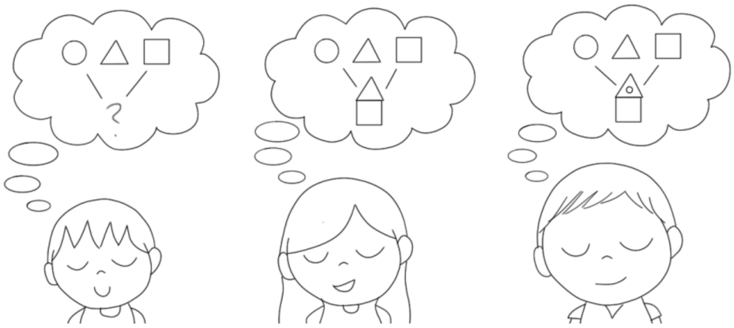
UNIVERSITY OF TWENTE.

# Exploring the Black Box of Critical Thinking: an Investigation into the Development among Elementary School Students




**Master thesis**

Merle Teders
Master Educational Science and Technology
Faculty of Behavioural, Management and Social Sciences
University of Twente
Enschede, the Netherlands

**1st Supervisor:** Dr. Tessa Eijsink
**2nd Supervisor:** Dr. Hannie Gijlers
**Date:** 13-03-2019

## Acknowledgement

I use this page to honour the people that supported me in all kinds of ways to complete this study. It would not have been possible to complete this master's thesis without the help and support of those special and important people in my life. First of all, of course, I would like to thank my supervisors Tessa and Hannie for their support, ideas and feedback. Especially at the start of the process, where this study was only a vague idea, founded by my own big ideas, Tessa and Hannie were a great help with narrowing down my ideas to a nice topic I was able to work with. Even though not all our meetings were equally nice or pleasant, I learned a lot from every single one of them.

Secondly, I would like to honour the help and support in other contexts, received from my boyfriend, family and friends. Even though the process of writing a thesis is (mostly) not their cup of tea, they did help me in other ways. Listening to my complaints about this process when times were tough, or telling me to start working on my thesis when 'studie ontwijkend gedrag (study evasive behaviour)' looked quite appealing! And although they might not feel as having been a great help, they sure were, especially on the moments where the road was really bumpy. Their objective view and sober eyes on the project, sometimes made it easier for me to take the turn to the smooth road.

It is safe to say that this process of writing the thesis, often did not go as planned. One of the benefits of this, is that I learned a lot about myself, my strong points, and weaknesses. Several classmates and students told me I was nuts when I was working on this study; defining my own project, writing my own instrument and conducting it 83 times. I must admit, that I have asked myself this question plenty of times as well, but in the end, I am glad I did it this way. It may have led to a difficult process, but it also made that I could investigate a topic I like, which shows quite a lot of my personality as well.

All in all, I am very proud looking at this document, because it displays a great deal of me. I hope you enjoy reading it as much as I do.

## Abstract

CT is a highly valued skill in the current information age. Therefore, schools are expected to teach CT during education. However, to successfully integrate a skill in education, a deep understanding of the skill is necessary. The level of understanding of CT is currently a black box regarding elementary school students. Therefore, the aim of this study was to enhance knowledge and understanding of CT development among elementary school students. To conduct this study, first CT was conceptualized targeting children. Then, an assessment tool was developed, based on our conceptualization. Eighty-three students, divided over three grades: 1$^{st}$ grade (n = 32), 3$^{rd}$ grade (n = 26), and 6$^{th}$ grade (n = 25), participated in this cross-sectional mixed-methods study. The age of the students ranged between six and thirteen. Results show statistical differences of CT ability during elementary education. Findings show that 1$^{st}$ graders were able to solve/start reasoning in a critical way, and employ some CT skills when solving cases, whereas 6$^{th}$ graders were able to use a wide variety of CT skills. More qualitative analyses revealed that differences arose concerning the used lines of reasoning of the students. The findings of this study indicate development in CT during elementary education and this knowledge can be used for the development of effective CT instruction for elementary education.



**Keywords:** critical thinking, elementary education, reasoning, mixed-methods, development

## Table of Contents

**Critical Thinking Ability during Elementary Education**

Critical thinking (CT) attracted much attention over the past decades and is viewed as "reasonable reflective thinking focused on deciding what to believe or do" (Ennis, 2015, p. 32). A high number of scholars provided insight in the added value of CT during the eighties. During this period, these scholars were often referred to as the "CT movement" (Ennis, 1993; Kuhn, 1999; Lipman, 1987; Paul, 1984). Currently, still a high number of scholars are investigating CT, whom specifically stress the importance of CT in the current information age, proven by several benefits (Doherty, Hansen, & Kaya, 1999). This added value of CT in our current information age caused CT to be categorized as a 21st century skill (Dede, 2010; Halpern, 2003; Kuhn, 1999; Lai, 2011; Partnership for 21st Century Skills, 2009; Saavedra & Opfer, 2012; Thayer-Bacon, 2000; Voogt & Roblin, 2010). CT is highly valued by employers, which caused CT to become a requirement for a lot of jobs (Harvey, Moon, Geall, & Bower, 1997; Kivunja, 2014). Scholars stress that CT is a widely applicable skill (Facione, 1990; Kuhn, 1999), which is associated with several benefits.

One benefit of CT is that people who are skilled in CT identify unreliable sources more easily compared to people unskilled in CT (Browne, Freeman, & Williamson, 2000; Cottrell, 2011; Dede, 2010). Especially in the information age where fake news can be found everywhere CT is a convenient skill to possess, because it helps identifying fake news. A second benefit of CT is that it is associated with several other skills. According to Cottrell (2011), an improve in CT performance positively affects the performance of skills, such as attention and observation, on-task focus, and analytic skills. Third, regarding education, it was found that people with CT skills performed better and for example, achieved higher grades (Facione, 2015).

This higher performance in education caused by CT, combined with the categorization of CT as a 21st century skill, led to the general accepted idea that CT is an important skill that should be taught at school (Dierking & Falk, 2016; Facione, 1990; Florea & Hurjui, 2015; Halpern, 1993, 2003; Kettler, 2014; Kuhn, 1999; National Research Council, 2012; Paul & Elder, 2005; SLO, 2017). Before the categorization of CT as a 21st century skill, CT was only part of the curricula of higher education (Halpern, 2003; Pithers & Soden, 2000). Therefore, elementary and secondary education did not provide CT instruction, which implies that CT instruction materials need to be developed. To develop effective instruction materials, educationalists agree that it is important to obtain a high level of understanding of a skill (e.g., age CT skills start to develop), which allows for aligning instruction with the cognition of students (Halpern, 2003; Kennedy, Fisher, & Ennis, 1991; Kuhn, 1999; Mayer, 2001; Timperley, 2008; Vygotsky, 1978). However, especially regarding elementary school students, CT literature is limited and remains a black box (Kennedy et al., 1991; Lai, 2011). Hence, enhancing knowledge and understanding of CT among elementary school students supports the development of effective CT instruction materials (Kuhn, 1999).

