT tests based on criteria such as clarity of definition and more commonalities between the targeted skills in the test across all tests, Tiruneh et al (2017) selected the Halpern Critical Thinking Assessment as best meeting the criteria and therefore they considered HCTA as a framework to design a domain-specific CT test in physics, the Critical Thinking test in Electricity and Magnetism (CTEM, Tiruneh et al., 2017). The target group is university students, and the test is designed to assess students' domain-specific CT skills in the five Halpern-defined skills. Answering the tasks requires the application of these thinking strategies and also of content knowledge acquired while attending the electricity and magnetism lecture at the university. The validity and reliability of the CTEM are measured by administrating the test to second-year university students whose major was mechanical engineering (N=45) and the results showed an acceptable validity and reliability of the test (see Tiruneh et al., 2017).

3.3.2.2 Qualitative approach

Analysis of student performance at different stages of implementation by analysing video and audio data, students' written works, and students' interviews provides insight into how their CT skills are developing. To this end, there are many qualitative data analysis methods (see LeCompte et al., 1992). One of the methods is the constant comparative method used in the development of grounded theory (Glaser & Strauss, 2017; Strauss & Corbin, 1998, 1990). This method uses inductive data analysis to develop conjectures about the development of students' CT skills. Based on a constant comparison between the conjectures, the conjectures are revised (Lobato, 2008, p. 184) and can finally describe the process of students' CT skill development. Employing the constant comparative method through open coding helps to establish accurate evidence of students' CT development and also provides theoretical insight into the learning process (Glaser & Strauss, 2017; Strauss & Corbin, 1998, 1990; Woods, 1992). A more detailed description of the constant comparative method, along with other approaches to qualitative data analysis, is covered in the next chapter (see section 4.2.3).

Furthermore, analysing teachers' perception on the effectiveness of CT instruction in improving students CT skills provides insight into how CT instruction might support teachers in teaching CT skills (Stokhof et al., 2019). To this end, qualitative analysis of data collected through interviews, questionnaires, or video-recordings of classroom activities is helpful.

In summary, qualitative analysis of data from students' work and teacher perceptions can provide insight into the relevance, effectiveness, and practicality of CT instruction in two directions: helping students learn critical thinking and helping teachers teach critical thinking. Besides that, quantitative data analysis helps to triangulate findings to strengthen the validity of data interpretation.

4 Design-based research

Design-based research (DBR) is a methodological framework (Bakker, 2018, pp. 6-8) in educational studies in which design and research are intertwined to bridge the gap between the practice and theory of education (Bakker, 2018; Barab & Squire, 2004; Collins et al., 2004; Design-Based Research Collective, 2003; McKenney & Reeves, 2019). It is referred to as a methodological framework because it is a “genre of flexibly using existing approaches for the purpose of gaining design-based insights and research-based designs” (Bakker, 2018, p. 7).

This chapter discusses the features of DBR, the steps to conduct DBR in educational studies to develop and refine instruction, relevant research tools, and evaluation methods. In addition, adherence to validity and reliability criteria to improve the quality of research is argued.

4.1 Design-based research (DBR) and its features

Design-based research (DBR) is defined as “a series of approaches, with the intend of producing new theories, artifacts, and practices that account for and potentially impact learning and teaching in naturalistic settings” (Barab & Squire, 2004, p. 2). However, the terminology which is used to describe this “series of approaches” is labeled differently but they are similar in nature. For example, the Design Based Research Collective (2003) uses the term *design-based research* while McKenney and Reeves (2019) use the term *educational design research* (for a historical overview see Bakker, 2018, pp. 23-45), but in both cases, they discuss the kind of research that happens in the educational context, to solve complex problems by using systematic approach including multiple iterations of design-enactment-analysis-redesign (Design Based Research Collective, 2003; McKenney & Reeves, 2019).

DBR takes place in real-world settings and interacts with practice from the beginning. The extent of this interaction increases over time and can contribute to the introduction of practical solutions (McKenney & Reeves, 2019). Empirical testing of these solutions contributes to the development of theories about teaching and learning (Bakker, 2018; Cobb et al., 2003; McKenney & Reeves, 2019). Therefore, DBR has a two-sided function: developing theoretical insights and presenting practical solutions to solve educational problems (Bakker, 2018; Barab & Squire, 2004; Collins et al., 2004; Herrington et al., 2007; McKenney & Reeves, 2019). Since these practical solutions emerge from theoretical claims about teaching and learning, implementing the design in a real setting and analyzing the data can contribute to the refinement or development of these theories (Design-Based Research Collective, 2003). There are some examples of best practices of DBR which indicate clearly this interplay between theory and practice (for finding inspiring examples see Bakker, 2018, p. 280; see also Collins et al., 2004; Design-Based Research Collective, 2003; McKenney & Reeves, 2019, p. 264).

Through the design of educational materials or environments as an integral part of research, DBR can help learners achieve specific goals, (Bakker, 2004, 2018; Design-Based Research Collective, 2003), especially in the domain-specific content with little prior research (Cobb et al., 2003). The novelty of the problem for which a solution is sought strengthens the decision to use DBR (Design-Based Research Collective, 2003).

For example, although teaching general and domain-specific CT through the development of domain-specific CT instruction is important, there are rarely concrete guidelines and principles for designing a high-quality instruction to reach these goals (*cf.* Sermeus et al., 2021; Tiruneh et al., 2018). However, the search for high-quality CT instruction, as discussed in section 3.2.2, has contributed to a proposal model for CT instruction that integrates Merrill's (2013) and Halpern's (1998) models, but using these models as a framework for constructing CT instruction in the abstract domain of particle physics is a novel and complex task. The novelty and complexity of the work therefore indicate that a systematic approach is needed to develop and optimize domain-specific instruction in an iterative cycle of design-enactment-analysis-redesign (Bakker, 2018; Collins et al., 2004; Design-Based Research Collective, 2003; McKenney & Reeves, 2019; Nieveen, 2010). The findings of the study may also contribute to the development of the theories for developing CT instruction.

4.2 Conducting DBR in education: Design and evaluation of an instruction

From the above discussion, the following characteristics can be attributed to DBR (*cf.* Bakker, 2018 following Cobb et al., 2003, pp. 17-18; Collins et al., 2004; McKenney & Reeves, 2019, pp. 12-18):

- DBR focuses on both theory and practice.
- DBR is in interaction with practice from the start and the scope of interaction increases over time.
- DBR has an iterative character, i.e. the design of the solution and its refinement form a cyclic process.

McKenney and Reeves (2019) incorporated these features into their developed generic model for the implementation of DBR in education, which is the result of a synthesis of existing models for DBR and is also inspired by theories of instructional design and curriculum development (p. 82). The generic model is represented in figure 4.1.

McKenney and Reeves (2019) introduce three main phases for DBR: analysis and exploration, design and construction, evaluation and reflection. As indicated in figure 4.1, these three main phases involve interaction with practice and contribute to the theoretical understanding and development of the intervention (McKenney & Reeves, 2019, p. 84). Furthermore, the iterative nature of DBR is represented by the arrows between different elements in figure 4.1.

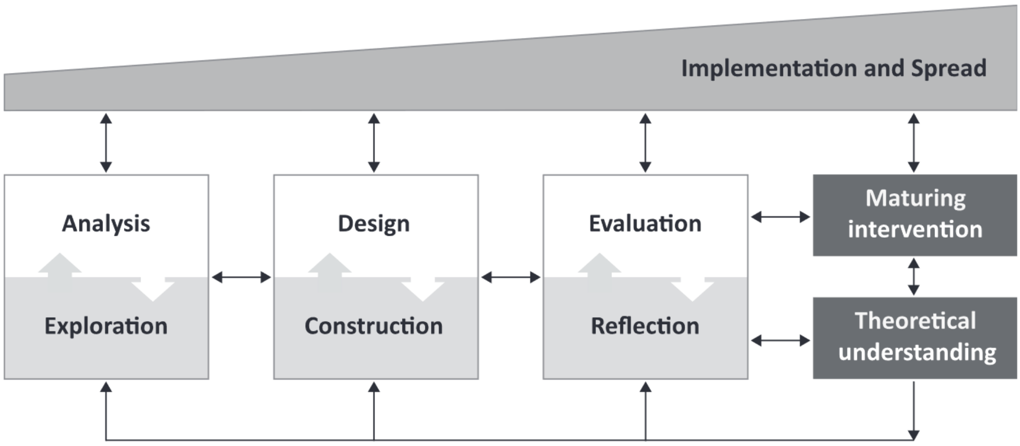


Figure 4.1: Generic model for conducting design research in education (McKenney & Reeves, 2019, p. 83).

In addition, McKenney and Reeves (2019) introduce a cyclic representation for the overall process of DBR to discusses that "within one larger study, several sub-studies take place, each with its own complete cycle of inquiry and sound chain of reasoning" (p. 15). Figure 4.2 shows this cyclic representation of a sample design-based research project.

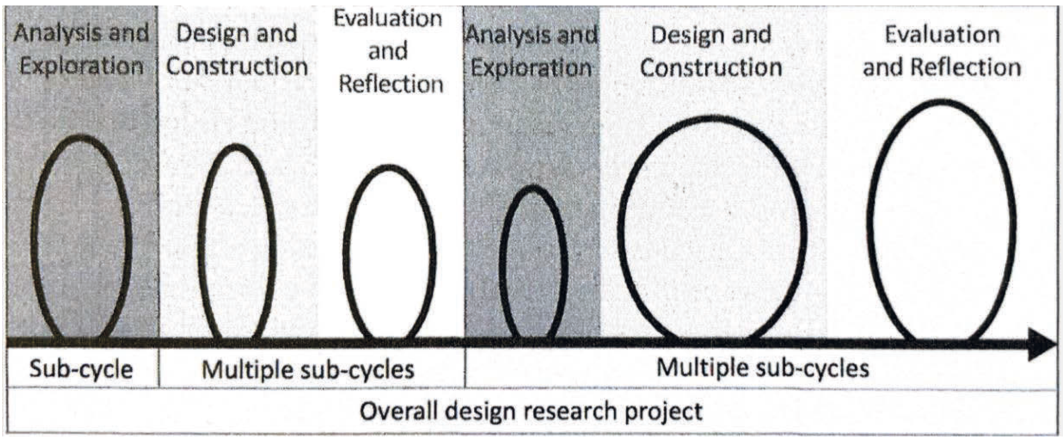


Figure 4.2: Single and multiple sub-cycles in one educational design research project (McKenney & Reeves, 2019, p. 84).

Figure 4.2 shows a research project included a total of six cycles, that each cycle is representing a core phase of DBR, and overall the research project is conducted in two studies. Loops of different sizes indicate variation in scope e.g., time, participation (McKenney & Reeves, 2019, p. 241). Assuming that the research project aims to design effective instruction for teaching CT, the first study involves designing, implementing, and evaluating the first version of instruction through the core phases of analysis and exploration, design and construction, and evaluation and reflection. Reflecting on what worked and what did not give the researcher(s) insight into the challenges, opportunities, and unanswered questions to focus on when conducting the second study of the research project to improve the quality of instruction.

DBR, however, does not always occur in the context of such macro-level analysis between different studies. For example, Cobb et al. (2001) report micro-level analysis in which they revise learning activities after each lesson and use the revised version in the next lesson. The scope of using DBR can vary depending on the goal and time.

In addition, both analytical and creative perspectives are needed when conducting core processes in DBR. While the analytical perspective promotes rationality, the creative perspective motivates openness (McKenney & Reeves, 2019, p. 92). That means to carry out DBR, researcher should play consciously two roles: detective and inventor (McKenney & Reeves, 2019, p. 89). These two perspectives are explicitly considered in the generic model (McKenney & Reeves, 2019), with an interplay between the two roles represented by the arrows inside each square (core phase) in Figure 4.1.

When conducting DBR to solve an educational problem, each core phase includes some activities. In the following, the relevant activities from both analytical and creative perspectives are discussed.

4.2.1 Analysis and exploration phase

The main goal of this phase is to identify and define the problem. In this phase, the analytical perspective uses literature review, context analysis, and needs assessment to define the problem, while the creative perspective finds out and learns how others have looked at and solved similar problems through professional meetings and networking

(Bakker, 2018, p. 60; Collins et al., 2004; Herrington et al., 2007; McKenney & Reeves, 2019, pp. 90-125). In addition, looking at DBR as an opportunity to enrich theory, identifying the gap between theory and practice at this stage can provide some initial insight into the direction in which a contribution to theory can be made (Bakker, 2018).

4.2.2 Design and construction phase

During design, the ideas emerging from the analysis and exploration phase are mapped to make them more practical, the decision is made on what should be designed to solve the problem (e.g. tasks, teaching-learning sequences, or a learning environment), and the initial design principles are formulated (Bakker, 2018, p. 60; McKenney & Reeves, 2019, pp. 126-160).

During construction, the solution is created and revised through a process of prototyping (McKenney & Reeves, 2019). These solutions can be an actual representation in physical form (e.g. a teacher's guide) or an indirect representation (e.g. process guidelines for a particular approach to teaching) (McKenney & Reeves, 2019).

In addition, this phase contributes to theory building by formulating and developing design principles, conjectures, and/or hypothetical learning trajectories (Bakker, 2018; McKenney & Reeves, 2019). These are research instruments that can be useful in different phases of DBR (Bakker, 2018). They are discussed in detail in the following section.

4.2.2.1 DBR and its contribution to the theory

As noted earlier, one of the characteristics of DBR is that it contributes to the development of teaching-learning theories and the development of designs (teaching-learning sequences, materials, learning environment) to support learning (Cobb et al., 2003). This contribution to theory building is achieved through the formulation and development of design principles, conjecture maps, and/or hypothetical learning trajectories (Bakker 2018; McKenney & Reeves, 2019).

Conjecture maps provide a theoretical model, design principles guide the design of a teaching learning environment, and hypothetical learning trajectories inform the implementation of teaching-learning sequences. All can be tested in the evaluation process.

Conjecture map

Conjecture map is a valuable research instrument in DBR. At the beginning of research, it plays the role of guidelines for conducting DBR systematically, and at the end, after its development during cycles of empirical research, it contributes to theory development, for example, by being formulated as design principles (Bakker, 2018; Collins et al., 2004; McKenney & Reeves, 2019; Sandoval, 2014).

The "Conjecture map" proposed by Sandoval (2014) illustrates how the theoretical ideas for achieving the desired outcomes should work together (Bakker, 2018; McKenney & Reeves, 2019; Sandoval, 2004, 2014). Conjecture mapping require an extensive literature review on the design of the solution for the educational problem (McKenney & Reeves, 2019; Sandoval, 2014) and provide general insights into for example how domain-specific instruction with specific features (content, teaching-learning sequences, and materials) might contribute to the development of students' CT skills.

Figure 3 shows a generalized form of a conjecture map illustrated by Sandoval (2014, p. 27) in the context of promoting scientific argumentation in elementary science. As it shown in figure 3, the conjecture map includes following elements (Sandoval, 2014):

- High-level conjecture is a general idea about how to support learning (first box from left in figure 3).
- Embodiment that includes the features of instruction like tools and materials, activity structures, participant structures, and discursive practices.

- Mediating process or mechanism that describes the process caused by the features of instruction and is assumed to result in the desired outcomes.
- Intervention outcomes

Furthermore, a conjecture map introduces two types of conjectures that relate the elements and can also be tested empirically (Bakker, 2018; McKenney & Reeves, 2019; Sandoval, 2014). The first are design conjectures that link explicitly the features of instruction to the mediating process to describe "if learners engage in this activity (task + participant) structure with these tools, through this discursive practice, then this mediating process will emerge" (Sandoval, 2014, p. 24). This type of conjecture can be tested by analysis of interaction through identifying the occurrence of an expected mediating process and tracing the process back to the features of the instruction (Sandoval, 2014, p. 24). In figure 4.3, the design conjectures are the arrows between embodiment of a design and the mediating process (Bakker, 2018, pp. 56-57).

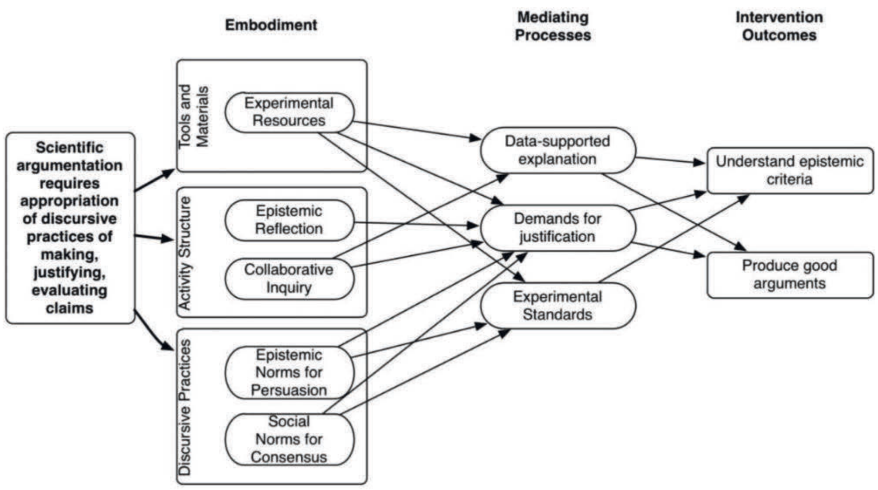


Figure 4.3: Illustration of a conjecture map in the context of promoting scientific argumentation in elementary science (Sandoval, 2014, p. 27).

The second type of conjectures, *theoretical conjectures* describe "if this mediating process occurs it will lead to this outcome" (Sandoval, 2014, p. 24). This type of conjecture can be tested by analyzing the interaction to trace the path from outcome to process, and also by assigning an appropriate measurement tool for the desired outcomes (Sandoval, 2014, p. 24). In figure 4.3, the theoretical conjectures are the arrows between mediating processes and outcomes (Bakker, 2018, p. 57).

Design principles

Design principles are conceptualized as guidelines (Bakker, 2018, p. 52), usually accompanied by reasons (*cf.* Van den Akker, 2013), and thus represent a contribution of DBR to theory (Bakker, 2018; McKenney & Reeves, 2019). Like conjecture maps, design principles evolve during iteration cycles, and at the end of the study they offer advice in an actionable form to explain where and when an education solution can solve similar problems (Bakker, 2018, pp. 49-54).

Design principles have hypothetical nature (Bakker, 2018) and predict under which circumstances with which design features, the learning outcomes can happen: "if <design feature D>, then <learning outcome L>" (Greeno, 2016, p. 635). Furthermore, design principles can also be empirically tested and therefore, evolve during cycles of iteration. For example, Kali (2008) describes how quantitative and qualitative analysis of data in her project on *involving students in evaluation process to give feedback to their peers* led to development of design principles (pp. 423- 438).

Van den Akker (2013, p. 67) formulates design principles in argumentative form grounded in scientific knowledge and practical experience as follows: "

- *If you want to design intervention X [for purpose/function Y in context Z]*
- *then you are best advised to give that intervention the characteristics C1, C2, ..., Cm* [substantive emphasis]
- *and to do that via procedures P1, P2, ..., Pn* [methodological emphasis]
- *because of theoretical arguments T1, T2, ..., Tp*
- *and empirical arguments E1, E2, ..., Eq*"

(Van den Akker, 2013, p. 67, emphasis in original).

To illustrate the above formula in more practical terms, Bakker (2018) presents an example in his book (pp. 49- 50) that is adapted from Vervoort's (2013) work with pre-service teachers:

> "if you want to design an intervention in which rich media cases are used to support student teachers putting innovative pedagogical insight into practice, you are advised to:
>
> - use exemplary cases that link practice to underlying theory, because this allows teachers to conceptualize feasible practice.
> - connect the use of rich media cases to a guiding task, because this encourages active learning and focuses the teachers' attention on the content of a case.
> - encourage dialogue about the content of the rich media cases, because case-discussion is essential for meaning making.
> - initiate a dialogue between a student teacher and a mentor teacher on exemplary cases about innovative pedagogical insights, because mentor teachers are able to make their practical knowledge explicit. This contributes to the interpretation of the cases and the application of the case content in student teachers' teaching practice."
>
> (Bakker, 2018, pp. 49-50, citations in the text left out).

Furthermore, it is important to note that design principles are suggestive and require local adaptation to be implemented in the new situation (Bakker, 2018; McKenney & Reeves, 2019, p. 20).

Hypothetical learning trajectory (HLT)

Both conjecture maps and design principles incorporate general ideas and specify important design features (Bakker, 2018), but when it comes to describe the sequential development of students' skills, the hypothetical learning trajectory (HLT) is a useful tool (Bakker, 2018; Simon, 1995). It is inspired by studying students' challenges regarding learning a domain-specific content or skill and predicts the way students getting support to reach learning goals. It includes three components in designing relevant tasks and activities toward achieving learning goals: the learning goal, the learning activities, and hypothetical learning process (Bakker & van Eerde, 2015; Simon, 1995).

In the literature, HLT is also referred to as "hypothetical teaching-learning trajectories" (*cf.* Bakker & van Eerde, 2015) in order to emphasize more the role of teachers in the student learning process. Simon was a teacher himself and did not feel the need to emphasize this when he used the term "hypothetical learning trajectory" (Bakker, 2018, p. 59).

Bakker and van Eerde (2015) describe the HLT as a research instrument that has different functions in the different phases of the DBR and is continuously evolving (p. 439). For example, while in the design phase the HLT is developed based on its three components and guides the design (of teaching-learning sequences and materials), in the implementation it guides enactment of the design, and in the evaluation phase it guides analysis of empirical results (Bakker & van Eerde, 2015, pp. 438-440).

However, due to the task-oriented nature of HLT, it takes too long to report the results of a DBR study. It is more popular to convert the results into conjecture maps for a research-oriented audience and/or into design principles for a practice-oriented audience (*cf.* Bakker & Smit, 2018).

4.2.3 Evaluation and reflection phase

When prototyping the solution to an educational problem (e.g. prototyping a goal-oriented domain-specific instruction) in the design and construction phase, it is the turn of empirical investigation and thus reflection on the findings (Bakker, 2018; McKenney & Reeves, 2019). The aim is to evaluate the quality of instruction in terms of achieving its objectives and thus, on the one hand, to improve the quality of instruction and, on the other hand, to contribute to theory (Bakker, 2018; Collins et al., 2004, p. 34; McKenney & Reeves, 2019; Sandoval, 2014).

Furthermore, evaluation of instruction can involve both formative and summative evaluation strategies (Collins et al., 2004, p. 39; Design-Based Research Collective, 2003; McKenney & Reeves, 2019, p. 166). For example, a formative evaluation is conducted if the focus of the evaluation is to identify potential opportunities to improve the design (Design-Based Research Collective, 2003; McKenney & Reeves, 2019), while a summative evaluation is conducted to assess the value of the design in achieving the desired outcomes (Collins et al., 2004; McKenney & Reeves, 2019).

In addition, the research question determines the evaluation approach and thus the evaluation criteria. For example, Bakker (2018) discusses that answering the research question of this type, "*How can a teaching-learning strategy with characteristics Ci support students to learn G?*" requires evaluating the instruction using criteria such as relevance, consistency, practicality, and effectiveness (pp. 104-106). These criteria as well as the relevant evaluation and data collection methods were already discussed in section 3.3.

The following section discusses the evaluation methods used to assess the effectiveness of instruction in achieving its goals under DBR, as well as how to meet validity and reliability criteria in evaluation.

4.2.3.1 Evaluation methods

Dependent on the goal of instruction, the retrospective analysis of data can include a different spectrum of methods. For example, investigating the instruction's goals in behavioral terms and the extent to which these were achieved, one compares the HLT created in the design and construction phase with empirical results (*cf.* Bakker & van Eerde, 2015, pp. 438-440; *cf.* Dierdorp et al., 2011). One uses the constant comparative method (*cf.* Bakker, 2004) or the case study (*cf.* Smit, 2013). These methods can be combined or applied individually, the important aspect is that the results are formulated in such a way that they can contribute to theory building. As always, it is also important to consider validity and reliability when conducting retrospective analyses to improve the quality of the research.

In addition, it is highly advisable to keep an open mind during data analysis to avoid narrowing to predefined study objectives (Bakker, 2018; Denscombe, 2014; McKenney & Reeves, 2019) and to see the interesting scenarios that might emerge from retrospective analysis and provide a new perspective on the study subject.

Constant comparative method

The constant comparative method is considered the core of qualitative analysis in the grounded theory approach and in other types of qualitative research (Boeije, 2002; Cho & Lee, 2014). It is based on comparing samples and looking for commonalities and differences in their behaviors, reasons, attitudes, and perspectives (Boeije, 2002) to connect codes and categories resulting from inductive data analysis and to develop a substantive theory (Cho & Lee, 2014; Glaser & Strauss, 2017). This is aligned with one of the goals of DBR, that is, to contribute to development of theory, and is therefore an appropriate method for qualitative data analysis in DBR (Cobb et al., 2003).

In the context of DBR, the constant comparative method looks for evidence of specific constructs in the intervention, or its implementation (McKenney & Reeves, 2019; Woods, 1992) to describe coherently the processes leading to the intended outcomes; namely to generate process-oriented explanation (Cobb & Gravemeijer, 2008; Glaser & Strauss, 2017). For example, Bakker and Gravemeijer (2004) used the constant comparative method in a DBR project in the context of mathematics to investigate *how informal reasoning about distribution can be developed in a technological learning environment*. They identified patterns of student answers that showed the evolving learning trajectory according to students' reasoning with the representation used (Bakker & Gravemeijer, 2004). Bakker (2018) explained the approach in the following way:

> "First, all transcripts were read and the video tapes were watched chronologically, episode-by-episode. With the HLT and research questions as guidelines, conjectures about students' learning and views are generated and documented, and then tested against the other episodes and other data materials (students' work, field note, assignments). More concretely, this testing implied looking for confirmation and counter-examples. The process of conjecture generating and testing was repeated for the whole data set. Seemingly crucial episodes were discussed with colleagues to test whether they agreed with our interpretation or could perhaps think of alternative interpretations. This process is called peer examination." (pp. 62-63)

In addition, Boeije (2002) introduced a systematic approach to the constant comparative method by clarifying the analysis activity in each step of the comparison (e.g. open coding, axial coding, triangulating), its aim, guiding questions, and results. Conducting qualitative data analysis using this step-by-step approach promised to increase the traceability and credibility of the analysis (Boeije, 2002, p. 406).

Furthermore, it is worth mentioning that the progress in the comparative process in the context of grounded theory requires theoretical sampling: that means the analysis of and reflection on previous data guides what data must be collected next in order to answer the arisen questions (Boeije, 2002; Glaser & Strauss, 2017). The comparison stops if adding new cases does not bring new information to the formed categories in the process of data analysis, the so-called saturated categories are formed (Boeije, 2002; Glaser & Strauss, 2017). In the context of retrospective analysis of data involving multiple cases, the saturated categories can be achieved by analyzing the data at different levels of analysis: within case analysis, within same data source analysis, and cross-case analysis (*cf.* Cho & Lee, 2014).

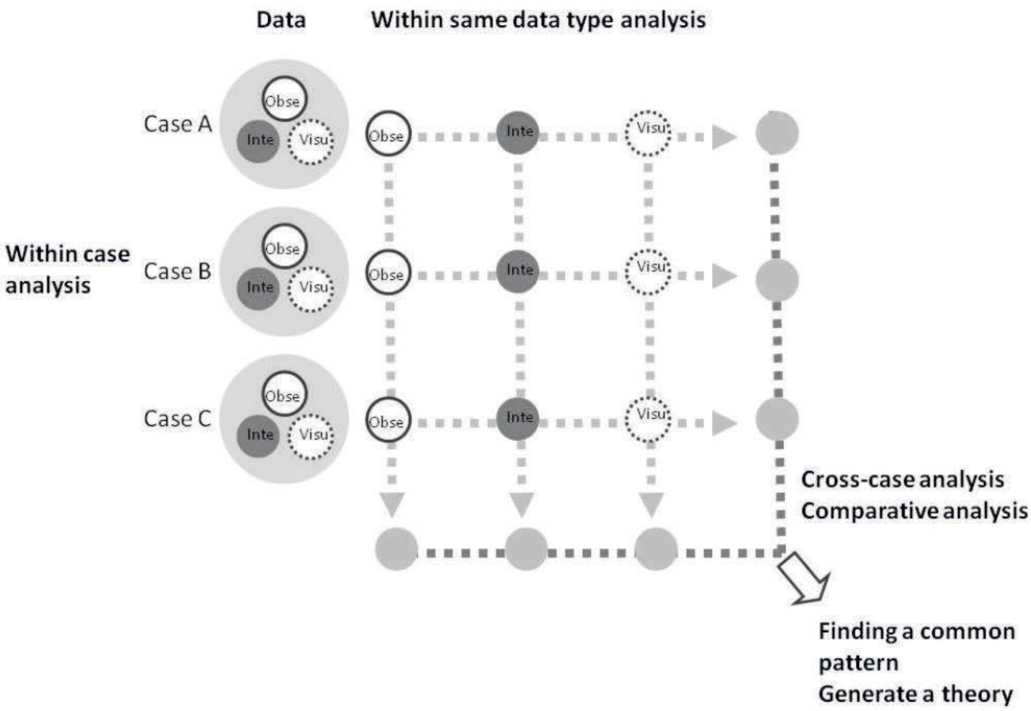


Figure 4.4: Data analysis procedure of within and cross-case analysis (Cho & Lee, 2014).

Since the procedures in the constant comparative method are similar to those of the qualitative content analysis method as both involve coding and identifying themes and patterns to analyze and interpret qualitative data both methods have to be distinguished (Cho & Lee, 2014).

Cho and Lee (2014) argued that the constant comparative method in the context of grounded theory approach focuses on finding relationships between categories to develop a theory that is grounded in the data, namely substantive theory, while qualitative content analysis focuses on extracting categories from the data and finding themes from categories to describe the meaning of the qualitative material. In other words, while in qualitative content analysis the information lies in the categories, in the constant comparative method the information lies (also) in the similarities and differences of the categories, especially in cross-case analysis. Furthermore, the constant comparative method uses an inductive approach in data analysis, while qualitative content analysis uses inductive or deductive approaches or a combination of both approaches (Cho & Lee, 2014).

So the constant comparative method is generally applied where still no theory exists (Cho & Lee, 2014; Glaser & Strauss, 2017). This dissertation focuses on the study of the processes leading to the development of CT in the context of particle physics. Although the literature discusses that CT involves attitude, knowledge, and some thinking skills (*cf.* Halpern, 2009, p. 5), it lacks a process-oriented explanation of CT development. Therefore, the goal is to develop a coherent explanation of the process of developing CT skills from the comparative analysis of different sources of data gathered during implementation of the antimatter course. Furthermore, constant comparison of the data enables to explain how each design element of CT instruction affects the development of CT skills.

Case study

Case study is one of the methods of qualitative data analysis in education research to examine a case (or cases) in-depth in its real-life context (Khan & VanWynsberghe, 2008; Yin, 2013). There is much literature on how to conduct a case study and how to improve the validity and reliability of the results (*cf.* Khan & VanWynsberghe, 2008; Yin 2009, 2012, 2013). However, the main goal here is to discuss the case study approach in the context of DBR.

Depending on the research question in DBR, a case study or/and a cross-case study can be conducted. For instance, if the variation among students' progress is to be investigated, a cross-case study is an appropriate method (*cf.* Smit, 2013), as it allows comparing the similarities and differences between cases (Khan & VanWynsberghe, 2008). On the other hand, when more practice oriented knowledge is needed that relates teaching and learning, an in-depth case study (or multiple case studies) is appropriate (Bakker & Smit, 2018; Yin, 2012). For example, Smit et al. (2016) report the results of a case study to explain how teachers can scaffold to achieve a specific teaching-learning goal.

However, generalizing the results of a case study analysis might be a concern (*cf.* Denscombe, 2014), Yin (1989) argues for analytic generalization and answers this concern as the following:

> "Case studies, like experiments, are generalizable to theoretical propositions and not to populations or universes. In this sense, the case study, like the experiment, does not represent "sample", and in doing a case study, your goal will be to expand and generalize theories (analytic generalization) and not to enumerate frequencies (statistical generalization)." (p. 21)

Comparing Hypothetical Learning Trajectory (HLT) with Actual Learning Trajectory (ALT)

As alluded before, HLT is a task-oriented research instrument that plays different roles in different phases of DBR (see section 4.2.2.1). While in the design phase, it guides developing teaching learning materials and activities, in the evaluation phase, it is used as an indicator to evaluate empirical results.

After implementing the instruction and collecting data, a comparison of students' actual learning during the different tasks (actual learning trajectory) with the HLT can help making sense of the data and on this basis discussing the quality of instruction (based on the quality criteria discussed in section 3.3). To this end data analysis matrix introduced by Dierdorp et al. (2011) is useful. This method helps to identify problematic sections in teaching-learning materials and consequently to improve the design (*cf.* Dierdorp et al., 2011). Figure 4.4 shows a generalized representation of data analysis matrix offered by Bakker and van Eerde (2015, p. 442) adopted from Dierdorp et al. (2011).

Hypothetical Learning Trajectory			Actual Learning Trajectory		
Task number	Formulation of the task	Conjecture of how students would respond	Transcript excerpt	Clarification	Match between HLT and ALT: Quantitative impression of how well the conjecture and actual learning matched (e.g., -, 0, +)

Figure 4.5: Data analysis matrix for comparing hypothetical learning trajectory (HLT) with actual learning trajectory (ALT) (Bakker & van Eerde, 2015, p. 442).

Insights into how conjecture map and actual learning align and why certain learning tasks do or do not occur help to improve instructional design in subsequent cycles (micro-level) and across subsequent implementations (macro-level) (Bakker & van Eerde, 2015).

4.2.3.2 Validity and reliability

As indicated earlier, DBR is used to solve a complex educational problem and, consequently, to contribute to an empirically grounded theory. The results of retrospective analysis of data help to achieve these goals. They can improve the quality of instruction at the micro and macro levels by examining the reasons why a learning task occurs or does not occur as well as making the knowledge useful to researchers and teachers by formulating the results in terms of design principles and conjecture maps.

However, an important question regarding the data analysis and drawing conclusions in the retrospective analysis is still open: the reliability of the data analysis and the validity of the conclusions. In the context of DBR, Bakker and van Eerde (2015) distinguish between internal and external validity and reliability and introduce some techniques to improve them (pp. 443-445). Table 4.1 provides an overview of the definition of validity and reliability and applicable techniques to improve them.

Table 4.1: Overview of validity and reliability in design-based research (DBR) in terms of definition and possible improvement techniques (adopted from Bakker & van Eerde, 2015, pp. 443-445).

Quality criteria	**Definition** (Bakker & van Eerde 2015, pp. 443- 445)	**Improvement Techniques**
Internal validity	"refers to the quality of the data and the soundness of the reasoning that has led to the conclusion."	- generating conjectures for specific episode and testing them for other episodes (searching for counter example) - triangulating data
External validity	"is interpreted as the generalizability of the results."	* Generalizability of the results is a criticism of DBR.
Internal reliability	"refers to the degree of how independently of the researchers the data are collected and analyzed."	- Collecting video- and audio data (using objective devices) - Peer examination (discussing critical episodes)
External reliability	"means that the conclusion of the study should depend on the subject and condition, and not on the researcher."	- Reporting research transparently in such a way that is clear how the research has been carried out and how the conclusions have been drawn from the data.

As shown in Table 4.1, the problematic part of the reliability and validity of DBR results is related to the generalizability of the results. This is one of the most common criticisms of DBR. In addition to concerns about the generalizability, there are also concerns about causality. In the following, both concerns and the ways in which design-based researchers attempt to overcome them are discussed.

Generalizability

As commented in Table 4.1, generalizability of the results is a criticism of DBR (Bakker, 2018; Design-Based Research Collective, 2003; Herrington et al., 2007; McKenney & Reeves, 2019). Design-based researchers respond to this critique by discussing analytic generalization and case-to-case generalization as forms of generalization applicable in DBR to improve the validity of the results (Bakker, 2018; Design-Based Research Collective, 2003; McKenney & Reeves, 2019).

Analytic generalization aims to develop theories (Design-Based Research Collective, 2003; Yin, 2013), while case-to-case generalization considers transferability of ideas from one study to another (Bakker & van Eerde, 2015; McKenney & Reeves, 2019). For example, Bakker and van Eerde (2015) present a new type of learning activity in mathematics classrooms to improve students' conception of distribution and discuss how it was successfully implemented in a new research project in another country (p. 444).

Furthermore, it is discussed that in case-to-case generalization, the validity of conclusion can be improved if the following aspects are taken into account: how comparable are the cases in external terms (e.g. type of intervention, type of school, type of learner)? what arguments were made in the first study and are they applicable in the new study? (McKenney & Reeves, 2019, p. 21).

Causality

In addition to the concern about generalizability, another criticism of DBR is the concern about causality. This criticism comes from both experimentalists and non-interventionist qualitative researchers who criticize claims about supporting learners in achieving specific educational goals with the voice, "how do you know this is due to your design?" Therefore, although DBR is all about causality (Bakker, 2018; Design-Based Research Collective, 2003), design researchers often avoid causal claims and present their research as descriptive or evaluative (Bakker, 2018, pp. 277).

In addition, design-based researchers discuss that the use of case studies or the constant comparative method in DBR clarifies differences and improvements and therefore does not require an experimental or control group (Bakker, 2018, pp. 277-280). Furthermore, triangulation of data from multiple sources provides “critical evidence to establish warrants for claims about why outcomes occurred” (Design-Based Research Collective, 2003).

4.3 Summary

In summary, design-based research is a genre of research used to solve novel and complex educational problems by designing the solution, implementing it in the real-world environment, analyzing the data, and using the results to improve the quality of the solution. In addition, the iterative nature of DBR gives researchers the opportunity to put their improvement ideas that emerged from the data analysis of one iteration into practice in the next iteration.

Furthermore, DBR contributes to theory development when results are communicatively formulated in the form of conjecture maps and design principles. In this case, the validity and reliability of the research enhances the power of sharing the results with researchers and teachers; the power of replicability by researchers and applicability by teachers.

In particular, in the context of physics education, DBR may be a promising approach if physics education research is to be successful in terms of its impact on teaching and its applicability by teachers. For example, Hake (2008) presents examples of research in physics education that are the result of development, testing, and refinement efforts over many cycles of application in a real-world setting and discusses how the results of these studies are being used to improve the effectiveness of teaching (pp. 493-508).

Part II

Empirical Study

5 Research questions

Although it is important to teach students to think critically, there is a lack of empirically tested guidelines for designing a CT instruction. Specifically, in the context of particle physics, there is no systematic research to design instruction for teaching critical thinking. The complexity of particle physics on the one hand and critical thinking on the other make the combination of the two concepts difficult to explore.

From the literature review, the combination of Merrill's (2013) First Principles of Instruction and Halpern's (1998) model of CT is assumed to improve students' critical thinking skills (see section 3.2.3). However, to implement this proposal in the context of particle physics and to explore the results, a systematic investigation is needed. To this end, the **main research question (MRQ)** is formulated as:

> ***MRQ****: How does a teaching-learning strategy about antimatter based on the First Principles of Instruction and Halpern's model of Critical Thinking support students of grades 10, 11 & 12 to learn general and domain-specific Critical Thinking skills?*

To answer this question, an evaluation study is required to assess the effect of the antimatter course in promoting students' general and domain-specific CT skills from two perspectives: students' perspectives and teachers' perspectives.

Considering the evaluation of the effectiveness of the antimatter course regarding **students' perspective** results in **two sub-research questions**:

> ***RQ 1.1****: What is the effectiveness of the antimatter course in improving students' general and domain-specific Critical Thinking skills?*

> ***RQ 1.2**: How do the students' general and domain-specific Critical Thinking skills develop (during the course) in relation to the teaching-learning activities and features of the antimatter course?*

The first sub-research question (RQ 1.1) aims to find evidence for the development of students' general and domain-specific CT skills in the antimatter course. The second sub-research question (RQ 1.2) aims to describe how course elements contribute to the development of students' general and domain-specific CT skills.

In addition, since the antimatter course is developed to be implemented in schools, the results of assessing how teachers implement it and perceive its effectiveness and practicality can be used to draw conclusions about the effectiveness of the developed course. Considering **teachers' perspective** results in the **second research question**:

> ***RQ 2**: What is the relevance, practicality, and perceived effectiveness of the antimatter course from the teachers' perspective with respect to promoting students' general and domain-specific Critical Thinking skills?*

Answering the above mentioned research questions helps to provide an empirically grounded model for CT skills development and improve knowledge about the nature of CT in a domain-specific context and the transfer of skills from a domain-specific context to a general context.

6 Design and methodology of the study

This research focuses on how to improve students' general and domain-specific CT skills in the particle physics context. This implies a need to design and develop a domain-specific instruction that supports improving students' general and domain-specific critical thinking skills. To systematically design such domain-specific instruction (teaching-learning strategy) and optimize it, the Design-based research (DBR) approach is adequate.

To evaluate the effectiveness of instruction in terms of student achievement in critical thinking (**to answer research question 1.1**), a qualitative analysis of student activities and interactions was required. This qualitative analysis was conducted using the constant comparative method to provide a coherent explanation of the process of developing students' CT skills by providing accurate evidence. To triangulate the findings about the effectiveness of the instruction obtained from the qualitative analysis, a summative evaluation was also conducted. To this end, it was important to develop (if needed) and administer a valid and reliable critical thinking assessment.

Furthermore, to argue how and why the design skeleton of the instruction contribute to the development of students' general and domain-specific CT skills (**to answer research question 1.2**), a qualitative analysis of the lessons was conducted.

In addition, to gain an insight to teachers' perspective on the relevance, practicality, and perceived effectiveness of the instruction (**to answer research question 2**), developing a teacher questionnaire was necessary.

In this chapter, the design and development of domain-specific instruction and relevant research instruments are discussed. Furthermore, the evaluation methodology and adherence to quality criteria are explained.

6.1 Design and development of instruction according to Design-based research

For conducting DBR, the generic model (McKenney & Reeves, 2019) introduced three main phases: analysis and exploration, design and construction, evaluation and reflection. To design a domain-specific instruction in particle physics for improving students' CT skills, each phase of the DBR sought to answer the following questions (*cf.* Bakker, 2018; *cf.* McKenney & Reeves, 2019):

Analysis and exploration: what are the challenges students face in applying CT skills in physics? What is already known about how to design domain-specific instruction in physics to improve students' CT skills? What is the need to design domain-specific instruction in particle physics?

Design and construction: what do improved CT skills in particle physics look like that should be established as learning objectives of instruction? What are the principles and criteria for designing domain-specific instruction in particle physics that would improve students' CT skills? how well were the design criteria met?

Evaluation and reflection: What challenges emerged after each implementation that provided guidance for improving the design of the instruction? What would improved design look like for the next implementation?

Figure 6.1 shows a general overview of the design process of the present study based on DBR approach to answer research questions in chapter 5.

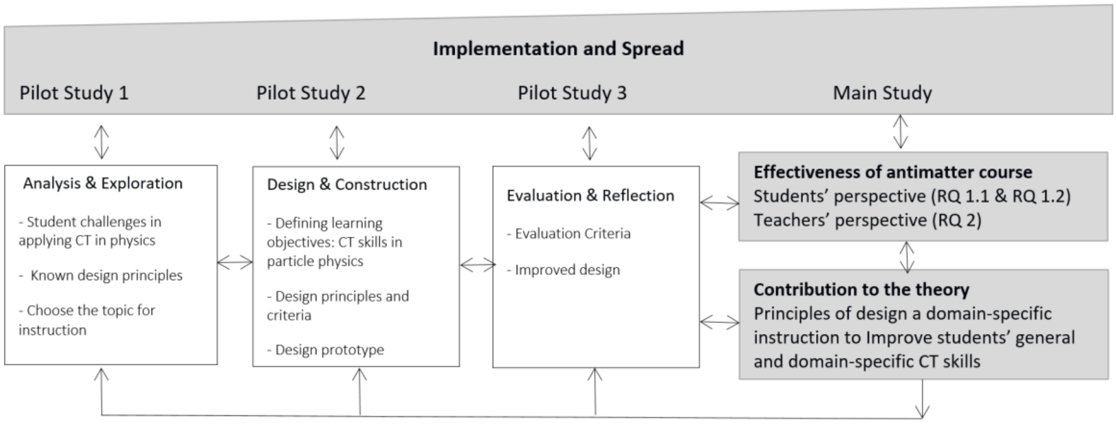


Figure 6.1: General overview of the design of the present study based on Design-based research approach (adopted from McKenney and Reeves, 2019).

6.1.1 Analysis and exploration phase

The literature review on students' challenges in thinking critically in physics (see section 2.4) showed that even if students have positive attitude toward learning physics as they see physics as an opportunity to think differently and critically (Levrini et al., 2016), they are not able to think critically when faced with a task that calls on these skills. Table 6.1 Shows students' difficulties in applying various CT skills in physics which are already in section 2.4 discussed. The paradox between students' attitudes toward learning physics and their performance on a physics task requiring CT indicates a lack of appropriate physics instruction to improve students' CT skills.

In addition, the literature review on designing domain-specific instruction in physics to improve students' CT skills pointed to efforts in electricity and magnetism (Sermeus et al., 2021; Tiruneh et al., 2016). Merrill's (2013) First Principles of Instruction model was used to design the course. Results showed the effectiveness of the instruction in improving domain-specific CT skills, but not general CT skills (Sermeus et al., 2021; Tiruneh et al., 2017; Tiruneh et al., 2018). The present study focused on a combination of Merrill's (2013) First Principles of Instruction model and Halpern's (1998) four-part model for teaching CT to design an instruction in particle physics to improve students' general and domain-specific CT skills (see discussion in section 3.2).