Although CT literature is limited regarding elementary school students, some scholars did investigate concepts similar to CT, such as metacognition or reasoning, and provided evidence for the development of CT (Kuhn, 1999; Lai, 2011). For example, children at the age of three were found able of asking for clarification (e.g., why is that?) according to D'Angelo (1971). Somewhere between three and five years of age, children start to identify expressions of others' point of view (Olson & Astington, 1993). Kuhn (1999) adds that this implies that children of this age can distinguish assertions from information. Moreover, these children are able to recognize correlations which help to find causal relationships (Shultz & Mendelson, 1975). Kuhn (1999) also found that four-year-olds can identify conclusions (although using made up evidence and arguments). Where young children tend to believe that everything is possible, this shifts when they reach the age of six (Kuhn, 1999; Lai, 2011; Perner & Wimmer, 1985). Children, when reaching the age of six, start to use theories to support their answer (Kuhn, 1999). Between the age of seven and ten, children are already aware of people's own motives to manipulate evidence or information (Heyman & Legare, 2005), which implies that they can judge the reliability of certain evidence. Moreover, it is found that young students can benefit from CT instruction as well as other students (e.g., secondary or higher education), when it is levelled to the cognitive development of these students (Ennis, 1993; Gelman & Markman, 1986; Kennedy et al., 1991).

At this point only little is known about the general CT performance of elementary school students and their development. Hence, the present study is a first attempt to provide insights into this matter. The added value of the results of this study are both scientific and practical of nature. Scientific value of this study arises from the contribution to CT literature regarding elementary education, and practical value derives from the data and results of this study, which can be used for developing effective CT instruction.

**Conceptualizing Critical Thinking Children**

To enhance this knowledge and increase insight in the development of CT, a clear and comprehensive conceptualization of CT for children is necessary. Currently, consensus regarding the conceptualization of CT is not reached yet, even though several scholars conceptualized CT (Lai, 2011; Massa, 2014). The complexity of the CT skill is, according to Massa (2014) the cause for the lack of consensus. Especially for children, a comprehensive conceptualization is unavailable, since existing frameworks are focussing on adults (e.g., Bailin et al., 1999; Cottrell, 2011; D'Angelo, 1971; Duron, Limbach, & Waugh, 2006; Ennis, 1962, 1991, 2011, 2015; Ford & Yore, 2011; Foundation for Critical Thinking, 2007; Halpern, 1998, 2003; Lipman, 1987; Paul, 1985; Washburn, 2010). Hence, in the context of this study, we chose to develop our own conceptualization of CT, based on existing frameworks, and focussing on children.

Therefore, in the current study, CT is conceptualized within a framework developed to target children (i.e. elementary school students), based upon some of the leading existing CT frameworks from Bailin et al. (1999), Ennis (2015), and Halpern (2003). These frameworks are chosen because of their popularity (measured by the number of citings), as well as their comprehensiveness (CT included problem till conclusion). For the creation of a new framework, the frameworks needed to be compared. When looking into these frameworks, similarities were discovered based on the elements included in the frameworks of Bailin et al., (1999), Ennis (2015), and Halpern (2003). Similar elements included providing and thinking of arguments, analyzing situations, judging information, conducting reasoning, and drawing conclusions. Differences between the frameworks were found within the structure of the frameworks. The first difference was that Ennis (2015) and Halpern (2003) focused on skills to conceptualize CT, whereas Bailin et al. (1999) conceptualized CT in terms of intellectual resources, which in their study were described as tools to use when conducting CT. Another difference is located in the constructs defined within each framework to categorize the elements of CT (as discussed in the overview of each framework in the following paragraphs).

To overcome the differences between the frameworks, all the elements of CT included in the framework of Bailin et al. (1999) were adapted to skills. Moreover, new constructs were created, so each skill could be categorized into one of the constructs. To create new constructs, the existing constructs and elements were analysed to find common ground. This ensured that the adapted constructs were as similar as possible to the existing constructs without excluding any of the skills of the existing frameworks. Based on this analysis, the following four constructs were identified: (a) analyse problems, (b) gather reliable information, (c) conduct reasoning, and (d) draw conclusions. Afterwards, the skills-list from each of the existing frameworks were rearranged to fit these new constructs. Lastly, to level the framework with the cognitive development of elementary school students, it was necessary to simplify or exclude some of the skills of the original frameworks, such as explanation of existing beliefs and definitions, and using numerical information and understanding diagrams. In Table 1, an overview and comparison between Bailin et al. (1999), Ennis (2015), and Halpern (2003) is made.

**Framework of Bailin et al.** Bailin et al. (1999) defined CT based on three key characteristics of CT, which are (a) “it is done for the purpose of making up one’s mind about what to believe or do”, (b) “the person engaging in the thinking is trying to fulfil standards of adequacy and accuracy appropriate to the thinking”, and (c) “the thinking fulfils the relevant standards to some threshold level” (p. 287). Bailin et al. (1999) consider CT as the sum of “intellectual resources” (as they explain, these are for example inquiry techniques and background knowledge). This implies that one can conduct CT when one is capable of using these intellectual resources, and that CT itself is not a skill. Although Bailin et al. (1999) did not consider these resources to be skills, they are similar to skills in the sense that one can learn how to use and apply these intellectual tools. According to Bailin et al. (1999), a list of skills would be insufficient to conceptualize CT, because it would neglect personality traits that correlate with CT (e.g., open-mindedness). Although this seems a fair point, such a correlation can exist both ways (e.g., open-minded people tend to conduct CT and vice versa).

Table 1

*Comparison of Existing Critical Thinking Frameworks*

| Constructs | Bailin et al. (1999) | Ennis (2015) | Halpern (2003) |
|---|---|---|---|
| Analyse problems | - Identify focus<br>- Ask for clarification<br>- Discover relevant information<br>- Guiding practices of inquiry<br>- Consider plausible alternatives | - Identify focus<br>- Ask/answer questions of clarification<br>- Identify/give explanatory hypotheses<br>- Seeking possible explanations | - Identify the goal<br>- Consider alternative solutions<br>- Recognizing gaps in information<br>- Selecting appropriate strategies<br>- Testing hypotheses |
| Gather reliable information | - Identify existing concepts, beliefs and values<br>- Identify conclusions<br>- Identify motives<br>- Acquire awareness of point of view of one's thinking<br>- Judge reliability of observation reports<br>- Judging credibility of statements<br>- Judge adequacy of moral/legal reasons | - Judge definitions<br>- Identify conclusions<br>- Identify motives<br>- Identify assumptions / equivocations<br>- Report a meaning<br>- Judge observation reports<br>- Judge credibility of a source<br>- Make and judge value judgments | - Recognize semantic slanting<br>- Seek out contradictory evidence<br>- Use numerical information<br>- Determining credibility of resources<br>- Analyzing arguments |
| Conduct reasoning | - Judge deductive reasoning<br>- Judge inductive reasoning | - Identify cause-effect relationship<br>- Deduce, judge deductions<br>- Make/identify generalizations | - Use probability judgments<br>- Conduct reasoning |
| Draw conclusions | | - Consider and reason from earlier gathered information<br>- Express a position on an issue | - Give reasons for choices<br>- Recall and combine relevant information<br>- Synthesize information from several sources |
| Others | - Background knowledge<br>- Make use of heuristics<br>- Habits of mind | - Designing experiments<br>- Integrate abilities in drawing a conclusion | - Generate a reasoned method<br>- Understand basic research principles<br>- Use matrices<br>- Reflect on conclusion<br>- Background knowledge |

Constructs Bailin et al. (1999) included in their framework are: background knowledge, operational knowledge of the standards of good thinking, knowledge of key critical concepts, heuristics, & habits of mind. Interesting is that drawing conclusions was not included as a construct in this framework, even though all the constructs consist of intellectual resources enhancing drawing conclusions. One possibility is that Bailin et al. (1999) considered drawing conclusions as the outcome of using intellectual resources, for which no other intellectual resources are necessary.

**Framework of Ennis.** Ennis (1962) was one of the first scholars investigating CT. With more than fifty years of CT research, he is considered an expert in this field. Ennis developed several CT frameworks and definitions (e.g., 1962, 1991, 2011, & 2015), and provided new information regarding CT ever since he started. The current definition Ennis uses for CT, is “reasonable reflective thinking focused on deciding what to believe or do” (2015, p. 32) and was first mentioned in an article of Ennis from 1991. This definition is adopted in this study, because of its general nature, which captures every aspect of CT. In Ennis his framework (2015) he translated personality traits into skills, divided over different constructs. The constructs in Ennis’s framework (2015) are: basic clarification, bases for a decision, inference, advanced clarification, supposition and integration, & auxiliary abilities. This last construct consists of skills which are not considered CT, but are related to conduct CT (e.g., problem solving and metacognition). Although Ennis pointed out the gap in CT literature regarding children and stressed the necessity of investigating CT among children (Kennedy et al., 1991), he left this topic untouched in his future research.

**Framework of Halpern.** According to Halpern (2003), the goal of CT is to use skills that increase the chance of reaching a desired outcome. Halpern (2003) defined CT as follows:

> CT [...] is used to describe thinking that is purposeful, reasoned, and goal directed – the kind of thinking involved in solving problems, formulating inferences, calculating likelihoods, and making decisions, when the thinker is using skills that are thoughtful and effective for the particular context and type of thinking task. (p. 6).

Halpern (2003) did not only conceptualize CT as a list of skills, but also questions to guide the thinking process. The main questions Halpern (2003) identified, are: “what is the goal?”, “what is known?”, “which skills will lead me to my goal?”, and “have I reached my goal?”. These questions are aligned with the defined skills in Halpern’s (2003) framework. Unlike other scholars, Halpern (2003) describes CT as a process which starts with the identification of a goal and ends with evaluating one’s conclusion.

**Proposed Framework** The identified constructs in Table 1 served as a base for the creation of our own framework in which the three discussed frameworks are combined. The construct *others* is a summary of elements that were included in the existing frameworks and are skills of other variables, such as metacognition or problem solving that are not relevant in the context of the proposed framework and study. The skills approach is adopted in this study, because the results can support the development of effective CT instruction by providing insights in CT development. Skills that were mentioned in at least two of the three frameworks were included in our framework, under the condition that they were relevant for elementary school students. As an outcome of the analysis of Table 1, Table 2 provides an overview of the adapted framework, constructs and constituent skills, which sets the boundaries of this study (what is CT, and what is not).

In line with Halpern (2003), CT is seen as a process in this study, more specifically a stagewise model which entails that one can only enter a next phase after completing the current one. Figure 1 displays the CT process as proposed in this study, as well as the difference between CT and non-CT. Each construct with its constituent skills will be discussed in the following paragraphs (skills are presented in italic font). An important note is that not all the skills are necessary for every problem. For example, if the only evidence found can be used to conduct deductions, one does not always need to conduct inductions regarding that specific problem (Halpern, 2003).

***Analyse problems.*** Analyse the problem helps to understand the problem, as well as thinking of ideas for a possible conclusion. A thorough analysis of problems starts with *identifying the focus* of the problem. Subsequently, it involves *considering alternatives*, which prevents closed-mindedness, and *creating hypotheses*, based on the information derived from the analysis combined with the most likely alternative. The hypothesis guides the CT process in the next phases. Hence, *consider alternatives* can be broader, because they do not need to be based on information derived from the problem (Bailin et al., 1999; Ennis, 2015; Halpern, 2003). This also entails *identifying gaps in problems*, to build a complete and clear picture of the problem, as well as to guide the CT process for the following phases.

***Gather reliable information*** After the analysis of a problem, a person conducting CT can start to gather reliable information, derived from several types of sources. Although personal communication often serves as a type of source as well, such information is not always reliable. Therefore, it is important that the receiver, considers the mindset of the person providing the information, which can be done by *identifying others' point of view*, *identifying motives* or *identifying opinions.* The mindset of a person can trouble the reliability of information he provides. Hence, CT subsequently consists of judging the reliability of a source (e.g., an opinion is less reliable than a fact). A person conducting CT should gather information derived from as many reliable sources as possible. Therefore, *judging reliability of information* and *observations* (e.g., are my/others' observations true?) is an important part of CT. To correctly judge the reliability of sources, it is important to be aware of the source. For example, observations can be subjective or biased. Identifying reliable sources is necessary to conduct reliable reasoning, which leads to a reliable conclusion (Bailin et al., 1999; Ennis, 2015; Halpern, 2003).

***Conduct reasoning.*** Reasoning forms the base of the final conclusion in the CT process, because the outcome of reasoning is the conclusion. Therefore, a great amount of CT involves reasoning (Bailin et al., 1999; Ennis, 2015; Halpern, 2003). Because reasoning is combining sources, one can only conduct reasoning after one has gathered information. Several types of reasoning exist, such as *conduct deductions*, *conduct inductions*, and *establishing causal relationships*, which all can be helpful to conduct CT. However, when the reasoning leads to inconclusive results, the person conducting CT needs to find additional reliable sources and conduct (other types of) reasoning again. This implies the possibility of going back and forth in the aforementioned construct and this construct.

***Draw conclusions.*** After passing the aforementioned phases, the person conducting CT can draw his own conclusion. To draw the best conclusion (e.g., most reliable, or complete), a person needs to found their conclusion by *recalling and combining sources* gathered during the aforementioned phases, and used to conduct reasoning. Lastly, the person also needs to be able to *state his own conclusion*. Without properly applying skills from the prior phases, the possibility exists that one ends up with a wrong conclusion (e.g. less reliable, or biased) (Bailin et al., 1999; Ennis, 2015; Halpern, 2003).

Table 2

*Adapted Critical Thinking Framework, Constructs & Constituent Skills*

| |
|---|
| Analyse problems |
| Identify the focus |
| Consider alternatives |
| Create hypotheses |
| Identify gaps in problems |
| Gather reliable information |
| Identify motives |
| Identify opinions |
| Identify others' point of view |
| Judge the reliability of information |
| Judge the reliability of observations |
| Conduct reasoning |
| Conduct deductions |
| Conduct inductions |
| Establish causal relationships |
| Draw conclusions |
| Recall & combine sources |
| State own conclusion |

**Figure 1.** Comparing Critical Thinking Process, and Non-Critical Thinking Process

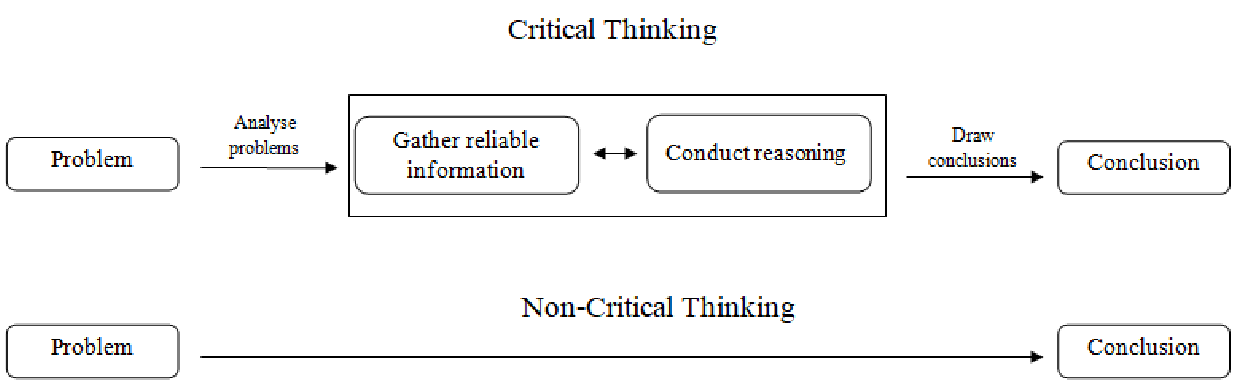



*Figure 1.* This model shows how critical thinking is a process that exist of different phases. It also shows that one needs to go through each phase, and might need to go back and forth between *gather reliable information* and *conduct reasoning* (due to inconclusive reasoning). The difference between critical thinking and non-critical thinking is also visualised.