Table 6.1: Students' difficulties in applying various CT skills in physics.

Critical thinking skill	**Students' challenges in physics** **(Lack of the ability of:)**
Verbal Reasoning	• Applying non-ambiguous (clear) terminologies
Argument Analysis	• Making coherent argument • Considering enough reasons to support conclusion
Thinking as Hypothesis Testing	• Identifying variables • Explaining the relationship between variables • Controlling variables
Likelihood and Uncertainty Analysis	-
Decision Making and Problem Solving	• Constructing meanings of the problem statement • Linking internally the meaning of the statements in the problem statement • Moving from more general to more detailed aspects of a problem

Criteria for choice of topic

In the domain of physics, the sub-domain of particle physics was chosen as the focus for designing a CT instruction. The embedding of particle physics into the German curriculum has been the focus of physics education in Germany for several years (see Passon et al., 2020). It is discussed that teaching particle physics not only provides students with domain-specific content knowledge (e.g. students learn that the proton and neutron are not elementary particles), but can enhance students' understanding of the nature of science, the history of science, scientific methods, and even personal development and responsibility in society (Wagner, 2020; Zügge, 2020; Pospiech, 2020). It is also discussed that since the terminologies of particle physics have entered society (e.g. the term antimatter appears in the science fiction book Angels & Demons (Dan Brown, 2000) and the movie of the same name (Sony Pictures, 2009)), it is important that students be able to evaluate the information they receive from the media and society and take a stand for and against it based on valid and reliable content knowledge (Wagner, 2020, pp. 4-5). However, there is a lack of concerted guidance and design principles for developing high-quality CT instruction in physics in general and particle physics in particular.

Particle physics covers a wide range of content (e.g. standard model, accelerators, detectors, interactions, antimatter). When choosing the topic for the design of the CT instruction, it was also important to consider the interest of students. Since critical thinking is an effortful task, the likelihood of engaging in such a challenging activity can be increased if students are interested in the topic.

The results of a study on the development of students' interest in particle physics (*cf.* Gedigk et al., 2017) showed that providing an authentic and challenging context in particle physics increases students' interest in engaging in higher-level knowledge in particle physics. There is also a special interest in the context of antimatter (personal communication with Gedigk), especially when it involves the series of open questions for scientists (Pospiech, 2020) such as "the Big Bang should have created equal amounts of matter and antimatter. Why then is there much more matter than antimatter in the universe?".

Moreover, studies modeling teachers' content knowledge in particle physics (*cf.* Oettle & Mikelskis-Seifert, 2020) showed that antimatter (and antiparticles) is considered by both teachers and particle physics experts to be the second most important topic that teachers need to learn in the context of particle physics in order to teach in schools in the future.

Therefore, considering antimatter as the topic of the CT instruction could fulfill the following criteria:

- The topic provides the opportunity for applying CT skills in an authentic context.
- The topic is interesting for students.
- The topic is relevant for teacher education.

6.1.2 Design and construction phase

In order to map the ideas derived from the literature for developing an antimatter course to improve students’ CT skills, the target group and learning objectives were defined. Furthermore, attempts were made to fulfill the criteria for a high quality intervention (see section 3.3.1) in the design of the teaching-learning sequences and the materials. In addition, the design principles were formulated.

6.1.2.1 Target group

The study was conducted in Germany. The German high school system (Gymnasium) includes two levels: secondary level 1 (from grade 5 to 10) and secondary level 2 (mostly grades 11 and 12; in some cases, includes grade 13).

Since teaching antimatter requires prior content knowledge (e.g. particles’ behavior in electric and magnetic fields) and this knowledge is covered in the secondary level 2 physics curriculum, the main target group was considered to be students of grade 11 and 12. However, in special cases, such as highly gifted students, grade 10 students could also be part of the target group. In this case, the necessary content knowledge was taught and practiced before teaching its practical application in the antimatter context.

Furthermore, to make the course feasible for teachers to implement during the school year, the course included 10 to 12 lessons based on the students' grade level.

6.1.2.2 Learning objectives

The overall learning objective of the antimatter course was to develop and improve students’ general and domain-specific CT skills on the basis of Halpern’s (2009, 2014) classification. Consequently, the observable outcomes of the application of these CT skills were identified as the target of the antimatter course in the development of students’ general CT skills and were also used as the basis for defining the domain-specific CT skills. For example, a general CT goal of the course was that students recognize the misuse of definitions in any context (school subjects or everyday life) and look for precise terminologies, and a domain-specific CT goal was defined as that students recognize the ambiguity of terms in the definition of "antimatter."

Table 6.2 gives an overview of some desirable outcomes of applying CT skills in a general context and also in the specific context of antimatter.

Table 6.2: General and domain-specific outcomes of applying critical thinking skills.

Critical thinking skills	**General outcomes of applying critical thinking skills (Halpern, 2009)**	**Domain-specific outcomes of applying critical thinking skills in antimatter**
Verbal Reasoning (VR)	○ Detecting misuses of definitions ○ Using precise terminologies ○ Using questioning and paraphrasing to comprehend text	• Interpreting the results of an experiment (Tiruneh et al., 2017) • Defining the term "mass" according to its relation to the "rest energy" • Detecting ambiguity of terms in the definition of "antimatter"
Argument Analysis (AA)	○ Identifying key parts of an argument like premises, reasons, and conclusion ○ Identifying counterargument ○ Making strong argument in which reasons support conclusion ○ Judging the credibility of an information source ○ Recognizing the missing components in an argument	• Identifying relevant information missing in an argument • Giving reasons to conclude that the moon is not composed of antimatter • Making a position for/against an argument and giving reasons • Making counterargument
Thinking as Hypothesis Testing (HT)	○ Recognizing the need for and using operational definitions (recognize the need to define a term and specify how to measure it) ○ Understanding the need to control variables in order to make strong causal claim ○ Being able to describe the relationship between any two variables as positive, negative, or unrelated ○ Checking for adequate sample size before drawing a conclusion	• Controlling variables to draw conclusion about the particle types in a cloud chamber • Identifying variables when hypothesizing the construction of the antimatter trap

Likelihood and Uncertainty Analysis (LU)	○ Estimating the probability of occurrence of an event ○ Using probability judgment to improve decision making ○ Understanding the limits of extrapolation (make estimations of future events) ○ Understanding the need for relevant and valid information for being able to conclude about the most probable event	• Predicting the probability of an event occurrence, but understanding the limits • Combining information to evaluate the probabilistic idea of "antimatter as fuel"
Decision Making and Problem Solving (PS)	○ Considering different sorts of solutions or alternatives ○ Selecting criteria to judge possible solutions ○ Recognizing the need to seek counterevidence ○ Examining the relevance of the procedures in solving problems (monitoring the decision making process) ○ Using analogies to solve problems	• Developing reasonable solution for "trapping antimatter" • Recognizing the feature of a problem and adjust solution plan accordingly (Tiruneh et al., 2017) • Deciding on the validity of a particular scientific law (explanation) when applied to a new situation • Identifying relevant information when solving a problem

6.1.2.3 Design of antimatter course

Design principles

Merrill's (2013) First Principles of Instruction model and Halpern's (1998) four-part model for CT instruction were integrated into the design of the antimatter course. Inspired by Van den Akker (2013), the following design principles were defined:

- Teaching general critical thinking skills explicitly at the beginning of domain-specific instruction, because this gives students a clear picture of the CT skills that the course will focus on (Andrews, 2015; Ennis, 1989; Halpern, 1998; Tiruneh et al., 2014).
- Valuing students' efforts to accomplish the CT tasks, and encourage them to think critically, because these improve students' motivation (external and internal) to think critically and consequently their disposition toward CT (Halpern, 1998; Hamby, 2015; Resnick, 1987). → **Halpern's Dispositional Component**
- Defining an authentic problem in the context of antimatter and, accordingly, plan teaching-learning sequences so that students acquire the knowledge and skills necessary to solve that problem, because a problem-centered instruction design based on authentic problem can engage students more in critical action to solve the problem (Barnett, 2015; Chinn & Malhotra, 2002; Merrill, 2013). → **Merrill's Problem-Centered and Application Principles & Halpern's Dispositional Component**
- Activating students' prior knowledge, because this provides a foundation for the newly acquired knowledge (Merrill, 2013). → **Merrill's Activation Principle**
- Providing opportunities in which students are demonstrated various examples of CT tasks to practice CT skills and to reflect on the structural aspects of the tasks, because this improves students' understanding of CT skills, increases their ability

to apply CT skills in a new context (Halpern, 1998; Merrill, 2013), and can develop their mental habits for using CT skills. → **Halpern's Skill Approach to CT, Structure-Training Activities, and Disposition Component & Merrill's Demonstration and Integration Principles**

- Providing discussion opportunities (by asking metacognitive monitoring, and critical questions) in which students reflect on their learning process before beginning the task, while working on the task, and after completing the task, because this helps students develop and evaluate their thought, be open-minded, and respect the other point of view (Barnett, 2015; Halpern, 1998; Merrill, 2013). → **Halpern's Metacognitive Monitoring & Merrill's Integration Principle**

Establishing these design principles helped in creating general ideas for designing the antimatter course. These ideas were discussed with physics education experts and particle physicists and, after minor revision, were formulated as follows:

In the antimatter course:

- General CT skills are explicitly taught in the first session of the antimatter course.
- Some scenes of the fiction movie "Illuminati" (German title of Angels & Demons) (2009) featuring an antimatter bomb are chosen as an authentic problem in the context of antimatter. The decision was made through brainstorming and discussion with particle physicists from IKTP (Institute for Nuclear and Particle Physics) at the Technische Universität Dresden, based on the criteria that the problem must be related to particle physics as well as to the students' everyday life, and requires CT skills.
- The outlines of the course are presented through a discussion with students when they are shown selected scenes from "Illuminati". During the discussion, students are asked to evaluate whether the information in the movie about antimatter is scientifically correct. In this way, students recognize the need for

content knowledge, and based on their requirements, the outlines of the course are established and presented.

- To fulfil Merrill's application and integration principles, in the last session of the antimatter course students are asked to analyse a scenario for a scene from "Illuminati" based on their newly acquired content knowledge and CT skills to correct the information about antimatter in this scene and present and discuss their work to the class.
- Five Halpern's CT skills are embedded in the antimatter course by designing worksheets and relevant tasks.
- In each lesson, discussion and reflection time is planned.

Figure 6.2 shows a schema of the antimatter course.

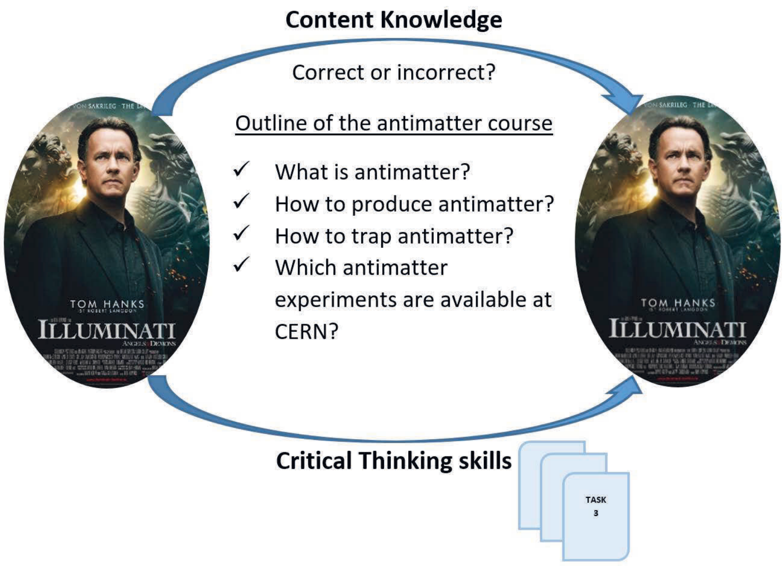


Figure 6.2: General overview of the antimatter course.

Source of picture: https://www.cineman.ch/movie/2009/AngelsAndDemons/

Design of materials (through the design of hypothetical learning trajectories)

To provide opportunities in which students practice CT skills in the antimatter context, appropriate worksheets as teaching-learning materials were designed. Hypothetical learning trajectories (HLTs) were used as an appropriate research instrument (Bakker & van Eerde, 2015) to guide the design of worksheets by specifying the learning goals, the learning activities, and a hypothetical learning process (Bakker & van Eerde, 2015; Simon, 1995; Simon & Tzur, 2004).

Once the content of the antimatter course was outlined, it was decided which CT skills a particular lesson might focus on as learning goals. The decision was inspired by reading scientific papers on the topic of each lesson. For example, the “positron discovery” lesson focused on teaching the Likelihood and Uncertainty Analysis skill, asking students to give alternative interpretations of the sign of the charged particle and the direction of particle motion in Anderson’s (1933) cloud chamber photograph and to evaluate their interpretations to make decision about the most probable interoperation (as Anderson did).

In addition, the students’ prior knowledge was analyzed, as this provides a basis for the hypothetical learning process and clarifies the decision about the specific learning activity. This also fulfills Merrill’s activation principle. Therefore, the physics curriculum for high schools in the Free State of Saxony in Germany (the state where the pilot studies were conducted) was analyzed. However, for the main study, it was necessary to look at the physics curricula of the other German states when the antimatter course was introduced in different schools in Germany.

Table 6.3 shows an example of generated HLT to design the “Anderson’s cloud chamber photograph” worksheet (see appendix H) which was used in “positron discovery” lesson and focused on developing likelihood and uncertainty analysis skill (Sadidi & Pospiech, 2019a; Sadidi & Pospiech, 2020).

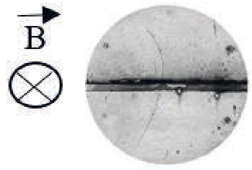
Table 6.3: Hypothetical learning trajectory (HLT) to design the “Anderson’s cloud chamber photograph” worksheet.

Students’ prior knowledge	Lorentz force, conservation of energy and momentum, relationship between radius of the track and momentum and energy of particles, and definition of range in particle physics.
Learning goals	Making multiple possible interpretations. Estimating the likelihood of occurrence of an event(s). Making decision about the most likely event. Recognizing the need for relevant and valid information to be able to conclude about the most likely event.
Hypothetical learning trajectory (process)	HLT 1: Students will apply their knowledge about the Lorenz force and the right-hand rule to **make different interpretations** about the sign of the charged particle and the direction of particle motion. HLT 2: Students will apply their knowledge about conservation laws to **evaluate their interpretations** and to **discuss the likelihood** of a certain type of particle causing the track. HLT 3: Students will **make decision about the most likely event**. HLT 4: Students will discuss and reflect on the activity to **realize the importance of having relevant and valid information for final conclusion**.
Tasks (for the worksheet)	This photograph is the Anderson’s photograph he took with the cloud chamber. The magnetic field lines point vertically in the page. **Task 1.** Write down all possible interpretations about the sign of the electric charge and the direction of motion of type of particle causing the track. **Task 2.** Which interpretation would be more likely? Argue for or against each of these interpretations, using conservation laws.

Figure 6.3 shows the “Anderson’s cloud chamber photograph” worksheet designed based on the hypothetical learning trajectory presented in table 6.3.

Worksheet 3: Anderson's cloud chamber photograph

This photograph is the Anderson's photograph he took with the cloud chamber. The magnetic field lines point vertically in the page.

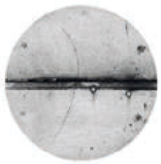
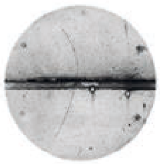
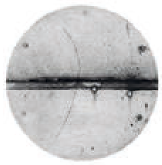
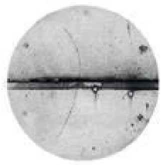


Anderson's cloud chamber photograph 1933, Phys.Rev.43 P.419

Task 1.

Write down all possible interpretations about the sign of the electric charge and the direction of motion of type of particle causing the track in Anderson's cloud chamber photograph by completing the table and drawing the particle's direction of motion in each copy.

(Note: All the following photos are copies of Anderson's cloud chamber photograph and they are the same.)

Interpretation regarding	 Copy 1	 Copy 2	 Copy 3	 Copy 4
sign of the electric charge				
direction of motion				

Task 2.1.
Which interpretation would be more likely?

Task 2.2.
Argue for or against each of these interpretations, using conservation laws.

Figure 6.3: "Anderson's cloud chamber photograph" worksheet.

Table 6.4 shows a general overview of the CT skills focused on each lesson and the corresponding tasks. It is important to note that some CT skills such as verbal reasoning and argument analysis are included in each lesson, as students are always encouraged to use precise terminology and draw a valid conclusion based on reasons. This also applies to decision making and problem solving skill. Table 6.4 lists the examples in which an explicit emphasis was put on CT skill through the design of a worksheet.

Table 6.4: Examples of embedded critical thinking (CT) skills in the antimatter course and corresponding CT tasks.

Critical thinking skill	Lesson or teaching-learning sequence	Domain-specific CT task
Verbal Reasoning (VR)	After completing the antimatter course students are asked to watch a scene from the movie "Illuminati" and analyze the scenario for that scene.	Worksheets 7 & 8 o Identifying the vague terms about antimatter in dialogues in the movie excerpt o Discussing why the terms are vague o Replacing the vague terms with precise ones when analyzing the scenario
Argument Analysis (AA)	Big Bang: The mystery of antimatter Students are asked to watch a video about Big Bang, ask questions, and systematically search to find answers to their questions, using the guidelines provided on worksheet 5 and then present their works to the class.	Worksheets 4 & 5 o Asking clear and precise questions o Gathering correct and sufficient information from valid sources o Drawing logical conclusions o Considering the assumptions made o Using correct and clear terms
Thinking as Hypothesis Testing (HT)	Trap Antimatter Students are asked to make hypothesis about the construction of antiparticle trap and evaluate their own proposal against existing systems.	Worksheet 6 o Making a list of variables that can be considered in constructing an antiparticle trap o Controlling the variables o Making hypothesis about the construction of an "antiparticle trap" o Evaluating the hypothesis about "antiparticle trap" by comparing it with "Penning trap"
Likelihood and Uncertainty Analysis (LU)	Positron Discovery Students are asked to provide alternative interpretations of Anderson's cloud chamber Photograph and assess their validity based on content knowledge and discussion of probability.	Worksheet 3 o Making different interpretations about the sign of a particle's electric charge and the direction of the particle's motion o Evaluating the interpretations and deciding on the most likely event

6.1.2.4 Quality criteria and evaluation methods

To evaluate the quality of the teaching-learning materials, the HLTs and the worksheets were discussed with physics educators and particle physicists at the Technische Universität Dresden based on the criteria of high quality intervention (see section 3.3.1). To this end, the relevance, consistency, expected practicality, and expected effectiveness of the tasks were discussed through the following evaluation methods (Sadidi & Pospiech, 2019b):

- Developer screening: The content of the tasks, their structure, their applicability in the real setting, and whether they could lead to the learning goals were discussed in regular meetings between the researcher and her supervisor and the first revisions were made to the HLTs and worksheets accordingly.
- Expert appraisal: the worksheets and the HLTs were discussed with a group of particle physicists and physics educators to evaluate their content, construct, applicability, and effectiveness. The experts "walked through" the materials, made notes on the HLTs and the worksheet, and provided feedback in the discussion round. This provided the improved version of teaching-learning materials.

The general ideas that emerged from the design principles, along with the worksheets formed the prototype of the antimatter course.

To check the practicality of the first draft of antimatter course, the course was implemented with 8 physics teacher students from the University College of Teacher Education Vienna (Pädagogische Hochschule Wien). Criteria such as the difficulty of the tasks and the time allotted for each task were established to determine whether the tasks and teaching-learning sequences could be used in a setting with high school students. Data were collected during the video recording of the lessons and discussion sessions with physics teacher students. The results of the data analysis provided an estimate of the time spent on each task and indicated how to enhance student understanding of the topic, e.g., by adding pictures, etc.

6.1.3 Evaluation and reflection phase

Once the first draft of antimatter course was developed, it was implemented with the real target group, high school students of grades 10, 11 and 12. In view of the cyclical and repetitive nature of DBR in the development of an intervention, several implementation phases were planned. After each implementation, the collected data were analyzed to identify the improvement points for the antimatter course. Figure 6.4 shows a chronological overview of the implementation phases in the present study.

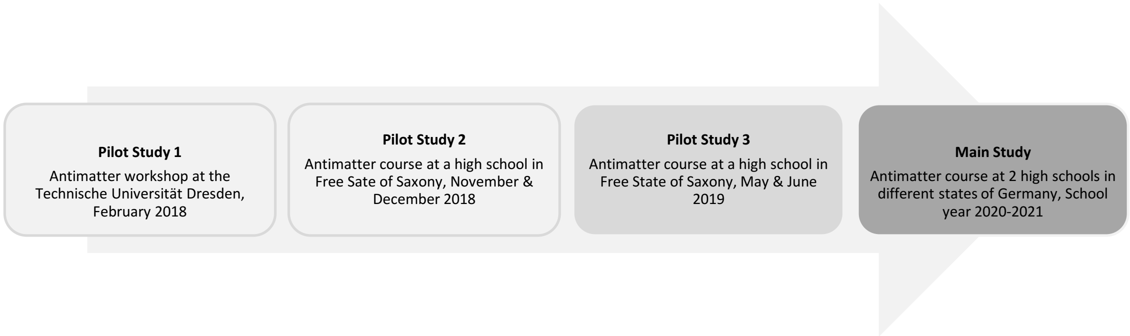


Figure 6.4: General overview of different implementation phases.

Table 6.5 also provides an overview of the grade and the number of students who participated in the study during different implementation phases, the data collection and the evaluation methods (see section 3.3.1).

As shown in table 6.5, the evaluation methods varied. As discussed in section 3.3.1, the goal of both the micro-evaluation and the tryout methods is to evaluate the actual practicality and actual effectiveness of the course, with the only difference being the setting in which the course is implemented. The first pilot study was implemented in the form of a 4-day workshop at the university and was tutored by the researcher. The second pilot study, on the other hand, was implemented in the school, but again was tutored by the researcher. However, in the third pilot study and also in the main study, the course was implemented in high schools and was tutored by the physics teachers in the schools.

Table 6.5: General overview of the grade and number of students participated in the study in different implementation phases.

Implementation phase		**Grade** (Number of students)	**Evaluation method**	**Data collection**
Pilot study 1		**10, 11, and 12** (N=4)	Micro-evaluation	audio, video, student work, pre- and post-tests, cognitive interview
Pilot study 2		**12** (N=28)	Micro-evaluation	audio, video, student work, pre- and post-tests
Pilot study 3		**10** (N= 10 highly gifted students)	Tryout	audio, video, student work, pre- and post-tests, interviews with some students after last lesson, audio recording of discussions with teacher at the beginning and after each session
Main study (3 classes)	Free State of Saxony	**10** (N= 9 highly gifted students)	Tryout	student work, pre- and post-tests, audio and video of some lessons, teacher questionnaire, student questionnaire
	State of Rhineland-Palatinate	**12** (N= 7)		student work, pre- and post-tests, teacher questionnaire, student questionnaire
		10 (N= 17 highly gifted students)		

In addition, the evaluation of the data after each implementation in pilot studies clarified the required adjustments. For example, after the second pilot study, adherence to the practicality of the course in the real class in terms of time convinced to remove a part of the lesson “what is antimatter?” that discussed the discovery of the quark to clarify the structure of matter. Moreover, there were some minor revisions in the structure of the worksheets regarding wording and adding or removing some questions that prepared the **final version of the antimatter course for the main study**.

Furthermore, analysis of the audio and video data and triangulation of the results with student work in the second pilot study helped to develop a teacher package that included the teaching-learning sequences, materials, step-by-step teaching guides, possible student questions during the lesson, additional content knowledge for teachers, and also suggested time for each teaching-learning activity. Data analysis from the third pilot study refined the teacher package and prepared the **final version of the teacher package for the main study**.

6.2 Evaluation of effectiveness of instruction

To evaluate the effectiveness of the antimatter course in improving students' general and domain-specific CT skills **qualitatively**, the constant comparative method was used to provide accurate evidence of the development of students' domain-specific and general CT skills and to present a coherent explanation of the process of developing students' CT skills (answer to research question 1.1). To triangulate the results of qualitative analysis, the **quantitative** data were collected and analyzed. To this end, two open-ended versions of Halpern Critical Thinking Assessment (HCTA) were assigned as pre- and posttests to measure the impact of the antimatter course to the students' development of general CT skills. To evaluate domain-specific CT skills, it was required to develop a specific Particle Physics Critical Thinking (PPCT) test. Both quantitative and qualitative data analysis methods are explained below.

6.2.1 Constant comparative method for qualitative evaluation

In analyzing the qualitative data, the constant comparison method (CCM) was used to firstly provide accurate evidence of the development of CT skills in the antimatter course and secondly to explain the processes of CT skill development. To this end, all types of data including video, audio, student work, student interviews/questionnaires, and teacher interviews/questionnaires were analyzed.

Since the teaching materials and teacher package used in pilot study 3 and the main study did not differ, it was decided to use the pilot study 3 data in addition to the main study to conduct the CCM. The most complete data of this type belonged to three classes. Data from three classes were analyzed, treating each class as a separate case.

Because the study involved three cases (see table 7.1), the CCM was applied in different level of analysis: within case analysis, within same data type analysis, and cross-case analysis. Figure 6.5 shows the data analysis procedure.

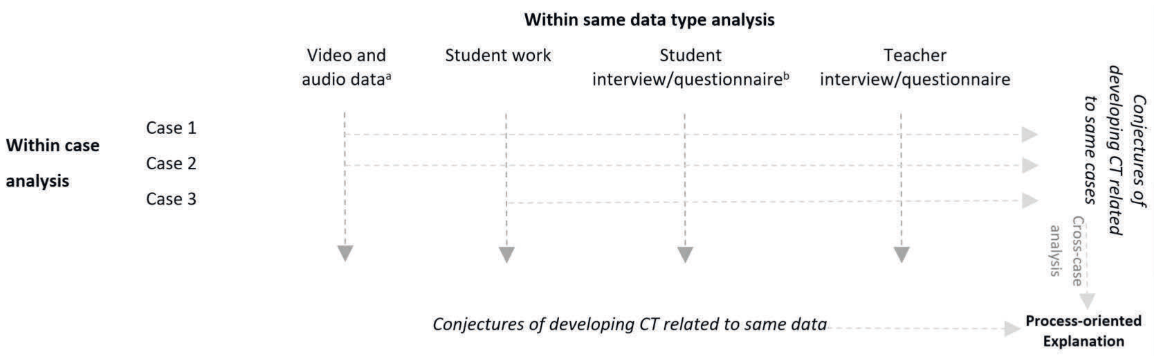


Figure 6.5: Data analysis procedure.

[a] Cases 2 and 3 were conducted during COVID 19. For case 2 the video and audio data of some lessons are available, for case 3 none.
[b] Interview and online questionnaire included the similar questions.

Inspired by the systematic approach for constant comparison introduced by Boeije (2002) the following steps and analysis activities are conducted:

6.2.1.1 Constant comparative method within case analysis

Each case included several groups of students (within a class), each group consisted of 2 to 3 students. The data available in each case were transcribed (if applicable), open coded, and compared within a single group and then between all groups.

In this procedure, each type of data was first treated separately and therefore conjectures about students' development of CT skills relating to a specific data type were formed. These were then compared to other lessons of the antimatter course to find confirmation and counterexamples.

After that, the conjectures assigned to each data type were compared in order to further develop, revise, and verify them. This served to triangulate the data in the study. It is obvious that the conjectures formed from audio and video data may differ from the student work. In the student discussion, the focus is on the process of developing CT skills, while in the worksheets, the focus is more on the results. For example, the problem of overconfidence in the context of likelihood and uncertainty analysis skill may not be identified in the worksheets, or the correct application of creative ideas based on content knowledge may not be inferred correctly from the discussion.

In this constant process of comparison, the conjectures were supported and stabilized by subsequent data. They were further developed into an explanatory construct for each case to explain the developmental pathways of CT skills. To improve the reliability of the results, the conjectures were discussed with colleagues.

An example of this type of analysis is presented below in the context of likelihood and uncertainty analysis (LU) skills as students discussed the tasks on the Anderson's cloud chamber photograph worksheet (worksheet 3, see appendix H). In this worksheet, students were asked to make various interpretations about the sign of the charged particle and the direction of particle motion, and to evaluate their interpretations using content knowledge and considering probability. Probability plays a particular role when students evaluate the interpretation that includes two tracks starting from the same point or two tracks positioned so that they appear as one track. In this case, considering the low probability of this interpretation occurring is the only way to explain why this interpretation cannot be considered the most likely event. Trying to explain this with content knowledge seems like falling into a faulty loop.

Comparison within a single group: students' discussion in the group while working on the CT tasks (worksheets) was transcribed and open coded. To this end, each individual task in the worksheet was considered as a fragment. For example, the three tasks in the LU worksheet: making interpretations, evaluating interpretations, and making decision about the most probable event were considered as three fragments. In each fragment, the processes students take to answer the task were coded. These codes were treated as initial conjectures about students' CT development. The consistency of coding in a single group discussion was examined by comparing codes between fragments. In addition, further comparison was made to find the common characteristics of the same codes (initial conjectures) to generate conjectures that are more general and describe the process of developing CT.

For example, in the first task of the LU worksheet, in the fragment of "making interpretations", when it was observed that students stopped after making two obvious interpretations, and stated that they could not find more interpretations the code (initial conjecture) was made that "students feel unconfident in going further in the task with probabilistic nature when the solutions are not obvious" This code was also attached to students' discussion when they evaluated interpretations in the second fragment. It was observed that students try to apply content knowledge to evaluate the probabilistic interpretation including two tracks originated at the same point or followed each other, without considering the probability. However, they could not give correct reasons to evaluate it based on the content knowledge, because just considering the probability could be a sound argument in the evaluation of that specific interpretation but students feel unconfident to think of that. Comparing these same codes showed the "overconfidence" problem defined by Halpern (2009) as that "people tend to be more confident in their decision about probabilistic events than they should be." The conjecture was formulated as "overconfidence prevents students to think of hidden solutions".

Comparison between all groups: after the open coding of each group discussion in the first step, a comparison was made between "corresponding" fragments from different groups that have been given the same code. This led to confirmation of conjectures or to their refutation.

Concluding CT development: the above procedures were repeated for all data types within a case. Confirmed conjectures explained the CT skill development process for each case.

6.2.1.2 Constant comparative method within same data type analysis

Once the codes and conjectures for each type of data within the cases were found, a comparison was made to finalize the confirmed conjectures applied to the same data type. It is apparent that the conjectures may vary depending on the type of data, e.g., more conjectures may emerge from the audio data compared to the student work, and also the conjectures in the audio data explain the process of developing CT skills, while on the worksheets the results may be observed. Combining all the information gives a complete picture of students' CT skill development because it links the processes students go through to the outcomes.

6.2.1.3 Constant comparative method in cross-case analysis

Once the conjectures from each case were found, the three cases were compared to identify similarities and differences between them that explain developing students' CT skills and check whether any common patterns in students' developing CT skills available.

6.2.1.4 Development of substantive process-oriented explanation

Finally, a comparison was conducted between the finalized conjectures within the same data and those from cross-case analysis to develop the process-oriented explanation of developing students' CT skills in particle physics context.

In this research, the quality of this type of qualitative analysis was adhered to as follows:

- applying the systematic approach introduced by Boeije (2002) enhanced the traceability of the CCM applied to this study.

- triangulating the results from multiple data sources and also multiple cases enhanced the trustworthiness of the study. It is worth noting that the cases in the study formed a homogeneous sample, they were all highly gifted students in grade 10, but in case 3 the students were younger. However, this would not affect the validity of the results, as the goal of the study was to generate a substantive process-oriented explanation grounded in the data, not a general theory. By specifying the context and the characteristics of students, the substantive theory will be valid.
- validity of the results was increased by considering counterexamples (implicit feature of CCM) and the use of peer reviews.

6.2.2 Description of instruments for quantitative evaluation

To triangulate the results of qualitative evaluation, quantitative evaluation of the effectiveness of the antimatter course was conducted. To evaluate general CT skills, Halpern Critical Thinking Assessment (HCTA) was assigned as pre- and post-tests. To evaluate domain-specific CT skills, it was required to develop a specific Particle Physics Critical Thinking (PPCT) test.

It is worth noting that in the present study using the DBR approach, comparison with non-treatment groups was not the aim of the study. As indicated earlier, the essence of DBR is to describe the achievement of learning objectives by providing accurate evidence, e.g., by using constant comparative method, rather than looking for cause and effect in relation to the experimental design (see discussion in Section 4.3.2.).

6.2.2.1 Halpern Critical Thinking Assessment: evaluation of general critical thinking

To evaluate the effectiveness of the course in terms of students' improvement in general CT skills, Halpern Critical Thinking Assessment (HCTA) (Halpern 2016) was used. HCTA is a standardized general CT test that includes 20 items (see section 3.3.2.1). The items focus on assessing desirable outcomes of applying CT skills in everyday life context (see table 6.2).

Besides the fact that HCTA follows a clear definition of CT skills and also covers the targeted CT skills discussed in different CT tests and therefore provides more commonalities between its targeted CT skills across various CT tests (Tiruneh et al., 2017), the following criteria were also used to guide the decision on the use of HCTA in the antimatter course:

- high validity and reliability of the test: HCTA is administrated to numerous divers samples and the results showed a high validity and reliability of the test (see Halpern, 2016; see also Verburgh et al., 2013).
- appropriateness of the test for the target group of the study: the HCTA is appropriate for respondents aged 15 years and older.
- including open-ended items to evaluate students' active application of CT skills: HCTA has open-ended form (S1 and S3).
- providing the opportunity to assess students' general CT skills before and after taking the course (*cf.* Adams & Wieman, 2011): HCTA has two versions of the open-ended form. These two versions focus on the same CT skills but in different scenarios that allow the HCTA to be assigned as pre-test and post-test.
- positive impression among students: the results of a validation study of general CT tests translated into Dutch, such as the Cornell Critical Thinking Test (CCTT) and the Halpern Critical Thinking Assessment (HCTA), show that HCTA has higher preference among students (Verburgh et al., 2013).

Since this study was conducted in Germany, it was needed to translate the HCTA into German. One item of the test related to verbal reasoning skill was removed because it does not fit the German culture. The next 19 items (in both S1 and S3 versions of the HCTA) were translated and revised by German colleagues.

To improve the validity of the translated version of the HCTA, the test was administered as a **pre-test** and **post-test** in the first pilot study. In a 4-day antimatter workshop at the Technische Universität Dresden, 4 high school students answered the tests and their answers were evaluated based on the corresponding scoring rubric of the test offered by Halpern (Halpern, 2016). Furthermore, the both tests were employed in the second pilot study. After each implementation, the problematic parts in the student responses were discussed with colleagues and revised to ensure clear wording and correct mention of the question (as determined by comparing the student responses to the scoring rubric).

Since the goal of this study was not to validate the German version of the HCTA in terms of psychometric properties in a quantitative data analysis (for a study on validating the Dutch version of the HCTA in terms of psychometric properties see de Bie et al., 2015), this level of validity in terms of clear wording to ensure that students correctly understood the mention of the item as in the English version was consistent with the goal of using the HCTA to assess students' general CT skills.

6.2.2.2 Design of Particle Physics Critical Thinking test: evaluation of domain-specific critical thinking

To evaluate the effectiveness of the antimatter course in improving students' domain-specific CT skills, it was necessary to develop a CT test in the domain of particle physics, Particle Physics Critical Thinking (PPCT) test. In the following sections, the steps for developing the items of the PPCT test and the attempts for improving its validity and reliability are discussed. Furthermore, a sample item and its corresponding scoring rubric are presented.

Item development and validation

the Halpern Critical Thinking Assessment (Halpern, 2016) as a standardized critical thinking test was used as framework for developing the domain-specific items of the PPCT test (see table 6.2).

The first version of PPCT contained 17 open-ended items. Furthermore, a relevant scoring rubric is defined to assess students' performance in the open questions of the PPCT test. To validate the content of the PPCT test, the content relevance and the content representation of each item were evaluated (Boateng et al., 2018). To this end, the criteria such as (a) the appropriateness of items in terms of assessment of the desired domain-specific outcomes, (b) the clarity of items wording, phrases, and images, (b) the accuracy of the information presented in the items based on scientific knowledge (*cf.* Boateng et al., 2018; Tiruneh et al., 2017) were defined. Accordingly, the following evaluation methods were used (Adams & Wieman, 2011; Boateng et al., 2018):

- Expert review: particle physicists and physics educators from the Technische Universität Dresden reviewed the items against the criteria and provided feedback.
- Small-scale paper-pencil administration: physics teacher students (N= 8) from the University College of Teacher Education Vienna (Pädagogische Hochschule Wien) answered the test after participating in the antimatter workshop and commented on the test.
- Validation interview: volunteer high school students (N= 4; including 2 males and 2 females) who had participated in the first pilot study were interviewed after the antimatter workshop. They were asked to think aloud while answering the items, and their answers were audio-recorded to evaluate whether students comprehend and answer to items the way was intended. During the interviews, an attempt was made to put the students in a comfortable mode of thinking aloud by first asking "icebreaker" questions. In addition, since the interview condition (environment) was intended to be very similar to that of the test administration, students were asked to express their thoughts as they worked through the test, but not to explain *how* they interpreted each item or *why* (*cf.* Adams & Wieman, 2011).

After each evaluation method, the items and the scoring rubric were revised, and finally 14 open-ended items were determined to form the PPCT.

Table 6.6 shows the distribution of items according to CT skills, their maximum and summed scores, and the percentage of the total test score (48 points) devoted to each skill. The item numbers in table 2 was retained from the first version of the PPCT, which included 17 items (items 12, 15, and 16 were removed after assessing their validity).

As already indicated, it is not possible to draw a sharp line between the different CT skills. They are interrelated. For example, one cannot argue convincingly (AA skill) without considering the importance of using correct terminology in his/her argument (VR skill). However, to categorize them, the main desirable outcomes of applying CT skills were focused on each item.

Table 6.6: Distribution of items in the Particle Physics Critical Thinking (PPCT) test according to critical thinking (CT) skills, their maximum and summed scores, and the percentage of the total test score devoted to each skill.

PPCT-CT skill	Item	Max Score	Sum Score	Percentage of Total Score
Decision Making & Problem Solving (DM & PS)	8	4	14	29%
	10	3		
	13	3		
	14	4		
Argument Analysis (AA)	3	2	12	25%
	4	3		
	5	3		
	17	4		
Likelihood & Uncertainty Analysis (LU)	7	4	8	17%
	9	4		
Verbal Reasoning (VR)	1	2	10	21%
	2	2		
	11	6		
Thinking as Hypothesis Testing (HT)	6	4	4	8%

Figure 6.6 shows a sample item of the PPCT test for evaluating Decision Making and Problem Solving skill. The item objective is that students develop a reasonable solution to solve a problem and identify relevant information when solving a problem.

You are curious about how much energy can be transformed in a pair annihilation.

First, consider an electron-positron pair. The minimum energy of annihilation of an electron-positron pair is 1.2 MeV.

Task 14.1:

Based on this calculation result, can you estimate the minimum amount of energy that would be transformed by annihilation of a proton-antiproton pair? □Yes □No

Task 14.2:

If Yes, explain what background knowledge you use to make your estimate.

Task 14.3:

If No, explain what other background information you need to arrive at your estimate.

Figure 6.6: Sample item of the Particle Physics Critical Thinking (PPCT) test.

Table 6.7 shows the corresponding scoring rubric for the sample item of the PPCT test shown in figure 6.6.

Table 6.7: Corresponding scoring rubric for the sample item of the Particle Physics Critical Thinking (PPCT) test.

Item weight: 4 points

The complete answer that we expect from students is:

To calculate the minimum energy of annihilation of a particle-antiparticle pair, we assume that particle and antiparticle are at rest (no kinetic energy) and the only energy they have is the rest energy. To calculate the rest energy, we only need the mass of particle and antiparticle (particle and its corresponding antiparticle have the same mass). Then we can use Einstein's equation $E=m\cdot c^2$ *to calculate the minimum energy of annihilation. Since the mass of the proton is almost 2000 times greater than that of the electron, we can conclude that the minimum energy of annihilation of the proton-antiproton pair is 2000 times greater than the minimum energy of annihilation of the electron-positron pair.*

Scoring rubric:

Task 14.1

 # Yes

Task 14.2

- Did the student answer indicate that "the minimum energy of annihilation of a proton-antiproton pair can be calculated by computing their rest energy using Einstein's equation". *# Yes: **2 points**, # No: **0 point**, # Less clear: **1 point***
- Did the student answer indicate that "the mass of the proton is almost 2000 times greater than that of the electron, therefore the minimum energy of annihilation of the proton-antiproton pair is 2000 times greater than the minimum energy of annihilation of the electron-positron pair." *# Yes: **2 points**, # No: **0 point**, # Less clear: **1 point***

 # No

Task 14.3

- Did the student mention they need to calculate the rest energy of the proton-antiproton pair and therefore they need to know the mass of proton? *# Yes: **4 points**, # No: **0 point**, # Less clear: mention only the concept of rest energy **or** mention only the need for mass without justification **2 points***

Quantitative analysis of the PPCT test

After the validation study discussed above, statistical tests were conducted to evaluate validity and reliability of the PPCT test. Therefore, the PPCT test was administrated to classes and the results were used to calculate the internal consistency of the test, its item difficulty and item discrimination, convergent validity, and the interrater reliability.

Administration of the PPCT test

The revised version of the PPCT test was administrated to German high school students of grades 12 and 10 (N= 53; 37% Female and 63% Male). The age of students was between 14 and 19 years old (M=16.2, SD=1.7) and the administration of the test took 45 minutes in average.

Since the PPCT test was assigned to different students in different high schools and also in different implementation phases, it was important to test the assumptions of normality and homogeneity of variances to be able to combine the groups for a larger sample size. To test the assumption of homogeneity of variances, Levene's F test was used by comparing the students' score on a pre-test on general CT skills (HCTA-S1). The result showed a satisfying homogeneity of variance ($p > .05$). Furthermore, a Shapiro-Wilk's test for the PPCT scores ($p > .05$) and a visual inspection of boxplot showed that the scores from different groups were normally distributed.

Internal consistency

Internal consistency of the items in the PPCT test was measured by calculating Cronbach's alpha to evaluate how closely related these 14 items as a group (Halpern 2016, p. 16) and judge how reliable the PPCT test is for data collection (Cohen et al., 2007). The analysis of the results (N=53) showed an acceptable internal consistency of the 14 items of the PPCT test, $\alpha = .73$ (Boateng et al., 2018, p. 13; Cohen et al., 2007, p. 506).

Inter-rater Reliability

To evaluate the inter-rater reliability, 17 students' answers were selected randomly and corrected independently by two raters using the same scoring rubric. The Cohen's Kappa coefficient was calculated for all 14 items and also for the total score. As it is shown in table 6.8, the Cohen's Kappa coefficient for all 14 items ranges from .71 to 1.00, and for the total score of the test, the calculation of Cohen's Kappa yields .75. These results suggested that the PPCT has adequate to high interrater reliability at both the item and total score levels.

Furthermore, a Shapiro-Wilk test ($p > .05$) and visual inspection of the boxplots indicated that the two rater scores were normally distributed. In addition, paired sample t test was calculated to investigate the effects of the raters on the mean scores of each item, and the results showed that the means of the items were not significantly affected by the two raters ($p > .05$) (*cf.* Tiruneh et al., 2017).

Table 6.8: Inter-rater reliability of the 14 items of the Particle Physics Critical Thinking (PPCT) test calculated for 17 students.

Item	1	2	3	4	5	6	7	8	9	10	11	13	14	17	Total score
Inter-rater reliability (n=17)	.89	.74	1.00	.75	.84	.71	.76	1.00	.90	1.00	.85	.82	1.00	1.00	.75

Item difficulty and item discrimination

Item difficulty is defined as "the degree of strength the respondent must possess in order to pass the item" (DeVellis, 2003, p. 140). Moreover, item discrimination is defined as "the degree that an item differentiates between respondents (Boateng et al., 2018, p. 9) and is attributed to the clarity of the item in classifying a response as a pass or fail (DeVellis, 2003, p. 142).

To calculate the item difficulty of the PPCT test, the formula suggested for open-ended items by the Evaluation and Examination Service of the University of Iowa was used (*cf.* Tiruneh et al., 2017).

$$P = \frac{fX - nX_{\min}}{n(X_{\max} - X_{\min})}$$

Where fX is the total score achieved by all students on an item, *n* is the number of students, $X_{\min}$ is the lowest possible item score, and $X_{\max}$ is the highest possible item score.

It is important to mention that the possible score for each of the items of the PPCT ranges differently, for example item 14 ranges from 0 to 4 while item 2 ranges from 0 to 2.

To calculate the item discrimination index, the highest and lowest 27% of the PPCT scores (for each n=15 students) were taken to form the upper and lower groups. Then the difficulty index of each item in upper group and lower group separately were calculated and finally were subtracted (*cf.* Tiruneh et al., 2017).

Table 6.9 shows an overview of item difficulty and item discrimination index of the 14 items of the PPCT test.

Table 6.9: Item difficulty and item discrimination index of the 14 items of the Particle Physics Critical Thinking (PPCT) test. (Remark: the item numbers were retained from the first version of the PPCT with 17 items).

PPCT-CT skill	Item	Item Difficulty	Discrimination Index
Decision Making & Problem Solving (DM & PS)	8	0.46	0.74
	10	0.11	0.16
	13	0.44	0.31
	14	0.13	0.12
Argument Analysis (AA)	3	0.41	0.43
	4	0.46	0.03
	5	0.33	0.23
	17	0.27	0.25
Likelihood & Uncertainty Analysis (LU)	7	0.21	0.37
	9	0.48	0.50
Verbal Reasoning (VR)	1	0.50	0.63
	2	0.08	0.17
	11	0.53	0.33
Thinking as Hypothesis Testing (HT)	6	0.15	0.22

To interpret the results in table 6.9, the criteria presented in table 6.10 were used. The results showed that all items seem to be in the range of very difficult to moderately difficult. The combination of a high level of content knowledge and the CT skills required to construct an answer make them high cognitive demand tasks (Clesham, 2013; OECD, 2017), requiring students to work with complex concepts and apply CT skills simultaneously. Therefore, such criterion-referenced assessment may be difficult and yet appropriate for certain test takers (*cf.* ETS, 2008; Tiruneh et al., 2017).

Table 6.10: Interpretation of difficulty and discrimination index (Sharma, 2021).

Item difficulty interpretation		**Item discrimination index interpretation**	
< 0.20	Most difficult	< 0.20	Not disriminating item
0.20-0.39	Difficult	0.20-0.29	Moderately discriminating item, fair item
0.40-0.59	Moderately difficult	0.30-0.39	Discriminating item, good item
0.60-0.79	Moderately easy	≥ 0.40	Very discriminating, very good item
0.80–0.89	Easy		
> 0.90	Easiest		

Moreover, the item discrimination calculation showed that most items can discriminate very well between the high and low performance levels, only items 2, 4, 10, and 14 have a very low discrimination index. Considering item 2, 10, and 14 as the most difficult items of the PPCT test, justified the results. But in case of item 4, as it is a moderately difficult item, the low discrimination index showed that this item is not appropriate for large-scale assessment and may need some modifications in formulation. Therefore, to calculate the convergent validity of the PPCT test, item 4 was removed from the results.

Convergent validity

To assess the extent to which the newly developed test (PPCT) can measure the same construct as HCTA, the convergent validity was calculated (Boateng et al., 2018, p. 14). To assess the convergent validity of the PPCT test, the Halpern Critical Thinking Assessment (HCTA) as a standard test was administered to the same participants after they completed the PPCT. The Pearson's correlation coefficient was calculated to compare students' performance on the 13 items of the PPCT (item 4 was removed because of the low discrimination index) to their performance on the HCTA. Since just 49 students completed both PPCT and HCTA, the correlation was calculated for 49 students. The result showed a positive correlation between the two sets of the scores ($r = .69$, $p < .001$, $N = 49$). It indicates that both tests are measuring the same construct and the PPCT has adequate convergent validity (*cf.* Tiruneh et al., 2017).

This quite high correlation verified that the general CT skills and outcomes were accurately reflected in the development of the domain-specific items of the PPCT tests. On the other hand, this may lead to the emergence of some questions, such as *what is the role of content knowledge* and *whether there is need to assess both general and domain-specific CT skills*. To answer these questions, a qualitative analysis of students' answers was conducted. This qualitative analysis illustrated the challenges students face in applying CT skills to particle physics and also showed the extent to which content knowledge influences student performance on CT tasks. Furthermore, the results were used to improve the quality of the construction of the PPCT test (*cf.* Fielding & Fielding, 1986; also Strauss & Corbin, 1998).

Qualitative analysis of open answers of the PPCT

The results of the statistical analysis showed that the PPCT test is a valid and reliable instrument for assessing students' CT skills in particle physics and therefore provides a sound basis for conducting a qualitative analysis of students' answers to its items.

To start with the qualitative analysis, first realizing the challenging CT skills for students was necessary. A hypothesis was made that a significant correlation between CT sub-skills in both tests (HCTA and PPCT) shows that students are more confident in applying CT skills from a general context to a domain-specific context. That means a good general CT skill might lead to a good performance in the domain-specific CT tasks.

Therefore, Pearson's correlation between different CT-subskills in both tests was calculated. The results showed that the both tests have adequate convergent validity on the total score, but in the level of subskills for VR, HT, and PS there is no statistically significant correlation between general and domain-specific skills (see table 6.11). However, for AA and LU a significant correlation at the level of 0.01 was observed.

On the other side, looking at the results of item difficulty in table 6.9, showed that the skills of VR, HT, and PS in the PPCT test include the most difficult items with low discrimination index while for AA and LU this is not the case. Therefore, a decision was made to conduct a qualitative analysis of students' answers to these most difficult items (items 2, 6, 10, and 14) to conclude about the students' challenges in applying CT to particle physics. Furthermore, since items 2, 10, and 14 have a low discrimination index, the qualitative analysis might reveal the problematic aspects of the items and would allow to improve the construction of the items if this is the case.

Table 6.11: Pearson's correlation between different CT-subskills in the PPCT test and the HCTA-S3 test.

Verbal reasoning	**Argument analysis**	**Thinking as hypothesis testing**	**Likelihood and uncertainty analysis**	**Decision making and problem solving**	**Total score**
.125 (p > .05)	.401 (p= .006)	-.215 (p > .05)	.40 (p= .007)	.124 (p > .05)	.69 (p < .001)

Guideline/Criteria for analyzing open answers

The goal was to identify to what extent students' performance in the PPCT test is related to the content knowledge and to what extent to the CT skills. Another goal might be to draw conclusions about which CT skills are difficult for students to apply and what challenges students have most in applying CT skills. The results will give educators and teachers who want to design lessons or develop a test some clues about which aspects of the CT skills are most difficult for students, why this is so, and what elements or concepts should be included in the lesson or intervention to achieve better results.

To this end, students' answers to the selected items (items 10, 14, 6, and 2) were analyzed from two perspectives: content knowledge and critical thinking. The content knowledge perspective assessed the correctness of the answers from a physics perspective, and the critical thinking perspective assessed the extent to which correct CT skills were applied to construct the answers. To draw valid conclusions about students' CT skills, the results of the analysis of student answers in PPCT were triangulated with the results of the analysis of student answers in HCTA.

Table 6.12 shows the guideline for conducting a qualitative analysis of students' answers in PPCT.

Table 6.12: Guideline for analyzing students' answers in the Particle Physics Critical Thinking (PPCT) test.

Aspect of evaluation	***Red thread question***	***code***	***Cod description***	***Indicator for***
Content knowledge (CK)	To what extent student write the correct answer based on the scoring guide?	Correct		How much did student learn in the course? Whether student' low performance in the task comes from CK or CT?
		Incorrect		
		Partly correct		
Critical thinking (CT)	To what extent student can apply the CT skills s/he learned in the course regardless of correctness or incorrectness of the answer?	Level 2	Desirable outcomes of applying CT skill and the criteria are **completely** fulfilled.	Whether student' low performance in the task comes from CK or CT?
		Level 1	Desirable outcomes of applying CT skill and the criteria are **partly** fulfilled.	
		Level 0	Desirable outcomes of applying CT skill and the criteria are **not** fulfilled.	

The results of qualitative analysis of students' answers (Sadidi & Pospiech, in preparation) showed that content knowledge plays an important role in students' performance in the PPCT test. Students might have a good level of general CT skills, but without enough content knowledge, they cannot perform well in the PPCT test. For example, one can identify the importance of using correct terminology but without enough content knowledge, he or she cannot put this desire to action or one can realize the importance of relevant information for solving the problem but without accessing to the information the problem cannot be solved.

6.3 Relating design skeleton facets to valued outcomes

Constant comparative method gave evidence of developing students' general and domain-specific CT skills. To investigate the relationship between the design skeleton facets of antimatter course with this development (answer to research question 1.2), an in-depth analysis of the results of constant comparative method was done. The goal was to check how the design facets of the course engender the desired processes and how their implementation contributes to the desired outcomes of the course. Figure 6.7 shows a schema of the aspects of investigation.

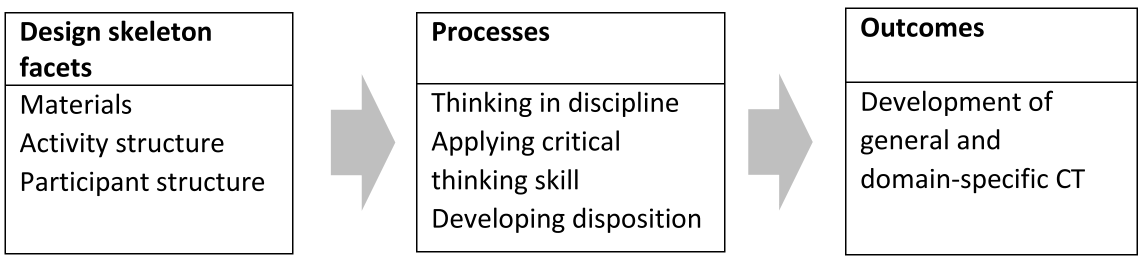


Figure 6.7: Schema of aspects of investigation in the qualitative data analysis of the lessons in antimatter course.

The design skeleton facets concern the materials designed for the course, the activities students have to do, and the participation roles, norms, and culture (McKenney & Reeves, 2019; Sandoval, 2014). The design skeleton facets aimed to engage students in the processes which are necessary for CT development such as thinking in discipline (applying domain-specific knowledge), applying critical thinking skill, and developing disposition toward critical thinking.

To demonstrate the connection between design facets, processes, and outcomes, conjecture map was used as a framework (*cf.* Sandoval, 2014; Wozniak, 2015). A back-tracking analysis of the results of the qualitative analysis of data (student work, discussion, questionnaire) was conducted to identify the co-occurrence of the processes students engaged in to develop their CT skills in relation to the facets of the design skeleton (*cf.* Philipsen et al., 2019; Sandoval, 2014).

Furthermore, the class activities were analysed in two levels of student-student interaction and teacher-student interaction to get insight of the participation roles, norms, and culture. Analysing video data and its triangulation with students' answers in the worksheets helped to identify co-occurrences and some connections in the data between design facets, processes, and outcomes inductively. The results will present in chapter 8.

6.4 Description of instrument for evaluating teachers' perspective

To be able to conclude about the effectiveness of the antimatter course in improving students' CT skills, both structure fidelity and process fidelity of its implementation (see section 3.3.1.2) were evaluated by developing a teacher questionnaire.

To evaluate process fidelity, the questionnaire addressed the teachers' perspective on the relevance, expected practicality, and expected effectiveness of the antimatter course in developing students' CT skills (*cf.* Stokhof et al., 2019). To this end, the following criteria and corresponding questions guided the design of the questionnaire (personal communication with Mckenney, 2019):

- Relevance: To get the teacher perspective on how the teaching-learning strategy in the antimatter course might improve general and domain-specific critical thinking skills, it is important to know:
    - What are their perceptions of the critical thinking skills (as manifested in the materials)?

- Practicality: To get the teacher perspective on the implementation of the antimatter course within the practical limitations of time, means and knowledge, it is important to know:
    - How do teachers experience teaching the lesson with available materials? Do teachers need additional tools/support to teach the lesson more effective?
- Effectiveness: To gain insight into the teacher perspectives on student outcomes, it is important to know:
    - What evidence would teacher normally use to monitor/assess student learning related to critical thinking in general and antimatter in particular?
    - What does the evidence that they currently have tell them about student outcomes?

To evaluate structure fidelity, some questions were also included in the questionnaire to evaluate whether the teachers were implementing the lessons as intended based on the design principles and the planned time. The decision was made because it was not possible to visit the classroom and make video recordings due to COVID 19.

The teacher questionnaire is presented in appendix E. Table 6.13 shows the indicators for structure and process fidelity of the antimatter course used to evaluate teachers' answers.

It is worth noting that the questionnaire was not so designed to guide the teachers to directly reference to a specific indicator of relevance, practicality, and effectiveness defined in the present study. The questions were formulated in general terms to avoid influencing teachers' perceptions. For example, instead of asking "What is for you the relevance of teaching likelihood and uncertainty analysis skill in the context of positron discovery in developing students' CT skills?", the question phrased as "Do you think that the designed activities in this lesson can contribute to improving students' CT skills?

please explain.”. In this case, teachers were given more space to think about different activities and were not locked into one particular aspect. Qualitative analysis of teachers’ responses revealed their views on the indicators listed in Table 6.13 and also provided clarity on how teachers perceive the other teaching-learning activities that are embedded in each lesson but might not be listed in this table.

Administration of the questionnaire

After redesigning the antimatter course during different pilot studies, the final version of the antimatter course was prepared for the main study. Furthermore, a teacher package was developed including the teaching-learning sequences, materials, step-by-step guidelines for teaching the lessons, possible students’ questions during lessons, the extra content knowledge for teachers, and also the suggested time for each teaching-learning activity. In addition, a power point presentation for teaching each lesson was prepared. The goal was to enable the teachers to implement the course alone.

The questionnaire was conducted in the main study. Teachers were asked to answer the questions after each lesson. Furthermore, the teachers were informed about the goal of the design of the course, the design principles of the course, the materials, and teaching sequences, and the goal of the questionnaire.

The questionnaires were collected and the data were analyzed qualitatively to conclude about teacher perceptions of the relevance, expected practicality, and expected effectiveness of the antimatter course. Furthermore, teachers’ responses to the structure fidelity questions showed to which extent teachers implement the suggested activities in the antimatter course as intended in terms of the content, the teaching method, and the time.

Table 6.13: Indicators for structure and process fidelity of antimatter course.

Lesson	***Structure fidelity***		***Process fidelity***		
Based on design principle(s)	*Adherence*	*Duration*	*Relevance* *(Perceived need for)*	*Practicality* *(Perceived ease to)*	*Effectiveness* *(Perceived outcomes of)*
Lessons 1 & 2: Teaching explicitly general critical thinking skills & Showing "Illuminati" movie					
Explicit teaching of general CT skills CT-structure-training activities Problem-centered principle Encourage critical thinking Encourage discussion and reflection	Teach general CT skills Encourage discussion and reflection on applying CT skills in everyday life Show scenes of "Illuminati" movie and outline the content of the antimatter course	-	Teaching explicitly general CT skills Using scene of movie to fulfill the problem-centered principle	-	Teaching explicitly general CT skills Using scene of movie to fulfill the problem-centered principle
Lessons 3 & 4 & 5: What is antimatter?					
CT-Structure-training activities Problem-centered principle Activation principle Encourage critical thinking Encourage discussion and reflection	Carrying out the teaching sequences as intended *including: introduction, activate students' prior knowledge, and teach Likelihood and uncertainty analysis skill in the unit positron discovery* Encourage discussion and reflection on Likelihood and uncertainty analysis skill	Amount of time spent on lessons 3 & 4 & 5	Teaching Likelihood and uncertainty analysis skill in the context of positron discovery Discussion and reflection	Teach Likelihood and uncertainty analysis skill in the context of positron discovery	Teaching Likelihood and uncertainty analysis skill in the context of positron discovery Discussion and reflection
Lesson 6: Annihilation & Pair production					
CT-Structure-training activities Problem-centered principle Activation principle Encourage critical thinking Encourage discussion and reflection	Carrying out the teaching sequences as intended *including: introduction, activate students' prior knowledge, and teach Thinking as hypothesis testing skill in the unit pair production* Encourage discussion and reflection on Thinking as hypothesis testing skill	Amount of time spent on lesson 6	Teaching Thinking as hypothesis testing skill in the pair production context Discussion and reflection	Teach Thinking as hypothesis testing skill in the pair production context	Teaching Thinking as hypothesis testing skill in the pair production context Discussion and reflection
Lesson 7: Big Bang					

CT-Structure-training activities Problem-centered principle Encourage critical thinking Encourage discussion and reflection	Carrying out the teaching sequences as intended *including: introduction, show a video clip about Big Bang, encourage students to write questions about the clip, and discuss and distribute questions between groups* Encourage CT for searching systematically for the answer Give feedback on students' presentation	Amount of time spent on lesson 7	Teaching Argument analysis skill in the context of a systematic search for answers to the questions raised by watching a video clip about Big Bang Discussion and reflection	Teach Argument analysis skill in the context of a systematic search for answers to the questions raised by watching a video clip about Big Bang	Teaching Argument analysis skill in the context of a systematic search for answers to the questions raised by watching a video clip about Big Bang Discussion and reflection
Lessons 8 & 9: Producing & Trapping antihydrogen					
CT-Structure-training activities Problem-centered principle Encourage critical thinking Encourage discussion and reflection	Carrying out the teaching sequences as intended *including: introduction, and teach Thinking as hypothesis testing skill in the unit trapping antihydrogen* Encourage discussion and reflection on Thinking as hypothesis testing skill	Amount of time spent on lessons 8 & 9	Teaching Thinking as hypothesis testing skill in the context of trapping antihydrogen Applying Thinking as hypothesis testing skill in the context of Bending Rods Discussion and reflection	Teach Thinking as hypothesis testing skill in the context of trapping antihydrogen Use Bending Rods example to transfer Thinking as hypothesis testing skill in the everyday life context	Teaching Thinking as hypothesis testing skill in the context of trapping antihydrogen Applying Thinking as hypothesis testing skill in the context of Bending Rods Discussion and reflection
Lesson 10: GBAR Experiment					
CT-Structure-training activities Problem-centered principle Encourage critical thinking Encourage discussion and reflection	Carrying out the teaching sequences as intended *including: introduction, and teach Thinking as hypothesis testing skill in the unit GBAR experiment* Encourage discussion and reflection on Thinking as hypothesis testing skill	Amount of time spent on lesson 10	Teaching Thinking as hypothesis testing skill in the context of trapping antihydrogen Discussion and reflection	Teach Thinking as hypothesis testing skill in the context of trapping antihydrogen	Teaching Thinking as hypothesis testing skill in the context of trapping antihydrogen Discussion and reflection
Lessons 11 & 12: Reviewing the learning content about antimatter & showing "Illuminati" movie & Application of antimatter in PET & Reflection on the course					
CT-Structure-training activities Problem-centered principle Encourage critical thinking Encourage discussion and reflection	Carrying out the teaching sequences as intended *including: introduction, and review the content knowledge and CT skills taught in the course* *Encourage CT skills in analyzing the scenario for the scene of "Illuminati" movie* Encourage discussion and reflection on students' presentation of new scenario Encourage CT skills in search about Positron Emission Tomography (PET) Encourage discussion and reflection on applying CT skills in physics and everyday life	Amount of time spent on lessons 11 & 12	Teaching Verbal reasoning skill in the context of analyzing the scenario for the scene of "Illuminati" movie Applying CT skills in search about Positron Emission Tomography (PET) Discussion and reflection	Teach Verbal reasoning skill in the context of analyzing the scenario for the scene of "Illuminati" movie Apply CT skills in search about Positron Emission Tomography (PET)	Teaching Verbal reasoning skill in the context of analyzing the scenario for the scene of "Illuminati" movie Applying CT skills in search about Positron Emission Tomography (PET) Discussion and reflection

7 Evaluation of effectiveness of the antimatter course

The effectiveness of the antimatter course was evaluated qualitatively by analyzing the video and audio data, student work, students' interview or questionnaire, and teachers' interview or questionnaire using the constant comparative method (see figure 6.5). In addition, since adherence to the design principles of the antimatter course in its implementation has an impact on the development of students' CT skills, the structure fidelity of the implementation of the course was evaluated. To this end, video and audio recordings of the classroom activities were collected, as well as a questionnaire for teachers. The data were analyzed with the focus on whether teachers implement the antimatter course as intended in the teacher package.

The following sections first characterize the participants, then discuss the structure fidelity of the implementation, and finally evaluate the effectiveness of the antimatter course in developing students' CT skills by presenting the results of the qualitative data analysis. In addition, based on the results a model for student CT skill development is proposed.

7.1 Participants

Since the teaching materials and teacher package used in pilot study 3 and the main study did not differ, it was decided to use the data from pilot study 3 in addition to the main study to evaluate the effectiveness of the antimatter course. This provided the opportunity to include more classes. Each class was treated as a separate case.

Considering that the main study was conducted during the COVID 19 pandemic, some data were missing. Students were home schooled and data had to be uploaded directly to the TUD cloud server by the students. Considering the complete data available, 3 cases were defined. All students were talented 10th grade students at schools in urban districts in different states of Germany. The only difference between case 1, case 2, and case 3 was that in case 1 and case 2 the antimatter course was taught as a separate course from the regular physics class and students could choose to take it voluntarily, whereas in case 3 it was integrated into the regular physics class. In addition, in case 1, the antimatter course was introduced in the regular school curriculum prior to the COVID 19 pandemic, whereas in cases 2 and 3, the implementation of the antimatter course coincided with the COVID 19 pandemic and students were mostly home-schooled. Furthermore, students in case 3 were younger.

Table 7.1 shows an overview of students' information. In all cases, the antimatter course was taught by the schools' regular physics teachers, and they voluntarily participated in the study.

Table 7.1: Overview of students' information.

		Case 1 (Pilot study 3) May & June 2019	**Case 2** (Main study) October 2020-February 2021	**Case 3** (Main study) September- December 2020
Special features		Face-to-face lessons before COVID 19, special course[b]	Combination of face-to-face and online lessons during COVID 19 (more in homeschooling), special course[b]	Combination of face-to-face and online lessons during COVID 19, regular physics class
Number of students		10	9	17
Female students (%)		30	22	18
Average age (SD)		16 (.47)	15.22 (.67)	13.88 (.70)
Average physics grade[a] (SD)		2.20 (.63)	1.78 (.67)	2.29 (1.10)
Available data	Audio and video of all lessons	#	X (some lessons)	-
	Student work	#	#	#
	Pre- and post-tests	#	#	#
	Student interview/questionnaire	#	#	#
	Teacher interview/questionnaire	#	#	#

[a]The grade can differ from 1.0 (best grade) to 5.0 (worst grade).

[b]Special course: Naturwissenschaftliches Profil

7.2 Evaluation of structure fidelity of implementation

Structure fidelity of the implementation of the antimatter course was evaluated by analyzing the data in each case using the adherence and duration indicators presented in table 6. 13. Since a different type of data was available in each case (see table 7.1), a different type of data analysis was consequently performed. For example, in case 1, the video data was observed and reviewed for the indicators listed in table 6.13. The results were also triangulated with the student work, both to improve the validity of the interpretations and to provide evidence of how adherence to the design principles would be reflected in the student work. These findings were incorporated into the interpretation of student work in case 2, in which the same teacher implemented the antimatter course, and were also used cautiously in the interpretation of student work in case 3, in which video data were not available. This improved the validity of the interpretation of the results. In addition, teacher responses to the questionnaire in cases 2 and 3 were used to triangulate the results.

7.2.1 Adherence

The results of data analysis in each case regarding the teachers' adherence to all the proposed activities is presented in Table 7.2. Lessons 1 and 2 are not included because these lessons were taught by the researcher and were therefore implemented as intended.

Data analysis showed that for **lessons 3, 4, and 5** the teaching sequences were implemented as mirrored in the teacher package in all cases. Triangulation of the results with the student work on worksheets 3 (see table 6.4) confirmed this interpretation. Most of students completed the tasks with correct reasoning and complete responses. To assess whether attention was paid to discussion and reflection on likelihood and uncertainty analysis skill in these lessons, the video data in case 1 and the explicit written reflection of the students in case 2 showed that this was the case, with the exception of case 3, where neither video data nor written reflection was available. No evidence of this aspect being fulfilled in class could be also found from the teachers' responses to the teacher questionnaire.

For **lesson 6**, data analysis showed that the teachers implemented the antimatter course as intended in cases 1 and 2. However, in case 3, many of the planned student activities described in the teacher packet were neglected by the teacher and the lesson was presented in the form of a lecture rather than encouraging student activities.

Lesson 7 focused on teaching students the steps of critical thinking in finding an answer to their questions in the context of Big Bang (see table 6.4). Data analysis showed that teachers conducted the teaching sequences as intended and also adhered to the principle of encouraging students to think critically while completing the task. It was also clear from the teachers' responses in the teacher questionnaire in case 2 and 3 that the opportunity for discussion and reflection on the tasks in this lesson was perceived very positively. However, it was not clear whether students in cases 2 and 3 also received feedback from the teacher when presenting their results to the class.

For **lessons 8 and 9**, and also for **lesson 10**, data analysis confirmed adherence to the structure fidelity of the antimatter course. Teachers' responses to the questionnaires in cases 2 and 3 also indicated a positive impression of discussion and reflection occasions planned in the teaching sequences for improving students' CT skills.

In **lessons 11 and 12**, search on the medical application of antimatter, positron emission tomography (PET), was unfortunately removed from the teaching sequences in all cases due to time constraints. In all cases, teachers explained PET themselves, and students did not have the opportunity to search for information and give a presentation about it in class. With exception of PET, in cases 1 and 2, all other teaching sequences and principles were implemented as intended. However, in case 3, the data analysis showed that most of the students did not write a new scenario for a scene in the movie "Illuminati" based on the correct content knowledge. The teacher also expressed in the questionnaire that the students were not satisfied with writing a new scenario. Although students discussed the unclear terms they had heard in the movie and made suggestions for correct terms, they rarely wrote a new scenario for the movie in the worksheets. This indicated a lack of motivation and perseverance in completing the task.

Moreover, it showed that they did not have enough encouragement and motivation for the task. However, this is an important principle of the antimatter course as it provided an opportunity to discuss and reflect on the CT skills learned in the lessons to solve the problem on which the course was built, but not enough attention was paid to its implementation in case 3.

In summary, the teachers were more adherent to the structure of the antimatter course in cases 1 and 2 than in case 3. If the lack of information about fidelity to the principles of the antimatter course in case 3 was not accounted for, the most obvious difference between cases 1 and 2 and case 3 was in two teaching situations: in the session on teaching pair production and in the final session on analyzing the scenario of a specific scene of the movie "Illuminati". In teaching pair production, the teacher neglected the guidance in the teacher package and taught the session by giving a lecture instead of encouraging students' activities, and in the final session, she failed to encourage students to complete the task on analyzing the scenario of the movie.

Table 7.2: Adherence to proposed activities in the antimatter course (+ adhered, - not adhered, 0 no information).

	Classroom activities	Case 1	Case 2	Case 3
Lesson 3 & 4 & 5	Carrying out the teaching sequences as intended *including: introduction, activate students' prior knowledge, and teach Likelihood and uncertainty analysis skill in the unit positron discovery*	+	+	+
	Encourage discussion and reflection on Likelihood and uncertainty analysis skill	+	+	0
Lesson 6	Carrying out the teaching sequences as intended *including: introduction, activate students' prior knowledge, and teach Thinking as hypothesis testing skill in the unit pair production*	+	+	-
	Encourage discussion and reflection on Thinking as hypothesis testing skill	+	0	-
Lesson 7	Carrying out the teaching sequences as intended *including: introduction, show a video clip about Big Bang, encourage students to write questions about the clip, and discuss and distribute questions between groups*	+	+	+
	Encourage CT for searching systematically for the answer	+	+	+
	Give feedback on students' presentation	+	0	0
Lesson 8 & 9	Carrying out the teaching sequences as intended *including: introduction, and teach Thinking as hypothesis testing skill in the unit trapping antihydrogen*	+	+	+
	Encourage discussion and reflection on Thinking as hypothesis testing skill	+	+	+
Lesson 10	Carrying out the teaching sequences as intended *including: introduction, and teach Thinking as hypothesis testing skill in the unit GBAR experiment*	+	+	+
	Encourage discussion and reflection on Thinking as hypothesis testing skill	+	+	+
Lesson 11 & 12	Carrying out the teaching sequences as intended *including: introduction, and review the content knowledge and CT skills taught in the course*	+	+	+
	Encourage CT skills in analyzing the scenario of the scene of "Illuminati" movie	+	+	-
	Encourage discussion and reflection on students' presentation of new scenario	+	+	-
	Encourage CT skills in search about Positron Emission Tomography (PET)	-	-	-
	Encourage discussion and reflection on applying CT skills in physics and everyday life	+	+	0
Total of observed activities (maximum is 16)		15	13	8

7.2.2 Duration

The antimatter course for grade 10 students was planned for 10 lessons (apart from the explicit teaching of CT skills and the showing of the movie in the first two lessons). Table 7.3 shows the results of the data analysis considering the duration of the work in each lesson. Data in cases 2 and 3 are mostly based on teachers' self-reports in answer to a corresponding question in the teacher questionnaire.

Teachers adhered to the to the time suggested in the teacher package with one exception. In case 3, a deviation was noted in the questionnaire for teaching lessons 3, 4, and 5. The teacher explained that due to the students' difficulties in converting the energy units from classical physics to the units of particle physics (eV), the implementation took more time. This might be explained by the fact that the students in case 3 were younger in comparison with cases 1 and 2 (see table 7.1).

Table 7.3: Adherence to the proposed time in each lesson of the antimatter course.

	Proposed time in Teacher Package (in minutes)	**Case 1**	**Case 2**	**Case 3**
Lessons 3 & 4 & 5	135	+	+	-[a]
Lesson 6	45	+	+	+
Lesson 7	45	+	+	+
Lessons 8 & 9	90	+	+	+
Lesson 10	45	+	+	+
Lessons 11 & 12	90	+	+	+
Total time	**495**	495	495	540

[a]The implementation took 180 minutes.

7.3 Results of qualitative data analysis

To obtain accurate evidence of the development of students’ domain-specific and general CT skills, the constant comparative method was used at the case level. This means that audio and video data (if applicable), student work, student interview/questionnaire, and teacher interview/questionnaire were analyzed for each case to develop conjectures about students’ development of CT skills. Conjectures describe student progress or challenges in applying CT skills developed based on data analysis. Comparing the conjectures developed for each case and finding commonalities (cross-case analysis) provided valid conclusions about the effectiveness of the antimatter course. Considering the conjectures at a meta-level and looking for categories that cover these conjectures can help build an empirically based model for CT development. Because the conjectures also address the challenges students face in applying CT skills in the context of particle physics, considering these challenges can guide the development of instructional design heuristics to develop effective domain-specific CT instruction.

It is worth noting that in the antimatter course, each lesson focused on a particular domain-specific CT skill (see table 6.4). However, to improve the validity of the conclusions about student CT skill development, relevant examples and counterexamples of the application of other CT skills besides the focused CT skill in a specific lesson were also considered.

The following sections examine the development of students’ CT skills in the context of antimatter by discussing each sub-skill first at the case level and then at the cross-case analysis level to reach a conclusion about the effectiveness of the antimatter course. The sub-skills discussed below are likelihood and uncertainty analysis (LU), verbal reasoning (VR), argument analysis (AA), and thinking as hypothesis testing (HT). The conjectures about the development of each sub-skill are presented and abbreviated by giving the abbreviation of the sub-skill plus the abbreviation of the word conjecture. For example, **LU-C1** describes the first conjecture in the likelihood and uncertainty analysis context. These conjectures were reviewed by one assistant and discussed with colleagues. The agreement was high.

7.3.1 Likelihood and uncertainty analysis skill in the positron discovery context

In the positron discovery lesson, worksheet 3 (Anderson's cloud chamber photograph worksheet, see appendix H) was used to support students' development of likelihood and uncertainty analysis (LU) skill. Students were asked to make various interpretations of Anderson's cloud chamber photograph about the sign of the charged particle and the direction of particle motion, and to evaluate their interpretations using content knowledge and considering probability (see figure 6.3). Probability plays a special role when students evaluate the interpretation that includes two tracks starting from the same point or two tracks positioned so that they appear as one track. In this case, considering the low probability of this interpretation occurring is the only way to explain why this interpretation cannot be considered the most likely event. Trying to explain this with content knowledge seems to get stuck in a vicious circle.

The general goal is that students learn to:

- Realize more possibilities while working on or thinking about the problem
- Evaluate different possibilities using probability judgment
- Decide about the most likely event

7.3.1.1 Results of evaluation of likelihood and uncertainty analysis skill- Case 1

Video and audio data

Case 1 (N=10) formed 4 groups in the class. Student discourse was analyzed in each individual task in the worksheet: making interpretations, evaluating interpretations, and making decision about the most probable event. Although evaluating different interpretations involves a decision-making process, in the worksheet a separation was made by asking two questions to explicitly emphasize on the goal of applying the LU skill by bringing the term “the most likely event” into students’ consciousness in order to increase the chance of transferring the skill to the other context e.g. everyday life.

Fragment 1: Realizing more possibilities

In the first task, students were given 4 copies of Anderson’s photograph and asked to make different interpretations about the sign of the charged particle and the direction of the particle motion causing the track. The overall goal of the task was that students learn to make more interpretations because finding hidden possibilities is not an easy task and requires students’ persistency.

Making two interpretations: To make interpretations the students applied the relevant content knowledge; the right-hand rule. Considering the direction of the magnetic field, they gave correct reasons to conclude about the first two interpretations: a positive particle moving upward from down to up or a negative particle moving downwards from top to down (see copy 1 and 3 in figure 7.1).

> Julia: *So it [the particle] can have minus, then it runs to the right, once counterclockwise, then maybe it can run clockwise, then/ (she draws a positive particle in the worksheet).*

Since the first two interpretations were very spontaneous and lacked complexity for the students, students were very positive in doing the task. Therefore, the first conjecture was formulated as following:

LU-C1: Students feel confident in concluding obvious solutions.

Making more interpretations: Going from 2 interpretations to 4 interpretations seemed a challenge. However, the students knew that more interpretations are possible (because they were given 4 copies) but they stopped after making 2 interpretations. In the following example, Martin has expressed that there are only two possible interpretations, and over time he has not changed his mind, although he has actively participated in the group discussion to find other interpretations.

00:01:54-5

> Martin: *I do not understand. (...) there are only two possibilities, aren't there?*

...

00:04:12-2

> Martin: *I actually still understand two [interpretations].*

Examples of this type showed that despite having elements such as motivation, time on task, and willingness to complete the task, students were unable to come up with more creative solutions because they felt unconfident and loss of control to go beyond the obvious solutions that closely matched their prior knowledge. This indicated the overconfidence problem defined by Halpern (2009) as that “people tend to be more confident in their decision about probabilistic events than they should be.” The relevant conjecture was formulated:

LU-C2: The overconfidence problem prevents students to think of hidden solutions.

Encouraging students to overcome "overconfidence" and to make more interpretations by the teacher helped them to progress on the task. The conjecture was made that:

LU-C3: Students make progress in finding hidden solutions when they are encouraged by the teacher to overcome their "overconfidence" and think of more solutions.

They used a brainstorming technique to find more interpretations and therefore, they applied content knowledge to reason about the type of the particle that could have caused that track. While brainstorming, some false reasons (misconceptions) such as “a neutral particle caused the curved track” showed up. Class discussion and reflection on the false reasoning helped the students to finalize more interpretations. The relevant conjecture was formulated that:

LU-C4: Students' creativity in finding hidden solutions is activated through brainstorming.

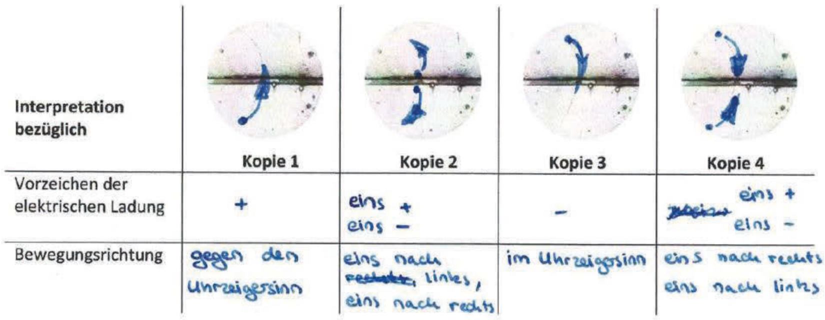


Figure 7.1: Julia's four interpretations in the first task on Anderson's cloud chamber photograph worksheet.

However, from the perspective of critical thinking, a creative thought is characterized as the one that offers relevant solutions, is reflective, and produces successful results (*cf.* Lipman, 2003, pp. 244-247). These criteria clarify the boundaries for classifying an interpretation as creative and identify a solution as creative that is correct based on required content knowledge in making interpretations; the right-hand rule. Therefore, the simple mentioning of 4 interpretations cannot be considered creative if they are not correctly based on the right-hand rule. Figure 7.2 shows an example of interpretations which cannot be considered a creative solution. Interpretations in copy 2 and 3 are not correct based on the right-hand rule. Therefore, the confirmation of LU-C4 was given by analyzing the students' answers on worksheet 3 (see student work section).

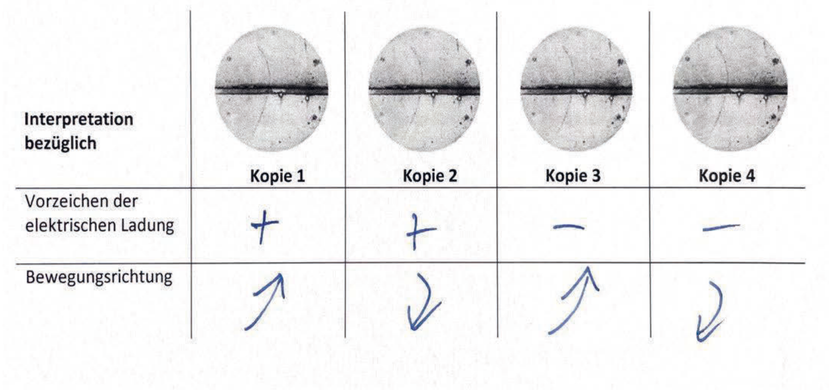


Figure 7.2: Example of an **uncreative** solution to the task involving different interpretations.

Disposition: During working on the task to make more interpretations, the students reflected on their answers, asked questions and participated in finding a correct answer.

> Marie: *It could also be an antineutron. That means it goes/ It has no sign at all. It just goes. Oh no.*

Moreover, the task was complex and required students' persistence to solve it. Two types of dialogues were identified as indicators of students' persistence to make more interpretations. Some students were motivated and enthusiastic to solve the task (e.g. Julia in the following discussion), and some were committed to do the task (e.g. Anton and Robert).

> Julia: *Why? Why? Wait wait. Now it hangs. I would say it should then/ (...) I would say or wait/ wait wait wait.*

And,

> Anton: *That's half of [interpretations].*
>
> Robert: *Half.*
>
> Anton: *Yes, but there is somehow more.*

In both examples, the students persisted in working on the task and they came up with 4 interpretations at the end. The relevant conjecture was formulated that:

LU-C5: Students' disposition (persistency and reflection) in CT tasks helps them perform more effectively in the task (findings correct reasons and completing the task).

Fragment 2: Evaluating different possibilities using probability judgment

In the second task, students were asked to evaluate their interpretations. The overall goal of the task was for students to apply content knowledge and to consider probability (when it applies) to evaluate their interpretations.

It is worth mentioning that in the positron discovery lesson students did not receive the first and the second task simultaneously. After completing the first task and an in-group and in-class discussion, students learned the necessary content knowledge e.g. calculating the radius of the track, relationship between radius of the track and energy of particle, and definition of range in particle physics. Next they receive the second task.

To evaluate the interpretations, students were expected to apply the law of conservation of energy. Along discussion and reflection in the group they gave correct reasons applying the law of conservation of energy to evaluate the interpretations.

Evaluating the obvious interpretations (copy 1 and 3 in figure 1) seemed not complicated (**confirmation of LU-C1**). The problem aroused in evaluating copy 4 in figure 1 when two particles, one positive and one negative, are coming to the centre of the photograph independently (hidden interpretation).

Two approaches were distinguished in the evaluation of this interpretation, which can be called a probabilistic event: some students reflected on the low probability of two events occurring simultaneously, while some others used content knowledge to argue against a probable event, but sometimes without good justification.

An example of first approach is Marie, when she considered the low probability of occurrence of this interpretation as a counterargument. She said:

Marie: *Why should the two particles meet at exactly the same point?*

In this case, Marie questioned the probability of occurrence of the two tracks caused independently by a positive and a negative particle pointing to the center of the photograph. This consideration must be used as an opportunity for an explicit discussion of likelihood and uncertainty analysis skill to make clear that "the co-occurrence of two or more independent events is less likely than the occurrence of either one alone." (Halpern, 2009, p. 187), whereas most people think just the opposite (Halpern, 2009, pp. 167-9). Although, a class discussion was planned that students discuss their ideas for evaluating this interpretation, but Marie did not dare to share this wonderful idea.

Observing this phenomenon contributed to an important heuristic for course design. However, after the students completed the task and discussed their results in class, the teacher made this point explicit to the students, but it has a greater learning effect if the students can think and figure it out by themselves. Therefore, while confronting students with the evaluation of a probable event, an explicit hint or question to compare the probability of two events occurring with the probability of one event occurring might help students better understand the application of the LU skill (**Heuristic 1**). In mathematics, students learn to calculate the probability of two events occurring by multiplying the probability of each event. The result of this calculation shows them the lower probability of two events occurring simultaneously compared to just one event. Here, they apply that skill in the mental level and see a bigger picture of using probability judgment to evaluate an interpretation.

The second approach was observed in a group where students discussed intensively applying the law of conservation of energy to argue against the interpretation in copy 4 in figure 1. They discussed that the energy of the particles decreases as they cross the plate. Therefore, an extra track of smaller radius for the negative particle is expected to be observed at the bottom of the plate as it passes through the plate from top to bottom. Since an extra track was not observed, it was assumed that the plate should completely stop the negative particle. Since this cannot occur (students did not mention the reason), this interpretation is unlikely.

The second approach could be interpreted either as a problem of overconfidence or as a lack of sufficient content knowledge. It was assumed that the lack of content knowledge might put students in an uncertain situation where they look for a cause that is hidden from them. However, this hypothesis was rejected because students in the class demonstrated a high level of conceptual understanding by discussing the law of conservation energy and the relation between the radius of the trajectory and the energy of the particle. Data analysis confirmed the overconfidence problem and showed that this group of students wanted to be more certain than it was necessary in an uncertain situation, and sometimes argued illogically to prove their ideas (e.g., that the plate cannot stop the negative particle without giving the reason) (**confirmation of LU-C2**).

Fragment 3: Making decision about the most likely event

Arguing for and against different interpretations guided students to make a decision about the most probable event. They all agreed that the copy 1 in figure 1 is the most probable event and argued that:

> *The positive particle loses energy passing through the lead plate, so the radius gets smaller.*

Furthermore, data analysis showed that students frequently used words such as “most likely”, “likely”, and “unlikely” which are related to applying probability in decision making. This showed that students were aware that there was a probabilistic scenario in the task, but in some cases the overconfidence problem prevented them from expressing their thoughts more clearly. A relevant conjecture was made that:

LU-C6: Using words such as “most likely”, “likely”, and “unlikely” accompanied by a correct reason (from content knowledge or probability perspective) shows that students developed the notion of reasoning in the context of LU.

Student work

To make a valid conclusion about students' progress in content knowledge and also in LU skill, students' answers in worksheet 3 were analysed. Data analysis showed that all students applied correctly high level content knowledge (law of conservation energy, relation between the radius of the trajectory and the energy of the particle) to evaluate interpretations and make decision about the most probable event using the word "most likely" (**confirmation of LU-C6**).

Furthermore, most students gave more than 2 interpretations, which were discussed above as hidden solutions. In order to label these solutions as creative, they were evaluated based on content knowledge (the right hand rule) and the results showed all the presented interpretations can be considered as creative ideas (**confirmation of LU-C4**).

Moreover, in evaluating the hidden creative interpretations, only one group addressed the low probability of these interpretations occurring as a counterargument by writing, "How is it possible for two particles to meet at exactly the same point?". Since discussing probability is important in the context of particle physics, embedding Heuristic 1 in the course might help make this explicit for students.

Student interview

Following the completion of the antimatter course, an interview was conducted to assess students' perceptions of the antimatter course, CT, and their ability to apply CT skills.

One of the first conjecture that emerged from the analysis of the video and audio data was that students' disposition in CT tasks helps them to complete the task more effectively (**LU-C5**). This is a self-evident explanation that students' persistence in working on a complex task and reflection on their own thought processes and those of others help them make progress on the task. Clearly, acting at this level of consciousness can be influenced by the student's perception of CT.

Case 1 students perceived CT valuable and applicable in everyday as well as academic life. However, students were aware that CT requires mental effort, but they expressed a desire to apply CT skills in everyday life (e.g. when discussing a topic or analyzing political arguments in elections) and also mathematics and physics when they are required to think logically. In addition, they expressed satisfaction that they were able to engage in critical thinking skills in the antimatter course. This positive perception of CT confirmed the interpretation of video and data on students' disposition (**confirmation of LU-C5**).

Teacher interview

Following implementing the lessons in each session, an interview was conducted to learn how the teachers perceived the course structure, materials, and CT.

Looking at teacher perceptions of the antimatter course in relation to student CT development revealed that the teacher perceived worksheet 3 as a good tool for developing students' LU skills. The students' efforts to make more hidden interpretations were appreciated by the teacher (**confirmation of LU-C5**), and the opportunity for discussion and reflection on the task was also perceived positively.

7.3.1.2 Developing students' likelihood and uncertainty analysis skill- Case 1

Given the above conjectures, the development of students' likelihood and uncertainty analysis skill is characterized by the integration of content knowledge, disposition, and the consideration of likelihood in decision making.

To realize obvious possibilities, students needed to reach a "threshold of comprehension" (Viennot & Decamp 2018) in particle physics context; the right hand rule. All students (N=9 students in 4 groups) reached this point and reasoned correctly to make the two obvious interpretations.

To realize hidden possibilities, students needed to overcome the "overconfidence problem". Brainstorming technique and appreciating students' efforts by the teacher, encouraged students' creativity to think of more hidden interpretations (considering two tracks instead of one). Furthermore, students' disposition was observed during working and finally 3 groups (N=7) made 3 and 4 interpretations.

To evaluate different possibilities, students needed to reach a higher level of comprehension in particle physics context; calculating the radius of the track, relationship between radius of the track and energy of particle, and the law of conservation of energy. Furthermore, they needed to argue the probability of occurrence of interpretations to improve the validity of their evaluation. All students (N=9) reached the required level of comprehension to argue about the probability of occurrence of obvious interpretations.