## Current Study

As discussed, CT literature regarding elementary school students is necessary for developing effective CT instruction, but current state of literature is limited. Available literature regarding CT in elementary school students derives from scholars who investigated one aspect of CT (e.g., reasoning) or investigated children of a certain age (e.g., only six-year-olds). Studies investigating the general CT ability of elementary school students during elementary education are non-existent, to our knowledge. Hence, the aim of this study is to provide insight in this matter. This information can be used for the creation of effective CT instruction for elementary education. Since CT is not taught as a subject in elementary education, this study investigates the natural differences in CT ability. Therefore, we seek to answer the following question in this study: To what extent does CT ability develop during elementary education? Answering this question demands a cross-sectional mixed-method approach that on the one hand provides data which shows CT ability of elementary school students from different grades, and on the other hand provides insights in differences in these ability levels. These results indicate CT development among elementary school students. To answer this question, four sub-questions are formulated.

a) To what extent does CT ability, constructs and constituent skills, improve during elementary education, comparing students of different grades?

b) To what extent does the performance of CT, constructs and constituent skills, differ concerning the pace in which these improve?

c) To what extent do differences occur between grades regarding CT processes, considering line of reasoning and argumentation?

d) To what extent is the CT framework and model, as developed for this study, similar to the existing literature?

CT is described as a process (see Figure 1), which implies a relationship between the performance on the different constructs and phases. This model is based on the existing literature of Bailin et al.

(1999), Ennis (2015), & Halpern (2003). However, since in this study, the participants are children, a confirmation of this model is necessary, because this the model might be different for children.

To provide answers for the other questions, CT will be assessed among elementary school students of different grade levels, which allows for gaining insights within and between grades. The nature of sub-question c is qualitative, because it demands in-depth information regarding the content of the answers students provide. Therefore, a mixed-method approach will be applied. To be able to generate criteria that describe CT, CT will be operationalized according to our proposed framework, as provided in Table 2. In this study, CT will be assessed by three tasks for which the student needs to draw conclusions based upon provided information. Due to this study being a master thesis project, with a certain time restriction, a cross-sectional study design will be used, assessing 1$^{st}$ till 6$^{th}$ grade students.

## Method

### Participants

All participants were students attending the same elementary school, which was chosen by the researcher by means of convenience and purposive sampling. Inhabitants of the region in which the school is located are mostly native Dutch. The school can be considered middle-sized, based on the amount of enrolled students. CT was not explicitly taught as a subject in the participating school. In total, 83 students participated, divided over three grades levels: 1st grade (n = 32), 3rd grade (n = 26), and 6th grade (n = 25), ranging from 6 to 13 years ($M_{age}$ = 8.93 years). Table 3 provides a complete overview of students' characteristics (gender and age). Dutch was the mother tongue of all participating students. No compensation was provided to the students. Parental consent was provided for all students prior to participation in this study.

Table 3

*Students' Characteristics*

| | Total | Grade 1 | Grade 3 | Grade 6 |
|---|---|---|---|---|
| Students | 83 | 32 | 26 | 25 |
| Boys | 35 | 13 | 14 | 8 |
| Girls | 48 | 19 | 12 | 17 |
| Age *M(SD)* | 8.93(2.14) | 6.75(.44) | 8.88(.65) | 11.76(.52) |
| Minimum | 6 | 6 | 8 | 11 |
| Maximum | 13 | 7 | 10 | 13 |

### Critical Thinking Assessment Elementary Education

Current existing CT assessments are not suitable for children under the age of ten, and do not cover CT as conceptualized in Table 2 (Ennis, 1993; Lai, 2011). For this study, an assessment tool was developed that allowed for a mixed-method approach (Critical Thinking Assessment Elementary Education, CTAEE). The tool should both, assess CT performance and provide insights in the CT process of the students.

**Structure of CTAEE.** The CTAEE consisted of three tasks which each revolved around a case. Each task assessed the CT framework described in Table 2. This led to a total of 14 items per task, respectively four items for *analyse problems*, five items for *gather reliable information*, three items for *conduct reasoning*, and two items for *draw conclusions*. Each task started with a description of the case and was followed by a series of questions which assessed and captured the CT process of the students (Ku, 2009; Kuhn, 1999). Considering the age of the 1st and 3rd graders, the researcher read out the CTAEE to all participants, precisely following the developed script. By reading it out, differences in reading skills were overcome. The CTAEE and its script were written in Dutch, because this was the mother tongue of the researcher and the students.

**Script.** A script was written which included the wording of the formulations that the researcher used to introduce and lead the task. The session started with the introduction of the CTAEE itself, where after each of the tasks were treated. Each task then started with an introduction of the case it revolved around. In this way, an equal procedure for each student was ensured. This script included additional explanatory materials as well.

***Introductory text.*** The session started with an introduction of the CTAEE tool. Students received information about the assessment, including the expected time span, that it was recorded, and what the procedure of the task was. Students got asked if they had any questions. Students did not receive information concerning the goal of the study.

***Tasks and Questions.*** After the introductory text, the assessment itself started, which consisted of three tasks. The three tasks of the instrument were parallel to each other, regarding their structure and elements (e.g., questions and evidence). Each task consisted of a case and was introduced by a short description of this case. This case description contained enough information regarding the case to understand the main problem, but left some blank spots, which the students had to fill or receive information about along the way. Thereafter, a series of questions regarding the case were asked, one by one, starting with identifying the focus of the case. The case of the first task involved a lost fidget spinner where the student had to find out what had happened with the fidget spinner (referred to as *fidget spinner-task*). The second task focused on a classmate claiming to have a chocolate allergy and refused to eat a treat, where the student had to judge, based upon provided evidence, if the classmate actually had this allergy (referred to as *allergy-task*). The last task focused on a new classmate with a harelip, where the student was questioned about their acceptance of this classmate, as well as their ideas of the integration of him in their class (referred to as *harelip-task*). Each of these tasks are based on a social issue to obtain data regarding the general CT performance and to minimize the influence of background knowledge among the youngest students (Bailin et al.,1999; Halpern, 2003; Lai, 2011).

The order in which the CT skills were assessed differed per task. On specific moments during the task, additional information concerning the case was included in the script as part of a question to ensure uniformity in the moment a student received certain information. Not all of the received information was reliable, and students had to judge the reliability of the provided information. Moreover, some questions included follow-up questions, for questions where students could provide two or more answers. Examples of questions and the skills assessed by these questions, are shown in Table 4.

***Additional explanatory materials.*** Concerning the allergy-task and the harelip-task, additional explanatory materials were included in the script. These materials were included to overcome biases caused by a lack of background knowledge. Regarding the allergy-task, the student got asked if he knew what an allergy was. When a student did not knew, additional explanatory material consisting of a short description of an allergy was read to the student. Regarding the harelip-task, a student got asked if he knew what a harelip was. Explanatory material consisted of a picture of a harelip which was provided to the student when he did not knew what a harelip was. In the script, the moment on which the researcher would ask this question, was integrated in the script, to ensure that each student received this information at the same moment.