When evaluating the hidden interpretations, students (N=7) took different approaches. Some discussed probability to argue against them, while the others tried to apply the higher level of content knowledge, to make counterarguments. The second group neglected the probabilistic nature of the interpretations and thus, as expected, could not argue logically (with reasoning) on some points to evaluate the hidden interpretations. This indicated the unconscious occurrence of the "overconfidence problem" (Halpern, 2009) and highlighted the importance of communicating the problem in the first stage when students stopped after making two obvious interpretations. Observing this phenomenon contributed to the second heuristic for course design. Discussing the "overconfidence" problem can help students to develop an attitude toward using probability judgment to improve decision making (**Heuristic 2**).

In order to make a decision about the most likely interpretation, given that it was one of the obvious interpretations, all students (N=9) correctly applied the content knowledge and used words such as "most likely" to argue for it. This showed that students developed the notion of reasoning in the context of LU.

Furthermore, students' **disposition** was observed through critical discussion on the ideas, reflection on their own line of reasoning and the others, and engage actively for a satisfactory level of understanding the content knowledge. During discussions some ideas were rejected and in the end, the ideas reported on the worksheet remained as accepted ideas by the group. The students' positive perception of CT also justified their disposition. In addition, teacher appreciation of students' effort confirmed students' disposition.

In summary, all students' achievement of the required content knowledge, observation of their disposition while working on the task, and discussion of the probability of a likely situation indicated the development of students' LU skill. It was found that helping students to overcome the overconfidence problem in the first task gives them courage to think out of the box and see more hidden possibilities. This is exactly the goal of the task to teach students to be brave enough, break their habitual mental patterns and look for more hidden possibilities while working on the problem.

However, an explicit discussion of the problem of overconfidence at this stage helps students to become aware of the problem and understand it at the reflective level (**Heuristic 2**). This awareness could improve students' understanding of the likelihood and uncertainty analysis skill and could help students to evaluate the hidden interpretations more effectively by considering the probabilistic nature of the interpretations.

7.3.1.3 Developing students' likelihood and uncertainty analysis skill- Case 2

The implementation of the positron discovery lesson in case 2 coincided with the strict lockdown regulations in the Free State of Saxony, so it was not possible to collect video and audio data directly by the researcher. The available audio data from this lesson included student discussions about the reason for placing a lead plate in the centre of the cloud chamber after working on the making interpretation task on worksheet 3.

Analysis of the audio data revealed that the students realized hidden interpretations and discussed the existence of two tracks moving toward the plate from different directions and, conversely, moving away from the plate in different directions. While trying to justify placing the plate in relation to these two tracks, the students reflected on their thoughts and asked reflective questions from the peers. This led them to consider the probability of these interpretations occurring as a counterargument and to think of one track running through the lead plate as one more possible interpretation.

It is worth noting that since the audio data was not directly related to the discussion of making and evaluating interpretations, the overconfidence problem, for example, cannot be realized in the discussion. However, it can be seen that while students were justifying the logic of placing a lead plate in the centre of the cloud chamber, they were discussing the interpretations already made in the first task and also the probability of occurrence of these interpretations. In addition, students' disposition was perceived in their discussion (**confirmation of LU-C5**).

Furthermore, analyzing students' answers on worksheet 3 (N=9) confirmed that the students applied the high level content knowledge to evaluate the interpretations and also decided correctly about the most probable event (**confirmation of LU-C6**).

In addition, analyzing students' answer in the online open-ended questionnaire (N= 7) showed that they perceived CT applicable in variety of contexts however they aware that applying CT is not easy and needs efforts (**confirmation of LU-C5**). In the teacher questionnaire, the teacher considered the tasks on worksheet 3 suitable for teaching the students the CT approach. The teacher also appreciated the students' efforts and discussion and reflection in finding and evaluating different possibilities as indications that the students' LU skill was developed (**confirmation of LU-C5**).

In addition to the data already mentioned, students in case 2 were asked to include their reflection on what they learned after completing the task on the worksheet. Analysis of the students' written reflection revealed that the students perceived the task as one that encouraged creativity and open-mindedness and demonstrated the scientific method of making hypotheses and weighing probabilities in evaluating them (**confirmation of LU-C5**).

In summary, students in case 2 recognized not only obvious interpretations, but also hidden interpretations that met the criteria for creative interpretations. To evaluate the interpretations, they applied a high level of content knowledge and also discussed the likelihood of these interpretations occurring. They also demonstrated a disposition to engage in CT by discussing and reflecting on their own ideas and those of others. Their disposition was also confirmed by the teacher and was validated by their positive perception of the CT tasks. Overall, they presented a developed model of LU skill.

7.3.1.4 Developing students' likelihood and uncertainty analysis skill- Case 3

In case 3, audio and video data were not available. However, analysis of student answers on the worksheet (N=14) showed that most students gave more than two interpretations that were correctly based on the right-hand rule and could be considered as creative solutions (**confirmation of LU-C4**). In addition, students applied a high level of content knowledge and also argued the probability of occurrence of these interpretations to evaluate them and decide on the most likely interpretation (**confirmation of LU-C6**).

In addition, the teacher evaluated students' discussion, reflection, and efforts in doing the tasks on worksheet 3 as an indicator for developing students' LU skill. This shows students' disposition when working on the tasks on worksheet 3 (**confirmation of LU-C5**).

In summary, a combination of a high level content knowledge, considering of the probability of occurrence of interpretations while evaluating them, and students' disposition toward CT demonstrated a developed model of likelihood and uncertainty analysis skill in case 3.

7.3.1.5 Discussion of developing likelihood and uncertainty analysis skill

Cross-case analysis

Constant comparison of the data in each case provided evidence of the development of content knowledge and notion of reasoning in the LU context, and demonstrated student disposition to critical thinking (see table 7.4). In all cases most students realized hidden possibilities and argued against them by considering probability. In addition, the audio and video data in case 1 provided a more detailed analysis of the processes students go through to move from obvious to hidden interpretations. This revealed the problem of overconfidence and highlighted opportunities to help students overcome overconfidence.

In conclusion, what can be inferred about the effectiveness of the antimatter course is that teaching the skill of **likelihood and uncertainty analysis** in the context of Anderson's cloud chamber photograph actively engaged students in brainstorming creative ideas in all cases and also enhanced their notion of reasoning by applying content knowledge and considering probability. The instruction also helped students overcome the overconfidence problem when looking for hidden interpretations.

Table 7.4: Conjectures about developing of Likelihood and Uncertainty Analysis skill observed in different cases.

	Case		
	1	**2**	**3**
Conjectures about developing of Likelihood and Uncertainty Analysis skill (LU-Cs)	**Approved by data type analysis**		
1. Feeling confident in concluding obvious possibilities	*Video and audio data*		
2. Overconfidence prevents thinking of hidden possibilities	*Video and audio data*		
3. Progressing in finding hidden possibilities when encouraged by the teacher to overcome overconfidence and think of more solutions	*Video and audio data*		
4. Brainstorming creative ideas adhered to relevant criterion	*Video and audio data* *Student work*	*Student work*	*Student work*
5. Disposition in CT tasks and positive perception of CT task	*Video and audio data* *Student interview* *Teacher interview*	*Audio data*[1] *Student questionnaire* *Teacher questionnaire*	*Teacher questionnaire*
6. Developed notion of reasoning in the LU context by using words such as "most likely", "likely", and "unlikely" accompanied by a correct reason (from content knowledge or probability perspective)	*Video and audio data* *Student work*	*Audio data*[1] *Student work*	*Student work*

[1]The audio data in case 2 included student discussions about the reason for placing a lead plate in the center of the cloud chamber after completing the making interpretation task on worksheet 3 task. Because the data are not directly related to the tasks on worksheet 3, C1, C2, and C3 could not be inferred from them.

7.3.2 Argument analysis skill in the Big Bang context

"Argument analysis skill is the skill that is needed to judge how well reasons and evidence support a conclusion" (Halpern, 2009) by considering counterevidence, stated and unstated assumptions, and the criteria of correctness, relevance, sufficiency, and validity (Halpern, 2009, p. 139).

The Argument analysis skill includes the following abilities while constructing or evaluating an argument:

- Identifying/giving premises
- Identifying/giving reasons
- Identifying/giving counterargument
- Identifying/giving conclusion
- Seeking criteria:
  - Correctness: are the premises, reasons, and conclusion correct?
  - Relevance: do the reasons support the conclusion?
  - Sufficiency: which and how many components are missing?
  - Validity: judging the credibility of an information source.

In lesson 7 (Big Bang lesson), worksheets 4 (Big Bang) and 5 (searching systematically) were used to support students' development of argument analysis (AA) skill. In worksheet 4 students were asked to watch a 4-minutes video about „Big Bang" and write their questions arising during watching the video. That was an individual task.

On worksheet 5, students were asked to work in groups using a step-by-step guide (instruction) to select a specific question from the questions they had found individually on worksheet 4, collect information to answer the question, and conclude about the answer. The step-by-step guidance (instruction) included:

1. Selecting a question as the main question to begin the systematic search

  - Defining a precise question
  - Considering the need to define sub-questions
2. Collecting information to answer the research question
  - Justifying the validity of information
  - Considering the need for more information
3. Making valid conclusion
  - identifying the reasons and premises
  - Justifying the logic of the conclusion
  - Considering counterarguments

In addition, based on the question and the information collected, different scenarios can emerge to draw valid conclusion:

→ Students may come up with a definitive answer to their questions. In this case, the conclusion included that answer and students were expected to provide reasons for that conclusion based on the information they gathered.

→ Students may recognize that there are different perspectives to answer the question. In this case, they should conclude that there are different perspectives to answer the question, and to support their conclusion, they should explain the different perspectives.

→ Students may not find an answer to the question because the question is open. In this case, they should conclude that there is no answer to the question, and to clarify the point, they should justify the openness of the question.

The data were analysed using constant comparative method to provide accurate evidence of the development of students' argument analysis skill. Furthermore, the processes that students go through to achieve the desired outcomes of applying argument analysis skill are discussed.

7.3.2.1 Results of evaluation of argument analysis skill- Case 1

Video and audio data

Selection of main question for the systematically search

To begin the systematic search, students were asked to select a precise question. In teaching general CT skills in the first session, students learned that a term is considered precise if it is correct and does not require further specification. "Correctness" refers to the meaning of the term based on content knowledge and in relation to the context in which it is used. Considering "further specification", the following example would be helpful: in the sentence "bake the cake in a hot oven" the term "hot" is not precise because it needs to be specified how hot the oven should be (Halpern, 2009).

Data analysis showed that when selecting a main question to begin the systematic search, the students considered the answerability of the question and the achievement of consensus among all members, in addition to the criterion of precise formulation of the question. This means that finally a question was selected that was asked by all members in their individual works and proposed in a clear formulation among the others. It shows a respectful atmosphere in the group and explains generally that in an open-minded and respectful atmosphere, students do not obey the rules firmly (e.g. only seeking criterion of precision), but they introduce some social values like achievement of consensus to their criteria (**General-C1**). Consensus-Seeking is one of the dispositions required for critical thinking (Halpern, 2009).

After finalizing the main question, the next task was to consider whether the question could be broken down into sub-questions. This led to a discussion about the definition and examples of sub-questions in the groups and eventually contributed to defining the sub-questions in some groups.

In addition, before beginning the search, students were asked to assume whether the answer to the question posed was a definitive answer or whether it included multiple perspectives. It was clear that students had no idea of the answer prior to the search, but the goal was to make students aware of the importance of considering multiple perspectives while searching for the answer since this is one of the students' challenges in making arguments (Garcia-Mila & Anderson, 2007; Zohar, 2007). Data analysis yielded explicit discussion of the importance of considering multiple perspectives in the groups, with the novelty of antimatter research cited as a reason.

Collect information to answer main question

Justifying the validity of information

In searching for the answer, students checked various websites as information sources. After noting the information, they were asked to justify the validity of the information. Most groups considered representing the similar information on different websites as a criterion for the validity of the information. They did not however take the credibility of the websites into account to strengthen their argument about the validity of the information. Observing this led to the conjecture that:

AA-C1: Students perceive "the number of sources" presenting the information as evidence of the "validity" of the information without examining the credibility of the information source.

Observing this phenomenon contributed to a heuristic to design the lesson. Explicit reference to the distinction between credible websites that have expertise and first-hand knowledge and "junk" websites that present copied information from other websites without citation of sources could improve students' sensitivity in using websites as the information source and eventually judging the validity of information (**Heuristic 3**).

Apart from the fact that students made the mistake of considering the "majority" as a criterion for validity, it was also observed that when judging the validity of information, students looked for counterarguments. The following discussion presents an example of this. In the group, the main question was: why was some matter left over after Big Bang? For the answer students considered the matter-antimatter asymmetry problem.

> Robert: *So, do you think this information is correct? Yes or no? Why do you think so? [the student reads the task on worksheet 5.]*
>
> Martin: *Yes.*
>
> Anton: *Probably.*
>
> Martin: *Two sources. Look, if scientists say: this is probable and this is on two pages, it must be true. Have you found another one yet? Or any other reasoning [which is different]?*
>
> Anton: *No, but it could be also simply that at the beginning not equally much matter and antimatter has originated.*
>
> Martin: *But we have said that matter and antimatter particles are always produced as a pair [counterargument].*

Martin applied content knowledge about pair production to provide a counterargument. This shows that confronting students with a reflective question that challenges their thinking (e.g. why do you think so?), when content knowledge is present, helps them reflect on their thinking process and possibly find counterexamples. Such an example demonstrates that students have reached the required level of content knowledge and also their persistence to engage in a reflection process.

Considering the need for more information

In searching for the answer and the reasons behind it, students noted that there are still open questions in antimatter research. When asked if they needed more information, they discussed the gaps in some places, asked relevant questions that if answered would close the gap, and expressed a desire to get more information to answer the question. Breaking down complex information into its components and asking relevant questions shows that:

AA-C2: Students adhere to the criterion of relevance in order to find sound reasons that support the conclusion (information to answer the question).

Making valid conclusion

After completing the task, each group presented the answer to their main question. It was expected that students make a sound argument based on their findings in which the conclusion is supported by the reasons.

Data analysis showed that students realized that besides the theoretical findings, the experimental results needed to allow them to make a valid conclusion. They discussed the open questions in their chosen antimatter research and argued the importance of further research. This led to the formulation of the conjecture that:

AA-C3: Students consider a conclusion is valid when the relevant information from different perspectives (e.g. theory and praxis) supports the conclusion. In other words, students recognize the missing components in the argument and understand that they are necessary to draw a valid conclusion.

Justifying logic of conclusion

To evaluate the logic of the conclusion, students referred to their searches and findings from various sources, again lacking an evaluation of the credibility of the information sources (**confirmation of AA-C1**).

Student work

The quality of the arguments students made was assessed by analyzing student responses on worksheet 5 to determine whether students used the correct terminology when writing reasons and conclusions and whether the reasons supported the conclusion. Data analysis showed that in all groups, students used correct terminologies in writing reasons and conclusions and also the reasons supported the conclusion (**confirmation of AA-C2**). Furthermore, the scientists' assumptions (premises) were realized correctly by students in term of relevance to the conclusion and the use of correct terminology (**AA-C4**). In addition, as in their discussion, the students considered the majority of the sources of information as criterion for evaluating the validity of this information (**confirmation of AA-C1**).

Student home work

To evaluate student development of AA skill, they were given a homework to choose a research topic from the list of proposed researches about antihydrogen (e.g. measuring its mass, trapping it, its spectrum, measuring fall acceleration) or by their choice and write a report.

Data analysis showed that students made valid conclusion by adhering to the criteria of

- correctness of the terminologies and the information when providing reasons, premises, and conclusion.
- relevance when giving reasons (**confirmation of AA-C2**).
- validity of the information by discussing the credibility of information sources (**counter example of AA-C1**).

Furthermore, the students expressed the need for more information when the information provided in the text was not clear to them (critical attitude toward incomplete explanation).

7.3.2.2 Developing students' argument analysis skill- Case 1

Argument analysis skill focuses on judgment about the process of reaching a conclusion. It includes the abilities of identifying the argument components (premises, reasons, conclusion), and evaluating them by applying criteria such as correctness, relevance, and validity.

Data analysis showed that students identified the argument components correctly. They searched different sources to find more reasons (information to answer the main question) and also considered scientists' premises. In addition, they adhered to the criteria of correctness and relevance by using correct terminology when formulating premises, reasons, and conclusions and by giving the reasons supported the conclusion.

However, in order to evaluate the validity of the reasons (the information), the credibility of the information source was neglected by the students in their group work. Discussing this point in class helped students to consider this point in their homework and discuss the credibility of the information source as an indicator of the validity of the information. Students also understood that they need more information and expressed this by asking relevant questions when they received information from the text that was new to them.

In summary, the students showed a developed AA skill by paying attention to the criteria of correctness, relevance, and validity when creating an argument. This was particularly evident in the analysis of their homework.

7.3.2.3 Developing students' argument analysis skill- Case 2

Student work

The students mostly formulated clear and precise questions from the beginning. Only in one group the main question was not precise, but as the students were asked to make sub-questions, they formulated the main question precisely.

When they searched for the answer, they found that their questions were still open and there were only some theories but no experimental evidence. Therefore, they all questioned the validity of the information due to the lack of experimental evidence (**confirmation of AA-C3**). They also expressed the desire for more information by asking relevant questions (**confirmation of AA-C2**). In addition, only one group (out of 3) correctly addressed the assumptions that the researchers made to answer the question. Other groups did not answer the respective question on worksheet 5 (**counterexample of AA-C4**).

Since the students did not find a concrete answer to their question due to the openness of the question and the divers perspectives, they did not formulate a conclusion. They did not consider the possibility of presenting the different perspectives in a way that would create an argument.

In summary, the lack of a conclusion in the student works does not allow any conclusions to be drawn about the development of AA skills in terms of writing a valid conclusion which supported by the reasons. However, data analysis showed that students in all groups used correct terminologies in writing the information gathered to answer the main question, but did not provide evidence of adherence to the relevance criterion. Since the students were unable to draw a conclusion, it was self-evident that they were also unable to recognize the scientists' assumption concerning the conclusion.

7.3.2.4 Developing students' argument analysis skill- Case 3

Student work

In student work, student realized that their questions are open questions. To judge the validity of information, however, some groups considered finding the information in internet as an indicator for validity of the information without considering the credibility of website (**confirmation of AA-C1**), some questioned the validity of information by discussing the lack of experimental evidence and the existence of only theories (**confirmation of AA-C3**). Moreover, like case 2, only one group (out of 4) correctly identified the assumptions that the scientists made to answer the question. The other groups did not answer the corresponding question on the worksheet (**counterexample of AA-C4**).

In case 3, the students did not draw a conclusion and justified this by the openness of the question. They did not consider the possibility of combining different perspectives to present a conclusion. This phenomenon was also observed in case 2.

This might indicate that students think of the conclusion as something solid and firm and do not consider the dynamic nature of a conclusion that focuses on what has already been found, and taking into account diverse perspectives (**AA-C5**). Observation of this phenomenon leads to a heuristic for teaching the skill of argumentation analysis. Explicit discussion of making an argument in case of openness of the answer should be planned (**Heuristic 4**).

Teacher questionnaire

The teacher appreciated worksheet 5 for teaching students the importance of credible sources of information in gathering valid information (**counterexample of AA-C1**) and mentioned:

> *Not being able to answer the "true big bang" question especially, "What came before?" frustrated the students a bit, but taught them a lot because you can't always just google an answer or again, you need multiple high quality sources. This was a clear learning progress!*

In summary, the failure to draw a conclusion in the student works suggests no conclusions about the development of AA skills in terms of writing a valid, reasoned conclusion. Regarding the openness of the questions students asked, they failed to gather relevant information and expressed their failure by writing, for example, "There is no good answer to the question". Although the teacher acknowledged the students' efforts in finding the answer, she described the class atmosphere as frustrating when students failed to meet their expectation of finding a definitive answer. This also confirms the importance of including heuristic 4 in the instructional design of the lesson.

7.3.2.5 Discussion of developing argument analysis skill

Cross-case analysis

Table 7.5 shows an overview of the conjectures derived from the data analysis. Constant comparison of the data in each case showed that in all cases students were actively working to answer their questions raised by watching the short video about the Big Bang. They recognized that some questions are still open in the field of antimatter research and that answering some questions requires multiple perspectives.

However, it was found that students in case 1 were better at drawing valid conclusions when they realized the open character of the question. They linked the information to conclude that "the question is still open" and then specified the aspects that need further research. In cases 2 and 3, this ability was not present. On the one hand, it shows a flaw in students' understanding of argumentation, namely that all components of argumentation, reasons and conclusion, must be confirmed information and cannot have an open character; on the other hand, it clarifies the importance of embedding Heuristic 4 in instruction to develop students' argument analysis skill when the question is still open.

Table 7.5: Conjectures about developing of Argument Analysis skill observed in different cases.

	Case 1	Case 2	Case 3
Conjectures about developing of Argument Analysis skill (AA-Cs)	**Approved by data type analysis**		
1. Evaluating the validity of information by considering "the number of sources" presenting the information without examining the credibility of information source	*Video and audio data* *Student work*		*Student work*
2. Adhering to the relevance criterion by breaking down complex information into its component parts and asking relevant questions in order to find sound reasons that support the conclusion	*Video and audio data* *Student work* *Student home work*	*Student work*	
3. Evaluating the validity of conclusion by testing its confirmation by theory and experimental work	*Video and audio data*	*Student work*	*Student work*
4. Identifying relevant assumptions to the conclusion	*Student work*		
5. Perceiving all components of argumentation, reasons and conclusion, as confirmed information, which cannot have an open character		*Student work*	*Student work*
Counterexample of AA-C1: Evaluating the validity of information by examining the credibility of information source	*Student home work*	*Student work*	*Teacher questionnaire*
Counterexample of AA-C4: Demonstrating difficulty to identify relevant assumptions to the conclusion		*Student work*	*Student work*

Furthermore, it was observed that students judged the validity of information based on "the number of sources" presenting the information without examining the credibility of the information source. This shows the importance of explicitly discussing this point with students, since it was observed that after the discussion, students corrected this fallacy and focused more on the credibility of information source to judge the validity of the information.

In conclusion, what can be inferred about the effectiveness of the antimatter course is that teaching students the skill of **argument analysis** in the context of finding answers to the questions that arose while watching a video about the Big Bang solved the majority fallacy in all cases and also enhanced their consideration of different perspectives in seeking to evaluate the validity of the information. However, it also revealed a new challenging aspect of applying argument analysis skill for students when the answer to the question is still open. Since this was not observed in case 1, no special treatment was considered to help students in cases 2 and 3 overcome this challenge. But it needs special attention.

7.3.3 Verbal reasoning in the context of analysing the scenario of scene of “Illuminati”

Verbal reasoning (VR) skill focuses on the language, how well the terms are defined and how precise they are. It is defined as the skill to identify and avoid persuasive or misleading information in a variety of contexts (Halpern, 2009, p. 79; *cf.* Tiruneh et al., 2018) in order to develop valid explanations (NRC, 2012; Viennot, 2001). A valid explanation is characterized as one that contains precise terms and complete information.

As discussed in the Argument Analysis skill, precise terms are characterized as correct terms based on content knowledge and in relation to the context in which they are used, as well as the fact that no further specification is required. Considering the correctness of a term in relation to the context protects the term from ambiguity, and considering the fact that the term is defined in such a way that no further specification is required protects the term from vagueness (Halpern, 2009, pp. 60-61). In addition, complete information is the result of discussing the need for more information.

In this study, the term "explanation" in the context of VR was taken from the German operational definition for the science curriculum (KMK, 2013) and means the application of content knowledge to communicate one's ideas (written or verbal) through justification. Although, there is a discussion about whether explanations need evidence (*cf.* McNeill & Krajcik, 2011), but it is not the aim of the present research to engage in this discussion.

To apply VR skill, one needs to

- Detect misuse of definitions (Recognize vague or/and ambiguous terms)
- Explain reason (give reason why the term is vague or/and ambiguous)
- Use precise terms
- Seek criterion "precision"
    - Correctness
        - how **correct** are the terms based on content knowledge?
        - are the terms **acceptable** in relation to the context presented?
    - Does the term need further specification?

As indicated earlier, the antimatter course was instructed such that in the last session, students should analyze the scenario for the scene from Illuminati seen in the first session based on their newly acquired content knowledge and CT skills (see figure 6.2). Students were asked to correct the information about antimatter in that scene and present and discuss their work to the class. To this end, in worksheets 7 and 8 students were asked first individually and then in group work to identify the vague and ambiguous terms about antimatter in dialogues in the movie excerpt, to give reason why the terms are vague or ambiguous, and to replace these terms with precise ones.

For example, the dialogue talked about 1 gram of antimatter stolen from CERN to destroy the Vatican. In the antimatter course, students learned in the lesson on the production of antihydrogen (Lesson 8) that CERN produces $5 \cdot 10^{-10} g$ of antihydrogen per year. Table 7.6 shows an example of student answers to the task of identifying vague and ambiguous terms, justifying why they are vague or ambiguous, and suggesting a precise term. The students' suggestion was evaluated based on the correctness and acceptability of the term, and also whether the term needs further specification. The term “small amount” is **correct** because the student defined the term by giving information about the yearly production rate of antimatter at CERN. It is also **acceptable** in this context because the term is used beside the information about the yearly production rate of antimatter at CERN, and a **further specification of the term** was given with this. Therefore, the term is **precise**.

Table 7.6: Sample of student answers on the "Illuminati" worksheet.

Vague/ambiguous term	Reason	Suggestion
1 gr Antimatter	*[producing 1 gr antimatter] would take 2 billion years.*	*Yearly $5 \cdot 10^{-10}$g of antimatter is produced at CERN. [My suggestion for making the excerpt precise is that instead of talking about 1 gr] in dialogues talk about a small amount [of antimatter].*

Regarding acceptability, it is important to distinguish between acceptability in context and acceptability in general. For example, the term "small amount" of antimatter **would not be considered acceptable** if used out of context without the information about the annual production rate of antimatter. This is because in everyday language the term "small amount" is often used to refer to the order of grams (e.g., 1 gram), but in this context it was referring to 10^{-10} grams.

7.3.3.1 Results of data analysis- Case 1

Video and audio data

Students’ discussion was analysed to find the strategies students use to replace a vague or ambiguous term with a precise term.

Identifying vague or ambiguous term

Data analysis showed that students recognize two types of vague or ambiguous terms:

- terms represented textually
- terms represented visually (visual input)

Even if VR skill reflects on the use of language, students' statements clearly showed that students treated visual input in the same way. For example, the movie depicted an antimatter trap with a different construction than what students had learned in the course. Students discussed the visual image of the trap shown in the movie and evaluated its shape and suitability for trapping antimatter.

This led to the conjecture that:

VR-C1: When combining text and visual tasks (e.g., analyzing the scenario for a scene from Illuminati while the scene is being shown), students reflect not only on the use of language, but also on the uses of visual input to identify vague concepts.

In addition, while discussing the vague terms, the students criticized incomplete explanation in the scenario and expressed their need for more information. Critical analysis of an explanation and identifying its inconsistency is considered crucial for conceptual understanding of a subject (Viennot & Decamp, 2020, p. 70). Observing this led to the conjecture that:

VR-C2: Students who take a critical attitude towards the incomplete explanation have at least reached a threshold of understanding the topic.

For example, in the following discussion, Anton criticizes the incomplete explanation in the scenario about the functionality of the antimatter trap.

> Anton: *I wrote down [as vague term] significant amounts [of antimatter].*
>
> Robert: *Ah, yes. Significant amounts.*
>
> Anton: *That's what they said. Just, we have a significant amount here. Then she [the actress] mentioned that it's flammable. (...) And how the storage is done. <u>But they just say it's in one of these traps, but [they did not say] how exactly it [the antimatter trap] works.</u>*

Using precise term

Terms represented textually

To replace the vague or ambiguous term with a precise one, students gave different suggestions. During discussion they reflected on each other ideas to find a correct and acceptable term. For example, in the following discussion, students try to find a precise term to replace “one gram antimatter”. They give suggestions, apply content knowledge to evaluate them, and challenge the suggestions by seeking an acceptable term.

> Marie: *… Well, I wrote that the film could be set in the future, where they've already made a lot of things [to produce 1 gram antimatter].*
>
> Anne: *Yes, but two billion years!*
>
> *[...]*
>
> Anne: *First a milligram [antimatter] and then through a million years …*
>
> Marie: *Well, you could also say first - I don't know - 20 years or so and then we'd be rid of at least a few of the [minus] powers of ten or something. I don't know.*
>
> Anne: *If we have 20 [years] now, then it's: one times ten to the minus eight.*
>
> Marie: *Exactly. Ten to the minus eight grams.*
>
> Julia: *Yeah, but that doesn't help you either. So that's almost nothing. I mean/*
>
> Marie: *[but it seems] something realistic.*

Julia: *I would say that the movie is played in the future and by then the technology is developed to make that [1 gr antimatter].*
Anne: *Okay.*

Marie: *Yes, [it] is good.*

[…]

Marie: *Or do you think that then in 20 years they would be able to make this one gram?*

Julia: *Well. It doesn't say now how far in the future/*

Marie*: Good.*

As evidenced by the discussion, the students applied content knowledge about the annual production amount of antimatter at CERN ($5\cdot10^{-10}$g), sought the acceptability of the term which would replace "one gram antimatter" by discussing the production amount of antimatter in 20 years and reducing some minus powers of ten. While thinking about using "the smaller amount of antimatter for example 10^{-8} g" in the scenario, they also discussed the amount of energy produced by this amount of antimatter and its power to destroy Vatican (as it pictured in the scene). They sought an acceptable term to preserve the consistency of the text.

They also reflected on each other ideas to provide precise information about production of antimatter. Realizing that "1 gram of antimatter" is a vague concept and cannot be achieved with the current technology at CERN, they suggested that the film be set in the future where the technology is developing. It is worth noting that the term "future" is vague out of context because it does not clearly state time, but in this context it is acceptable because students gave reasons based on content knowledge and sought criteria to justify the term. Observing this led to the conjecture that:

VR-C3: To replace a vague term with a precise one, students apply content knowledge, seek criteria of correctness of the term based on the content knowledge and its acceptability regarding to the context, and reflect on each other ideas.

This criterion-based performance when accompanied by reasoned justification, as observed in case 1, demonstrates students' critical thinking skill (Lipman, 2003). Furthermore, the discussion of the energy concept, especially since it is not directly mentioned in the scenario, shows a high level of content knowledge in which students make a mental connection between rest energy ($E=m\cdot c^2$) and its destructive power by comparing the amount of energy that destroyed Hiroshima.

Terms represented visually

As alluded to before, the students evaluated the visual image of antimatter trap shown in the movie and suggested an appropriate visual representation of the antimatter trap. They applied content knowledge about a Penning trap and discussed that the antimatter trap represented in the movie needs magnets around. This level of sensitivity to replace not only the vague concepts but also the vague visualisation, especially since it is not required in the task, shows students' persistence in completing the task and also their mastery of content knowledge. The relevant conjecture was formulated that:

VR-C4: Students demonstrate persistence by conscious attention to all details to replace vague terms in dialogues and visualizations with precise terms.

In class discussion

A class discussion was planned to collect students' ideas about the terms that were substituted in the analysis of the scenario during the group work. On the interactive whiteboard, the scenario was presented and a student was asked to write down the proposed new terms. The students reflected on the suggested terms to find precise terms (**confirmation of VR-C3**) and finally a consensus was reached among the class and a new scenario for this particular movie scene was determined.

Student work

In both individual and group work (worksheets 7 and 8), students identified vague terms represented textually and visually (**confirmation of VR-C1**). They correctly applied content knowledge to justify why the terms were vague and replaced them with precise terms, e.g., instead of the term "explosion" of matter and antimatter when they contact, students used the term "annihilation" or instead of the term "cylinder" that used to explain the antimatter trap, students used the precise name "Penning trap" (**confirmation of VR-C3**).

7.3.3.2 Developing students' verbal reasoning skill- Case 1

Data analysis showed that the students identified vague terms textually and visually in the scene of "Illuminati" movie, applied content knowledge to reason why they are vague and to replace them with precise ones. During group work and in class discussion, they reflected on each other ideas and challenged the suggested terms by asking reflective questions. Furthermore, they showed persistency in completing the task by making efforts in giving different suggestions and seeking criteria of correctness and acceptability. This shows a developed model of verbal reasoning skill that integrates content knowledge, critical disposition, and criterion-based performance.

Regarding the acceptability criterion of a term, it is important to emphasize on the context in which the term is presented again. This means that in some cases adherence to the acceptability criterion was not evident from the students' answers on the worksheet and further analysis of the students' discussion was required. For example, the term "future" was considered vague based on student responses on the worksheet, but it is considered an acceptable term based on analysis of student discussion when accompanied by reasoning and application of content knowledge. This shows the importance of data triangulation.

7.3.3.3 Developing students' verbal reasoning skill- Case 2

Video and audio data

Like case 1, students realized the vague terms textually and visually. They reflected not only on the language used in the scenario, but also the representation of antimatter trap (**confirmation of VR-C1**). In addition, the students criticized the incomplete information in the scenario, for example the functionality of antimatter trap (**confirmation of VR-C2**).

To replace the vague term with precise ones, the students applied content knowledge, sought criteria to formulate clear terms, and also reflected on their own and the others suggested terms (**confirmation of VR-C3**). They applied a high level of content knowledge and also showed persistence in completing the task by demonstrating sensitivity and attention to all details and both textual and visual information (**confirmation of VR-C4**).

In addition, it was observed that students used the internet website cautiously when searching for a precise term and used the credibility of the information source as a criterion for evaluating the correctness of the term (**counterexample of AA-C1**). This is a lesson learned from the task searching systematically in the context of Big Bang, and it helped them overcome the "majority fallacy".

Student work

Analysis of students' responses showed that they identified vague terms not only textually but also visually (**confirmation of VR-C1**). The students correctly reasoned why the terms were vague and suggested the precise terms (**confirmation of VR-C3**). Only in one individual work was the vague term "decay" used to explain the annihilation of matter and antimatter when they come into contact with each other. However, the term was not included in the group work. This could be due to the fact that during the group work and reflection on the proposed terms, it was determined that this term was not correct and therefore was not listed on the group worksheet.

In summary, data analysis showed that the students applied content knowledge to analyse the scenario, to justify why some terms were vague, and to replace them with precise terms. The students' effort to find precise terms by reflecting on their own and others' ideas and adhering to the criteria of correctness and acceptability was observed and showed a development in their verbal reasoning skill.

7.3.3.4 Developing students' verbal reasoning skill- Case 3

Student work

The data analysis showed that the students applied content knowledge correctly and identified the vague terms in the dialogue of the movie scene (**confirmation of VR-C1**). For example, they criticized the realism of the production of a significant amount of antimatter discussed in the scenario, and also the term "force" that was incorrectly used instead of "energy".

However, in the next task, where they were asked to find a precise term instead, only one group (N=3 students) correctly applied their content knowledge to find precise terms. In the other groups (N= 16 students), the vague terms were not replaced with precise terms. However, students in these groups showed their understanding of the need for more information and also of the need for precise definition of terms only through such superficial responses as "the information needs to be explained in detail" and "the term needs to be defined more precisely" (**confirmation of VR-C2**). This suggests that while there was an awareness of the need for more information and the need to define a term precisely, the students were not interested in acting on that awareness. The teacher clarified this point in the teacher questionnaire as well.

Teacher questionnaire

The teacher explained that the students did not perceive the task in worksheets 7 and 8 positively, as they criticized the task by mentioning that not much (in terms of content) is mentioned that makes sense in a pseudoscientific dialogue in an entertainment movie. This could be the reason why the students did not make effort to make suggestions when looking for precise terms and were satisfied with answers like "the information needs to be explained more precisely" and "the term needs to be defined more precisely".

In summary, although, the students recognized the vague terms by giving correct reason, they did not complete the task. Data analysis showed that students understood general CT skills such as the need to define terms accurately, but it failed to show evidence of applying this concept in the antimatter context. Since the students were sceptical of completing the task as the teacher stated, a conclusion about the development of students' VR skill is not possible.

7.3.3.5 Discussion of developing verbal reasoning skill

Cross-case analysis

Table 7.7 shows an overview of the conjectures derived from the data analysis. It was observed that in all cases, students achieved a high level of content knowledge as they recognized vague and ambiguous terms and correctly reasoned why they were not precise. They also expressed a critical attitude toward the incomplete explanation.

Because students in case 3 did not complete the task on worksheet 8, a conclusion about the development of students' VR skills in this case was not possible. However, students in cases 1 and 2 showed persistency in completing the task. They actively engaged in discussion about the vague and ambiguous terms and reflected on the suggested term and adhered to the correctness and acceptability criteria to find precise terms.

In conclusion, what can be inferred about the effectiveness of the antimatter course is that teaching **verbal reasoning** skill in the context of analyzing the scenario of a scene of "Illuminati" enhanced students' abilities in identifying vague concepts and also in adhering to correctness and acceptability criteria when applying content knowledge to find precise terms. Although these abilities were observed only in cases 1 and 2, they can be generalized because in case 3, as discussed in section 7.2, the teacher did not adhere to the structure of the antimatter course in which this activity was covered.

Table 7.7: Conjectures about developing of Verbal Reasoning skill observed in different cases.

	Case		
	1	2	3
Conjectures about developing of Verbal Reasoning skill (VR-Cs)	**Approved by data type analysis**		
1. Identifying vague concepts in the use of language and visual input	*Video and audio data* *Student work*	*Video and audio data* *Student work*	*Student work*
2. Demonstrating a critical attitude towards the incomplete explanation	*Video and audio data* *Student work*	*Video and audio data* *Student work*	*Student work*
3. Applying a high level content knowledge by linking various concepts and adhering to the correctness and acceptability criteria while reflecting on the suggested terms to find precise terms	*Video and audio data* *Student work*	*Video and audio data* *Student work*	
4. Demonstrating persistence by conscious attention to all details to replace vague terms in dialogues and visualizations with precise terms	*Video and audio data* *Student work*	*Video and audio data* *Student work*	

7.3.4 Thinking as hypothesis testing in the antimatter trap context

Thinking as hypothesis-testing (HT) skill is a mental checking of the thinking process when it involves in hypothesizing about possible solutions and examining their consequences based on content knowledge. A hypothesis is defined as possible explanation about the relationship between variables that predicts a particular outcome based on limited evidence (Halpern, 2009, 2014; Klahr et al., 2001; Norris & Ennis, 1989; NRC, 2012). To test its validity in mental level, it is important to understand the need to control variables, and also the need to operationalize the variables (Halpern, 2009).

The skills used in thinking as hypothesis-testing have much in common with the research methods that scientists use when they study phenomena in their academic field; namely scientific reasoning (Halpern, 2009, p. 156). During a discovery process, scientists generate hypotheses, design and perform experiments, and evaluate evidence (Klahr et al., 2001, p. 78). One difference, however, is that thinking as hypothesis-testing remains at the mental level, while scientists put their thinking into action. Table 7.8 shows the general commonalities between these two skills (Halpern, 2009, p. 159; Klahr et al., 2001, pp. 79- 81).

Table 7.8: Scientific Reasoning skills *vs* Thinking as Hypothesis Testing skills.

Scientific Reasoning skills	**Thinking as Hypothesis Testing skills**
Generating hypotheses	• Identifying variables • Being able to describe the hypothetical relationship between variables
Designing & executing experiments	• Understanding the need to control variables in order to make a strong causal claim • Recognizing the need for and using operational definition of terminologies used in a claim
Evaluating evidence	• Checking for adequate sample size • Testing the validity of hypotheses

To develop students' thinking as hypothesis testing skill, a worksheet in the context of trapping antimatter was designed. Before offering students the worksheet, they learned about trapping matter by watching a video about the quadrupole ion trap or Paul trap (TU Dresden, 2019) and discussing it. In the "Trapping Antimatter" worksheet students were asked to think of the changes which should be made in the construction of Paul trap to be able to trap antimatter.

Making a plan for the construction of an antimatter trap was considered as hypothesis generation. Here identifying variables and controlling them play a role. Students should identify that by changing from matter to antimatter, one variable of "type of particles in terms of particles and antiparticles" is varied. In this case, to hypothesize about the construction of antimatter trap, the other variables like mass and charge of (anti)particles should be kept constant (control variables). Using content knowledge about annihilation of matter with antimatter, they should make a proposal for a device to trap antimatter. To evaluate their hypothesis, they should compare the proposed antimatter trap with the Penning trap used at

CERN to trap electrically-charged antimatter (*cf.* Amole et al., 2014). Table 7.9 shows the domain-specific Thinking as hypothesis testing skills in the context of trapping antimatter.

Table 7.9: Domain-specific thinking as hypothesis-testing skills in the context of trapping antimatter.

Domain-specific thinking as hypothesis-testing skills	
Generating hypotheses	• Making a plan (a hypothesis) for the construction of an antimatter trap • Justifying the hypothesis based on content knowledge • Identifying the variables that cause a change in the design of the antimatter trap compared to the Paul tarp
Designing experiments	• Understanding the need to control variables in order to construct the antimatter trap • Drawing a sketch of the experimental setup and describing the functionality of each part
Evaluating evidence	• Testing the validity of the hypothesis by checking the proposed antimatter trap against existing systems

7.3.4.1 Results of data analysis- Case 1

Data were analyzed to realize the strategies students used to generate hypotheses and test their hypothesis in the context of trapping antimatter.

Video and audio data

Generating hypotheses

By watching and discussing a short video about Paul trap, students gained knowledge about trapping electrically-charged particles in the electric field, using AC-Voltage. They also observed that in the video a brush was used to place the electrically-charged particles in the Paul trap.

To hypothesize about the construction of an antimatter trap, students applied this background knowledge in the context of antimatter to identify what changes need to be made in the construction of the Paul trap. For example, students proposed to add vacuum to the Paul trap construction to avoid annihilation of antimatter on contact with air. They understood that change from matter to antimatter would cause a change in the design of the Paul trap.

Furthermore, data analysis showed that students were concerned about "the charge of antiparticle in terms of a neutral or an electrically-charged antiparticle" when making a plan for the construction of antimatter trap. They understood that trapping electrically-charged antiparticles can simplify the construction of an antimatter trap since in this case the electrically-charged antiparticle can be trapped between the electrodes connected to a AC voltage (as in the Paul trap). Observing this kind of discussion led to the formulation of the conjecture that:

HT-C1: Students consider unconsciously the concept of control variables. They try to keep the variable "charge of (anti)particle" constant while changing matter to antimatter (varying one variable).

To make the concept of identifying variables and controlling them explicit in the next study, worksheet 6 was redesigned for case 2 and case 3 by adding explicit questions to identify variables and control for them. The results are discussed later in the relevant sections.

Designing (thought) experiments

Recognizing the need for and using operational definition

Once students applied content knowledge to hypothesize about the construction of the antimatter trap, they were asked to draw a sketch of the proposed antimatter trap and explain the functionality of its different parts. Data analysis showed that students defined the vacuum explicitly while doing the task. It showed an unconscious applying of operational definition strategy which is considered as one of the outcomes of applying thinking as hypothesis testing skill (**HT-C2**). An operational definition is "an explicit set of procedures to recognize and measure a construct" (Halpern, 2009). For example, Martin defined the vacuum as following:

> Martin: *Well. In the/between the electrons there must be* ***vacuum*** *or at least* ***an extremely low air pressure****, so that (...) the antimatter does not come into contact with normal matter.*

Since the antimatter course consistently emphasized the use of precise terminology, it was natural for students to clarify the terms they used. In other words, the unconscious application of the operational definition is a conscious application of the definition of a precise terminology strategy as verbal reasoning or general CT skill. Therefore, above HT-C2 is reformulated as:

HT-C2: Students define terms precisely.

Understanding the need for more information

While discussing the construction of the antimatter trap, students realized that they need more information about the method of getting antimatter into the trap. In the following example, Lars and Richard discuss this requirement in different groups.

> Lars: *How do you get the antimatter in there [in the trap] or do you just wait?*

> Richard: *(…) Well, how do I get the [anti]particles in then [in the trap]. I have to get them in. It's as easy as I make it here [in Paul trap] with the spores, that I can take a brush and transport them in there? That will be difficult (...) So I have to think about it carefully: How can I realize that I get the particles there?*

Richard's statement shows that although he intuitively arrived at an answer based on his prior knowledge of getting particles into the Paul trap (with a brush), before reaching a conclusion he reflected on his answer and understood the need for more information. This shows that he has initiated a critical thinking process (*cf.* Kryjevskaia et al, 2021). This indicates that students who reflect on their intuitive solution based on their prior knowledge and experiences are the ones who engage in a process of critical thinking (*cf.*

Kryjevskaia et al., 2021), even though this may pose many challenges to achieving the desired outcomes, but can at least ensure that there is a disposition toward CT (**HT-C3**).

Evaluating hypothesis

To evaluate the mental representation of the antimatter trap (the hypothesis), a comparison of students' proposal and the Penning trap was planned. Unfortunately, due to time restriction (shortened lessons due to the summer heat) this phase was skipped by the teacher.

Student work

Looking at students' answers on the worksheet showed that they applied content knowledge and reasoned correctly to justify their hypothesis about the construction of the antimatter trap and also explain the functionality of its different parts (**confirmation of HT-C2**). Since students learned about Paul trap and its functionality, that seemed not too complicated for them to conclude about the obvious change in the construction of Paul trap to trap antimatter, adding vacuum. This confirms the conjecture in likelihood and uncertainty analysis skill that student feel confident in concluding obvious solutions (**confirmation of LU-C1**).