Table 4

*Example Questions Task, plus Assessed Skill*

| Case and information | Questions from the case | Assessed skill |
|---|---|---|
| Fidget spinner-case: Tom brings his fidget spinner to | What could have happened to the fidget spinner? | Consider alternatives |
| school, but during the break, he discovers that it is gone. | What do you think has happened? And why? | Create hypotheses |
| Another classmate has the exact same fidget spinner, but says she got it during the weekend, although she seems suspicious. | What else do you need to know, to find out what happened with the fidget spinner? | Identify gaps in problems |
| Allergy-case: You brought treats to school, because it is | What could, according to Jan, be a reason for Inge to refuse to eat the cupcake? | Identify motives |
| your birthday. When you want to hand out a chocolate cupcake to Inge, she refuses it, | Imagine Inge tells Jan about her allergy for chocolate. Would Jan change his mind about Inge's behaviour? And why? | Identify others' point of view |
| and says she cannot eat it, due to her allergy. Jan thinks Inge is just behaving appallingly, when she can, again, not eat something. He does not believe Inge. | You want to know more about chocolate allergy. You come across a webpage which says that chocolate is bad for everyone and that everybody gets sick from eating chocolate. Do you think that this is true, and why? | Judge the reliability of information |
| Harelip-case; Stijn is a new | Why doesn't Bart want Stijn to participate? | Identify motives |
| classmate and he has a harelip. During the break, Stijn wants to play soccer with you and | Bart states that everyone with a harelip is weird. Assuming that Bart is right, does that make Stijn weird as well? And why? | Conduct deductions |
| Bart. Bart does not want Stijn to participate. | What could you do, to make Bart like Stijn? | Establish causal relationships |
| Later in the case, Bart and Stijn had several nice moments together, playing | You want to know if Bart changed his opinion about Stijn. Do you think Bart changed his mind? | State own conclusion |
| soccer and other games. | What could have caused Bart changing his mind about Stijn? | Recall & combine sources |

*Note.* These examples are translated from Dutch to English. Therefore, it could have happened that the phrasing in the table differs from the original questions. Moreover, not all evidence is provided in these examples to answer each question.

**Pilot test.** A pilot test was conducted that provided insights in the cognitive level of the participants and helped to improve the instrument. Six students conducted the task, aged between five and twelve years. Then, the instrument was adapted based upon questions that arose during the assessments and observations of the researcher that indicated which questions were too difficult for the target group. Thereafter, a second round of pilot tests were conducted, where another six students conducted the task (aged between six and twelve years). Only textual changes were carried out, using easier synonyms for example, which led to the final instrument.

Cronbach's alpha analysis of the tasks, revealed that one item of the allergy-case negatively affected the internal consistency of the skill *state own conclusion*, as part of *draw conclusions*. Excluding this items improved the alpha from .59 to .63. Hence, this item was excluded from further analyses. This exclusion led to only two items assessing the skill *state own conclusion* and, therefore, was left out of further analyses. The total amount of items became 41, and the allergy-case consisting of 13 items. The internal consistency of the instrument was .89. The calculated internal consistency of *conduct reasoning* ($\alpha$ = .54) is on the low end. Since CT is a process, the internal consistency could be affected by the performance of the students on earlier phases. Table 5 provides an overview of the exact levels of internal consistencies. The Pearson correlation between the different cases and constructs of the task can be considered fair (Cicchetti, 1994) (see Tables 6 and 7).

Table 5

*Calculated Internal Consistency, as Measured by Cronbach's Alpha*

| | n items | α |
|---|---|---|
| Overall | 41 | .89* |
| Fidget spinner | 14 | .70 |
| Allergy | 13 | .69* |
| Harelip | 14 | .78 |
| Analyse problems | 12 | .78 |
| Gather reliable information | 15 | .70 |
| Conduct reasoning | 9 | .54 |
| Draw conclusions | 5 | .63* |

*Note.* * after deletion of one item.

Table 6

*Correlation Coefficients between Constructs, as Calculated with Pearson's Analysis*

| | Analyse problems | Gather reliable information | Conduct reasoning | Draw conclusions |
|---|---|---|---|---|
| Analyse problems | - | .69 * | .53* | .63* |
| Gather reliable information | - | - | .59* | .66* |
| Conduct reasoning | - | - | - | .70* |
| Draw conclusions | - | - | - | - |

*Note.* * = significant correlations (p <.001 for each).

Table 7

*Correlation Coefficient between Cases, as Calculated with Pearson's Analysis*

| | Fidget spinner | Allergy | Harelip |
|---|---|---|---|
| Fidget spinner | - | .69* | .69* |
| Allergy | - | - | .74* |
| Harelip | - | - | - |

*Note.* * = significant correlations (p <.001 for each).

**Procedure**

The data was gathered between the 14th and 25th of may 2018. The students completed the CTAEE in one session during school hours. Data gathering took place at the students' school, for which an empty and quiet room was offered. The researcher led the sessions herself and were recorded with a voice recorder. The students completed the task individually and the sessions took around 20 minutes per participant.

**Research design.** In this study, a mixed method study with a cross-sectional design was used, including three units of analysis, respectively the three grades. This design, combined with the used instrument, allows for comparing grades, but simultaneously gaining insights in the CT processes of the students as well. The independent variable in this study was the grade of the students, whereas the dependent variable were students' performance of CT and its constructs and constituent skills. The three tasks were administered according a randomized complete counterbalanced design to minimize sequence effects (e.g., tiredness or lack of attention). There were six possible orders in which the cases were proposed, respectively 1-2-3, 1-3-2, 2-1-3, 2-3-1, 3-1-2, 3-2-1. All these orders were used in similar rates which resulted in each order being used 13 or 14 times.

**Transcribing and Coding**

The CTAEE assessed student's CT performance and captured the CT process. The recordings of the voice recorder were transcribed verbatim by the researcher and imported in ATLAS.ti. Since this study demands a mixed-method approach, a codebook was developed including two codes with specific characteristics; *skill-codes* for the quantitative part and *argument-codes* for the qualitative part. Per question, several skill- and argument-codes were developed and included in the codebook.

**Skill-codes.** The purpose of these codes was to label answers as correct or incorrect, which provides insights in a student's CT performance of the assessed CT framework (as described in Table 2). Determining an answer was correct or incorrect was based on whether or not the student showed the required skill. Based on the framework (described in Table 2), criteria were written per question what answers received which codes. In the coding scheme, a description was provided to describe the type of answers students had to provide to receive a certain code, accompanied with an example answer. An example of skill-codes is provided in Figure 2.

**Figure 2.** Example Skill-codes.

**Question 1.07. Would Bram talk differently about what happened when he would like Tom? And why? (Use codes 1.07 of the code list)**

**Skill-codes:**
1.07 Understands influence of relationship: assign this code when the student provides an answer which shows that the student does understand the influence of the relationship. *Example answer: 'Yes because you don't talk like that about a friend.'*
1.07 Does not understand influence of relationship: assign this code when the student provides an answer which shows that the student does not understand the influence of the relationship. *Example answer: 'No because Bram is a bully.'*

*Figure 2.* This figure shows an example of possible skill-codes for an item.

**Argument-codes.** Argument-codes served the purpose of capturing the CT process of the students, which is based on their use of arguments. In contrast with the skill-codes these codes are used to gain insights in the CT process of students, instead of assessing students. Therefore, the development of these codes was data driven, and based on received answers, more specifically the argumentation students used. The developed codes categorized the most common types of argumentation per question. Answers that did not belong to the common types of argumentation, got labelled *others*. An example of argument-codes is provided in Figure 3.

**Coding procedure.** The coding was carried out by two coders. The researcher, and first coder, included the code list, including skill- and argument-codes in ATLAS.ti. Then, the codebook, including the skill-codes and argument-codes, was explained to, and discussed with the second coder, till both coders reached mutual understanding and agreement. Because of the script that was used to conduct the CTAEE, only the answers provided by students needed to get coded. The first coder coded each transcript (n = 83) and the second coder coded 21.69 % of the transcripts (n = 18).

**Figure 3.** Example Argument-codes.

**Question 1.07. Would Bram talk differently about what happened when he would like Tom? And why? (Use codes 1.07 of the code list)**

**Argument-codes:**
1.07 Answer: be kind/helpful: assign to all answers in which the student mentions that he/she would show positive behaviour. *Example answer: 'Yes than I would help searching.'*
1.07 Answer: could lost: assign to all answers in which the student mentions that he/she thinks the fidget spinner got lost. *Example answer: 'No Tom just lost it.'*
1.07 Answer: others: assign to all answers in which the student mentions something different than the aforementioned codes. *Example answer: 'You will understand when you talk..'*

*Figure 3.* This figure shows an example of possible argument-codes for an item.

The coding procedure was two-staged, due to the quantification of the skill-codes. The first stage, served to provide every answer with a skill- and argument-code. In the second stage, the coders had to value the answers in SPSS in line with the received skill-code. Possible values were *0* and *1*, where *0* was provided to incorrect answers, and *1* was provided to correct answers.

The CTAEE's maximum score was 44 points, divided over the constructs; *analyse problems* (max. = 15), *gather reliable information* (max. = 15), *conduct reasoning* (max. = 9), and *draw conclusions* (max. = 5). Each skill got assessed three times, and students could receive a maximum of three points for each skill (with the exception of *state own conclusion*, due to the deletion of one of the assessing questions). The skill *considering alternatives* is an exception, because the student could receive two points for items assessing this skill, respectively *0* when a student failed to mention one alternative, *1* when a student mentioned one alternative, and *2* when a student mentioned two or more alternatives. This exception was chosen because a difference in quantity equalled difference in performance of this skill.

Analyses of the quantified skill-codes, included statistical analyses in SPSS (to compare scores within and between grades), whereas analyses of the argument-codes, included calculation of frequency-tables in ATLAS.ti. Figure 4 shows an example of the elements of the coding scheme. The inter-coder agreement concerning the skill-codes, reached .90 (Cohen's Kappa) and .99 for the argument-codes, as measured with Krippendorff's α, which can both be considered high.

**Figure 4.** Example Coding Scheme.

**Question 1.13. After you have heard all of this, what do you think has happened to the fidget spinner of Tom? (Use codes 1.13 of the code list)**

**Skill-codes:**
1.13 Logical conclusion: assign this code when the student provides a logical conclusion, bearing in mind the provided storyline. (In SPSS: value *1*) *Example answers: 'Kim took it during the break.'* or *'Tom lost it.'*
1.13 Illogical conclusion: assign this code when the student provides an illogical conclusion, bearing in mind the provided storyline. (In SPSS: value *0*) *Example answer: 'It was thrown over the hedge.'*

**Argument-codes:**
1.13 Answer: Kim took it: assign to all answers in which the student mentions that he/she thinks Kim took the fidget spinner. *Example answer: 'Kim took it during the break.'*
1.13 Answer: got lost: assign to all answers in which the student mentions that he/she thinks Tom lost the fidget spinner. *Example answer: 'No Tom lost it.'*
1.13 Answer: others: assign to all answers in which the student mentions something different than the aforementioned codes. *Example answer: 'Someone destroyed it, and hid it.'*

*Figure 4.* This figure shows an example of the coding scheme for one item of the task.

# Results

## Relation between Consecutive Constructs of the Critical Thinking Process

Our first principle of interest was to confirm the model as displayed in Figure 1. To confirm the model, the relationship between the consecutive phases needed to be calculated. Regression analyses were calculated to reveal these relations. Significant relations were found between each construct and its follow-up construct ($R^2$ ranging between .35 and .48). Construct *analyse problems* showed a significant relationship with *gather reliable information*, $R^2 = .47$, $F(1, 81) = 72.98$, $p < .001$. *Gather reliable information* was found responsible for the total score on *conduct reasoning*, $R^2 = .35$, $F(1, 81) = 42.91$, $p < .001$. Lastly, regarding *conduct reasoning* and *draw conclusions*, similar results were found, $R^2 = .48$, $F(1, 81) = 75.77$, $p < .001$.

## Critical Thinking Performance, Constructs and Skills

Most of the formulated sub-questions were based on the differences in CT ability of the participants. Table 8 displays the means and standard deviations for each grade, per construct and skill. Differences in overall CT ability between grades were compared using the overall CTAEE scores, calculating an ANOVA (one-tailed) with Bonferroni corrections and showed significant differences between all grades: grade 1-3, $F(2,81) = 43.81$, $p < .001$; grade 1-6, $F(2,81) = 43.81$, $p < .001$; grade 3-6, $F(2,81) = 43.81$, $p < .001$.

To more specifically locate differences between CT performance, MANOVA tests with Bonferroni corrections were calculated for each of the included constructs and constituent skills (see Table 2). Outcomes showed significant differences between the scores of 1$^{st}$ and 6$^{th}$ grade (all $p$s $\leq .05$) for each construct and constituent skill, with 6$^{th}$ grade scoring higher than 1$^{st}$ grade. Table 8 shows an overview of calculated significant differences (see Appendix A for an overview of the between subjects test scores with Bonferroni corrections). The amount of variance in scores related to grade, was calculated by regression analysis. Regression analyses calculated significant predictions for each of the constructs and constituent skills (see Table 9 for an overview). In the upcoming paragraphs, CTAEE score differences within each grade will be discussed. Figures 5 till 7 show the cumulative mean for several constructs, which visualizes the differences in mean scores of each skill. For the construct *draw conclusions* such a table is missing, because only one skill could be investigated within that construct.

Within each grade, pairwise comparisons between every combination of constructs and skills were calculated to identify differences in mean scores. Outcomes of pairwise comparisons of the constructs are included in the text, whereas outcomes of pairwise comparisons of skills are included in Appendix B. These outcomes reveal significant differences in constructs and skills within a grade. Distribution tables were calculated for each construct and skill to provide insight in the variance of students' scores per constructs and constituent skill (provided in Appendix C).

**Grade 1.** The overall mean score of grade 1 for the CTAEE was 47.14 % (*SD* 16.94), which corresponds with 20.66 out of 44 points. Table 8 shows that for most of the skills, students achieve a score of 30 % or higher ($M \geq .30$). Skills with lower performance rates were *identify the focus* ($M = .13$, $SD = .18$) and *identify gaps of the problem* ($M = .23$, $SD = .26$).

***Pairwise comparisons.*** Results of pairwise comparisons between constructs showed that 1st graders performed better on *conduct reasoning* compared to all other constructs: analyse problems – conduct reasoning, $t = -4.25$, $p <.001$; gather reliable information - conduct reasoning, $t = -2.10$, *p* .044; conduct reasoning – draw conclusions, $t = 2.66$, *p* .012. Another difference was found between *analyse problems* and *gather reliable information*, $t = 3.45$, *p* .002. Pairwise comparisons of the skills, see Appendix B showed a difference between *identify the focus* and *identify gaps in problems* and the other skills, with a lower performance for aforementioned skills.