Furthermore, students showed their concern about considering the trapping of electrically-charged antiparticles in simplifying the construction of antimatter trap by explicitly asking the question about trapping neutral or electrically-charged antiparticles on the worksheet. This shows the unconscious application of the concept of control variables (**confirmation of HT-C1**).

7.3.4.2 Developing students' thinking as hypothesis testing skill- Case 1

Data analysis showed that students applied content knowledge and reasoned correctly for the obvious solution, adding vacuum in the construction of Paul trap. They unconsciously identified the variables such as "type of particle in terms of matter or antimatter", and "charge of (anti)particle" and discussed the need for control variables by planning the construction of antimatter trap for electrically charged antiparticles. Furthermore, they clarified their definition of vacuum while drawing a sketch of antimatter trap.

They understood the need for more information consciously by expressing their need to find information about how the antimatter will be placed in the trap. They discussed the different possibilities of getting antimatter into the trap in the group and also in the class, reflect on their ideas, and finally find an answer to their question, when the teacher explained penning trap apparatus for trapping antimatter.

In summary, however, students applied the content knowledge to hypothesize the construction of antimatter trap, they did not consciously discuss the concept of variables and controlling them. Observing this phenomenon showed the importance of an explicit reference to these concepts by the teacher to make students aware of the strategies they are using unconsciously and to bring these concepts to the conscious level (**Heuristic 5**).

To embed heuristic 5 in worksheet 6 for the next implementations in cases 2 and 3, explicit questions about identifying variables and controlling them were added to the worksheet. In addition, a general task, the “Bending Rods”, was added to allow students to transfer the skill of thinking as hypothesis testing from a domain-specific to a general context, also in physics.

The “Bending Rods” task defined by Piaget (*cf.* Halpern, 2009, pp. 148-150) asks students to think about which factors affect rod flexibility (see figure 7.3). The task gives students the chance to apply the principles of identifying variables and controlling them.

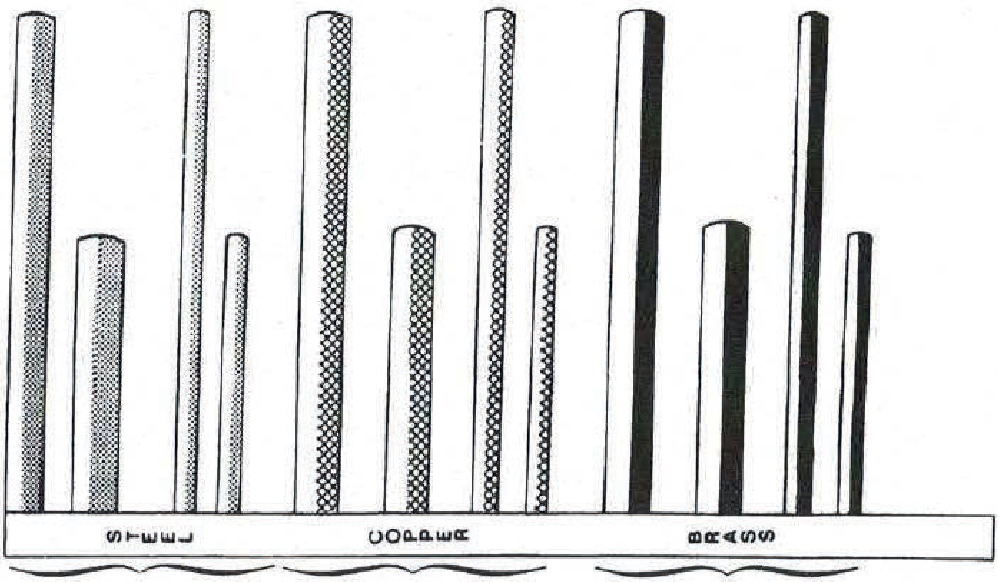


Figure 7.3: bending Rods (Halpern, 2009, p. 149).

In addition, lack of information about the processes students engage in when evaluating their hypotheses made it clear that this phase should not only be explicitly highlighted in the teacher package, but also discussed with teachers in advance of the next implementations.

7.3.4.3 Results of data analysis- Case 2

Student work

As indicated earlier, worksheet 6 was redesigned by embedding questions about identifying variables and controlling them, and adding the “Bending Rods” task (see appendix K).

Domain-specific thinking as hypothesis testing skill

Identifying variables

Analyzing students’ answers on the worksheet showed that the students identified the variables that influence the construction of the antimatter trap correctly (**HT-C4**). All three groups of students (N=7) mentioned that the mass and the charge of antimatter affect its trapping. In addition, the magnitude of magnetic field (by one group) and the temperature (by one another group) are mentioned as variables that influence the construction of antimatter trap.

Controlling variables

Students also realized correctly which of the mentioned variables bring changes in the construction of antimatter trap in comparison with Paul trap. They applied content knowledge and reasoned correctly that for example since the mass of particle and antiparticle is the same, mass causes no change in the construction of antimatter trap in comparison with Paul trap. To discuss about the role of the variable “charge”, students classified charge into the neutral and the electrically charged antiparticles. They reasoned that the electrically charged antiparticles can be trapped using the AC-Voltage like Paul trap but the neutral antiparticle cannot be. They suggested then to control the charge in term of considering electrically charged antiparticles not neutral antiparticles to be able to trap them applying the concept of AC-Voltage like Paul trap.

Looking at the answers of students who addressed the magnitude of magnetic field and the temperature as variables, makes something interesting clear. Although the temperature and strength of the magnetic field must be controlled when trapping antimatter, meaning that they must be kept constant during the experiment, here, in the context of hypothesizing the antimatter trap by considering the construction of the Paul trap, it became apparent that the students misunderstood the concept of control variables.

Control variable in the context of hypothesizing the antimatter trap by considering the construction of Paul trap means that the variables must be kept constant in both constructions (both experiments) as students discussed the variable “charge” and decided to consider trapping electrically charged antiparticles. In this sense, thinking of controlling the magnitude of magnetic field and the temperature means that using the same magnitude of magnetic field and temperature in both constructions which is false.

Paul Trap works in the room temperature, while for trapping antimatter the temperature of a few Kelvin is needed. This shows a misunderstanding about the concept of control variable between two experiments. Although they reasoned correctly that by controlling temperature, the atoms will have less energy and move less, they did not understand the concept of controlling variables between two experiments. The findings indicated students' difficulty in understanding and applying the concept of control variables in hypothesis building and testing (**HT-C5**) which is also confirmed by literature (*cf.* Boudreaux et al., 2008).

Hypothesizing the construction of antimatter trap

Students reasoned correctly that by changing from matter to antimatter (varying a variable), vacuum must be added to the construction of the antimatter trap. In addition, students applied their prior knowledge of the behavior of electrically-charged antiparticles in the magnetic field (as discussed for positron in the cloud chamber) and realized that the magnetic field is required to keep the electrically-charged antiparticle in a circular motion to avoid contact of the antiparticle with the sides of the trap, which is covered by a vacuum. They also added magnets in their sketches.

Considering that adding a vacuum is the most plausible change in the construction of the Paul trap, adding magnets can be regarded a less likely change for students to consider. This demonstrates a high level of content knowledge that allows students to think of hidden solutions. It might also show students' persistency in completing the task (**confirmation of HT-C3**).

Testing hypothesis

Although most students (N= 5 out of 7) considered the magnets when sketching the antimatter trap, these students surprisingly neglected the role of the electric field and did not add electrodes to the construction of antimatter trap. They only realized the error when asked to compare their proposal to the Penning trap. It could show that the students were thinking of either a magnetic or an electric field and were unable to imagine that both fields existed at the same time. This might be explained by grade level, since the electromagnetic filed is covered in the upper levels (grades 11 and 12) not in grade 10.

General thinking as hypothesis testing skill

Analysis of students’ responses to the "bending rods" task showed that students correctly identified the variables affecting the flexibility of the rods (e.g. length, diameter, and material). This explains that an explicit reference to the concept of identifying variables in the antimatter trap context helped students to pay conscious attention to the concept (**confirmation of HT-C4**).

However, they showed a lack of understanding of the concept of control variables, as only in one group (out of 3 groups) did students explicitly explain the strategy of finding the answer by comparing rods where two variables were held constant, or rather controlled, and one variable was allowed to change. This confirms students’ difficulty in understanding and applying the concept of control variables in hypothesis building and testing (**confirmation of HT-C5**).

Student reflection

The students perceived the task positively and specifically mentioned that by working on the task they learned how to generate hypothesis by identifying variables and finding the relationship between them and how to test them (**confirmation of HT-C3**). They distinguished the hypothetical nature of the task and stated that by comparing their hypothetical antimatter trap device to the Penning trap, they were able to revise their hypotheses and understand how the Penning trap works. However, they did not explicitly refer to the concept of control variables (**confirmation of HT-C5**).

They stated that thinking as hypothesis testing is applicable in everyday life when evaluating claims, in Physics and Biology lessons where hypotheses are to be made and tested by experiments, and also more generally when considering whether it will rain tomorrow.

They also mentioned that they learned to look at things in an unbiased way and that it is important to consider different perspectives. They also stated that one should use precise terminology when formulating hypotheses. These show students' development of **general CT skills.**

7.3.4.4 Developing students' thinking as hypothesis testing skill- Case 2

Data analysis showed that students applied content knowledge and reasoned correctly to hypothesize the construction of antimatter trap. Not only did they think of adding a vacuum to the construction of the Paul trap, but they also added magnets, which is based on their prior knowledge of how (anti)particles behave in a magnetic field, and that is exactly the case with the Penning trap. Since this idea is not very obvious, it shows the high level of students' content knowledge and might also explain their persistence in solving the task. However, data analysis showed that while students considered the magnetic field, they neglected the role of the electric filed in antimatter trapping. This might be explained by the grade of the students, as the concept of electromagnetic field was not familiar to the students.

Analysis of the students' reflection on the task showed that the students understood how to make a hypothesis by identifying the variables and formulating a relationship between them, but they did not understand the concept of control variables when testing the hypothesis. Although the concept was explicitly highlighted on the worksheet by asking a question about the variables to be controlled and the reason for it, the students did not seem to have grasped the concept correctly. This shows that the concept of control variables is a difficult concept for students and needs special attention in teaching this skill. This also confirms the findings of studies on students' difficulties in understanding the concept of control variables (*cf.* Boudreaux et al., 2008). However, apart from this difficulty, students' reflection showed their awareness of **general CT skills** such as considering different perspectives, treating things without bias, and the importance of using precise terminologies.

7.3.4.5 Results of data analysis- Case 3

The following findings are the results of analysis of data from 13 students who completed and returned the worksheet.

Student work

Domain-specific thinking as hypothesis testing skill

Identifying variables

Analysis of students' answers on the worksheet revealed a wide spectrum of thoughts about variables that according to students' opinion influence the construction of antimatter-trap (e.g. charge of antiparticle, its mass, magnitude of magnetic field, magnitude of electric field, and the length of trapping apparatus). This shows that the students were actively engaged in solving the task, as also mentioned by the teacher in the questionnaire (see teacher questionnaire section) (**confirmation of HT-C3**).

Controlling variables

Despite students' engagement in the task, lack of enough content knowledge hindered most students to reason correctly whether the mentioned variables can bring changes in the construction of antimatter trap in comparison with Paul trap. For example, only 3 students out of 9 students which addressed the charge as variable, classified charge into the neutral and the electrically charged antiparticles, and applied the concept of control variable correctly by considering trapping electrically charged antiparticles, not neutral antimatter.

Like case 2, students showed a huge difficulty understanding the concept of control variable. As alluded before, some of students (N=7) mentioned the "length of trapping apparatus" as a variable. All of them mentioned that this variable must be controlled and reasoned "more antiparticles can be captured at the same time or the antiparticles can be kept farther away from the electrodes". This shows that students view the control variable as having the authority to change the variable to satisfy their desire (e.g.

to capture more antiparticles), rather than understanding that the variable must be held constant in two experiments. This could be due to the use of the term “control”, which is used in everyday life in connection with authority. Observing this in cases 2 and 3 led to the conjecture that students’ prior understanding of the term “control” from everyday life influences their understanding and application of the concept of control variables in the experimental context. This is referred to here as **communication fallacy** in the context of control variable and explains students’ difficulty in applying the concept of control variables (**confirmation of HT-C5**).

Hypothesizing the construction of antimatter trap

Although 9 students in the identifying variable task referred to the magnitude of magnetic field as one of the variables, only 2 students added the magnet to their proposed sketch about antimatter trap apparatus. Furthermore, among all students only 2 students suggested adding vacuum to the construction of antimatter trap. Most of students neglected the problem of annihilation when the antimatter comes in contact with matter (here air). This might be explained by one of the common students’ misconception that “there is air between particles” (*cf.* Schecker et al., 2018) that consequently can be interpreted by students that “there is air between antiparticles”. Students in case 3 were not aware that air consists of (matter) particles and its contact with the antimatter contributed to the annihilation.

Testing hypothesis

Although the penning trap was presented to students in the task of testing hypothesis includes the vacuum, most of students did not clarify this point when asked to compare their proposal to the Penning trap.

General thinking as hypothesis testing skill

Analysis of students’ responses to the “Bending Rods” task showed that all students (N=13) correctly identified the variables (e.g. length, diameter, and material) (**confirmation of HT-C4)**, but they did not clearly state the procedure to find out the factors that affect the flexibility of the rods. Only 5 students correctly applied the concept of control variables to explain the procedure, and the others had difficulty understanding and applying the concept (**confirmation of HT-C5**).

Teacher questionnaire

The teacher perceived the antiparticle trapping task positive and expressed that the students actively participated in the discussion. The teacher found the video on trapping matter (Paul trap) very motivating for the students and mentioned that the discussion on antiparticle trapping was very interesting for students (**confirmation of HT-C3**).

7.3.4.6 Developing students' thinking as hypothesis testing skill- Case 3

Although, students engaged in the task actively, as the teacher mentioned and also observed from the wide spectrum of variables mentioned by them, lack of enough content knowledge hindered students to progress in the task. Furthermore, students showed difficulty in understanding the concept of control variable in the context of antimatter trap. The **communication fallacy** was observed among most of students which means students interpret the term control variable as having authority to change the variable based on their desire instead of getting the variables constant between 2 experiments. The communication fallacy originated in the students' understanding of the term "control" in everyday life.

Switching from the complex context of trapping antimatter to a less complicated context, the "bending rods" task, did not solve the problem of understanding and applying the concept of control variable. In the "bending rods" task, only 5 students applied the concept of control variable correctly. This indicates that students have difficulty understanding the concept of control variable regardless of context, which is consistent with other studies (*cf.* Boudreaux et al., 2008; Viennot & Decamp, 2020).

7.3.4.7 Discussion of developing thinking as hypothesis testing skill

Cross-case analysis

Data analysis showed that in all cases, students were actively engaged in CT process by discussing and reflecting on the task. In cases 1 and 2, students applied their content knowledge to hypothesize about the construction of the antimatter trap, but in case 3, students' lack of content knowledge prevented them from making progress on the task. For example, in case 3, most students did not propose a vacuum in the construction of antimatter trap to avoid the antimatter contact with the air, which might be explained by the common students' misconception that that "there is air between particles" (*cf.* Schecker et al., 2018). This shows the importance of addressing this misconception in the class and correcting it by planning appropriate activities to help students to progress in the trapping antimatter task (**Heuristic 6**).

Although reflection on the process of identifying variables and controlling them failed in case 1, the redesign of the worksheet provided students in cases 2 and 3 with the opportunity to engage with these concepts. Students performed well in identifying variables, but they still did not understand the concept of control variables. This shows that students are able to cope with the challenge of identifying variables when confronted with an explicit question to think about, but the concept of control variables is a challenging concept that requires special attention when designing instruction (Heuristic 5). Because students' difficulty in understanding and applying the concept of control variables is not a challenge unique to particle physics but has been reported in other studies (*cf*. Boudreaux et al., 2008; Viennot & Decamp, 2020), heuristic 5 can be extended to a general task in (many) physics classes. The new insights gained in this study about the **communication fallacy** in the context of control variables, that students interpret the concept of control variables as having the authority to change the variable to their preference rather than holding the variable constant between two experiments, could be helpful in planning discussions to correct students' misunderstanding of the concept of control variables.

In conclusion, what can be inferred about the effectiveness of the antimatter course is that teaching the skill of **thinking as hypothesis testing** in the context of trapping antimatter developed students' ability to identify variables and also helped them transfer this skill to the other context (Bending Rods task). Although the instruction encouraged students to apply the concept of control variables in the context of antimatter, it did not succeed in transferring this skill to the other context. This indicates that special treatment is needed to allow the transfer of the skill of control variables to other contexts.

Table 7.10: Conjectures about developing of Thinking as Hypothesis Testing skill observed in different cases.

Conjectures about developing of Thinking as Hypothesis Testing skill (HT-Cs)	**Case** 1	2	3
	Approved by data type analysis		
1. Unconscious applying of the concept of identifying variables and controlling them	*Video and audio data* *Student work*		
2. Conscious applying of defining terms precisely (operational definition strategy)	*Video and audio data* *Student work*		
3. Demonstrating disposition to engage in CT process	*Video and audio data*	*Student work* *Student reflection*	*Student work* *Teacher questionnaire*
4. Conscious attention to the concept of identifying variables when an explicit reference to this concept is present		*Student work* *Student reflection*	*Student work*
5. Demonstrating difficulty in understanding and applying the concept of control variables despite an explicit reference to the concept (**communication fallacy**)		*Student work* *Student reflection*	*Student work*

7.4 Conclusion on the effectiveness of antimatter course

Summarizing all the results discussed above, it can be concluded about the effectiveness of the antimatter course that if the teacher adheres to the structure of the antimatter course, the course can help students to develop the skills of likelihood and uncertainty analysis, argument analysis, verbal reasoning, and thinking as hypothesis testing. In addition, a comparison between the findings with the previous reports of students' difficulties in applying CT skills in physics discussed in section 2.4 clarifies the extent to which the antimatter course was effective (see table 7.11).

In addition to developing domain-specific CT skills, the antimatter course helped students develop general CT skills, such as defining precise terminologies and considering the credibility of the information source when evaluating the validity of information.

However, there are still some challenges that the course could not address. Some of them are the new findings of this study, such as making an argument when the answer to the question is still open, and some have already been reported, such as the application of the concept of control variables.

Table 7.11: Successful aspects of antimatter course in developing critical thinking with regard to reported students' challenges in literature.

Critical thinking skill	**Students' challenges in applying CT skills in physics** *Lack of the ability to:*	**Successful aspects of antimatter course regarding reported challenges**
Verbal Reasoning (VR)	• Apply non-ambiguous (clear) terminologies	• Identifying vague and ambiguous terminologies by evaluating their adherence to precision criterion • Seeking criteria to find precise terminologies
Argument Analysis (AA)	• Make coherent argument • Consider enough reasons to support conclusion	• Seeking the "relevance" criterion to find reasons that support conclusion
Thinking as Hypothesis Testing (HT)	• Identify variables • Explain the relationship between variables	• Conscious attention to the concept of identifying variables when an explicit reference to the concept is present • Generating hypothesis by explaining the relationship between variables
Likelihood and Uncertainty Analysis (LU)		• Realizing hidden possibilities in a probabilistic situation • Arguing for and against different possibilities by considering probability of their occurrence • Developing notion of reasoning in the LU context by using words such as "most likely", "likely", and "unlikely" accompanied by a correct reason

7.5 Development of critical thinking skills: a model proposal

The conjectures describe the processes students in the antimatter course go through to achieve the desired outcomes of applying CT skills. These desired outcomes were defined as follows:

- Develop a valid explanation using precise terminologies and complete information
- Draw a valid conclusion supported by acceptable and consistent reasons
- Generate a hypothesis and evaluate its validity
- Realize more hidden possibilities and decide on the most probable event

To build an empirically based model for CT development, the conjectures were reviewed to find categories that covered them. In the search for categories, it appeared that three categories, “thinking in discipline”, “critical thinking skill”, and “disposition” fit the results well. These three categories extracted from the data analysis correspond to developed critical thinking. This is also consistent with Russell’s definition of CT, which sees CT as involving attitude, knowledge of facts, and some thinking skills (Russell, cited in Halpern, 2009, p. 5).

In the following, a model for each CT skill is first offered separately, and then an attempt was made to combine all models to find a fundamental model for CT development. Furthermore, student challenges in applying CT skills are assigned a hazard symbol.

7.5.1 Model proposal on developing likelihood and uncertainty analysis skill

Figure 7.4 shows the model for developing students’ likelihood and uncertainty analysis skill. Students are constantly applying their content knowledge as they identify possibilities to decide on the most likely event. Recognizing obvious possibilities or solutions does not seem to be a challenge; the challenge begins with looking for hidden possibilities. The problem of overconfidence prevents students from thinking of hidden solutions. To stay on track and not give up, students need to overcome their overconfidence. They also need persistence in searching for hidden possibilities. Brainstorming produces creative solutions which can be considered hidden possibilities. Once they have identified the hidden possibilities, they should evaluate all the possibilities to decide on the most likely event. Some of the possibilities may be discarded because of the content knowledge, while others may be discarded because of the low probability of their occurrence. After completing the evaluation process, students decide on the most likely event.

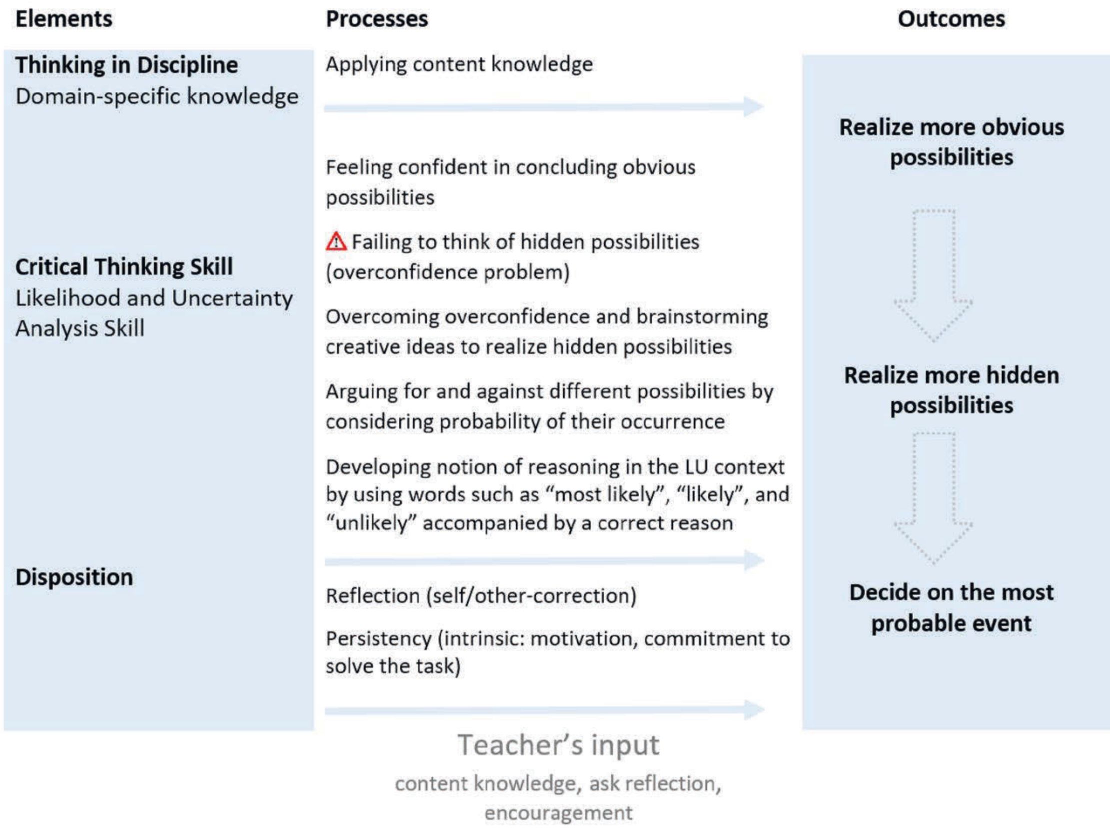


Figure 7.4: Model on developing students' likelihood and uncertainty analysis skill.

The teacher also plays an important role as students go through the process. By encouraging students to think of more hidden solutions, organizing the brainstorming situation, asking reflective questions, and appreciating students' efforts, the teacher helps students stay on task and complete it.

7.5.2 Model proposal on developing argument analysis skill

Figure 7.5 shows the model for developing students' argument analysis skill. The study of the process students go through to reach a valid conclusion shows that in looking for reasons on the internet, students judge the validity of the reasons based on the number of sources that present those reasons, without checking the credibility of the source of information. In other words, they consider the "majority" as a criterion for validity. However, explicit reference to the distinction between credible websites and "junk" websites can correct the "majority fallacy" and enhance students' argument analysis skill.

Once finding reasons which include complex information, students break the information down to its components and ask relevant questions to clarify the reasons and improve their comprehension of complex information. Analyzing complex information and asking thoughtful and relevant questions demonstrates a progression of learning in the specific context (Halpern, 2009, pp. 67-68). It also shows students' disposition toward understanding a complex information. This improves students' skill to evaluate whether the reasons support the conclusion and improves adherence to the criterion of relevance.

Furthermore, once a conclusion is reached, students evaluate its validity by considering different perspectives. For example, in the context of antimatter research, when students realize that in some topics there are only theories and no experimental work available, they question the validity of the conclusion. This sensitivity to the validity of conclusions shows they have developed argument analysis skill.

However, in making argument, a misconception was distinguished among some students. Students perceive an argument as something solid that its components, reasons and conclusion, must contain certain information without any open and probabilistic character. This shows the "overconfidence problem" discussed in likelihood and uncertainty analysis skill that students tend to be most confident in probabilistic situations because it gives them a sense of control over uncertain events (Halpern, 2009, pp. 164-165).

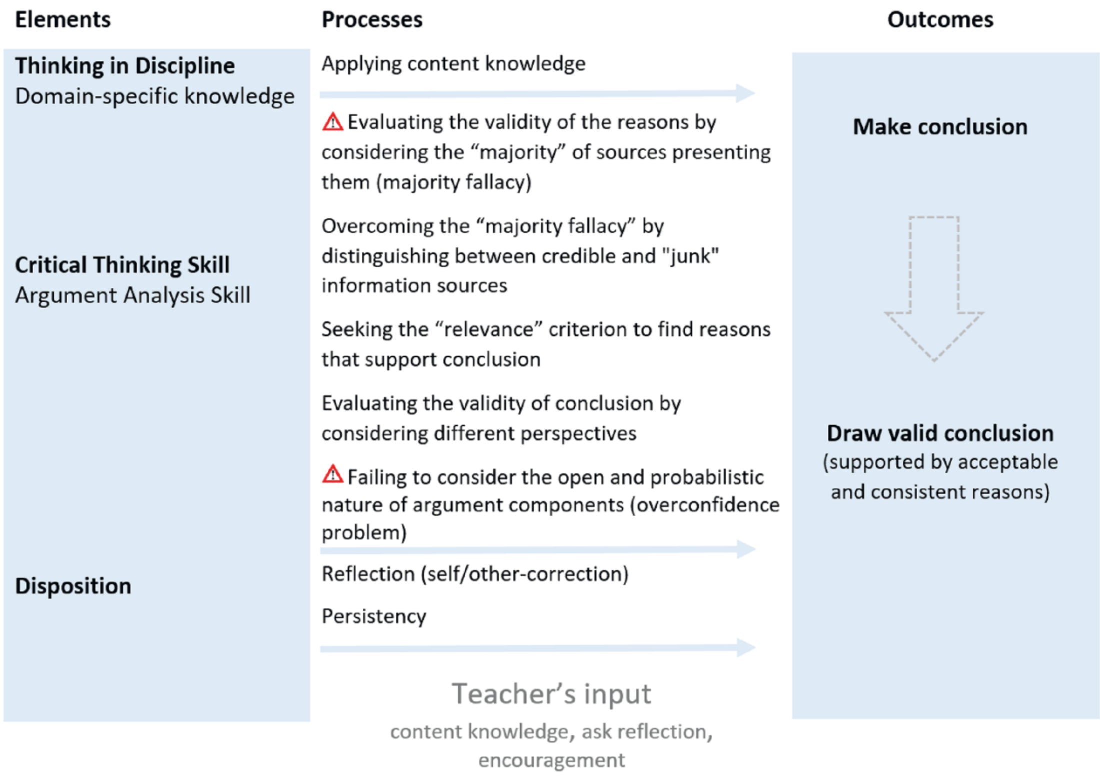


Figure 7.5: Model on developing students' argument analysis skill.

7.5.3 Model proposal on developing verbal reasoning skill

The development of verbal reasoning skill is the result of advanced content knowledge and a critical attitude toward incomplete, vague, and ambiguous explanations. By examining the terms used in an explanation, students transform their understanding of content knowledge from a deep structure in mind to a surface structure in words and phrases (*cf.* Halpern, 2009, pp. 47-48). It shows an intertwining of content knowledge and CT skills. By consciously paying attention to the correctness and acceptability of terms and completeness of information in an explanation, they show disposition toward CT. Figure7.6 shows a model on developing verbal reasoning skill.

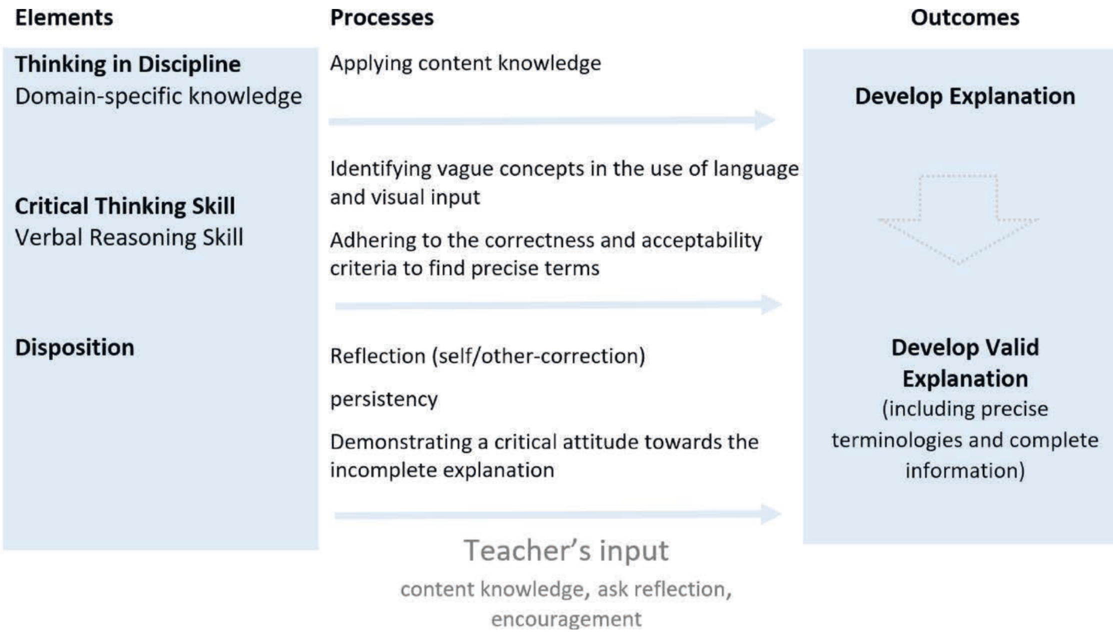


Figure 7.6: Model on developing students' verbal reasoning skill.

7.5.4 Model proposal on developing thinking as hypothesis testing skill

Figure 7.7 shows the model for developing students' thinking as hypothesis testing skill. To generate a hypothesis, students apply content knowledge to identify variables and describe the hypothetical relationship between them. Furthermore, they apply content knowledge to define terms precisely, which in the context of Thinking as hypothesis testing is referred to as "applying operational definition strategy". Developing this ability requires a sensitivity to precise definition of terms that can be triggered by the course and demonstrates students' disposition to engage in the CT process.

The concept of identifying variables can be understood by students when there is an explicit reference. However, this is not the case with the concept of control variables. Students might unconsciously apply the concept of control variables in making hypothesis (e.g. discussing trapping electrically charged antiparticles instead of neutral ones), understanding and applying the concept in the conscious level seems a challenge. Interpreting the concept of control variables as using authority to change variables according to preference instead of keeping them constant, which could result from the understanding of the term “control” in daily life, referred to as the **communication fallacy** in this study, could explain this difficulty. This needs special attention to develop students Thinking as hypothesis testing skill.

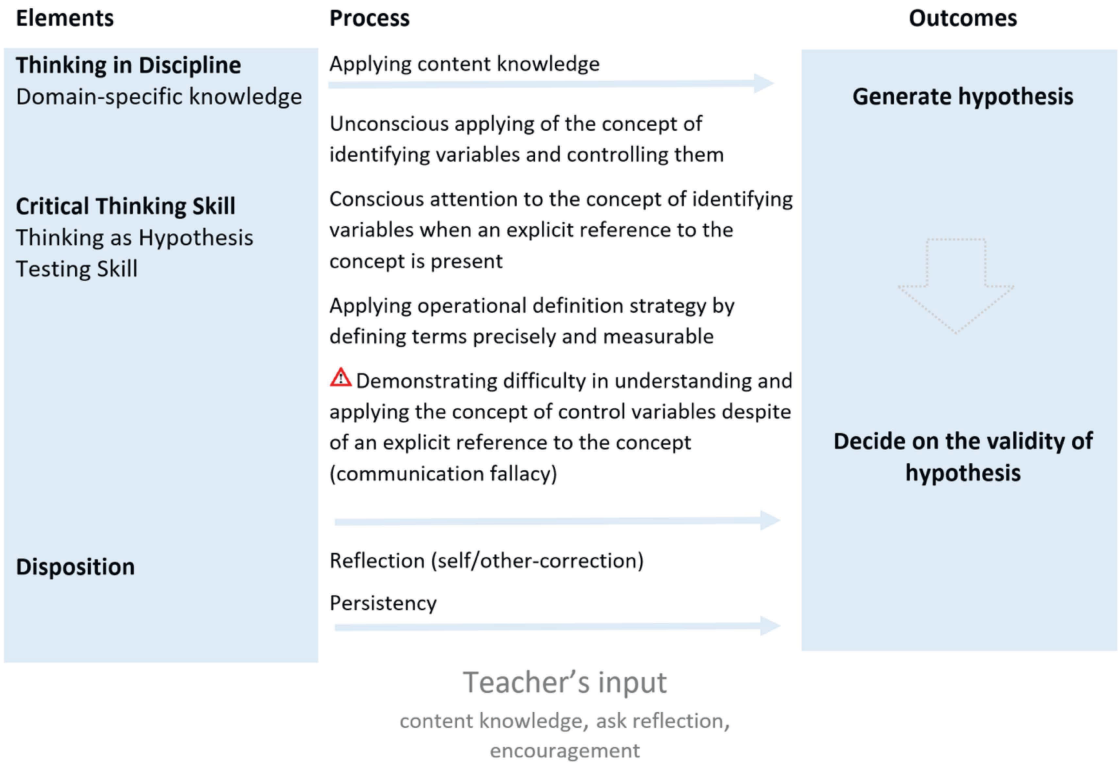


Figure 7.7: Model on developing students’ thinking as hypothesis testing.

7.5.5 An underlying model on developing critical thinking

Figure 7.8 shows an overview of the combination of all the above models for all four CT skills. To establish an underlying pattern from this joint model across all four skills, critical thinking, as discussed, is the process of consciously evaluating the thinking process and the results of the thinking process (Halpern, 2009). To engage in this evaluation process, students required tools; they applied criteria such as precision, relevance, and validity, as well as procedures such as variable identification and operational definition. Regarding this, the detailed processes of applying CT skills can be divided into two main categories: criteria-based evaluation and application of appropriate procedures. Table 7.12 shows an overview of the application of CT skills classified under these two categories.

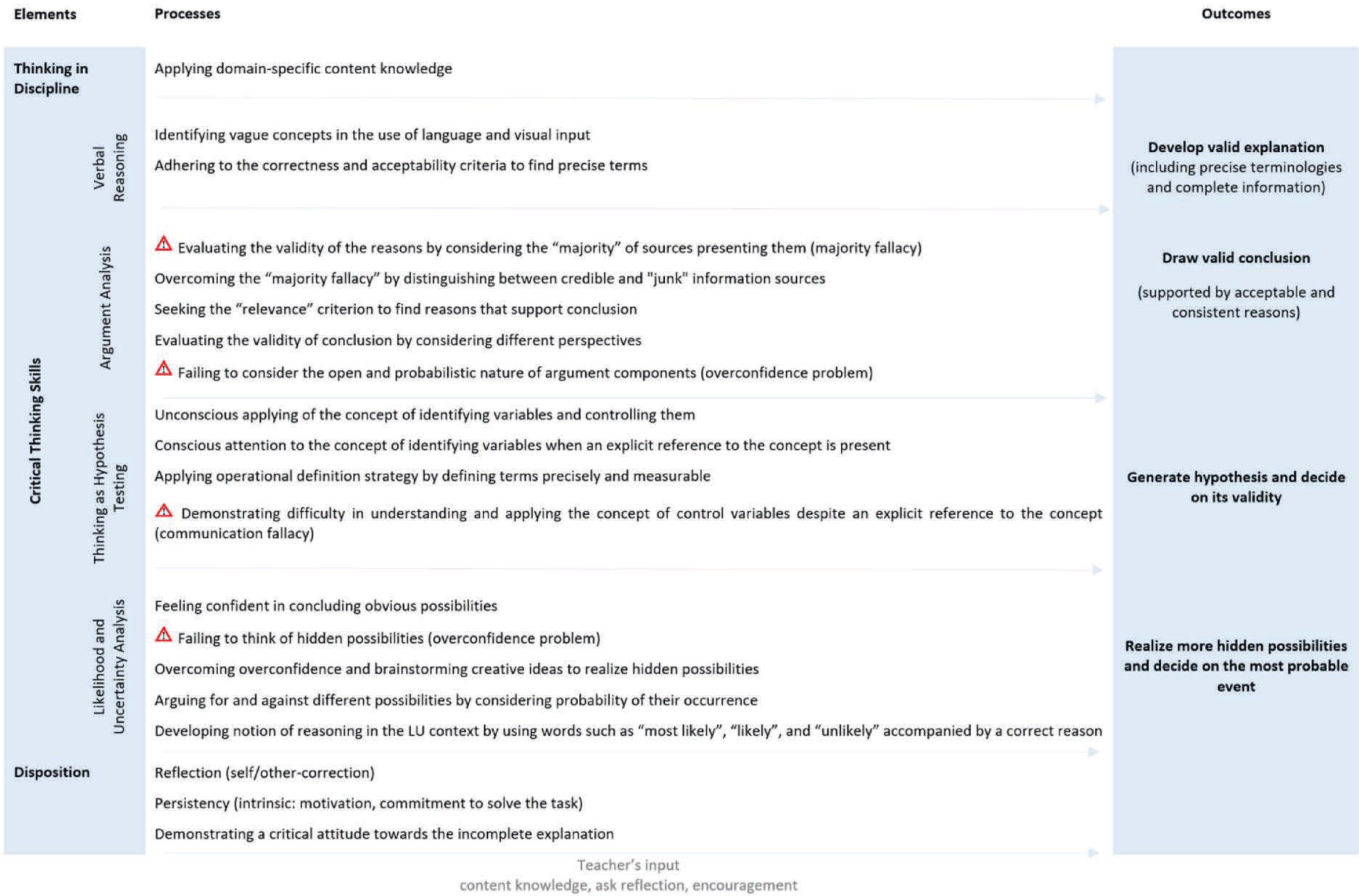


Figure 7.8: Model on developing students’ critical thinking skills in the antimatter context

Table 7.12: Overview of the application of critical thinking skills classified under two categories of criterion-based evaluation and applying appropriate procedures.

Main Category	Processes included in the category
Criterion-based evaluation	**Seeking precision criterion** - Identifying vague and ambiguous terminologies - Finding precise terminologies **Seeking validity criterion** - Considering credibility of information source - Considering different perspectives **Seeking relevance criterion** - Finding reasons that support conclusion **Seeking sufficiency criterion** - Identifying need for more information
Application of appropriate procedures	**Argument analysis skill** - Considering the open and probabilistic nature of argument components **Thinking as hypothesis testing** - Applying the concept of identifying variables - Applying the concept of control variables - Applying operational definition strategy by defining terms precisely and measurable - Considering all the (obvious) results when testing hypothesis **Likelihood and uncertainty analysis** - Considering all obvious and hidden possibilities in a probabilistic situation by overcoming overconfidence - Arguing for and against different possibilities by considering probability of their occurrence - Developing notion of reasoning in the LU context by using words such as “most likely”, “likely”, and “unlikely” accompanied by a correct reason

Thus, an underlying pattern (see figure 7.9) that applies to both domain-specific discipline and daily life might be as follows:

Inputs such as content knowledge, skills, and disposition are combined to engage students in the evaluation process in which they apply criteria and appropriate procedures based on the task, apply content knowledge, and show disposition toward CT to achieve the desired outcomes.

Teacher guides students in this process by providing content knowledge, asking reflective questions, and encouraging students to persist in the effortful process of critical thinking and appreciating their efforts.

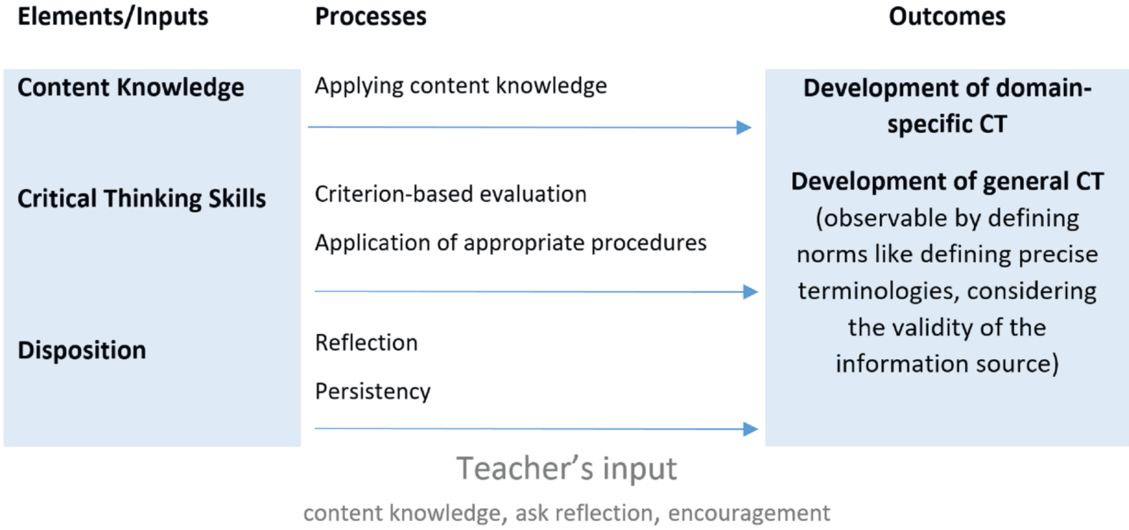


Figure 7.9: The fundamental model on developing students' critical thinking.

It is also worth noting that it is not necessary to master content knowledge to be able to think critically in context. A threshold understanding is required to engage a thinker in the process of critical thinking, which varies from student to student (Viennot & Decamp, 2020, p. 70). The critical thinker is aware that there is an opportunity to develop an understanding of content knowledge as he/she engages in the process of critical thinking, and that this is an interplay between understanding the need for more information (querying content knowledge) and thinking critically about the knowledge to determine if it fits the evaluation criteria.

8 Relating design skeleton facets to valued outcomes

The results of the data analysis in chapter 7 investigated the processes students in the antimatter course went through to solve CT tasks and the challenges they faced (see figure 7.8). It was found that students applied domain-specific knowledge, applied CT skills, and demonstrated a disposition by discussing and reflecting on the tasks. In addition, finding an underlying model for student CT skill development (see figure 7.9) led to a categorization of CT skill application into two main categories: criterion-based evaluation, and application of appropriate procedures.

This chapter examines how the design skeleton facets of the antimatter course triggered these processes that contribute to the development of general and domain-specific CT (answer to research question 1.2). Figure 8.1 shows a schematic of the aspects under investigation. A conjecture map was used as a framework to illustrate the results of this investigation (see Sandoval, 2014; Wozniak, 2015).

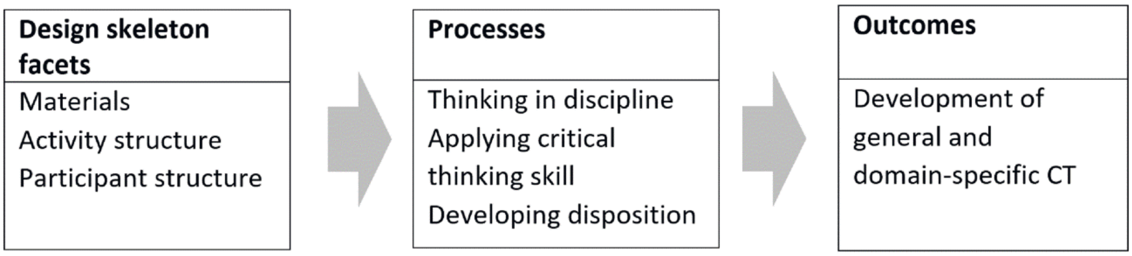


Figure 8.1: Schema of aspects of investigation in the qualitative data analysis of the lessons in antimatter course.

8.1 Design skeleton facets of antimatter course

In the antimatter course the design skeleton facets include the materials, the students' activities, and the participant structure including their perception of CT, culture, and their roles. Development of the design skeleton facets was inspired by the design principles discussed in chapter 6.

8.1.1 Materials

The *worksheets* (see table 6.4) were developed to allow students to practice CT skills in the antimatter context and to reflect on the structural aspects of the tasks. Additional to the domain-specific CT worksheets, three worksheets were created to explicitly focus on the application of general CT skills.

In addition, *a scene from the movie "Illuminati"* was chosen to define an authentic problem around which the course sequences were defined. Students had to apply the content knowledge and the CT skills acquired in the course to solve the problem.

8.1.2 Activity structure

Students engaged in two main activities in the antimatter course: group work and class discussion. These both activities were so structured to encourage discussion and reflection.

Group work: students worked in the small group on the domain-specific CT tasks in which they were expected to apply content knowledge, dispositions, and CT skills. However, some instructional sequences included individual work to prepare students for group work, such as watching a video about the Big Bang and writing the questions on worksheet 4, but these few individual tasks can be considered an introduction to group work.

Class Discussion: students discussed their findings with the class and reflected on the tasks. An explicit discussion was also planned to get students to reflect on the content knowledge and CT skills learned in each session.

8.1.3 Participant structure

Participant structure refers to how students participated in the activities (Sandoval, 2014; Wozniak, 2015). The way they participate in the course is influenced by their perception of CT, their culture, and roles (*cf.* Ryu & Sandoval, 2012; Sandoval, 2014).

Perception of CT tasks: students' perceptions of CT tasks influence their performance in the antimatter course. Based on their perceptions, they define epistemic norms related to the criteria of critical thinking (*cf.* Cobb, et al., 2001; Ryu & Sandoval, 2012). For example, the norm "the terminologies must be defined precisely" is an epistemic norm developed by students in the antimatter course and was observed in their discussions.

Participant culture: classroom culture has an impact on student performance in CT tasks. For example, Cobb and colleagues discuss how participation norms and culture influence students' reasoning in mathematics (Cobb et al., 2001). In addition, Ryu and Sandoval (2012) argue that classroom culture affects the development of students' argument skills in science classroom.

In the antimatter course, however, it was not necessary to clarify and negotiate participation culture with students. For example, there was no need to negotiate with students that they should respect each other's ideas and require them to do so because the students presented this culture from the beginning.

Participation role: participant roles are aligned with the structure of the activity (Ryu & Sandoval, 2012). In the antimatter course, students should collaborate with peers as they complete group tasks and reflect on their learning path in terms of content knowledge and CT skills as they discuss these planned sequences in each lesson.

Moreover, teacher-student interaction affects students' performance in the course. Encouraging students to discuss and reflect on the tasks and also acknowledging their efforts in completing CT tasks contribute to defining values that characterize class discussion such as being actively involved at all times (Cobb et al., 2001).

Bringing the above descriptions into the framework of the conjecture map, the following scheme for the design of the antimatter course emerges (see figure 8.2). In the next section, the results of the search for evidence for the mapping of the process to the design facets are discussed.

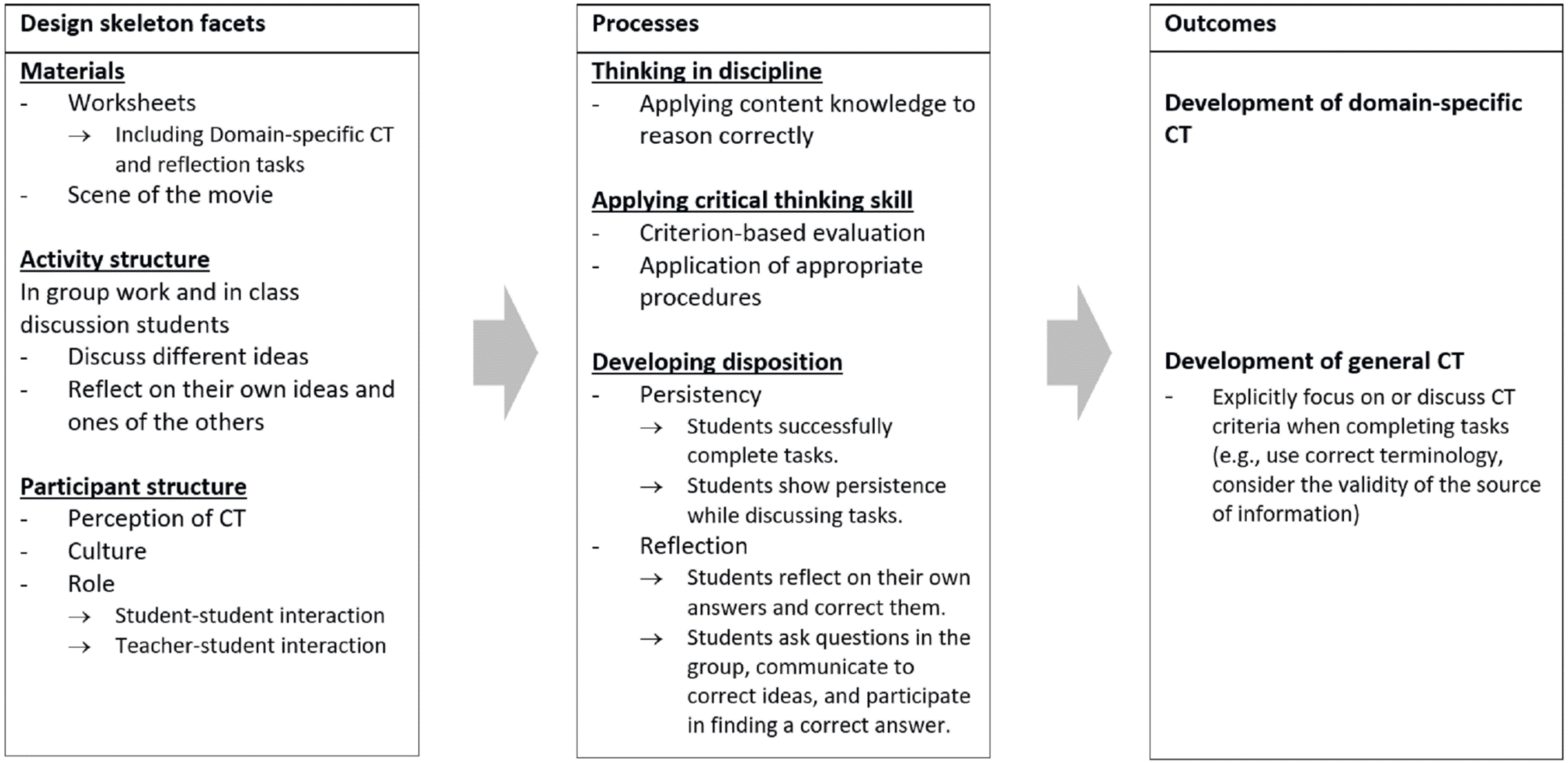


Figure 8.2: Schema of design of antimatter course to promote domain-specific and general CT using conjecture map framework.

8.2 Relation of design skeleton facets to valued outcomes

How did the design features influence the processes?

This question explains the extent to which the high quality processes were inspired by the design facets. High quality processes were defined as the high level performance in applying content knowledge and CT skills, as well as developing disposition toward CT.

To attribute the high quality of the processes to the design facets of the antimatter course, the interviews and questionnaires of the teachers and students, as well as video and audio data were analyzed. These findings have already been discussed in detail in various contexts in chapter 7. Here, a general overview is provided on the basis of students' and teachers' statements.

8.2.1 Materials

The domain-specific CT tasks in worksheets were perceived positively by students in developing their CT skills. For example, one student appreciated the tasks on worksheet 3 and also considered the structure of the worksheet to be a novel approach that allows students to discover the solution. The student stated the following:

> *"I think [in the antimatter course] we went into all areas of critical thinking. The best part for me was when we had to figure out for ourselves what kind of particles were on the pictures [cloud chamber photograph in worksheet 3], because this method is not used that often. I found it very interesting to try to come up with our own solutions. I also think that this course made you realize what methods you use subconsciously to think critically. And that you can use them more consciously and therefore better."*

Or one other student stated:

> *"Well, I think: Basically, I already knew that [CT]. So I mean, that/ I just think it's something natural that everyone does in principle, which you just forget. (...) And that's why it was quite good to think about it again and to deal with it [in the antimatter course]. It somehow drew our attention to it."*

In addition to the positive impression of dealing with the domain-specific tasks in the antimatter course, students made specific reference to the use of some scenes of the movie "Illuminati" in teaching:

> *"I liked the reference to the movie Illuminati. I found it exciting to learn what antimatter is based on [discussing] the information and descriptions in the movie."*

Moreover, although some students in case 3 expressed dissatisfaction with too many tasks in the course (see section 7.3.1), analyzing the teacher questionnaire showed that students were very motivated and actively engaged in working on the LU, AA, and HT worksheets. Only the VR worksheet that asked students to analyze the scenario of a scene from "Illuminati" was not well received by case 3 students, who criticized the task on the grounds that one cannot expect logical information in a pseudoscientific dialogue in an entertainment movie. However, this is a CT reflection and worthy of recognition, but it was not what was expected.

Teachers also appreciated the well-structured worksheets and the discussion and reflection opportunities offered in the worksheets. They explicitly referenced to the materials of the antimatter course as a means for improving students' critical thinking skills. For example, the tasks in worksheet 3 (Anderson's Photograph), were perceived positively by the teachers and they stated that:

> *"Task 1 on worksheet 3, in particular, is appropriate to help students understand the procedure [of applying likelihood and uncertainty*

> *analysis skill]. For this purpose, tasks 2.1 and 2.2 on this worksheet can also be referred to."*

> *"On the basis of the trajectory analysis of the particle [in worksheet 3], which later proved to be an antiparticle, it was possible to develop students' likelihood and uncertainty analysis skill."*

Furthermore, the materials were characterized by the teachers as interesting and motivating for students. For example, the teacher in case 3 stated:

> *"The discussion of the traps [Paul trap and Penning trap in worksheet 6] was interesting and comprehensible for many students. The video [of Paul trap] was especially motivating. The research on the GBAR experiment was also quite good and students found good sources."*

Moreover, re-watching the scene of "Illuminati" and the analysis of its scenario were perceived positive by the teachers.

> *"Re-watching the movie is a good idea to show students their learning progress."*

To summarize, the materials were perceived positively by students and teachers in developing CT skills.

8.2.2 Activity structure

Analysis of the video and audio data revealed that students actively engaged in discussion and reflection on their own ideas and ones of others while working on the tasks. They also expressed satisfaction at being involved in discussions, saying, for example,

> *"[in the antimatter course] a lot with discussion and little with strict copying. That way you can clear up possible misunderstandings better than just accepting it as it is."*

Or stating such as:

> *"I liked the fact that we talked a lot and exchanged ideas. I think that improved the learning environment."*

Teachers also valued the students' participation in the discussion and reflection as evidence of the development of their CT skills. For example, one teacher stated:

> *"The detailed, small-step guide in worksheet 5 was appropriate for engaging students in discussions about their approaches [of applying argument analysis skill] in research tasks."*

8.2.3 Participant structure

Perception of CT: Students' positive perception of CT (see students' quotes in the material section) justified their efforts and persistence in completing the tasks, which were appreciated by the teachers, as was also observed in the video and audio data. This showed a willingness and positive attitude towards CT. Students also defined some epistemic norms such as defining terms precisely and addressed them constantly in the discussions and reflections.

Culture of participation: Analyzing the video and audio data showed that the culture governed in the antimatter class was structured by students themselves. Students respected each other ideas while discussing, and looked for consensus while working on the tasks.

Role: Due to the structure of activities students collaborated with each other to complete the tasks. They also reflected on their own ideas and ones of the others.

In addition, data analysis revealed the teacher's role in encouraging students to continue working on the tasks and to overcome challenges such as overconfidence. Teachers also asked reflective questions and appreciated students' efforts in completing CT tasks.

Bringing all facets of the design skeleton together to structure the antimatter course, the following statements express students' positive perceptions of the antimatter course and demonstrate its success:

> *"I think through the [antimatter] course we learned to be more aware of our critical thinking. It [the antimatter course] wasn't just focused on performance and grades. It wasn't just about learning as much material as possible, but about thinking and that can be applied in many more areas in life, so the lessons were better and more fun [than the regular lesson], I think."*

> *"I enjoyed learning about antimatter because you get to deal with topics you wouldn't normally come in contact with. In addition, my critical thinking skills have improved as a result of dealing with it."*

> *"Also it [The antimatter course] was very interesting and I have/ We have to write a scientific paper the next school year and that [the course] has also motivated me that I probably write my scientific paper in the direction of so particle physics."*

In summary, the above evidence explains that the antimatter course engaged students actively in dealing with CT tasks and discussion and reflection. Bringing these results in the conjecture map framework, resulted to the following Illustration (see figure 8.3). The dark arrows reflect the developmental path of CT skills in relation to design facets.

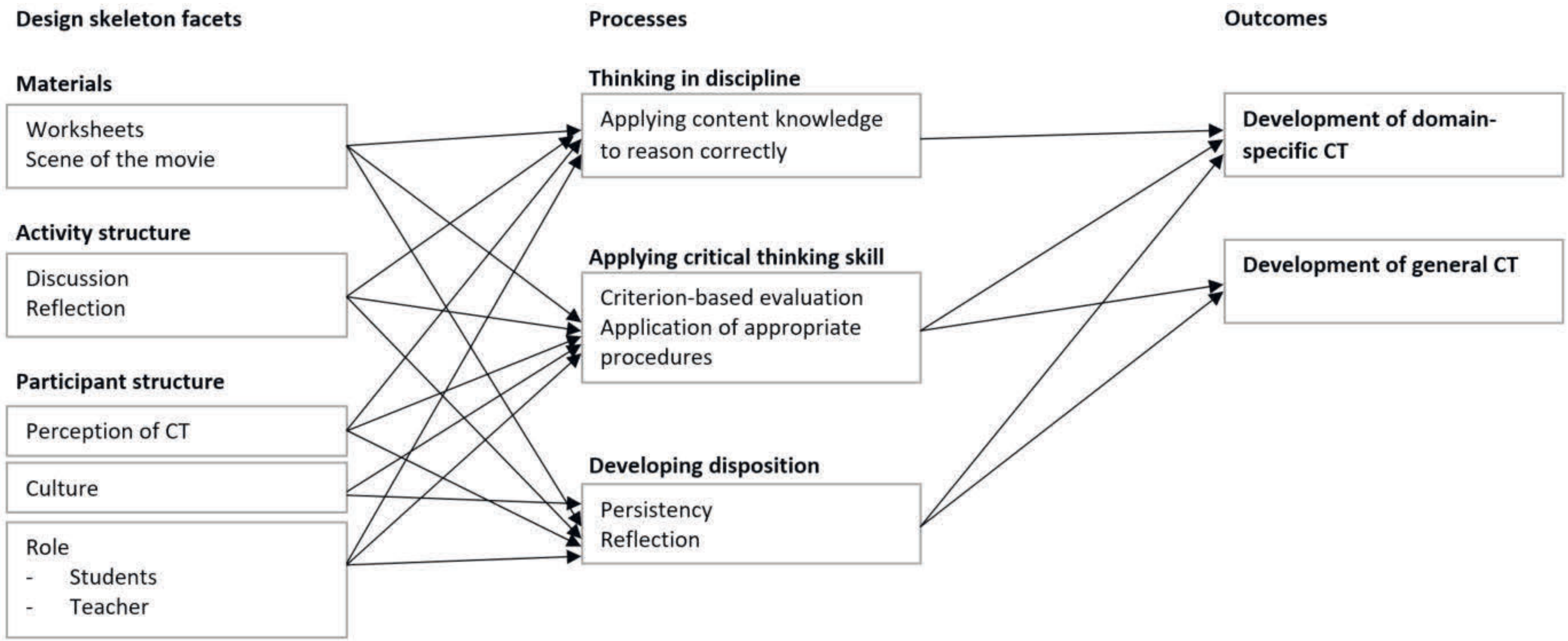


Figure 8.3: Illustration of the results of data analysis to relate the design skeleton facets to the processes and outcomes, using conjecture map framework.

To develop an underlying pattern of this illustration in figure 8.3, it can be discussed that the developed materials define the activity structures as well as the role of students (*cf.* Ryu & Sandoval, 2012). Because of this connection between the materials, the activity structure, and the role of the participants, they can be considered as a single entity. Furthermore, students' perceptions of the CT tasks and their culture of participation influence how they interact with the materials and their engagement in the desired processes. Figure 8.4 shows an underlying pattern of the findings illustrated in figure 8.3.

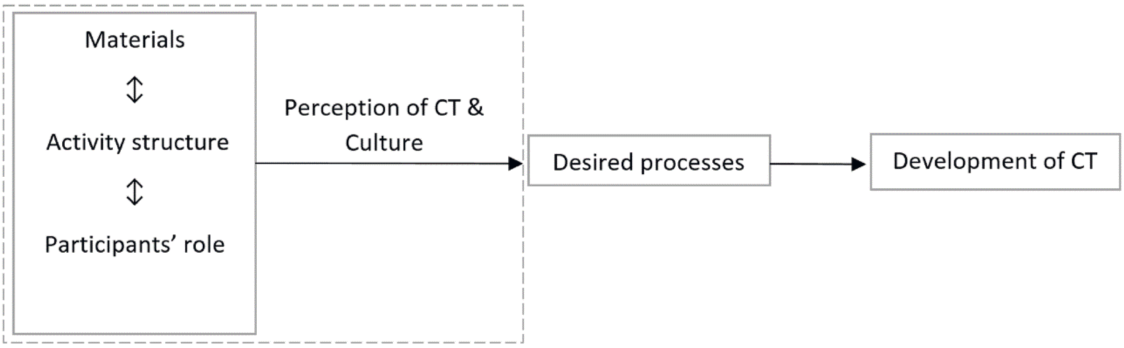


Figure 8.4: Underlying pattern of findings illustrated in figure 8.3.

Therefore, in addition to developing effective materials, two prerequisites are needed to develop students' CT skills: first, students' positive perceptions of CT, and second, creating a safe classroom culture where students are encouraged to share and discuss ideas without fear of being judged or ridiculed. Creating this safe classroom culture is possible through student culture, which is influenced by school culture, although teachers can also influence it by encouraging students to share their ideas, acknowledging students' efforts, and showing openness and respect for different ideas.

In the antimatter course, general CT skills were explicitly taught in the first session and values such as open-mindedness, respect for other ideas, and consideration of others' perspectives were discussed. This might help improve classroom culture. However, it is beyond the scope of this thesis to discuss whether this explicit teaching of CT influences student perceptions and changes classroom culture.

8.3 Conclusion and discussion

To investigate how the design skeleton facets of the antimatter course engendered the processes which are necessary to develop CT, video and audio data, and students' and teachers' interview and questionnaire were analysed. It was observed that the developed materials based on the design principles of the antimatter course were perceived positively by students and teachers and also encouraged the discussion and reflection in the class. This positive perception of both the CT tasks and the antimatter topic motivated students to work through the tasks. For example, the tasks on worksheet 3 required persistence to find hidden interpretations of the sign and charge of the particles in Anderson's cloud chamber photograph. The explicit reference to this worksheet by teachers and students as a means of developing students' likelihood and uncertainty analysis skill proves students' persistency. Altogether, the findings prove the effectiveness of the design of the antimatter course in developing students' CT.

Despite the successful aspects of the design of the antimatter course to engage students in processes that develop CT, some desired processes were not successfully addressed in the course. As alluded to before, although the overconfidence problem was addressed in the particular teaching sequence on the likelihood and uncertainty analysis skill, it was observed again in the context of argument analysis, where students did not consider the open and probabilistic nature of argument components. Furthermore, students showed difficulty in applying the concept of control variables in the other contexts, even though it was explicitly mentioned in the antimatter course (see chapter 7). This highlights the need for further treatment and explicit reference to these concepts in different contexts to raise students' awareness of the overconfidence problem and improve their understanding of the concept of control variables.

9 Teacher perception of the antimatter course

This chapter discusses the teachers' perceptions of the relevance, practicality, and effectiveness of the antimatter course (answer to research question 2). As described in chapter 6, a teacher questionnaire was designed and completed after each lesson in cases 2 and 3 to collect teachers' opinions about the relevance, practicality, and effectiveness of the lessons in developing students' CT skills. In case 1, an interview was conducted with the teacher after each lesson to gain insight into her perception on these aspects. The results are discussed in the following sections.

9.1 Participants

The teachers in case 1 and case 2 were the same person who conducted the course before COVID19 (May & June 2019) and during COVID 19 (October 2020- February 2021) at the same school. She had almost 35 years of teaching experience and had already participated in particle physics workshops at CERN, DESY, and Technische Universität Dresden (TU Dresden).

The teacher in case 3 implemented the antimatter course during COVID 19 (September-December 2020) in two classes at the same school, grades 10 and 12. She had 12 years of teaching experience and had already participated in a particle physics workshop at CERN.

Both schools were located in an urban district in different states of Germany. In both schools, the 10th grade students were highly gifted students. However, in cases 1 and 2, the teacher taught a small group of nearly 10 students in a special course, while the teacher in case 3 taught a larger number of students, nearly 19, in regular physics class (see table 6.5). Furthermore, the teacher in case 3 taught the antimatter course in grade 12, with almost 6 students. Therefore, the teacher perception of the relevance, practicality, and effectiveness of the lessons in case 3 is the result of the teacher experience of implementing the course in two classes.

The teachers voluntarily participated in the study by receiving information about the study through presenting it at the IKTP Teachers' Roundtable (Institute for Nuclear and Particle Physics at TU Dresden), and also through an advertisement in a newsletter distributed among teachers in Germany by Netzwerk Teilchenwelt (a community of particle physicists communicating particle physics research with young people and teachers, with project leadership at TU Dresden).

The goal of the antimatter course was communicated to teachers in advance. Furthermore, teachers were provided with a teacher package, including the teaching-learning sequences, materials, step-by-step instructions, possible student questions during lessons, additional content knowledge, and also a suggested time for each teaching-learning activity.

9.2 Perceived relevance

Table 9.1 summarizes teachers' perception of relevance of the developed teaching-learning activities in the antimatter course for developing students' CT skills. Data analysis showed that in **lessons 1 and 2** all teachers perceived the explicit teaching of general CT skills and the use of scenes of "Illuminati" to fulfill the problem-centered principle of the antimatter course as relevant. They mentioned that teaching explicitly general CT skills make students aware of the skills and improve the chance of applying these skills consciously in everyday life. They also appreciated the use of examples from everyday life to discuss with students after presenting each skill.

They also expressed their desires to apply these skills in their physics classes. As one teacher said:

> *"Teaching explicitly CT skills motivated me to think about that and getting ideas to teach this in my physics class, for example, with making and testing hypothesis for the experiments."*

In addition, one teacher stated that teaching explicit CT skills has given her a new perspective on CT skills, that they are obvious skills that everyone should know. She mentioned:

> *"And that [explicitly teaching of CT skills] was exciting for me, so to speak, because I had not yet seen it from that point of view. And there are so many self-evident things that were actually mentioned, where it might be good to make them clear to someone or to make them aware of them."*

As indicated, this unit was taught by the researcher, however, in the interview for case 1, a desire was expressed for the teaching of general CT skills to by the teacher herself.

In **lessons 3, 4, and 5**, the teachers considered teaching of likelihood and uncertainty analysis skills in the context of positron discovery relevant for developing students' CT skills. The teachers explicitly referred to the designed tasks in Andersons' photograph worksheet as a good means for this purpose. In addition, the planned opportunities for discussion and reflection in these lessons were perceived relevant to student CT development by the teachers in cases 1 and 2, but it was not clear from the teacher responses in case 3 whether the teacher appreciated this point or not.

In **lesson 6**, in cases 1 and 2, the teachers appreciated the relevance of teaching thinking as hypothesis testing in the context of pair production for the development of students' CT. However, in case 3, the teacher criticized that in this lesson too much was presented by the teacher and the student activities were too short, so she did not consider this part relevant to the development of students' CT skills. However, given the steps described in the teacher package for this lesson, this criticism could be rejected and it was clarified that the teacher in lesson 6 had not implemented the lesson as intended in the teacher package (evidence for process fidelity discussed in section 7.2.1). Regarding the discussion and reflection planned in this lesson, only in case 1 the teacher explicitly referred to discussion and reflection on the tasks as a means of developing CT skills.

Table 9.1: Perceived relevance.

Relevance indicators of the antimatter course	Perceived relevance		
	Case 1	**Case 2**	**Case 3**
Lessons 1 & 2			
Teaching explicitly general CT skills	+	+	+
Using scene of movie to fulfill the problem-centered principle	+	+	+
Lessons 3 & 4 & 5			
Teaching Likelihood and uncertainty analysis skill in the context of positron discovery	+	+	+
Discussion and reflection	+	+	0
Lesson 6			
Teaching Thinking as hypothesis testing skill in the pair production context	+	+	-
Discussion and reflection	+	0	-
Lesson 7			
Teaching Argument analysis skill in the context of a systematic search for answers to the questions raised by watching a video clip about Big Bang	+	+	+
Discussion and reflection	+	+	+
Lessons 8 & 9			
Teaching Thinking as hypothesis testing skill in the context of trapping antihydrogen	+	+	+
Applying Thinking as hypothesis testing skill in the context of Bending Rods	NA[a]	0	0
Discussion and reflection	+	+	+
Lesson 10			
Teaching Thinking as hypothesis testing skill in the context GBAR Experiment	+	+	+
Discussion and reflection	+	+	+
Lessons 11 & 12			
Teaching Verbal reasoning skill in the context of analyzing the scenario for the scene of “Illuminati” movie	+	+	0
Applying CT skills in search about Positron Emission Tomography (PET)	-	-	-
Discussion and reflection	+	+	+

[a] Not Applicable

In **lesson 7**, all teachers appreciated the relevance of teaching of argument analysis skill in the context of Big Bang and also discussion and reflection for developing students' CT skills. For example, the teacher in case 3 stated:

> *"Critical questioning of statements and acquisition of professional background knowledge serve the acquisition of further competencies regarding critical thinking."*

In **lessons 8 and 9**, the topic of trapping antihydrogen and the associated teaching-learning sequences were considered relevant to the development of students' CT skills. The teachers appreciated the clear procedure for forming a hypothesis and testing it in these lessons as an appropriate way to discuss the steps with students. However, teachers' responses in case 2 and 3 did not indicate whether or not they considered the Bending Rods example relevant. This was not applicable in case 1 because, as mentioned earlier, the example was included in the lessons for cases 2 and 3.

In addition, the GBAR Experiment and discussion and reflection in **lesson 10** were perceived relevant for developing students' CT skills in all cases. The teacher in case 3 explicitly refereed to the possibility of search on GBAR experiment for students as a good means for improving students' CT skills.

In **lessons 11 and 12**, all teachers eliminated the positron emission tomography (PET) search task due to lack of time, neglecting the task's potential for developing students' CT, as it gives students the opportunity to apply CT skills in the context of medical application of antimatter. In addition, in cases 1 and 2, the teachers considered the task "analysis of the scenario for the scene of "Illuminati"" and the discussion and reflection as relevant, while in case 3, no information was found about the consideration of its relevance. Instead, the teacher in this case referred to the teaching sequence planned in these lessons to review the CT skills applied in the context of antimatter as a relevant means of developing students' CT skills. She also appreciated the relevance of discussion and reflection provided by the implementation of this teaching sequence.

9.3 Perceived practicality

Teachers' perceptions of the ease of implementing the teaching-learning activities in the antimatter course by available teacher package and materials are summarized in table 9.2. The indicators offered in this table are the CT-Structure-training activities developed to engage students in applying CT skills in the antimatter context. However, in addition to these aspects, the difficulties reported by teachers in implementing the lessons were also presented.

As mentioned earlier, lessons 1 and 2 were in all cases taught by the researcher in attendance or via videoconference. Therefore, this stage was not considered relevant to perceived practicality, because here the focus is on the teacher's implementation of the lessons.

In **lessons 3, 4, and 5**, all teachers expressed that the teaching concept was logically thought out and the teaching likelihood and uncertainty analysis skill was well structured in the context of positron discovery. They all perceived the lessons as practical, and the teachers in cases 1 and 2 reported no difficulties in implementation. The only challenge reported by the teacher in case 3 in grade 10 was the conversion of units, as in particle physics the unit of mass had to be converted from kg to ev/c^2. To improve student's understanding it was planned that students do this themselves rather than present this information as given knowledge. As mentioned earlier, the students in case 3 were younger (see table 7.1), which might explain their difficulty in converting units, as reported by the teacher.

Lesson 6, teaching pair production and annihilation, were perceived practical by all teachers. The teachers did not report any difficulty in implementing the lesson. As part of annihilation lesson, a comparison was made between the energy produced by the annihilation of 1 gram of antimatter and 1 gram of matter and the energy of the explosion at Hiroshima, which was specifically mentioned by the teacher in case 3 as an interesting teaching-learning activity for students. In this case, however, the teacher reported that the students had difficulty converting the units of energy needed to make this comparison.

In **lesson 7**, all teachers considered the teaching of argument analysis skill using step-by-step guidance offered by worksheet 5 practical. They also appreciated the practicality of the worksheet 5 in other classes as the teacher in case 1 said:

> *"Worksheet 5 with the step-by-step guide to research is good and useful in regular classes, good for discussion in other classes."*

Lessons 8 and 9, were also perceived practical by all teachers. For example, teacher in case 3 mentioned in the questionnaire:

> *"These lessons with the teacher package were very good to teach. The students learned a lot here."*

She also described the lessons on trapping antihydrogen with the materials provided as interesting, motivating, and comprehensible to the students. She wrote:

> *"There was much student activity possible here. The discussion of the traps was interesting and comprehensible for many students. The video with the lycopodium seed [Paul Trap] was especially motivating."*

Furthermore, since these lessons include the "Bending Rods" task, even though the practicality of this task was not explicitly mentioned by the teacher, it can be concluded from the teachers' responses in cases 2 and 3 that the teachers considered the "Bending Rods" task to be practical and implemented it without difficulty. Otherwise, a critical reflection would be included in answers.

In **lesson 10**, teaching thinking as hypothesis testing in the context of GBAR Experiment was perceived practical by all teachers. In **lessons 11 and 12,** teachers in all cases found analyzing the scenario for the Illuminati scene to be practical and did not express any difficulties in teaching these lessons with the provided materials. However, eliminating the search task on PET due to time constraints questioned the practicality of this task.

Table 9.2: Perceived practicality.

Practicality indicators of the antimatter course	**Perceived practicality**		
	Case 1	**Case 2**	**Case 3**
Lessons 3 & 4 & 5			
Teach Likelihood and uncertainty analysis skill in the context of positron discovery	+	+	+
Lesson 6			
Teach Thinking as hypothesis testing skill in the pair production context	+	+	+
Lesson 7			
Teach Argument analysis skill in the context of a systematic search for answers to the questions raised by watching a video clip about Big Bang	+	+	+
Lessons 8 & 9			
Teach Thinking as hypothesis testing skill in the context of trapping antihydrogen	+	+	+
Use Bending Rods example to transfer Thinking as hypothesis testing skill in the everyday life context	NA[a]	+	+
Lesson 10			
Teach Thinking as hypothesis testing skill in the context GBAR Experiment	+	+	+
Lessons 11 & 12			
Teach Verbal reasoning skill in the context of analyzing the scenario for the scene of “Illuminati” movie	+	+	+
Apply CT skills in search about Positron Emission Tomography (PET)	-	-	-

[a] Not Applicable

9.4 Perceived effectiveness

Teachers’ perceptions of the effect of the antimatter course on developing students’ CT are summarized in table 9.3. In **lessons 1 and 2**, teaching explicitly CT skills was perceived by all cases effective in developing students’ CT skills. The teacher in case 1 said:

> *"It was good that in teaching explicit CT skills, the students were shown some examples from everyday life. This taught them that these skills are not only applicable in physics, but also in everyday life."*

Furthermore, in teaching explicitly CT skills, students were provided with a worksheet (worksheet 1) on which they could practice the CT skills they had learned. Students were asked to describe a situation in which they had to think critically in school or in everyday life, describe their reaction, and then reflect on whether they applied CT skills in that situation.

To assess students' learning of general CT skills, students were given back a copy of their responses on this worksheet in the last session (Lessons 11 and 12) and asked to reflect on their responses and justify whether they needed to consider any CT skills in addition to those already mentioned on the worksheet. The teacher in case 2 explicitly perceived this activity as effective in developing students' CT skills since it engages students in the reflection on their thinking process. She wrote:

> *"Referring back to their own answers [on worksheet 1] in the last session was also a good way to help students understand their own progress [in CT]."*

In **lessons 3, 4, and 5**, teaching likelihood and uncertainty analysis skill in the context of positron discovery was perceived effective in developing students' CT skills. In all cases, teachers expressed that students were actively engaged in this activity. For example, in case 3, the teacher expressed the following statement as evidence of student CT development:

> *"On the basis of the trajectory analysis of the particle [in worksheet 3], which later proved to be an antiparticle, a lot [in CT] could be achieved."*

However, this statement did not clarify what evidence the teacher used to assess the development of students' CT skills, but at least showed that there were satisfactory activities observed by the teacher that she mentioned "a lot could be achieved". In addition, in case 2, the teacher expressed that students learned that:

> *"That one should not make only one possible hypothesis [solution], but should always look for others [hidden solutions]. And that only after thorough research and consideration of the various hypotheses [solutions] can be decided which has the greatest probability."*

In **lesson 6**, teaching thinking as hypothesis testing in the context of pair production was considered effective in developing students' CT skills by the teachers in cases 1 and 2. However, the teacher in case 3, criticized the effectiveness of this lesson by referring to the amount of teacher lecture and the few student activities. As indicated earlier, this was a misunderstanding of the teacher in this lesson, as the teacher package presented the lesson in a different way.

In **lesson 7**, in all cases, students' activities in systematically searching for answers to the questions raised after watching a video clip about Big Bang were considered as effective in developing CT skills. For example, in case 2, the teacher referred to the students' discussion while using the step-by-step guide on worksheet 5 as evidence of students' development of CT skills. In case 3, the teacher also mentioned:

> *"The unanswerable question about the "true big bang" especially: "What was before?" frustrated the pupils a bit, but brought a lot, because you can't always just google an answer or even here you need several high quality sources. This was a clear learning progress!"*

In **lessons 8 and 9**, the teachers used students' discussion about trapping matter and antimatter by using worksheet 6 as an evidence for improving students' CT skills. In addition, in case 2, the teacher used students' critical reflection on their use of thinking as hypothesis testing skill in the school as an evidence that students learned CT skills.

Table 9.3: Perceived effectiveness.

Effectiveness indicators of the antimatter course	Perceived effectiveness		
	Case 1	Case 2	Case 3
Lessons 1 & 2			
Teaching explicitly general CT skills	+	+	+
Using scene of movie to fulfill the problem-centered principle	+	+	+
Lessons 3 & 4 & 5			
Teaching Likelihood and uncertainty analysis skill in the context of positron discovery	+	+	+
Discussion and reflection	+	+	0
Lesson 6			
Teaching Thinking as hypothesis testing skill in the pair production context	+	+	-
Discussion and reflection	+	0	-
Lesson 7			
Teaching Argument analysis skill in the context of a systematic search for answers to the questions raised by watching a video clip about Big Bang	+	+	+
Discussion and reflection	+	+	+
Lessons 8 & 9			
Teaching Thinking as hypothesis testing skill in the context of trapping antihydrogen	+	+	+
Applying Thinking as hypothesis testing skill in the context of Bending Rods	NA[a]	0	0
Discussion and reflection	+	+	+
Lesson 10			
Teaching Thinking as hypothesis testing skill in the context GBAR Experiment	+	+	+
Discussion and reflection	+	+	+
Lessons 11 & 12			
Teaching Verbal reasoning skill in the context of analyzing the scenario for the scene of "Illuminati" movie	+	+	-
Applying CT skills in search about Positron Emission Tomography (PET)	0	0	0
Discussion and reflection	+	+	+

[a] Not Applicable

In **lesson 10**, the teacher in case 3, particularly, referred to students' search for GBAR Experiment and their attention to using credible information sources as evidence of students' CT skill development.

In **lessons 11 and 12**, the re-watching of the scene of "Illuminati" and the analysis of its scenario were perceived effective in cases 1 and 2. For example, in case 2, the teacher stated:

> *"Re-watching the movie is a good idea to show students their learning progress."*

However, in case 3, the teacher critically reflected on the use of the scenario and suggested using some selected statements from the movie to improve the effectiveness of the lessons. Although the scenario did not include too much text, the teacher suggested that:

> *"Concrete questions or statements from the movie would be good here. In other words, really well-chosen quotes that can be contradicted with the newly acquired knowledge or embedded in a context. Also, a summary at the end would be useful; in the form: "How do I approach a problem in a structured way?" or even earlier "How do I recognize questionable facts that I should investigate more closely?"."*

She used also the students' critical reflection on the analysing the scenario as an evidence for developing their CT skills and stated:

> *"Funnily enough, the students' criticism referred directly to the movie, because why should I worry about the (pseudo)scientific "babble" of an entertainment movie? Here the pupils assume anyway that not much (in terms of content/subject matter) is said that makes sense, since the audience does not expect this. It might be possible to take advantage of this and start even earlier: Can I expect from an*

entertainment movie technically correct dialogues to take place? And then start your series of lessons?!"

As alluded before, in none of the cases did the teachers conduct the instructional sequence in the context of PET, so there was no information on the effectiveness of this part.

In addition to the teachers' statements about the effectiveness of aforementioned activities, they reflected in the questionnaire on the effectiveness of the antimatter course in general. For example, the teacher in case 2 commented that:

"The opportunity to apply the critical thinking steps and discuss them with students makes the work effective."

She also appreciated the planned discussion and reflection phase in each lesson that gives students the opportunity to reflect on the CT skills taught in that particular lesson, and cited the high level of student engagement in the discussion and reflection as an indicator of the effectiveness of the antimatter course.

Furthermore, in all cases teachers commented that the lessons were very exciting and provided an opportunity to show students the application of various physics concepts in the subject of particle physics (e.g. electricity and magnetism in trapping antimatter, special relativity in describing the rest energy of particles).

9.5 Conclusion and discussion

A general overview on the results presented in table 9.1, 9.2, and 9.3, shows that the teaching-learning activities designed in the antimatter course were perceived by the teachers as relevant and effective to improve students' CT skills. Furthermore, they were considered practical since teachers could implemented them with available materials, without them reporting any undue difficulty in the implementation. However, some aspects needed to be addressed to improve the relevance, practicality, and effectiveness of the antimatter course.

Initial attention should be given to embedding Positron Emission Tomography (PET) in the course by devoting an additional lesson to this topic. This lesson clarifies the purpose of antimatter research at CERN by illustrating to students a medical application of antimatter. Furthermore, it provides students with the opportunity to apply the CT skills while searching on PET and presenting their results to the class.

The next issue is the difficulty reported by the teacher for students, especially the younger ones, in converting the units of energy and mass from the SI-units to the units applicable in particle physics (e.g. ev/c^2 for mass). Although the teacher suggested deleting this part, the values associated with mastering the concept of rest energy make this suggestion impractical. It is necessary to give younger students more time to engage in the activity in order to understand the concept. In addition, it is crucial to encourage students to persist and to appreciate their efforts as these also are part of the design principles of the antimatter course.

In addition, the fact that no information was provided about the relevance, practicality, and effectiveness of the “Bending Rods” example presented in the trapping antimatter lesson reveals that teachers did not recognize the importance of this task. However, the students’ work showed that this part was treated in the class but not highlighted by the teachers. This might be due to the fact that the topic of trapping antimatter was so fascinating that the teachers paid special attention to it when filling out the questionnaire. Such a finding illustrates the need to explicitly emphasize the importance of the task for teachers.

Apart from these aspects, the high level of relevance, practicality, and effectiveness of the teaching-learning activities in the antimatter course perceived by teachers suggests that the teacher package and materials are of good quality and can be used by teachers in schools even in regular physics classes.

10 Triangulation of findings

To triangulate the qualitative findings on the effectiveness of the antimatter course, a comparison is made with the quantitative results. This chapter discusses the results of the quantitative data collected by administrating Halpern Critical Thinking Assessment as pre- and post-tests to evaluate general CT skills and the PPCT test to evaluate domain-specific CT skills.

The discussion of the quantitative results addresses two goals. The first goal is to show how the quantitative results reflect the findings on the effectiveness of the antimatter course obtained through qualitative data analysis (see chapter 7). Furthermore, as discussed in section 7.2, a difference in the implementation of the antimatter course was found between cases. For example, the structural fidelity assessment showed lower results in case 3 compared to cases 1 and 2. The second goal, therefore, is to show how the quantitative results reflect the differences in course implementation.

10.1 Participants

From 36 students participating in cases 1, 2, and 3, only 21 students completed the pre- and post-test of HCTA for general CT and the PPCT for domain-specific CT (see table 10.1). Most of the data from the main study were missing because students were in home schooling due to the COVID 19 pandemic and data had to be uploaded directly by the students in the TUD-Cloud server. Table 10.1 shows an overview of students' information and also the results of evaluation of adherence to structure fidelity of implementation by teacher in each case.

Table 10.1: Overview of students' information and adherence to structure fidelity of the antimatter course by teacher in each case.

	Case 1	Case 2	Case 3
Number of students	10	9	17
Number of students (with pre-post-HCTA and PPCT)	9	4	8
Female students (%) (completed pre-post-HCTA and PPCT)	33	50	0
Adherence to structure fidelity of implementation by teacher (maximum of proposed activities is 16)	15	13	8

10.2 Evaluation of general critical thinking skills

To evaluate development of students' general CT skills, the results of student performance on the HCTA pretest and posttest were compared. In cases 1 and 2, the mean value of the pre- and posttest was so close that no significant improvement was to be expected. In case 3, the mean value had decreased (see table 10.2).

Due to the small sample size in each case, it was not expected to draw a conclusion about the effects of the antimatter course on students' general CT skills. The results from the literature, although using a much larger sample, also did not confirm any significant improvement in the acquisition of general CT skills observed when the pretest and posttest were administered.

For example, in the context of electricity and magnetism, a comparison of mean scores on the pretest and posttest of HCTA in both the treatment and control groups showed that the treatment setting with domain-specific CT instruction did not result in a statistically significant improvement in acquisition of general CT skills between the pretest and posttest scores (*cf.* Sermeus et al., 2021; Tiruneh et al., 2016). However, the average score improved between the pretest and posttest, but it was assumed to be a test-retest effect (*cf.* Tiruneh et al., 2016).

Moreover, similar results have been observed in psychology studies in which the mean scores on the pretest and posttest of a general CT test were similar in both the control and treatment groups and, furthermore, no improvement was observed between the pretest and posttest, which was more evident specifically in high-achieving students (*cf.* Williams et al., 2004). Therefore, given that the sample of the present study was highly gifted students, it is not surprising that no significant improvement in general CT skills was observed.

Table 10.2: Descriptive statistics for several cases on the Halpern Critical Thinking Assessment (HCTA) pre- and posttest and the Particle Physics Critical Thinking (PPCT) test scores and the p-value reported from the Shapiro-Wilk test for normality.

	Case 1		**Case 2**		**Case 3**	
Test	Mean±SD	*p*-value Shapiro-Wilk test for normality	Mean±SD	*p*-value Shapiro-Wilk test for normality	Mean±SD	*p*-value Shapiro-Wilk test for normality
HCTA Pretest*	46.67 ± 8.77	.133	55.75 ± 5.32	.666	38.63 ± 5.78	.132
HCTA Posttest*	46.22 ± 4.41	.454	55.25 ± 7.50	.565	33.50 ± 4.87	.341
PPCT**	22.00 ± 3.67	.962	21.75 ± 3.77	.976	10.63 ± 5.45	.988

* Total HCTA score is 75. ** Total PPCT score is 45.

However, to explain case 3's performance on the HCTA test and the decline in the mean on the posttest, a reference to the students' responses in the online questionnaire might be helpful. Students in case 3 explicitly referred to the many questions in the HCTA test and expressed exhaustion about answering the test, whereas this was not the case in cases 1 and 2. Discomfort with answering the HCTA in case 3 could explain a lower mean score on the HCTA posttest than on the HCTA pretest (*cf.* Williams et al., 2004).

10.3 Evaluation of domain-specific critical thinking skills

The results of administrating the PPCT test at the end of the antimatter course were reported for each case in Table 10.2. It was found that students in cases 1 and 2 gained better results than students in case 3. Comparing this finding with the result of the evaluation of structure fidelity of the implementation of the antimatter course (see table 7.2) showed that the teachers in cases 1 and 2 adhered more to the structure of the antimatter course than in case 3.

This result indicated that the implementation of the course as designed and reflected in the teacher package correlates with better results in the PPCT. However, due to the small number of students in each case, it was not possible to make a causal statement, but a positive correlation between the implementation of the antimatter course as intended with students' results in the PPCT test was observed. This highlights the impact of how the teacher implements the antimatter course on student performance in the PPCT.

10.4 Conclusion and discussion

The results of conducting the HCTA as pre- and posttests in the antimatter course showed no development of students' general CT skills. This finding is also consistent with other studies on designing domain-specific instruction to improve students' CT skills in physics and using HCTA to assess the development of general CT (*cf.* Sermeus et al., 2021; Tiruneh et al., 2016). However, the results of the qualitative data analysis did provide evidence for the development of students' general CT skills (see section 7.5). This emphasizes the importance of using both qualitative and quantitative methods of analysis to provide valid and reliable interpretations about CT development and the actual impact of domain-specific CT instruction (*cf.* Behar-Horenstein & Niu, 2011; Ennis, 1993, p. 186; McMillan, 1987, p. 15; Merrill, 2013, p. 371; Spicer & Hanks, 1995; Williams, 2005, p. 173).