***Performance distribution.*** Distribution tables (see Appendix C) concerning grade 1, show a broad range of scores (some students achieve a lot of points, whereas others achieve scores of zero points). For each of the constructs, over 75 % of the 1st graders scored at least 30 % of the points: analyse problems, 78.1 %; gather reliable information, 78.1 %; conduct reasoning, 90.6 %; draw conclusions, 78.1 %.

**Grade 3.** The overall mean score of grade 3 for the CTAEE was 65.28 % (*SD* 12.64), which corresponds with 28.54 out of 44 points. Table 8 shows that 3rd graders scored more than half of the points ($\geq .50$) for most of the constructs and constituent skills. Skills with lower performance scores, were *identify the focus* ($M = .44$, $SD = .25$), *identify others' point of view* ($M = .45$, $SD = .30$) and *conduct deductions* ($M = .35$, $SD = .24$).

***Pairwise comparisons.*** Pairwise comparisons of the constructs showed that students achieved higher scores for *draw conclusions* than *analyse problems* ($t = -2.30$, *p* .030). Similar comparisons regarding skills (see Appendix B), showed that the mean scores for the skills *identify the focus*, *identifying others' point of views*, and *conduct deductions* were significantly lower than the scores for other skills.

***Performance distribution.*** Distribution tables in Appendix C show that for each of the constructs a big difference was visible between *analyse problems* and the other constructs, when comparing the share of students scoring 50 % or higher: analyse problems, 50 %; gather reliable information, 88.5 %; conduct reasoning, 84.6 %; draw conclusions, 84.6 %.

**Grade 6.** The mean score of grade 6 for the CTAEE was 79.60 % (*SD* 6.01), corresponding with 34.72 out of 44 points. Table 8 shows that 6th graders score 70 % or higher ($M \geq .70$) for each construct and most of the constituent skills, with the exception of the skills *identify others' point of view* ($M = .65$), *judge reliability of information* ($M = .65$), and *conduct deductions* ($M = .52$).

***Paired comparisons.*** Pairwise comparisons of the constructs revealed that 6th graders achieved a higher mean score for *draw conclusions*, compared to scores of other constructs: analyse problems - draw conclusions, $t = -2.10$, *p* .046, gather reliable information – draw conclusions, $t = -2.80$, *p* .010, conduct reasoning – draw conclusions, $t = -2.29$, $p$ .031. Similar comparisons of the skills (see Appendix B), revealed that the mean scores of *create hypotheses* and *conduct deductions* were significantly lower than the scores for other skills.

***Performance distribution.*** Derived from the distribution tables in Appendix C, it was revealed that at least halve of the 6th graders managed to score 80 % or higher for each of the constructs: analyse problems, 60 %; gather reliable information, 56 %; conduct reasoning, 68 %; draw conclusions, 84 %. Moreover, none of the 6th graders scored 50 % or lower for each of the constructs.

Table 8

*Means, Standard Deviations & Differences, Constructs & Constituent Skills*

| | Grade 1 *M (SD)* | | Grade 3 *M (SD)* | | Grade 6 *M (SD)* |
|---|---|---|---|---|---|
| Analyse problems | .38 (.14) | < | .61 (.16) | < | .78 (.11) |
| Identify the focus | .13 (.18) | < | .44 (.25) | < | .77 (.23) |
| Consider alternatives | .56 (.20) | = | .67 (.20) | = | .77 (.16) |
| Create hypotheses | .44 (.32) | < | .68 (.24) | = | .73 (.19) |
| Identify gaps in problems | .23 (.26) | < | .58 (.31) | < | .85 (.22) |
| Gather reliable information | .48 (.20) | < | .66 (.13) | < | .77 (.12) |
| Identify motives | .49 (.27) | < | .65 (.28) | = | .76 (.26) |
| Identify opinions | .62 (.32) | = | .74 (.32) | = | .89 (.21) |
| Identify others' point of view | .39 (.33) | = | .45 (.30) | < | .65 (.33) |
| Judge the reliability of information | .39 (.30) | < | .63 (.27) | = | .65 (.25) |
| Judge the reliability of observations | .53 (.34) | < | .81 (.30) | = | .87 (.19) |
| Conduct reasoning | .55 (.22) | = | .65 (.15) | < | .77 (.11) |
| Conduct deductions | .32 (.22) | = | .35 (.24) | < | .52 (.27) |
| Conduct inductions | .66 (.31) | < | .81 (.19) | = | .91 (.15) |
| Establish causal relationships | .68 (.29) | = | .80 (.30) | = | .89 (.16) |
| Draw conclusions* | .47 (.26) | < | .70 (.22) | < | .86 (.15) |
| Recall & combine sources | .39 (.26) | < | .64 (.27) | < | .81 (.19) |
| Total score | .47 (.17) | < | .65 (.13) | < | .80 (.06) |

*Note.* As calculated by MANOVA tests with Bonferonni corrections. A difference was calculated for each assessed construct and skill, between grade 1 and 6. < means significant difference; = means no significant difference; * due to the exclusion of one of the items for further analyses, the skill *state own conclusion* cannot be included for further investigation.

Table 9

*Calculated Influence Grade on Critical Thinking, Constructs & Constituent Skills*

| | Constant | Grade | $R^2$ | *F* | *p* |
|---|---|---|---|---|---|
| Total score | .31 | .16 | .52 | 87.92 | <.001 |
| Analyse problems | .19 | .20 | .59 | 117.71 | <.001 |
| Identify the focus | -.20 | .32 | .61 | 124.83 | <.001 |
| Consider alternatives | .46 | .11 | .18 | 17.69 | <.001 |
| Create hypotheses | .31 | .15 | .19 | 18.43 | <.001 |
| Identify gaps in problems | -.07 | .31 | .50 | 80.25 | <.001 |
| Gather reliable information | .35 | .14 | .36 | 44.80 | <.001 |
| Identify motives | .36 | .14 | .15 | 14.67 | <.001 |
| Identify opinions | .47 | .14 | .14 | 13.17 | <.001 |
| Identify others' point of view | .24 | .13 | .11 | 9.59 | .003 |
| Judge the reliability of information | .28 | .14 | .15 | 14.09 | <.001 |
| Judge the reliability of observations | .39 | .17 | .20 | 19.62 | <.001 |
| Conduct reasoning | .44 | .11 | .23 | 23.67 | <.001 |
| Conduct deductions | .21 | .10 | .10 | 8.74 | .004 |
| Conduct inductions | .54 | .13 | .17 | 16.19 | <.001 |
| Establish causal relationships | .57 | .11 | .11 | 9.97 | .002 |
| Draw conclusions | .28 | .20 | .37 | 47.26 | <.001 |
| Recall & combine sources | .18 | .22 | .36 | 44.97 | <.001 |