In addition, the decrease in the mean score of students in case 3 on the post-HCTA indicated that students' perceptions of the CT tasks influence their performance (*cf.* Halpern, 1998; Williams et al., 2004). This demonstrates the importance of disposition as a component of CT (see section 7.6). Students in case 3 did not positively perceive the tasks in the HCTA and did not make an effort to answer the items.

However, the results of administering the PPCT test in different cases indicated a correlation between teachers' high structural fidelity and students' better performance in the PPCT test. This supports the effectiveness of the antimatter course in developing students' domain-specific CT skills gained by qualitative analysis.

Part III

Conclusion

11 Summary and discussion

This work focuses on the development and the evaluation of a teaching unit in particle physics to promote students' general and domain-specific critical thinking (CT). The work is an empirical effort inspired by the background theories of CT and its teaching and has been conducted using the Design-Based Research (DBR) approach. The empirics of the work and its contribution to the theory of CT instructional design are summarized and discussed below. In addition, the limitations of the study are reflected upon.

11.1 Empirical study

This work was founded on Halpern's (2009, 2014) definition of CT, focusing on 5 CT skills such as Verbal Reasoning (VR), Argument Analysis (AA), Thinking as Hypothesis Testing (HT), Likelihood and Uncertainty Analysis (LU), and Decision Making and Problem Solving (PS). The measurable outcomes of applying each skill in everyday life (general CT skills) were contextualized in the particle physics context and defined domain-specific CT skills of this work (see section 6.1.2.2). An antimatter course was designed for students in grades 10, 11, and 12 to enhance their general and domain-specific CT skills (see section 6.1.2.3). Using DBR, the antimatter course was developed by analyzing video and audio data of the lessons and student works collected in pilot studies. The final version of the course was implemented in different schools by regular physics teachers before and during COVID 19 (see section 7.1). Data of three cases were analyzed quantitatively and qualitatively to answer the main research question of the work.

How does a teaching-learning strategy about antimatter based on the First Principles of Instruction and Halpern's model of Critical Thinking support students of grades 10, 11 & 12 to learn general and domain-specific Critical Thinking skills?

The main research question was subdivided into the following research questions, the answers to which have already been discussed in chapters 7, 8, and 9. In the following, the answers to the individual sub-questions are summarized and discussed, and the results were finally used to answer the main research question.

RQ1.1 *What is the effectiveness of the antimatter course in improving students' general and domain- specific Critical Thinking skills?*

Regarding the development of general CT skills, the qualitative analysis of the data showed that students developed some general norms over the course, such as defining precise terminologies and finding credible sources of information. They also expressed the need for more information while working on tasks and for considering different perspectives. This shows a development in the students' general CT skills.

Regarding the development of domain-specific CT skills, data analysis showed that the antimatter course engaged students in the process of applying CT skills to achieve the desired outcomes of the course (see section 7.6.5). For example, students sought the precision criterion to identify vague and ambiguous terminologies when analyzing the scenario of a scene from the "Illuminati" movie, or used a strategy such as identifying variables when hypothesizing about the antimatter trap. Combining these skills with applying content knowledge and demonstration of disposition toward CT led to the conclusion that students were developing domain-specific CT skills (see figure 7.9). Furthermore, a comparison of the skills students developed during the antimatter course with the previous reports of students' difficulties in applying CT skills in physics (see table 7.11) clarified the extent to which the antimatter course was effective.

However, some challenges in applying CT skills in the specific tasks were also identified, such as the "communication fallacy" in which students interpret the concept of control variables as using authority to change variables according to preference instead of keeping them constant.

RQ1.2 *How do the students' general and domain-specific Critical Thinking skills develop (during the course) in relation to the teaching-learning activities and features of the antimatter course?*

The materials, the activity structure, and the participant structure including students' perception of CT, their culture of participation, and roles in teaching-learning sequences, constituted the design facets of the antimatter course. The design principles discussed in section 6.1.2, inspired the design of materials and the structure of the teaching-learning activities and therefore defined the students' role in the course. The course was so designed to engage students actively in discussion and reflection on CT tasks. Furthermore, the materials were perceived by the teachers to be well-structured and motivating. The teachers also valued students' active participation as an indicator for the development of CT skills. Moreover, students' effort and persistence in completing the tasks showed their willingness and positive attitude toward CT in particle physics. The data shows that the design of the antimatter course triggered the processes of applying content knowledge and applying critical thinking skills, as well as the development of dispositions that collectively correspond to developed CT.

However, the success of the antimatter course cannot be discussed independently of students' perceptions of CT and CT tasks, as well as their culture of participation in CT instruction. In the antimatter course, students perceived the materials positively in developing their CT skills, and also appreciated the opportunity of being involved in discussion and reflection. In addition, their culture of participation in the course, being open and respectful of each other's ideas, plays an important role in the success of the antimatter course.

Overall results indicate that the antimatter course was successful and that the design principles discussed in chapter 6 worked well.

RQ2 *What is the relevance, practicality, and perceived effectiveness of the antimatter course from the teachers' perspective with respect to promoting students' general and domain-specific Critical Thinking skills?*

In general, teachers in all three cases perceived the teaching-learning activities in the antimatter course as relevant to the development of students' CT skills. They pointed to the designed tasks in the worksheets and the planned opportunities for discussion and reflection in the course as a good means of developing students' CT skills. The only lesson that was not considered relevant was positron emission tomography (PET), which was eliminated all teachers due to time limitations.

All teachers found the implementation of the antimatter course practical with the available teacher package and materials. The only challenge in implementing the course was the conversion of mass and energy units from SI to particle physics units (ev for energy and ev/c^2 for mass), which was especially problematic in case 3. The fact that the students in case 3 were younger compared to the others might justify their difficulties in this area. In addition, eliminating the research task on the PET in all cases due to time constraints challenged its practicality.

The teaching-learning activities in the antimatter course were perceived effective in promoting students' CT by all teachers. Teachers used students' engagement in discussing and reflecting on the CT tasks, their questions for more information about the topics discussed, and their critical reflection on the application of CT skills in everyday life as evidence of the development of their CT skills. However, the task "analyzing the scenario for a scene of "Illuminati" movie" for teaching verbal reasoning skill was not considered effective by the teacher in case 3. She suggested using some selected statements from the movie that contradicted the students' newly acquired knowledge instead of this task. Although, the specific task of analyzing the scenario also had the character that the teacher desired, but it could not satisfy the teacher.

Overall, the teaching-learning activities in the antimatter course were mostly perceived to be relevant, practical, and effective in developing students' CT skills. However, some aspects should be reconsidered and possibly structured differently. For example, allocating one extra lesson to PET might improve the chance of it being addressed by teachers. Also, restructuring the activities to convert energy and mass units from SI-units to particle physics units might facilitate the students' work with these units.

Answer to the main research question

Based on the findings from the answers to the three sub-questions discussed above, the main research question of this study can be answered: *How does a teaching-learning strategy about antimatter based on the First Principles of Instruction and Halpern's model of Critical Thinking support students of grades 10, 11 & 12 to learn general and domain-specific Critical Thinking skills?*

The antimatter course engaged students in CT structure activities in which they were trained step-by-step to apply CT skills to solve tasks in the context of particle physics. The course was structured so that students had the opportunity to discuss and reflect on the tasks as a group and as a class. The students' positive perception of the course facilitated their engagement in the course and demonstrated their willingness to do CT, even if it took considerable effort. The teachers not only provided content knowledge, but also encouraged students to think critically and discuss and reflect on the tasks, acknowledged students' efforts, and asked reflective questions. Their positive perception of the course also facilitates their adherence to the structure and process of the antimatter course.

Overall, the design of the Antimatter course was able to support students' development of general and domain-specific CT skills, especially those that were raised as challenges in developing students' CT skills in physics (see table 7.12). However, there are still some challenges that need more attention, such as transferring the skill of controlling variables from particles physics context to the general context or vice versa, and overcoming overconfidence when arguing about open questions in particle physics.

11.2 Contribution to theory

The domain-specific CT instruction discussed in this study is an effective instruction for developing students' general and domain-specific CT skills. This can contribute to theory in several ways: the 6 principles used to develop the CT instruction can contribute to the theory of instructional design, which deals with building understanding and promoting thinking skills (*cf.* Moseley et al., 2005). Also, the evaluation of its effectiveness yielded some considerations for assessment, and the study of the processes in which students were involved led to a model for CT development. These are discussed in the following.

11.2.1 Theory of instructional design

The 6 principles are 1) Explicit teaching of general critical thinking skills, 2) Problem-centered principle, 3) Activation principle, 4) CT-Structure-training activities, 5) Encouraging discussion and reflection on tasks, and 6) Encouraging critical thinking. These 6 principles can be discussed at 3 meta-levels: fundamental principles, instruction planning principles, and instruction performing principles.

Fundamental principles: Explicit teaching of general critical thinking skills and Problem-centered principle

These two principles can be considered fundamental principles for designing effective CT instruction. The **explicit teaching of general critical thinking skills** helps students to get a clear picture of CT, clarifies for them the focus of the lesson, and makes them aware of the skills they might use to some degree unconsciously. Explicitly teaching CT skills helps students to become aware of these skills, learn how to apply the skills to achieve the goal, and show them that CT requires effort and persistence. This also contributes to defining some CT norms for students that accompany them during domain-specific CT lessons, such as defining terms precisely.

The **problem-centered principle** presents some criteria for defining a problem, such as authenticity, relevance to students' daily lives, and the necessity of applying CT skills to solve the problem. When defining a problem, the teaching-learning sequences have to be planned so that students acquire the knowledge and skills necessary to solve that problem. To make the course meaningful to students and help them understand why they are learning these topics, and to improve student participation in the course, the course can be outlined through a discussion with students about the content knowledge they need to solve the problem. Each lesson can begin by addressing the relationship between content knowledge and the central problem of the course.

Instruction planning principles: Activation principle and CT-Structure-training activities

The **activation principle** and **CT-Structure-training activities** are interrelated. The core message of these principles is to provide opportunities for students to be presented with various examples of CT tasks in order to practice CT skills and reflect on the structural aspects of the tasks, as well as to activate students' prior knowledge if applicable. The results of this study on students' challenges in applying CT skills in the context of particle physics provide guidance for the design of CT activities. For example, observing the problem of overconfidence in teaching likelihood and uncertainty analysis skills, as well as in teaching argument analysis skills for open questions, indicates that particular attention needs to be paid to this challenge in planning instruction.

Instruction performing principles: Encouraging discussion and reflection on tasks and Encouraging critical thinking

The principles of **encouraging critical thinking** and **encouraging discussion and reflection** on CT tasks relate to the teacher's role in domain-specific CT instruction. Valuing students' efforts to accomplish the CT tasks motivates students to stay on tasks and overcome the challenges, such as overconfidence when making a decision in a probabilistic situation. Furthermore, encouraging discussion and reflection on the tasks helps students evaluate their thinking process and overcome some fallacies in their thinking. For example, discussing the "majority fallacy", whereby students evaluated the validity of information based on the majority of sources presenting it, helped students distinguish between credible and "junk" information sources in the tasks that followed.

In summary, clarifying CT skills and their desired outcomes and defining an authentic problem that requires these skills, provide a framework that guides the planning and performing of CT instruction. The teaching-learning sequences can be systematically designed so that students acquire content knowledge and the CT skills needed to solve the problem. When performing CT instruction, appreciating students' efforts and also giving them space to discuss and reflect on their newly acquired knowledge and skills can engage students actively in the CT instruction.

11.2.2 Evaluation of critical thinking instruction

The domain-specific instruction in particle physics in this work improved both general and domain-specific CT skills. The inductive qualitative analysis revealed that students applied content knowledge and CT skills and demonstrated disposition in completing the tasks. Furthermore, they developed some norms during the course, such as defining terms precisely and considering different perspectives, that describe the development of students' general CT skills.

However, developing students' general CT was not supported by quantitative data obtained by implementing the HCTA as a pretest and posttest, which is also consistent with the results of other studies (*cf.* Sermeus et al., 2021; Tiruneh et al., 2017). This indicates that domain-specific CT instruction can also develop some general CT skills that are not visible and cannot be captured by HCTA. Therefore, to draw a full conclusion about the effectiveness of instruction in developing general and domain-specific CT skills, qualitative analysis is also needed.

11.2.3 Model on developing critical thinking

In the literature, the combination of attitude, content knowledge, and thinking skills is defined as CT (*cf.* Halpern, 2009, p. 5). In terms of thinking skills, a long list is provided for each CT skill (*cf.* Halpern, 2009; Norris & Ennis, 1989). In this work, the processes students went through to solve the domain-specific CT tasks were analyzed. From the detailed processes, a model emerged that can be used as a simple model for teaching domain-specific CT skills (see section 7.5.5). This model presents two general categories for applying CT skills: criterion-based evaluation and application of appropriate procedures. Dividing CT skills into these two main categories, rather than just listing the skills, simplifies understanding of the concept of CT and may improve the chance of its application by students.

Based on this model, in a given situation requiring CT, students must first recognize that they need to think of procedures and criteria as two main categories. Then they must distinguish which criteria and procedures apply to the given situation. They should ask *what criteria must be met* and *what procedures can be used to achieve the goal*. Asking these questions guides the thinking process and provides an entry point for critical thinking. The consistent application of these questions increases the chance that critical thinking becomes a mental habit for the thinker.

11.3 Limitations

In order to properly interpret the results of the study, the limitations of the study are reflected upon. The first limitation of the study is the limited number of cases for which the qualitative and qualitative data analysis was conducted. The COVID-19 pandemic limited the possibility of conducting the antimatter course in more schools. Although more teachers declared their voluntary participation in the study, the confrontation with COVID-19 limited the study to three cases. In addition, 5 classes were approached at the beginning of the main study, but due to the difficulty of switching between face-to-face and online lessons, two classes dropped further participation in the study.

In addition, some data were missing because due to home-schooling students were required to upload their work to the TU Cloud via a link. Furthermore, the limited participation of the researcher in the antimatter course conducted in cases 2 and 3 via COVID-19 prevented the collection of video and audio data of all lessons for these cases.

Another limitation concerns the lack of diversity of the cases. All cases were 10^{th} grade highly gifted students, which might explain their motivation and high level of performance in the antimatter course. Moreover, this concerns the saturated sample size discussed in grounded theory. However, it was not possible to diversify the sample size, but the constant comparative method that examines a common pattern between different groups in each case and the cases themselves could support the concept of saturated samples.

12 Outlook

The theoretical and empirical results of this work provide a basis for implications for CT teaching in physics and other subjects in schools and also inform future research perspectives, which are explained in the following sections.

12.1 Implications for teaching critical thinking

Teaching the skills that are applicable in a variety of contexts, including everyday life is the goal of education. CT as a key competence of the 21st century (Halpern, 2009; Lipman, 2003; NCR, 2008; see also review in Saleh, 2019) is one of one of the skills that needs to be transferred from school or university subjects to everyday life (Halpern, 2009; Lipman, 2003). The development of CT in schools needs well-designed and effective instruction, in which students gain content knowledge, apply CT skills, and develop their disposition toward CT. The design principles offered in this work for developing an antimatter course can be transferred to other contexts and can help teachers to design effective domain-specific CT instruction in other topics. In the following, the key aspects of application of these principles are elaborated to simplify their application to the other contexts.

The first principle, **explicit teaching of general critical thinking skills**, focuses on elaborating the targeted CT skills and their desired outcomes in everyday life. The proposed model for developing CT (see section 7.5.5), in which critical thinking skills are divided into two main categories, criterion-based evaluation and application of appropriate practices, facilitates the teaching of CT for teachers. It also simplifies the application of CT for students. This helps students look at CT skills from a meta-level rather than drowning in a long list of CT skills. This also directs students' thinking process.

The second principle, the **problem-centered principle**, introduces an authentic problem around which the course revolves, which means that during the course, students acquire the knowledge and skills needed to solve the problem. The course ends with the students going back to the problem and solving it. Merrill (2013, pp. 299-323) offers practical guidelines for designing a problem. The antimatter course was concerned with analyzing a scenario from the "Illuminati" movie, and at the beginning of the course students were asked to judge whether the information about antimatter presented in the scenario was scientifically correct. This exposed students to one of the rules of critical thinking: *To make a judgment about something, you need relevant and correct information*. For example, to judge the accuracy of the information in the movie about antimatter, one needs to know what antimatter is, how it can be produced, and so on. Discussion of the information needed to judge the correctness of the information presented in the scenario led to an outline of the course.

As indicated earlier, these two principles provide a framework for the course as they clarify the desired CT skills and content knowledge covered in the course and can guide the planning of the course, including the definition of learning objectives, the design of materials, and the design of teaching-learning sequences, as well as the implementation of the course.

When planning the CT instruction, the next two principles, the **activation principle** and **CT-structure-training activities**, play a role. In the antimatter course, the design of materials was guided by Hypothetical Learning Trajectories (see section 6.1.2) and also analyzing students' prior knowledge. An example of a hypothetical learning trajectory and the worksheet based on it for teaching the skill of likelihood and uncertainty analysis can be found in section 6.1.2. In the antimatter course, an attempt was made to teach each CT skill in the context of an authentic problem, e.g., likelihood and uncertainty analysis in the context of positron discovery, thinking as hypothesis testing in the context of trapping antimatter using the Penning trap. This was found to be very motivating by the teachers running the course and was also positively perceived by the students.

In implementing CT instruction, the two principles of **encouraging critical thinking** and **encouraging discussion and reflection** on CT tasks play a role. In addition to the quality of the task that maintains motivation to complete the task (Resnick, 1987, p. 43), the teacher's appreciation of the students' efforts in thinking critically is important. For example, one result of the principle of encouraging critical thinking in the antimatter course was that students were helped to overcome overconfidence when thinking about the hidden possibilities in an uncertain situation. Furthermore, encouraging in-class discussion and reflection helped students, for example, to recognize their misconceptions of the content (sated by students) and to reflect on the structural CT aspects of the task and think about transferring the skills to the other context (stated by teachers).

A very detailed practical application of the above design principles was presented in a teacher package developed to teach the antimatter course using the CT approach. The teacher package can be used by other teachers who wish to teach CT in particle physics, and may also motivate the design of other domain-specific CT instructions. It should be noted, however, that the well-written teacher package might not be sufficient to successfully implement CT instruction. Experience in implementing the antimatter course in various schools suggests that good teacher preparation for each lesson is needed, in which teachers review the package and materials (including worksheets and classroom PowerPoint presentations) to gain an understanding of the CT skills and how they can be achieved through this lesson.

In addition to the design principles discussed above, a number of new aspects of the challenges students face in applying CT skills in the context of particle physics, and possibly more generally in applying these skills in other contexts, became apparent when looking at the processes students went through in completing each CT task. These led to heuristics that might improve the effectiveness of domain-specific CT instruction (see also section 7.3) and are summarized below:

- In the context of the likelihood and uncertainty analysis (LU) skill, while confronting students with the evaluation of a probable event, an explicit hint or question to compare the probability of two events occurring with the probability of one event occurring might help students better understand the application of the LU skill.
- In the context of the likelihood and uncertainty analysis skill, discussing the “overconfidence” problem can help students to develop an attitude toward using probability judgment to improve decision making.
- In the context of using the internet to search, explicit reference to the distinction between credible websites that have expertise and first-hand knowledge and "junk" websites that present copied information from other websites without citation of sources could improve students’ sensitivity in using websites as the information source and eventually judging the validity of information.
- In the context of the argument analysis (AA) skill, explicit discussion of making an argument in the case of open questions might improve students’ AA skill.
- In the context of thinking as hypothesis testing, an explicit reference to the concepts of identifying variables and controlling them is important to make students aware of these strategies they might use unconsciously.
- In the context of thinking as hypothesis testing, special attention must be paid to the concept of controlling variables to allow the transfer of the skill to the general context.

In addition to the design principles developed in this study, a CT test for particle physics was also developed. The validity of reliability of the Particle Physics Critical Thinking (PPCT) test was also demonstrated. The development of the PPCT inspired by the CTEM (Tiruneh et al., 2017) may motivate other CT tests in other physics topics.

12.2 Future research

This work in physics education research can be considered relevant to further research on the following aspects:

The design principles offered by this work in the development of the antimatter course can be used for the systematic design of other domain-specific CT instructions to investigate how they can be applied in another context.

In addition, this study identified the transfer problem in applying the concept of controlling variables from particle physics to a general context through the design of appropriate tasks. Further research is needed to address other challenges students face in transferring CT skills from a domain-specific context to a general context and vice versa. This will clarify some aspects in the long-term discussion about the nature of CT, whether domain-specific and/or general, and the extent to which transfer can occur.

Furthermore, this study illustrated several other benefits of the antimatter course beyond promoting students' CT. For example, the likelihood and uncertainty analysis task (worksheet 3) was described by students as a creative task. This could lead to further research on the connection between different forms of thinking: critical thinking and creative thinking, and investigating whether developed critical thinking leads to developed creative thinking.

Another example of the added value of the course is the potential of the antimatter course in addressing the nature of science. The challenge for students to draw conclusions when the question is still open shows the potential of linking CT instruction to students' understanding of the nature of science.

In addition, this study introduced an underlying model (see figure 7.9) to facilitate the teaching and learning of CT by dividing CT skills into two main categories: criterion-based evaluation and application of appropriate procedures. The extent to which this model is successful from the perspective of teachers and students needs to be investigated.

Furthermore, professionalization of teachers for teaching CT in schools is necessary. Research on the professional development program for teachers is necessary so that teachers themselves can design effective CT lessons. In addition, research on the development of university courses for prospective teachers is relevant.

Appendix

Research instruments

Note: The research instruments were originally developed in German. A translation into English is presented here to ensure consistency in the language of the dissertation.

A Student information and prior knowledge questionnaire

TECHNISCHE UNIVERSITÄT DRESDEN

Bereich Mathematik und Naturwissenschaften, Fakultät Physik, Professur Didaktik der Physik

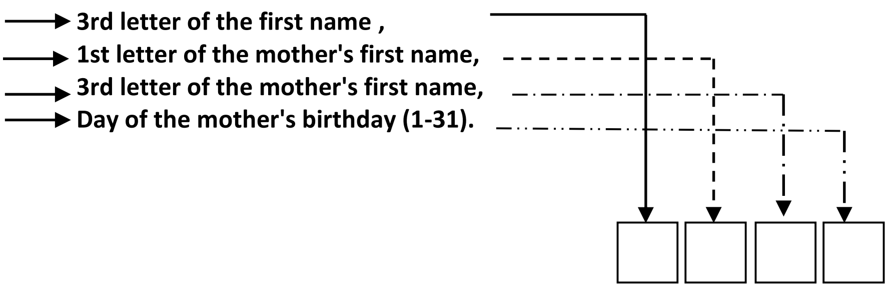
Dear Student,

At the TU Dresden a scientific study on the topic "Development and Evaluation of a teaching unit in particle physics to promote students' critical thinking" is carried out. For this purpose, a teaching unit on elementary particle physics (especially on antimatter) has been developed. To test the effectiveness of this teaching unit, **we need your help!**

Participation in the study is voluntary and completely anonymous!

Please read each question carefully and answer as accurately as possible.
Give the answers that seem most appropriate to you!
As an answer, make a cross in the fields
or write your answer in the lines provided.

All questionnaires in this course remain anonymous, but we must be able to clearly assign them to each other for the evaluation. This is made possible by a recognition code, which is derived from the following information. Please create your code according to the following rules:



Date: ____ . ____ . _______   Location: ____________________

Thank you for your assistance!

Farahnaz Sadidi (Technische Universität Dresden, 2018)

Questions about yourself

Please tick the boxes that apply to you, or write your answer on the lines provided!

1. You are ☐ female ☐ male

2. How old are you? ________ Years

3. What grade level are you in? ________

4. Do you attend a special physics course?
(Multiple answers possible)

☐ Science profile
☐ Inclination course in natural science
☐ Basic course physics
☐ Advanced physics course
☐ Physics study group
☐ Other, namely: ______________________________

____

5. What grade did you have on your last report card (or mid-year information) in physics?

☐ 1	☐ 2	☐ 3	☐ 4	☐5	☐ 6

Are you satisfied with this grade?

☐ yes	☐ no

Questions about the state of your prior knowledge on specific topics

What would best describe your knowledge of the following topics:

	I do not know anything about it yet	I have a rough idea what this is	I can explain to others what it is
Pair production	☐	☐	☐
Annihilation	☐	☐	☐
Dark Matter	☐	☐	☐
Quark	☐	☐	☐
Antimatter	☐	☐	☐
$E=mc^2$ (equivalence of mass and energy)	☐	☐	☐
Antimatter confinement	☐	☐	☐
Standard Model of Particle Physics	☐	☐	☐
Elementary building blocks of matter	☐	☐	☐
Positron Emission Tomography	☐	☐	☐

In what context have you heard the following terms before? ***(Multiple answers possible)***

	Term never heard before	In physics class	In movies/ novels	In the news	In science programs	Other namely
Pair production	☐	☐	☐	☐	☐	____________
Annihilation	☐	☐	☐	☐	☐	____________
Dark Matter	☐	☐	☐	☐	☐	____________
Quark	☐	☐	☐	☐	☐	____________
Antimatter	☐	☐	☐	☐	☐	____________
$E=mc^2$ (equivalence of mass and energy)	☐	☐	☐	☐	☐	____________
Antimatter confinement	☐	☐	☐	☐	☐	____________
Standard Model of Particle Physics	☐	☐	☐	☐	☐	____________
Elementary building blocks of matter	☐	☐	☐	☐	☐	____________
Positron Emission Tomography	☐	☐	☐	☐	☐	____________

Thank you very much for your cooperation!

B Particle Physics Critical Thinking (PPCT) test

Dear Student,

At the TU Dresden a scientific study on the topic "Development and Evaluation of a teaching unit in particle physics to promote students' critical thinking" is carried out. For this purpose, a teaching unit on elementary particle physics (especially on antimatter) has been developed. To test the effectiveness of this teaching unit, **we need your help!**

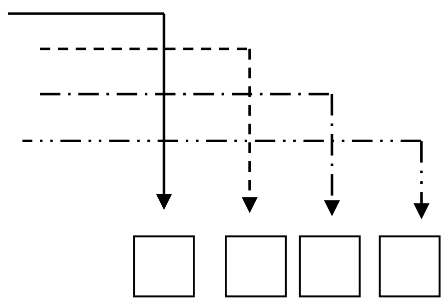
Participation in the study is voluntary and completely anonymous!

Please read each question carefully and answer as accurately as possible.

All questionnaires in this course remain anonymous, but we must be able to clearly assign them to each other for the evaluation. This is made possible by a recognition code, which is derived from the following information. Please create your code according to the following rules:

Anonymization

→ **3rd letter of the first name ,**
→ **1st letter of the mother's first name,**
→ **3rd letter of the mother's first name,**
→ **Day of the mother's birthday (1-31),**



Date and place of the answering the test:

Date: ____ . ____ . _______ Location: ____________________

Thank you for your assistance!

Farahnaz Sadidi (Technische Universität Dresden, 2019)

Imagine that you are invited by the Institute of Nuclear and Particle Physics (IKTP) to CERN (Conseil Européen pour la Recherche Nucléaire) near Geneva. You are excited and ready to learn about particle physics.

Figure 1: Location of the underground circular accelerator LHC at CERN

Source: https://cds.cern.ch/record/1295244, Photo by Maximilien Brice.

On the plane you will read about the development of the atomic model from Democritus (Greek philosopher) to today.

You can read the following:

See next page!

In 1909, Ernest Rutherford observed in scattering experiments with alpha particles on thin gold foils that the fewest alpha particles arrive at detector 2, while most alpha particles arrive at detector 1, see schematic in Figure 2.

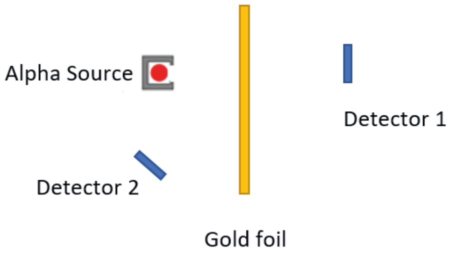


Figure 2: Schematic of the Rutherford experiment.

He interpreted this to mean that there is a small, heavy nucleus in the center of the gold atoms in the foil, which is orbited at large distances by the comparatively light electrons, see Figure 3.

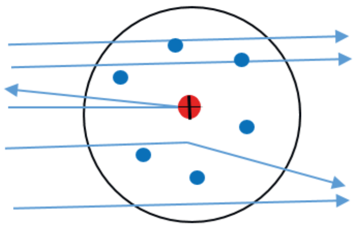


Figure 3: Schematic of Rutherford's atomic model.

Imagine that another scientist has developed an alternative model, as shown in Figure 4.

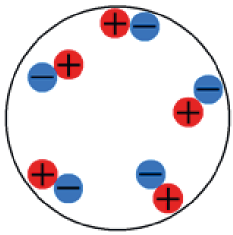


Figure 4: Schematic of an alternative atomic model.

Task 1:

Describe in detail your arguments against this alternative model, based on Rutherford's experimental data.

Tip: You can make your own drawings to support your thoughts ☺

Now you are at CERN.

Figure 5: Exhibition hall from CERN https://www.dawn.com/news/1114339, Photo by AFP.

You attend lectures to learn more about your favorite topic: Antimatter.

They are already looking forward to the program:

Date	Time	Activity
26.02.2021	8:00-9:30	Lecture: Antimatter 1
	10:00-11:30	Lecture: Antimatter 2
	13:00-14:00	Visit CERN exhibition
	14:30- 15:30	Lecture: GBAR Experiment
	16:00-17:30	Visit S'Cool-LAB

Before the first lecture, you talk to your friend about the famous Einstein equation:

$E = m \cdot c^2$.

E: Energy m: Mass c: Speed of light

In this context, your friend claims that mass can be converted into energy.

Task 2.1:

Do you agree with this statement?

○ Yes, I agree.

○ No, I don't agree.

Task 2.2:

Explain your reason.

After this pleasant discussion, the first lecture on antimatter begins.

Rolf Landua explains how a lot of kinetic energy can be used to create heavy matter and antimatter: "When a huge amount of energy is concerted on a tiny volume, new heavy particles and antiparticles can be created."

He also says, "Andy Murray, winner of the Olympic gold medal in tennis, transfers a lot of energy to the tennis ball-much more than you need to create a billion protons and antiprotons. But in fact, he can't generate any particles or antiparticles in his match."

Figure 6: Andy Murray.

The tennis ball example confuses you.

Task 3:

How do you have to argue logically to make both of Rolf's statements compatible?

After the first part of Rolf's lecture, you will have a break.

Figure 7: Time for a coffee!

You are fascinated by the antimatter topic and discuss with your friend the possibility of finding antimatter in the universe. Someone in the group overhears your interesting discussion and asks to join in. He asks you, "Is the moon made of antimatter?"

Figure 8: Neil Armstrong's footprint on the moon, photo: NASA.

Task 4.1:

What is your answer?

○ Yes.

○ No.

Task 4.2:

Give reasons for your answer

In the lecture, Rolf compares the collision of energetic particles with the collision of two strawberries. Using the resulting fruit, he explains that completely new particles and antiparticles are created in the particle collision, see Figure 9.

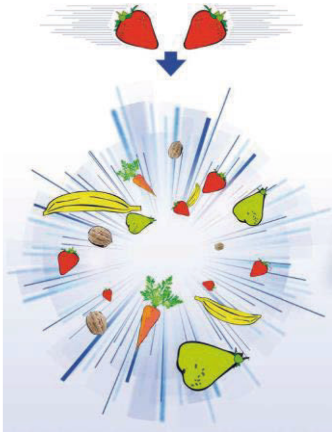


Figure 9: Analogy to particle collision, source: Netzwerk Teilchenwelt.

After Rolf's lecture, you surf the Internet to find more information about particle physics. You wonder about the following brochure, especially the text ("Smash particles together to look inside them …what will you find?").

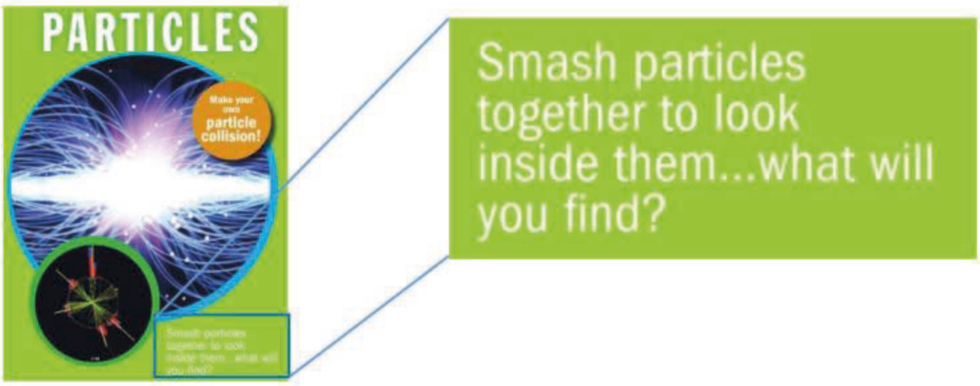


Figure 10: Brochure, source: Institute of Physics (IOP).

Task 5.1:

Evaluate this brochure text in light of Rolf's presentation and "position" yourself on it.

Task 5.2:

Give reasons for your judgment.

It's time to visit the CERN exhibition.

Figure 12: CERN exhibition room, source: CERN

In the exhibit, you see a picture of particle tracks from a bubble chamber and you discuss with your friend which type of particle would leave the spiral tracks (highlighted in blue), see Figure 13.

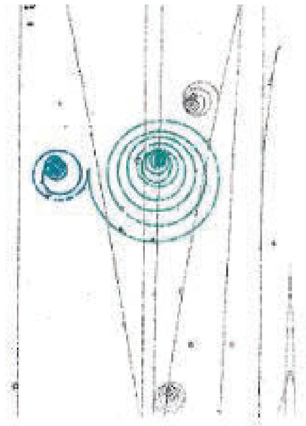


Figure 13: Particle tracks from a bubble chamber, source: Schmidt, 2015, p.44

Task 6.1:

What assumptions must you make before you can interpret the photograph in terms of particle sorts? Name these assumptions.

Task 6.2:

Do you need additional information to support your conclusion regarding particle sorts?

○ Yes. ○ No.

Task 6.3:

If yes, formulate questions regarding additional information needed.

Task 6.4:

If no, describe your thought process from the photograph (Figure 13) to the conclusion regarding particle sorts.

It is time to attend another lecture. This time it's about antimatter in the gravitational field. Interesting, isn't it?

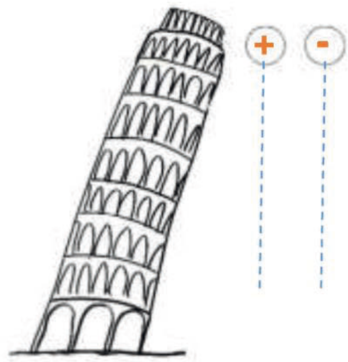


Figure 14: Leaning Tower of Pisa.

In this talk you will hear about the GBAR experiment. GBAR is designed to measure the gravitational acceleration ($\overline{\boldsymbol{g}}$) of neutral anti-hydrogen atoms in the gravitational field of the Earth. An experimental measurement uncertainty of 1% is expected.

You ask: If we were to drop a neutral anti-apple above the earth, would it fall up or down?

You ask your friend. He says, "Since neutral anti-hydrogen falls down in the GBAR experiment, we can conclude that the anti-apple would also fall down."

Task 7.1:

Would you agree with your friend?

○ Yes, I agree.

○ No, I don't agree.

Task 7.2:

Justify.

Figure 15: Anti-apple and Earth.

You are invited to the S'Cool-LAB at CERN. A full-day program with hands-on experiments on particle physics awaits you.

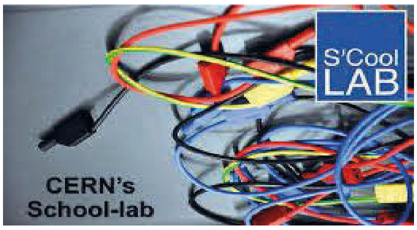


Figure 16: S'Cool LAB, source: CERN.

They want to build a Paul trap and experience how to trap particles.

Julia Woithe introduces the program: "Today you will explore antimatter. In doing so, you will find out how to catch particles with a Paul trap. Your particles today are not elementary particles like electrons, but very large particle systems, so-called Lycopodium spores."

Figure 17: Lycopodium spores.

After the introduction, perform the experiment and successfully capture spores. Hooray!

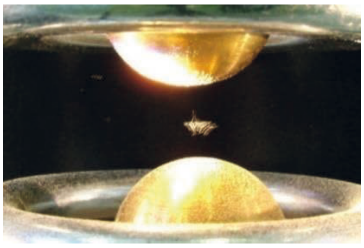


Figure 18: Capture of spores in Paul trap, photo: Göran Tronicke.

Suddenly you get the idea of trapping antimatter instead of spores, e.g. antiprotons.

Task 8.1:

What changes would you make in the experiment if you wanted to capture antiprotons?

Task 8.2:

Justify each proposed change.

After a long day at work, you talk to your friends about using antimatter as an alternative to fossil fuels.

One of your friends says, "Today we learned that when 0.5 g of antimatter and 0.5 g of matter annihilate, at least an amount of energy equal to $9 \cdot 10^{13}$ J is released. That's a lot of energy."

Another friend tells you about Star Trek: "Star Trek is about powering a spaceship with the energy of the pair annihilation of matter and antimatter."

Figure 19: Spaceship.

They are all excited and want to watch this part of the series to find out how they do it.

The idea in the movie is to travel to Sirius with this spaceship at half the speed of light (0.5 c).

You doubt this and try to calculate the energy needed to accelerate the spaceship from 0 to half the speed of light. According to your calculation, one needs an energy of $2.6 \cdot 10^{24}$ J.

Task 9.1:

What questions do you need to ask yourself to assess whether ideas described in the film are scientifically possible?

Question1:

Question2:

Task 9.2:

If you have further questions, write them down as well.

After watching Star Trek, you see one of your friends reading an article. Excited, he tells you, "Look, it says here that scientists have discovered how to convert light into matter and antimatter. Maybe we can do that, too."

You ask, "How?"

He replies, "We can use visible light, we have it everywhere."

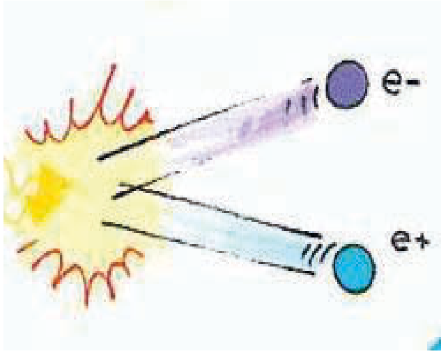


Figure 20: Pair production.

They have doubts about whether visible light can be used to form electron-positron pairs.

So think about how you can answer this question.

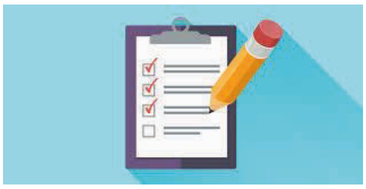


Figure 21: Answer sheet.

Task 10:

What steps should you take to answer the question? Write down these steps. (You do not need to do calculation).

Question: Can we create electron-positron pairs with visible light?
☐ ☐ ☐ ☐

You sit in the lobby of the hostel and talk with your friends about your exciting day and what you learned.

One of your friends wants to know how you actually define antimatter.

Philipp: "Antimatter consists of antiparticles. These have the same mass as particles of ordinary matter, but have a different electric charge."

Maria: "Antimatter consists of antiparticles. Antiparticles have the opposite electric charge of particles."

Eric: "Antimatter consists of antiparticles. The antiparticles have the same mass as particles and an equal but opposite electric charge."

Now it's your turn: you, too, are to define antimatter.

Task 11.1:

Evaluate your friends' definitions beforehand. If necessary, name any problems.

Philip:

Mary:

Eric:

Task 11. 2:

Who do you agree with?

Task 11.3:

Justify.

You heard in a lecture that scientists at CERN are discussing how to build a larger collider (let's call it XLHC) to find out more about the origin of the universe and its mysteries. In the image below, you see the LHC as a blue circle and the proposed circular collider as a gray dashed circle.

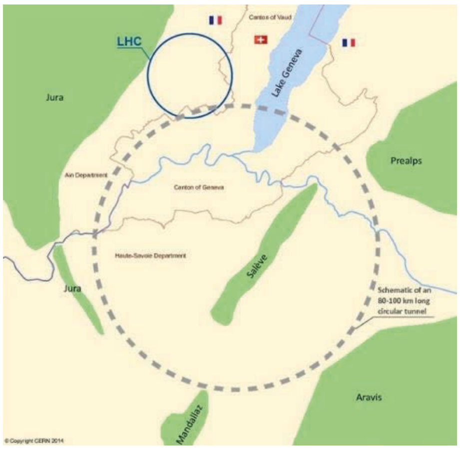


Figure 22: The proposed collider (XLHC), source: CERN.

The LHC has a 27-kilometer-long accelerator ring with superconducting magnets that greatly increase particle energy. It took almost 30 years to build.

The new collider (XLHC) is expected to have an 80-100 kilometer ring.

It is estimated that construction will take 20 years. Then it could be raised for the first time.

Task 12.1:

Do you think the scientists' plan is realistic and timely?

- Yes.
- No.

Task 12.2:

Justify.

You are chatting with your friend about antimatter. He asks, "If we could convert all matter particles into antimatter particles, would we notice differences?" He continues, "I mean, if we were antimatter people living in an antimatter world, would we feel differences?

Task 13.1:

How do you answer?

○ Yes, you notice the difference.

○ No, you don't notice any difference.

○ I'm not sure.

Task 13.2:

Explain your answer. You can refer to a law of nature or use an analogy.

You are curious how much energy can be converted in a pair annihilation.

First, consider a pair of **positron and electron**. For this, you calculate a minimum energy of 1.2 MeV.

Task 14.1:

Based on this calculation result, can you estimate how much energy at least comes out of a pair of **proton and antiproton**?

- o Yes.
- o No.

Task 14.2:

If yes, explain what background knowledge you use to arrive at your estimate.

Task 14.3:

If No, explain what other background information you need to arrive at your estimate.

You know that CERN can produce about $5 \cdot 10^{-10}$g of antihydrogen every year.

Imagine you are a CERN director and you want to increase the production rate of antihydrogen.

You have announced a competition. Now there are two proposals on your table. The first proposal claims that they can increase the production rate of antihydrogen to **5 g** per year. The second proposal states that they can increase the production rate of antihydrogen to **10 g** per year. Both proposals cost the same.

Task 15.1:

What questions would you ask the representatives of the proposals when assessing the feasibility of each? Formulate an accurate and complete question in complete sentences.

Question 1:

Question 2:

Task 15.2:

If you have further questions, write them down as well.

You have learned that CERN scientists are working on an antiproton cell experiment (ACE) to target cancer cells with antiprotons. It is still an experiment and it would take another decade to develop its clinical application.

Suppose a physician from hospital X wants to test the clinical success of a cancer treatment with antiprotons. He plans to experimentally treat three sick mice with the new method. During the therapy all mice die.

He then proclaims, "Antiprotons cannot be used for treatment. It kills people."

Task 16.1:

Do you agree with the doctor from hospital X?

- Yes, I agree.
- No, I don't agree.

Task 16.2:

Justify.

You are leafing through the regional newspaper. A blatant headline catches your eye:

"Antimatter, the ideal energy source! Without by-products, without radiation, without health risks."

They decide to write a letter to the editor to correct the headline.

List two counterarguments to the current headline.

Task 17:

Counterargument 1:

Counterargument 2:

C Student questionnaire

Questionnaire after the antimatter course

First, please enter your code according to the following rules: - 3rd letter of your own first name, - 1st letter of mother's first name, - 3rd letter of mother's first name, - day of mother's birthday (1-31).

1. Provide any content about antimatter that you remember.
2. How well can you explain to your classmates the content you mentioned in the last question about antimatter?
3. Describe what was easy to understand in the course about antimatter. Give reasons.
4. Describe what was difficult to understand in the course about antimatter. Give reasons.
5. Explain what unanswered questions about antimatter you still have?
6. List any critical thinking skills that you feel have developed during the course.
7. How well do you think you can use these critical thinking skills?
8. Explain and justify the assessment you made in the last question about your critical thinking skills.
9. Was this class different from regular classes?
10. Describe how it differed from regular classes.
11. Describe and give reasons for what you enjoyed in class.
12. Describe and give reasons for what you did not like in class.
13. How can we improve the quality of the course? List your suggestions.

D Teacher information

Dear Teacher,

At the TU Dresden a scientific study on the topic "Development and Evaluation of a teaching unit in particle physics to promote students' critical thinking" is carried out. For this purpose, a teaching unit on elementary particle physics (especially on antimatter) has been developed. To test the effectiveness of this teaching unit, **we need your help!**

Participation in the study is voluntary and completely anonymous! Please answer the following questions.

1. How many years have you been employed as a teacher?
2. Have you taught in secondary school before? □Yes □ No

   If Yes,

   Advanced course: □ every year

   □ almost every year

   □ occasionally

   □ never before

   Basic course: □ every year

   □ almost every year

   □ occasionally

   □ never before

3. Have you attended training courses on particle physics (at CERN, DESY, IKTP, ...)?

   □ Yes □ No

   If yes, which ones?

4. Have you taught in the field of particle physics (projects, AG, regular classes, ...)?
   □ Yes □ No

   If yes, which ones?

Date and place of the answering questionnaire:

Date: Place:

Thank you for your assistance!

Farahnaz Sadidi (Technische Universität Dresden, 2019)

E Teacher questionnaire

Dear Teacher,

Below are some questions about the practicality, relevance and effectiveness of the teaching unit. Your complete answer will help us a lot in evaluating the above variables (aspects)of the teaching unit. We ask for your assistance with your complete response.

Date: **Teacher Package Number:** **Class:**

1. were you able to carry out the teaching sequences in terms of content and method as suggested the teacher package? Please explain what you liked and justify your changes if necessary.

2.

a) Do you think that the amount of time allocated to each phase of teaching sequences was enough?

b) For teaching which part of this lesson did you need more time than allocated time in the teacher package? Why?

c) For teaching which part of this lesson did you need less time than allocated time in the teacher package? Why?

3. do you think that the activities designed in this lesson are fundamentally appropriate for improving students' critical thinking skills? Please explain.

4. What was your experience teaching this unit using the available teacher packets and materials? Please explain.

5. did you have any difficulties or problems in carrying out the lesson? Please describe.

6. did the teaching and learning activities in this lesson prove to be effective in improving students' critical thinking skills? Please elaborate.

7. do you have evidence to show that students are learning some critical thinking skills? Please describe the evidence.

Dear sincere thanks for your support.

Antimatter course materials

F Worksheet 1: Critical Thinking

You have already learned about critical thinking. Here I would like you to apply your newly acquired knowledge and answer the following questions.

Task A1

Describe a situation (or question or problem) in school or everyday life in which you had to think critically.

Task A2

Describe how you proceeded in this situation. Give reasons for your actions. Explain your thought processes.

For the work partner(s):

Task B1
Read and summarize your classmate's answers to questions **A1** and **A2.**

Task B2

B2.1 Indicate which critical thinking skills your classmate(s) used. Give reasons for your assignment.

B2.2 Indicate what other critical thinking skills could be used.

Task A3

Indicate which critical thinking skills you used. Give reasons for your assignment.

G Worksheet 2: Illuminati

Please watch the movie clip from "Illuminati" and answer the following questions.

Task 1

Describe in **keywords** what you think this part of the film is about.

Task 2

Now we want to assess whether the information in the film is scientifically correct. The answers to the following two questions might be helpful for this:

2.1. Indicate in key words what information was present in the film.

2.2. Indicate what information you need to assess whether the information in the film is scientifically accurate.

Task 3

Arrange the required information you mentioned in **Task 2.2.** in a logical order. Which knowledge do you need first?

For example: in order to judge the accuracy of the information in the film, we should

- first know ....
- second to know ....
- third to know ....
- ....
- ....

H Worksheet 3: Anderson’s cloud chamber photograph

Anderson took this photo with the help of a cloud chamber (see Figure 1). The magnetic field B is perpendicular to the plane of the sheet.

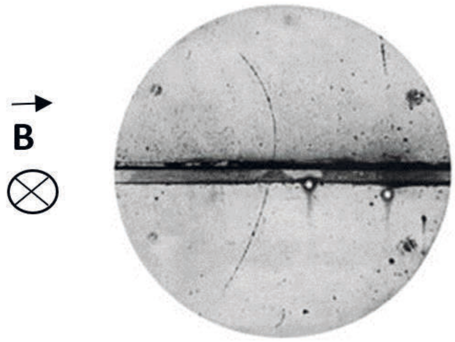
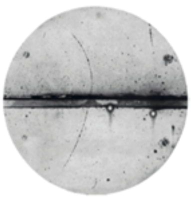
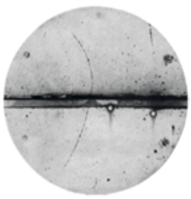
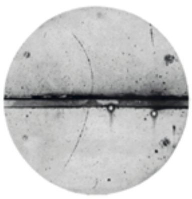
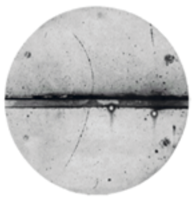


Figure 1: Anderson's cloud chamber photograph 1933, Phys.Rev.43, p.419.

Task 1.

Give a total of four different interpretations of the traces in Anderson's cloud chamber image, completing the table and drawing in the direction of motion of the particle in each copy.

(Note: All of the following photos are copies of Anderson's cloud chamber shot and they are the same.)

Interpretation regarding	 **Copy 1**	 **Copy 2**	 **Copy 3**	 **Copy 4**
Sign of the electric charge				
Direction of movement				

Task 2.1.

Indicate which of the four interpretations you gave in Task 1 is most likely?

Task 2.2.

For each interpretation, use the conservation laws to explain the arguments **for** or **against** each possible interpretation.

I Worksheet 4: Big Bang

You will be shown a video on the subject of the "Big Bang". Write down questions that arise for you during the video.

Question 1:

Question 2:

Question 3:

Question 4:

....

J Worksheet 5: Search systematically

First: Enter your code and the codes of all other students working in a group with you:

Code ....

Code ....

Code ....

You have already watched the Big Bang video and noted down some questions about it. Now the answers to your questions are to be found.

With your work partners, choose one of the previous questions (or two related questions) to begin your research on this topic.

1. Note your selected question here:

....

....

1.1 Is the question clear and precise?

□ Yes. Continue with 1.2.

□ No. Discuss the wording of your question. Then note down your revised question:

....

1.2 Is it possible to break the question down into sub-questions?

□ Yes. Discuss possible sub-questions in your group and note the ones that arise in the process developed sub-questions:

Main question:

Sub question 1:

Sub question 2:

□ No, continue with 1.3.

1.3 What do you think about the answer to this main question? (Tick off)

□ The question has a definite answer.

□ The answer to this question has something to do with one's own opinion.

□ The answer contains different points of view.

Start searching and gathering information or data now to answer your question(s).

2. indicate in bullet form what information you have gathered to answer your main question?

2.1 Do you believe that this information is correct?

□ Yes. Continue with 2.2.

□ No. Give reasons for your assessment.

2.2 How do you know that this information is correct?

2.3 Do you need more information?

□ Yes. Specify what information you need.

□ No. Continue with the next question.

3. state your conclusions (answers to the question).

3.1 Is the conclusion logical?

□ Yes. How do you know that?

□ No. Give reasons for your assessment.

3.2 Are there any other conclusions you should consider?

□ Yes. Note these.

□ No. Go to 4.

4. What exactly have scientists taken for granted? What assumptions did you make to reach this conclusion?

K Worksheet 6: Trapping antimatter

You have already learned how to trap matter (electrically charged Lycopodium spores) with the "Paul trap". Now we want to trap antiparticles.

1. Create a list of variables that can be considered in the construction of an antiparticle trap.

Variables	

2. How can these variables cause changes in the construction of the antiparticle trap compared to the Paul trap? Justify.

Variable	**Causes a difference in the design of the antiparticle trap compared to the Paul trap?**	**Justification**

3. Consider the variables that cause changes in the design of the antiparticle trap. Which of them can be controlled? Why do you need to control this variable?

Variables that make a difference in design	Can you control the variable?	Why do you need to control the variable?

4. Now hypothesize how the antiparticle trap should be constructed. Sketch the "antiparticle trap" you propose.

5. The next task is to compare your proposed antiparticle trap in Task 4 with the existing antiparticle trap at CERN. Look at the picture of the Penning trap, compare it with your proposed antiparticle trap. Describe differences and similarities.

Reflection question:

In the task above, you learned the skills of making hypotheses and designing experiments to test them. Please list the steps and justify the steps in a way that the skills can be applied to other problems.

-

-

-

-

Please apply these steps to the following task .*

Now imagine that you are given a long horizontal bar with 12 rods standing vertically on it. Each bar is made of either brass, copper or steel. The bars come in two lengths and two thicknesses. Figure 1 shows the structure.

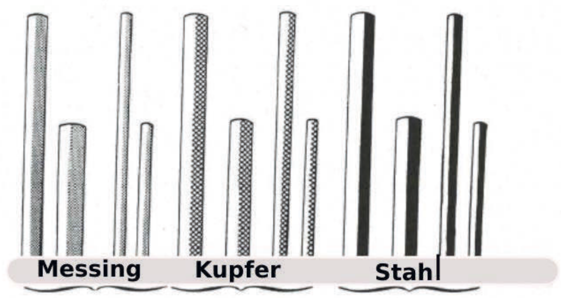


Figure 1: Brass, copper, and steel rods in two lengths and two thicknesses (Halpern, 2009, p. 149).

Your task is to find out which of the rods bends more than the others. You can test this by pressing down on each rod to see how much it bends. What do you need to do to determine which factors are important for bendability? You can make as many comparisons as you like. Write down your answer. Give reasons for your answer.

* Task according to Piaget, adopted from Halpern, 2009, p. 149.

L Worksheet 7: Individual work “Illuminati”

They were shown a scene from "Illuminati" in the first session of the antimatter course. You were asked to judge whether the information in the film was scientifically correct.

In doing so, you have learned that you need to be well informed about the topic at hand if you want to judge the correctness and accuracy of information or statements (movie scene, information from the newspapers or on the websites, and many other situations in everyday life).

You and your classmates in the course were interested in learning more about the following topics:

What is antimatter?
How can we create antimatter?
How can we "capture" antimatter?
What happens when antimatter comes into contact with matter?
During class, we discussed the above topics and now I would like you to take another critical look at the film clip and answer the following questions:

1) Which statements in the dialogs are vague and unclear? Give brief reasons for your assessment.

2) Give suggestions to make the information clearer and more precise. For the suggestions you mentioned, give the correct information about antimatter in this movie clip.

M Worksheet 8: Group work “Illuminati”

First: Enter your code and the codes of all other students working in a group with you:

Code ....

Code ....

Code ....

In your group, discuss the unclear information and your suggestions for clarifying/correcting it. Collect all the group's ideas and write down the unclear information and your suggestions here in key words.

unclear information	Proposal for clarification/correction

Based on your suggestions (clear information), rewrite the following scene.

Scene:

Vittoria Vetra:	Der Zylinder enthält eine extrem brennbare Substanz: Antimaterie. Wir müssen ihn sofort aufspüren oder den Vatikan evakuieren!
Cdr. Richter:	Mit Brandstoffen kenne ich mich ganz gut aus, Frau Vetra. Von Antimaterie habe ich dabei noch nie etwas gehört.
Vittoria Vetra:	sie wurde auch noch nie zuvor in entsprechender Menge generiert. Ein Weg, den Ursprung des Universums zu studieren... ...Die Antimaterie schwebt frei, im Vakuum einer Hülle aus Nano-Verbundstoffen mit Elektromagneten. Aber wenn sie ihren Schwebezustand verliert und mit Materie in Kontakt kommt, zum Beispiel dem Zylinderboden, zerstören sich die beiden entgegengesetzten Kräfte.
Cdr. Richter:	Und wie könnte sie den Schwebezustand verlieren?
Vittoria Vetra:	Wenn die Batterie leer ist, was kurz vor Mitternacht passiert.
Cdr. Richter:	Wie schlimm wäre diese Zerstörung?
Vittoria Vetra:	Katastrophal. Eine grelle Explosion mit der Kraft von fünf Kilotonnen.
Robert Langdon:	Der Vatikan wird verschlungen vom Licht.

Scenario from Illuminati

new scene with your suggestions:

List of tables

List of figures

References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research*, *85*(2), 275-314.

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of educational research*, *78*(4), 1102-1134.

Adams, W. K., & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. *International journal of science education*, *33*(9), 1289-1312.

Aizikovitsh-Udi, E., & Amit, M. (2011). Developing the skills of critical and creative thinking by probability teaching. *Procedia-Social and Behavioral Sciences*, *15*, 1087-1091.

Alazzi, K. F. (2008). Teachers' perceptions of critical thinking: A study of Jordanian secondary school social studies teachers. *The social studies*, *99*(6), 243-248.

Allegretti, C. L., & Frederick, J. N. (1995). A model for thinking critically about ethical issues. *Teaching of Psychology*, *22*(1), 46-48.

Amole, C., Andresen, G. B., Ashkezari, M. D., Baquero-Ruiz, M., Bertsche, W., Bowe, P. D., ... & ALPHA Collaboration. (2014). The ALPHA antihydrogen trapping apparatus. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 735, 319-340.

Anderson, C. D. (1933). The positive electron. *Physical Review*, *43*(6), 491.

Anderson, T., & Soden, R. (2001). Peer interaction and the learning of critical thinking skills. *Psychology Learning & Teaching*, *1*(1), 37-40.

Andrews, R. (2015). Critical thinking and/or argumentation in higher education. In *The Palgrave Handbook of Critical Thinking in Higher Education* (1st ed., pp. 89-153). New York: Palgrave Macmillan.

Bakker, A., & Gravemeijer, K. P. (2004). Learning to reason about distribution. In *The challenge of developing statistical literacy, reasoning and thinking* (pp. 147-168). Springer, Dordrecht.

Bakker, A., & Eerde, D. V. (2015). An introduction to design-based research with an example from statistics education. In *Approaches to qualitative research in mathematics education* (pp. 429-466). Springer, Dordrecht.

Bakker, A. (2018). Design Research in Education: A Practical Guide for Early Career Researchers (1st ed.). Routledge. https://doi.org/10.4324/9780203701010

Bakker, A., & Smit, J. (2018). Using hypothetical learning trajectories in design research. In *Design Research in Education* (pp. 255-271). Routledge.

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The journal of the learning sciences*, *13*(1), 1-14.

Barak, M., & Dori, Y. J. (2009). Enhancing higher order thinking skills among inservice science teachers via embedded assessment. *Journal of Science Teacher Education*, *20*(5), 459-474.

Barnett, R. (2015). A curriculum for critical being. In *The Palgrave handbook of critical thinking in higher education* (pp. 63-76). Palgrave Macmillan, New York.

Behar-Horenstein, L. S., & Niu, L. (2011). Teaching critical thinking skills in higher education: A review of the literature. *Journal of College Teaching & Learning (TLC)*, *8*(2).

Ben-Chaim, D., Ron, S., & Zoller, U. (2000). The disposition of eleventh-grade science students toward critical thinking. *Journal of Science Education and Technology*, *9*(2), 149-159.

Black, S. (2005). Teaching students to think critically. *The education digest*, *70*(6), 42.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in public health*, *6*, 149.

Boeije, H. (2002). A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Quality and quantity*, *36*(4), 391-409.

Boudreaux, A., Shaffer, P. S., Heron, P. R., & McDermott, L. C. (2008). Student understanding of control of variables: Deciding whether or not a variable influences the behavior of a system. *American Journal of Physics*, *76*(2), 163-170.

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive psychology*, *20*(4), 493-523.

Brown, D. (2000). Angels and Demons/Dan Brown. *NY: Pocket Books*.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science education*, *86*(2), 175-218.

Cho, J. Y., & Lee, E. H. (2014). Reducing confusion about grounded theory and qualitative content analysis: Similarities and differences. *Qualitative report*, *19*(32).

Choy, S. C., & Cheah, P. K. (2009). Teacher perceptions of critical thinking among students and its influence on higher education. *International Journal of teaching and learning in Higher Education*, *20*(2), 198-206.

Clesham, R. (2013). Good Assessment by Design. https://www.pearson.com/content/dam/one-dot-com/one-dot-com/uk/documents/about/news-and-policy/good-assessment-by-design-report.pdf (Retrieved on 08/08/2020)

Cobb, P., & Gravemeijer, K. (2008). Experimenting to support and understand learning processes. In *Handbook of design research methods in education* (pp. 67-95). Routledge.

Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational researcher*, *32*(1), 9-13.

Cobb, P., Stephan, M., McClain, K., & Gravemeijer, K. (2001). Participating in classroom mathematical practices. *The journal of the Learning Sciences*, *10*(1-2), 113-163.

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

Collins, A., & Joseph, D. K. Bielaczyc. (2004). *Design research: Theoretical and methodological issues. Journal of the Learning Sciences*, *13*(1), 15-42.

Davies, M. (2013). Critical thinking and the disciplines reconsidered. *Higher Education Research & Development*, *32*(4), 529-544.

Davies, M., & Barnett, R. (Eds.). (2015). *The Palgrave handbook of critical thinking in higher education*. Springer.

Davies, W. M. (2006). An 'infusion'approach to critical thinking: Moore on the critical thinking debate. *Higher Education Research & Development*, *25*(2), 179-193.

de Bie, H., Wilhelm, P., & van der Meij, H. (2015). The Halpern critical thinking assessment: Toward a Dutch appraisal of critical thinking. *Thinking skills and creativity*, *17*, 33-44.

Demir, S. (2015a). Evaluation of Critical Thinking and Reflective Thinking Skills among Science Teacher Candidates. *Journal of Education and Practice*, *6*(18), 17-21.

Demir, S. (2015b). Perspectives of Science Teacher Candidates Regarding Scientific Creativity and Critical Thinking. *Journal of Education and Practice*, *6*(17), 157-159.

Denscombe, M. (2014). *The Good Research Guide: For Small-scale Social Research Projects*. McGraw-Hill Education (UK).

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational researcher*, *32*(1), 5-8.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.) Sage publications.

Díaz, C., & De La Fuente, I. (2007). Assessing students' difficulties with conditional probability and Bayesian reasoning. *International Electronic Journal of Mathematics Education*, *2*(3), 128-148.

Dierdorp, A., Bakker, A., Eijkelhof, H., & van Maanen, J. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning*, *13*(1-2), 132-151.

Doyle, W., & Ponder, G. A. (1977). The practicality ethic in teacher decision-making. *Interchange*, *8*(3), 1-12.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science education*, *84*(3), 287-312.

Duron, R., Limbach, B., & Waugh, W. (2006). Critical thinking framework for any discipline. *International Journal of Teaching and Learning in Higher Education*, *17*(2), 160-166.

Elder, L., & Paul, R. (2012). *The thinker's guide to analytic thinking: How to take thinking apart and what to look for when you do*. The Foundation for Critical Thinking.

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational leadership*, *43*(2), 44-48.

Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational researcher*, *18*(3), 4-10.

Ennis, R. H. (1990). The extent to which critical thinking is subject-specific: Further clarification. *Educational researcher*, *19*(4), 13-16.

Ennis, R. H. (1993). Critical thinking assessment. *Theory into practice*, *32*(3), 179-186.

Ennis, R. H. (1997). Incorporating critical thinking in the curriculum: An introduction to some basic issues. Inquiry: Critical Thinking across the disciplines, 16(3), 1-9.

Ennis, R. H. (2015). Critical thinking: a streamlined conception. In *The Palgrave Handbook of Critical Thinking in Higher Education* (1st ed., pp. 31-47). New York: Palgrave Macmillan.

Etkina, E., & Planinšič, G. (2015). Defining and developing "critical thinking" through devising and testing multiple explanations of the same phenomenon. The Physics Teacher, 53(7), 432-437.

ETS (Educational Testing Service). (2008). Guidelines for constructed-response and other performance assessment. https://www.ets.org/content/dam/ets-org/pdfs/about/constructed-response-guidelines.pdf (Retrieved on 05/07/2017)

European Commission, Directorate-General for Research and Innovation. (2015). *Science education for responsible citizenship: report to the European Commission of the expert group on science education*, Publications Office. https://data.europa.eu/doi/10.2777/12626

Evagorou, M., & Dillon, J. (2011). Argumentation in the teaching of science. In *The professional knowledge base of science teaching* (pp. 189-203). Springer, Dordrecht.

Facione, N. C., & Facione, P. A. (1996). Externalizing the critical thinking in knowledge development and clinical judgment. *Nursing Outlook*, *44*(3), 129-136.

Fielding, N. G., & Fielding, J. L. (1986). Linking qualitative data. *Linking Data: The Articulation of Qualitative and Quantitative Methods in Social Research*, 41-53.

Ford, M. (2008). Disciplinary authority and accountability in scientific practice and learning. *Science Education*, *92*(3), 404-423.

Fox, R. B. (1962). Difficulties in developing skill in critical thinking. *The Journal of Educational Research*, *55*(7), 335-337.

Garcia-Mila, M., & Andersen, C. (2007). Cognitive foundations of learning argumentation. In *Argumentation in science education* (pp. 29-45). Springer, Dordrecht.

Garside, C. (1996). Look who's talking: A comparison of lecture and group discussion teaching strategies in developing critical thinking skills.

Gedigk, K., Kobel, M., & Pospiech, G. (2017). Development of Interest in Particle Physics as an Effect of School Events in an Authentic Setting. *Scientia in educatione*, *8*.

Gelder, T. V. (2005). Teaching critical thinking: Some lessons from cognitive science. *College teaching*, *53*(1), 41-48.

Glaser, B. G., & Strauss, A. L. (2017). *The discovery of grounded theory: Strategies for qualitative research*. Routledge.

Greeno, J. G. (2016). Cultural-historical activity theory/design-based research in Pasteur's quadrant. *Journal of the Learning Sciences*, *25*(4), 634-639.

Hake, R. R. (2008). Design-based research in physics education: A review. *Handbook of Design Research Methods in Education*, 493-508.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American psychologist*, *53*(4), 449.

Halpern, D. F. (2009). *Critical thinking across the curriculum: A brief edition of thought & knowledge*. Routledge.

Halpern, D. F. (2014). *Thought and Knowledge: An Introduction to Critical Thinking*. Taylor & Francis

Halpern, D. F. (2016). *Halpern Critical Thinking assessment: Manual.* Vienna, AT: Schuhfried GmbH.

Hamby, B. (2015). Willingness to inquire: The cardinal critical thinking virtue. In *The Palgrave handbook of critical thinking in higher education* (pp. 77-87). Palgrave Macmillan, New York.

Hartini, S., Liliasari, S., Sinaga, P., & Abdullah, A. G. (2021, March). An investigation of critical thinking skill of pre-service physics teacher in the case of fission and fusion reactions. In *Journal of Physics: Conference Series* (Vol. 1806, No. 1, p. 012031). IOP Publishing.

Hashweh, M. Z. (1996). Effects of science teachers' epistemological beliefs in teaching. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *33*(1), 47-63.

Haviz, M., & Maris, I. (2020). Measuring mathematics and science teachers' perception on thinking and acting in 21st-century learning. *Journal for the Education of Gifted Young Scientists*, *8*(4), 1319-1328.

Henderson, J. B., MacPherson, A., Osborne, J., & Wild, A. (2015). Beyond construction: Five arguments for the role and value of critique in learning science. *International Journal of Science Education*, *37*(10), 1668-1697.

Herrington, J., McKenney, S., Reeves, T., & Oliver, R. (2007, June). Design-based research and doctoral students: Guidelines for preparing a dissertation proposal. In *EdMedia+ Innovate Learning* (pp. 4089-4097). Association for the Advancement of Computing in Education (AACE).

Holyoak, K. J., & Morrison, R. G. (Eds.). (2005). *The Cambridge handbook of thinking and reasoning* (Vol. 137). Cambridge: Cambridge University Press.

https://www.kmk.org/fileadmin/Dateien/pdf/Bildung/Auslandsschulwesen/Kerncurriculum/Auslandsschulwesen-Operatoren-Naturwissenschaften-02-2013.pdf

Idol, L., & Jones, B. F. (Eds.). (1991). *Educational values and cognitive instruction: Implications for reform*. Routledge.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, *80*(4), 237.

Kali, Y. (2008). The design principles database as a means for promoting design-based research. In *Handbook of design research methods in education* (pp. 423-438). Routledge.

Kennedy, M., Fisher, M. B., & Ennis, R. H. (1991). Critical thinking: Literature review and needed research. In L. Idol & B.F. Jones (Eds.), *Educational values and cognitive instruction: Implications for reform (pp. 11-40)*. Hillsdale, New Jersey: Lawrence Erlbaum & Associates.

Kennedy, M., Fisher, M. B., & Ennis, R. H. (1991). Critical thinking: Literature review and needed research. *Educational values and cognitive instruction: Implications for reform*, *2*, 11-40.

Khan, S., & VanWynsberghe, R. (2008, January). Cultivating the under-mined: Cross-case analysis as knowledge mobilization. In *Forum: Qualitative Social Research* (Vol. 9, No. 1, p. 34). Institut für Qualitative Forschung.

Klahr, D., Chen, Z., & Toth, E. E. (2001). Cognitive development and science education: Ships that pass in the night or beacons of mutual illumination? In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 75–119). Lawrence Erlbaum Associates Publishers.

KMK (2013). Operatoren für die naturwissenschaftlichen Fächer (Physik, Biologie, Chemie) an den Deutschen Schulen im Ausland.

KMK (2020). Bildungsstandardsim Fach Physikfür die Allgemeine Hochschulreife, Beschluss der Kultusministerkonferenz vom 18.06.2020, https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2020/2020_06_18-BildungsstandardsAHR_Physik.pdf (Retrieved on 23/12/2020).

Kowalczyk, N., Hackworth, R., & Case-Smith, J. (2012). Perceptions of the use of critical thinking teaching methods. *Radiologic technology*, *83*(3), 226-236.

Kryjevskaia, M., Heron, P. R., & Heckler, A. F. (2021). Intuitive or rational? Students and experts need to be both. *Physics Today*, *74*(8), 28-34.

Kuebli, J. E., Harvey, R. D., & Korn, J. H. (2008). Critical thinking in critical course: principles and applications. *Teaching Critical Thinking in Psychology: A Handbook of Best Practices (2008)*, *137*.

Lai, E. R. (2011). Critical thinking: A literature review. *Pearson's Research Reports*, *6*, 40-41.

Larkin, J. H., & Reif, F. (1979). Understanding and teaching problem-solving in physics. *European journal of science education*, *1*(2), 191-203, DOI: 10.1080/0140528790010208

Lauer, T. (2005). Teaching critical-thinking skills using course content material. *Journal of College Science Teaching*, *34*(6), 34.

Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, *15*(1), 11-24.

Lawson, A. E. (2004). The nature and development of scientific reasoning: A synthetic view. *International Journal of Science and Mathematics Education*, *2*(3), 307-338.

LeCompte, M. D., Millroy, W. L., & Preissle, J. (1992). The Handbook of Qualitative. *Research in Education*.

Lesh, R., Kelly, A., & Yoon, C. (2008). Multitiered design experiments in mathematics, science and technology education. *Handbook of design research methods in education: Innovations in science, technology, engineering and mathematics learning and teaching*, 131-148.

Levrini, O., De Ambrosis, A., Hemmer, S., Laherto, A., Malgieri, M., Pantano, O., & Tasquier, G. (2016). Understanding first-year students' curiosity and interest about physics—lessons learned from the HOPE project. *European Journal of Physics*, *38*(2), 025701.

Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory into practice*, *32*(3), 131-137.

Lipman, M. (2003). *Thinking in education*. Cambridge University Press.

Lipman, M. (2008). Interview on Philosophy for Children. https://p4c.com/julie-winyard-interviews-matthew-lipman/ (Retrieved on 23.12.2020).

Lobato, J. (2008). Research methods for alternative approaches to transfer: Implications for design experiments In A. Kelly & R. Lesh. *Design research in education*.

Lynch, S. (2005). *Fidelity of implementation in implementation and scale-up research designs: Applications from 4 studies of innovative curriculum materials and diverse populations*. Paper presented at the American Educational Research Association (AERA), Montreal, CA.

McBride, R., & Bonnette, R. (1995). Teacher and at-risk students' cognitions during open-ended activities: Structuring the learning environment for critical thinking. *Teaching and teacher Education*, *11*(4), 373-388.

McKenney, S., & Reeves, T. C. (2019). *Conducting educational design research*. Routledge.

McMillan, J. H. (1987). Enhancing college students' critical thinking: A review of studies. *Research in higher education*, *26*(1), 3-29.

McNeill, K. L. (2009). Teachers' use of curriculum to support students in writing scientific arguments to explain phenomena. *Science Education*, *93*(2), 233-268.

McNeill, K. L., & Krajcik, J. S. (2011). Supporting Grade 5-8 Students in Constructing Explanations in Science: The Claim, Evidence, and Reasoning Framework for Talk and Writing. *Pearson*.

McNeill, K. L., Katsh-Singer, R., González-Howard, M., & Loper, S. (2016). Factors impacting teachers' argumentation instruction in their science classrooms. *International Journal of Science Education*, *38*(12), 2026-2046.

McNeill, K. L., González-Howard, M., Katsh-Singer, R., & Loper, S. (2017). Moving beyond pseudoargumentation: Teachers' enactments of an educative science curriculum focused on argumentation. *Science Education*, *101*(3), 426-457.

McPeck, J. E. (1990). *Teaching critical thinking: Dialogue and dialectic*. Routledge.

McPeck, J. E. (2016). *Critical thinking and education*. Routledge.

Merrill, M. D. (2002). First principles of instruction. *Educational technology research and development*, *50*(3), 43-59.

Merrill, M. D. (2013). First principles of instruction: Identifying and designing effective, efficient and engaging instruction. Hoboken, NJ: Pfeiffer.

Miller, M. A. (1992). Outcomes evaluation: Measuring critical thinking. *Journal of Advanced Nursing*, *17*(12), 1401-1407.

Moore, T. (2004). The critical thinking debate: how general are general thinking skills?. *Higher Education Research & Development*, *23*(1), 3-18.

Moseley, D., Baumfield, V., Elliott, J., Higgins, S., Newton, D. P., Miller, J., & Gregson, M. (2005). *Frameworks for thinking: A handbook for teaching and learning*. Cambridge University Press.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American journal of evaluation*, *24*(3), 315-340.

Nieto, A. M., & Valenzuela, J. (2012). A study of the internal structure of critical thinking dispositions. *Inquiry: Critical Thinking Across the Disciplines*, *27*(1), 31-38.

Nieveen, N. M. (1997). Computer support for curriculum developers: A study on the potential of computer support in the domain of formative curriculum evaluation.

Nieveen, N. (1999). Prototyping to reach product quality. In *Design approaches and tools in education and training* (pp. 125-135). Springer, Dordrecht.

Nieveen, N. (2010). Formative Evaluation in Eduational Design Research. Dalam Tjeer Plom and Nienke Nieeveen. *An Introduction to Educational Design Research*, 9-35.

Nieveen, N., & Folmer, E. (2013). Formative evaluation in educational design research. *Design Research*, *153*, 152-169.

Norris, S. P. (1986). Thinking about critical thinking: Philosophers can't go it alone. Philosophy of Education Society.

Norris, S. P., & Ennis, R. (1989). *Evaluating Critical Thinking: The Practitioners' guide to teaching thinking series*. Midwest Publications

Noula, I. (2018). Critical Thinking and Challenges for Education for Democratic Citizenship: an ethnographic study in primary schools in Greece. *Educação & Realidade*, *43*, 865-886.

NRC (National Research Council). (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

OECD (2017). PISA 2015 science framework. https://www.oecd-ilibrary.org/docserver/9789264281820-3-en.pdf?expires=1670067320&id=id&accname=ocid70029272&checksum=12D6D16A02AEF11376E8B8636390BFF1(Retrieved on 31/08/2020).

Oettle, M., Mikelskis-Seifert, S. (2020). Lehrkräftefachwissen in der Teilchenphysik Finale Delphi-Ergebnisse. In Habig, S. (2020). Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen. Gesellschaft für Didaktik der Chemie und Physik Jahrestagung in Wien 2019.

Ogunleye, A. O. (2009). Teachers and students perceptions of students problem-solving difficulties in physics: Implications for remediation. *Journal of College Teaching & Learning (TLC)*, *6*(7).

P21 Partnership for 21st Century Learning. (2019). Framework for 21st Century Learning Definitions. https://static.battelleforkids.org/documents/p21/P21_Framework_DefinitionsBFK.pdf (Retrieved on 10/08/2022).

Papadouris, N., Vokos, S., & Constantinou, C. P. (2018). The pursuit of a “better” explanation as an organizing framework for science teaching and learning. *Science Education*, *102*(2), 219-237.

Passon, O., Zügge, T., & Grebe-Ellis, J. (2020) Kohärenz im Unterricht der Elementarteilchenphysik. Springer Spektrum, Berlin, Heidelberg.

Paul, R. W. (1993). Critical Thinking: What Every Person Needs to Survive. *A Rapidly Changing World. 38rd edition revisited*, 97-98.

Paul, R. W., Elder, L., & Bartel, T. (1997). Teachers of Teachers: Examining: Preparation for Critical Thinking. Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, IL, March 24-28.

Perkins, D. N., Allen, R., & Hafner, J. (1983). Difficulties in everyday reasoning. In W. Maxwell (Ed.), *Thinking: The frontier expands* (pp. 177–189). Hillsdale, New Jersey: Lawrence Erlbaum & Associates.

Philipsen, B., Tondeur, J., McKenney, S., & Zhu, C. (2019). Supporting teacher reflection during online professional development: a logic modelling approach. *Technology, pedagogy and education*, *28*(2), 237-253.

Pithers, R. T., & Soden, R. (2000). Critical thinking in education: A review. *Educational research*, *42*(3), 237-249.

Pospiech, G. (2020). Teilchenphysik in der schule: Ein fachdidaktischer Blick auf die Elementarteilchenphysik. *Unterricht Physik*, *180*, 9- 13.

Reddy, M., & Panacharoensawad, B. (2017). Students Problem-Solving Difficulties and Implications in Physics: An Empirical Study on Influencing Factors. *Journal of Education and Practice*, *8*(14), 59-62.

Redish, E. F., & Hammer, D. (2009). Reinventing college physics for biologists: Explicating an epistemological curriculum. *American Journal of Physics*, *77*(7), 629-642.

Reed, J. H., & Kromrey, J. D. (2001). Teaching critical thinking in a community college history course: Empirical evidence from infusing Paul's model. *College Student Journal*, *35*(2), 201-201.

Resnick, L. B., & Science National Research Council (US). Committee on Research in Mathematics. (1987). Education and learning to think.

Rini, D. E. P., & Prabawanto, S. (2021, March). Mathematics teachers’ perception of learning models associated with high school students’ critical thinking skills. In *Journal of Physics: Conference Series* (Vol. 1806, No. 1, p. 012080). IOP Publishing.

Ritchhart, R., & Perkins, D. N. (2005). *Learning to Think: The Challenges of Teaching Thinking*. Cambridge University Press.

Ryu, S., & Sandoval, W. A. (2012). Improvements to elementary children's epistemic understanding from sustained argumentation. *Science Education*, *96*(3), 488-526.

Sadidi, F., & Pospiech, G. (2019a). Teaching Critical Thinking in the physics classroom: High school students think about antimatter. In *Journal of Physics: Conference Series* (Vol. 1287, No. 1, p. 012063). IOP Publishing. https://iopscience.iop.org/article/10.1088/1742-6596/1287/1/012063

Sadidi, F., & Pospiech, G. (2019b). Teaching Critical Thinking in the Physics Classroom: Formative assessment of a content-specific instructional design. In C. Maurer (Hrsg.), Naturwissenschaftliche Bildung als Grundlage für berufliche und gesellschaftliche Teilhabe. Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Kiel 2018. (SS. 281- 284).

Sadidi, F., & Pospiech, G. (2020). Developing students' critical thinking: likelihood and uncertainty analysis in particle physics. In Levrini, O. & Tasquier, G. (Eds.), Electronic Proceedings of the ESERA 2019 Conference. The beauty and pleasure of understanding: engaging with contemporary challenges through science education, Part 5/5 (co-ed. *Nikos Papadouris & Italo Testa*), (pp. 657-665). Bologna: ALMA MATER STUDIORUM – University of Bologna. 978- 88-945874-0-1978-88-945874-0-1

Sadidi, F., & Pospiech, G. (in preparation). Evaluating students' critical thinking: Design of a critical thinking assessment in particle physics.

Saleh, S. E. (2019). Critical thinking as a 21st century skill: conceptions, implementation and challenges in the EFL classroom. *European Journal of Foreign Language Teaching*.

Sandoval, W. A. (2004). Developing learning theory by refining conjectures embodied in educational designs. *Educational psychologist*, *39*(4), 213-223.

Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the learning sciences*, *23*(1), 18-36.

Schecker, H., Wilhelm, T., Hopf, M., Duit, R., & Fischler, H. (2018). *Schülervorstellungen und Physikunterricht*. Springer Berlin Heidelberg.

Sermeus, J., De Cock, M., & Elen, J. (2021). Critical thinking in electricity and magnetism: assessing and stimulating secondary school students. *International Journal of Science Education*, 1-21.

Shaheen, N. (2016). International students' critical thinking–related problem areas: UK university teachers' perspectives. *Journal of Research in International Education*, *15*(1), 18-31.

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for research in mathematics education*, *26*(2), 114-145.

Simon, M. A., & Tzur, R. (2004). Explicating the role of mathematical tasks in conceptual learning: An elaboration of the hypothetical learning trajectory. *Mathematical thinking and learning*, *6*(2), 91-104.

Smit, J. (2013). *Scaffolding language in multilingual mathematics classrooms*. Utrecht University.

Smit, J., Bakker, A., Van Eerde, D., & Kuijpers, M. (2016). Using genre pedagogy to promote student proficiency in the language required for interpreting line graphs. *Mathematics Education Research Journal*, *28*(3), 457-478.

Songer, N. B., & Gotwals, A. W. (2005, April). Fidelity of implementation in three sequential curricular units. In *Annual Meeting of the American Educational Research Association, Montreal, Canada*.

Spicer, K. L., & Hanks, W. E. (1995). Multiple Measures of Critical Thinking Skills and Predisposition in Assessment of Critical Thinking. Paper presented at the Annual Meeting of the Speech Communication Association (81st, San Antonio,TX, November 18-21, 1995).

Stokhof, H., De Vries, B., Bastiaens, T., & Martens, R. (2019). Mind map our way into effective student questioning: A principle-based scenario. *Research in Science Education*, *49*(2), 347-369.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: grounded theory procedures and techniques,* CA: Sage publications.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research techniques* (pp. 1-312). Thousand oaks, CA: Sage publications.

Tarr, J. E., & Jones, G. A. (1997). A framework for assessing middle school students' thinking in conditional probability and independence. *Mathematics Education Research Journal*, *9*(1), 39-59.

Tarr, J. E., & Lannin, J. K. (2005). How can teachers build notions of conditional probability and independence?. In *Exploring Probability in School* (pp. 215-238). Springer, Boston, MA.

Temel, S. (2014). The effects of problem-based learning on pre-service teachers' critical thinking dispositions and perceptions of problem-solving ability. *South African journal of education*, *34*(1), 1-20.

Tiruneh, D. T., De Cock, M., & Elen, J. (2018). Designing learning environments for critical thinking: Examining effective instructional approaches. *International journal of science and mathematics education*, *16*(6), 1065-1089.

Tiruneh, D. T., De Cock, M., Weldeslassie, A. G., Elen, J., & Janssen, R. (2017). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. *International Journal of Science and Mathematics Education*, *15*(4), 663-682.

Tiruneh, D. T., Verburgh, A., & Elen, J. (2014). Effectiveness of critical thinking instruction in higher education: A systematic review of intervention studies. *Higher Education Studies*, *4*(1), 1-17.

Tiruneh, D. T., Weldeslassie, A. G., Kassa, A., Tefera, Z., De Cock, M., & Elen, J. (2016). Systematic design of a learning environment for domain-specific and domain-general critical thinking skills. *Educational Technology Research and Development*, *64*(3), 481-505.

Van den Akker, J. (2013). Curricular development research as specimen of educational design research. *Educational design research*, 53-70.

Verburgh, A., François, S., Elen, J., & Janssen, R. (2013). The assessment of critical thinking critically assessed in higher education: A validation study of the CCTT and the HCTA. *Education Research International*, *2013*.

Viennot, L. (2001). *Reasoning in physics: The part of common sense*. Springer Science & Business Media.

Viennot, L., & Décamp, N. (2018). Activation of a critical attitude in prospective teachers: From research investigations to guidelines for teacher education. *Physical Review Physics Education Research*, *14*(1), 010133.

Viennot, L., & Décamp, N. (2020). *Developing critical thinking in physics*. Springer International Publishing.

Wagner, W. (2020). Quarks und Feynman-Diagramme: Basiswissen im Schulunterricht?. In *Kohärenz im Unterricht der Elementarteilchenphysik* (pp. 3-11). Springer Spektrum, Berlin, Heidelberg.

Wendland, M. W., Robinson, C., & Williams, P. A. (2015). Thick critical thinking: Toward a new classroom pedagogy. In *The Palgrave handbook of critical thinking in higher education* (pp. 153-168). Palgrave Macmillan, New York.

Williams, R. L. (2005). Targeting critical thinking within teacher education: The potential impact on society. *The Teacher Educator*, *40*(3), 163-187.

Williams, R. L., Oliver, R., & Stockdale, S. (2004). Psychological versus generic critical thinking as predictors and outcome measures in a large undergraduate human development course. *The Journal of General Education*, 37-58.

Wilson, C. D., Miles, C. L., Baker, R. L., & Schoenberger, R. L. (2000). Learning Outcomes for the 21st Century: Report of a Community College Study.

Woods, P. (1992). Symbolic Interactionism: Theory and Method. In the handbook of qualitative research in education. M. LeCompte, W. Millroy, &J. Preissle.

Wozniak, H. (2015). Conjecture mapping to optimize the educational design research process. *Australasian Journal of Educational Technology*, *31*(5).

Yin, R. K. (1989). Case study research: Design and methods, revised edition. *Applied Social Research Methods Series*, *5*.

Yin, R. K. (2009). How to do better case studies. *The SAGE handbook of applied social research methods*, *2*(254-282).

Yin, R. K. (2012). Case study methods. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 141-155). American Psychological Association.

Yin, R. K. (2013). Validity and generalization in future case study evaluations. *Evaluation*, *19*(3), 321-332.

Zembal-Saul, C., Munford, D., Crawford, B., Friedrichsen, P., & Land, S. (2002). Scaffolding preservice science teachers' evidence-based arguments during an investigation of natural selection. *Research in Science Education*, *32*(4), 437-463.

Zohar, A., Degani, A., & Vaaknin, E. (2001). Teachers' beliefs about low-achieving students and higher order thinking. *Teaching and Teacher Education*, *17*(4), 469-485.

Zohar, A. (2007). Science teacher education and professional development in argumentation. In *Argumentation in science education* (pp. 245-268). Springer, Dordrecht.

Zügge, T. (2020). Mit moderner Physik zum mündigen Bürger?. In *Kohärenz im Unterricht der Elementarteilchenphysik* (pp. 55-77). Springer Spektrum, Berlin, Heidelberg.

Acknowledgements

The development of the concept of critical thinking in the context of particle physics and all its accompanying research, as well as this dissertation would not be possible without the enormous support of advisors, colleagues, friends, and especially my family.

My gratitude goes to my thesis supervisor, Prof. Dr. Gesche Pospiech, for her continuous support and guidance through this project. Her unfailing encouragement, and also the freedom she gave me to pursue my goals were always greatly appreciated. I also sincerely thank my second supervisor, Prof. Dr. Renee' Schwartz for all her time and constructive feedback.

My special thanks go to Prof. Dr. Michael Kobel who, despite his busy schedule, took the time to help me develop the particle physics content of this dissertation with clear and concise feedback. In this way, my colleague Philipp Lindenau also helped me a lot when it came to discussing the content. In addition, Dr. Monica Dunford gave me a lot of help during her stay in Dresden with her encouragement as well as feedback on the particle physics content of the dissertation. I would also like to thank Dr. Frank Beier for his time for intensive discussion about the data analysis method.

My thanks also go to my colleagues at the Department of Physics Education at TU Dresden for their suggestions regarding content and methods and their generous time for discussions, especially Sascha, Marie, Antje, Kerstin, Wiebke, Erik and Thomas. I would also like to thank Sylvia and Göran for their support and good humor, which made the working atmosphere here very pleasant.

I am very grateful to the teachers for devoting their time to implement the antimatter course in the schools before and also during COVID 19, and to all the students who participated in my study. I am grateful for the opportunity provided by the "IKTP Stammtisch", which put me in contact with physics teachers at the regular meetings at the Institute of Nuclear and Particle Physics, and I would like to thank Uta, Michael and Philipp for their great support.

For the support in data processing and quality assurance of the evaluation, my thanks go to Nico, and students Eva, and Tom. I would also like to thank Melanie Kaiser (SLUB), whose guidance on word processing helped me greatly in presenting this dissertation in this format.

I am very grateful for the support from the European Social Fund (ESF) and Free State of Saxony for funding the project. Furthermore, I would like to thank the Graduate Academy of TU Dresden, and the ERASMUS Staff Mobility, which funded my travel to conferences and visits to other educational researchers to discuss my work and expand my vision. In this regard, I would like to give special thanks to Prof. Dr. Susan McKenney, Dr. Nienke Nieveen, and Dr. Arthur Bakker, who gave me constructive feedback in a pleasant Dutch atmosphere.

I wish to give my thanks to my parents, Tahereh and Gholamreza, for their lifelong support and backing. Sadly, my father has already passed away, but I am sure that his soul sees my success and is proud of me. A big thank you also goes to my lovely sisters Mahnaz and Behnaz, who have always been very supportive and encouraging, especially during the difficult times of homesickness. I also thank my friends Laura, Nasim, and Maedeh for their endless support and belief in me.

To my ray of sunshine, Iliya, your understanding, patience and encouragement during these years make you a boy who stands out for his great maturity. Thank you, you are truly extraordinary!

To the love of my life, Shahram, your unconditional love and encouragement have been a constant source of inspiration and learning. I am infinitely grateful for your support in leaving your homeland and helping me follow my dreams in Germany. This work would not have happened without you. I thank you with all my heart and dedicate this book to you.

Statement of Authorship

I, Farahnaz Sadidi, hereby certify that I have authored this PhD thesis entitled *Development and evaluation of a teaching unit in particle physics to promote students' critical thinking* independently and without undue assistance from third parties. No other than the resources and references indicated in this thesis have been used. I have marked both literal and accordingly adopted quotations as such. There were no additional persons involved in the intellectual preparation of the present thesis. I am aware that violations of this declaration may lead to subsequent withdrawal of the degree.

Dresden, 12/12/2022