**Figure 5.** This figure shows mean scores per grade for each skill within the construct *Analyse problems*

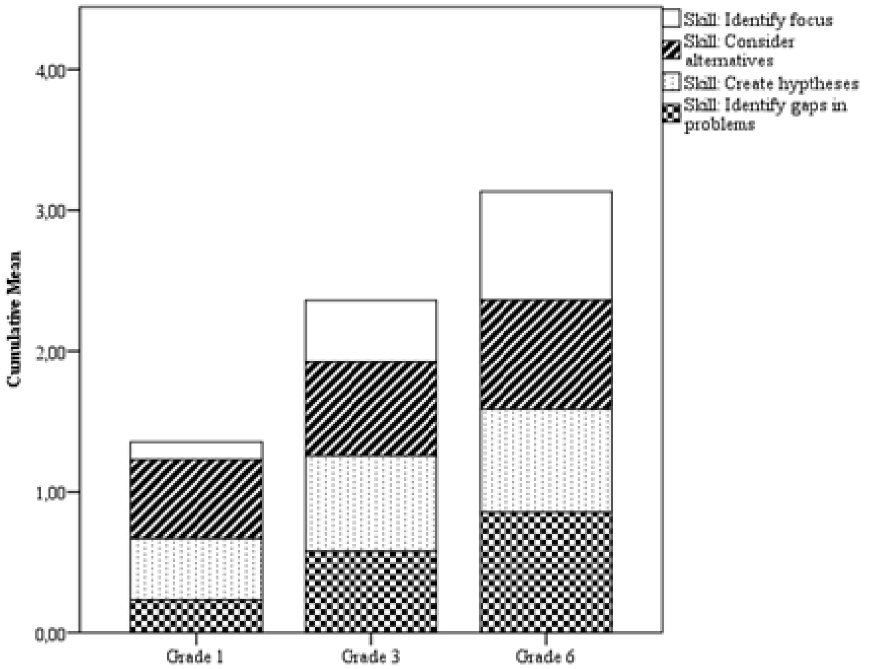



*Figure 5.* Y-axis shows the cumulative score for *Analyse problems*, X-axis shows this score per grade.

**Figure 6.** This figure shows mean scores per grade for each skill within the construct *Gather reliable information*

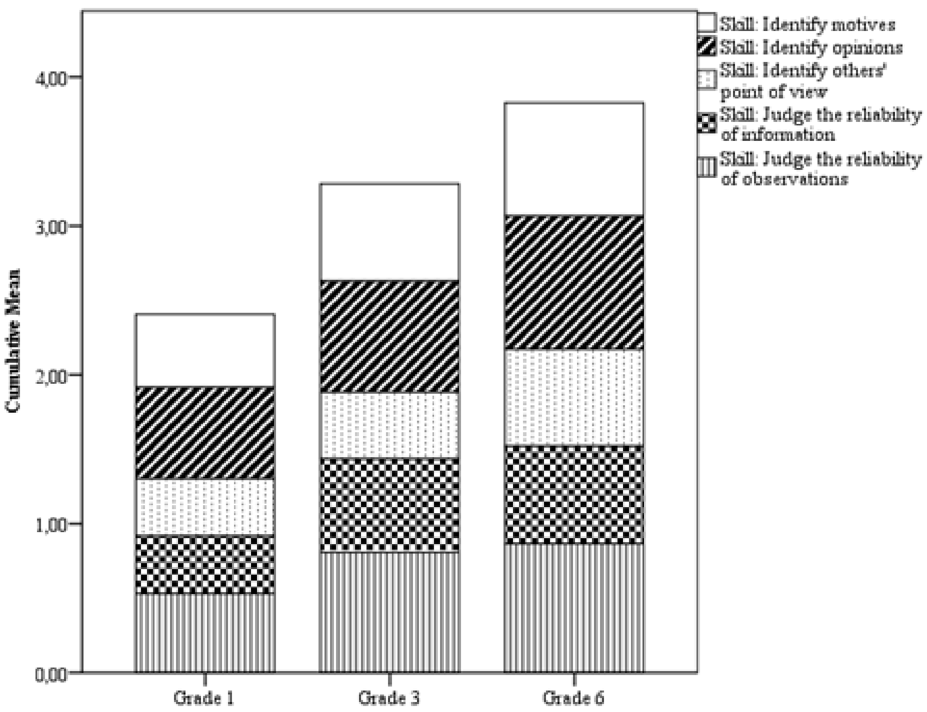



*Figure 6.* Y-axis shows the cumulative score for *Gather reliable information*, X-axis shows this score per grade.

**Figure 7.** This figure shows mean scores per grade for each skill within the construct *Conduct reasoning*

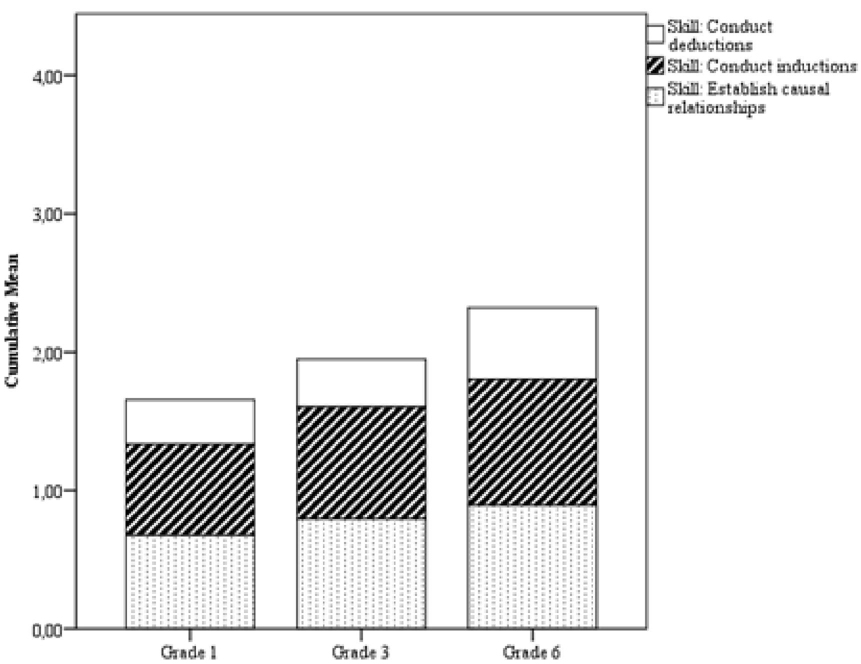



*Figure 7.* Y-axis shows the cumulative score for *Conduct reasoning*, X-axis shows this score per grade.

**Students' Line of Reasoning and Argumentation**

To gain insight in students' line of reasoning and argumentation, frequency tables were calculated and analysed in ATLAS.ti. The answers of the students were the input for these analyses and show students' use of information, which can identify differences in argumentation, and explain calculated findings of quantitative analyses. A certain level of background information concerning the theme of the three tasks included in the CTAEE is necessary to understand the results. The first task involved a lost fidget spinner where the student had to find out what had happened with it (was it stolen or lost?). The second task addressed a classmate who claimed to have a chocolate allergy and refused to eat a treat, where the student had to judge, based upon provided evidence, if the classmate actually had this allergy. The last case focused on a new classmate with a harelip, where the student was questioned about their acceptance of this classmate, as well as their ideas of the integration of him in their class.

**Grade 1.**

***Line of reasoning.*** Regarding the fidget spinner-task, most students made a shift between their hypothesis and conclusion. At the start of the task, 53.13 % of the students considered the fidget spinner getting lost as an alternative, although only one student drew this conclusion. One student (3.13 %) identified a suspect in the task when identifying the focus, whereas in the end, 40.63 % of the students identified a suspect. However, 34.38 % of the students drew a conclusion irrelevant to the task. Table 10 provides examples of three students and their line of reasoning within this task. At the start of the allergy-task, most of the students assumed the main character was allergic for the treat (75 %). Some students (25 %) were able to think of another alternative, which for each of them was that the character might not like the taste of the treat. For the harelip-task, 81.25 % of the students did not know what a harelip was and made use of the additional information. The harelip itself was mostly identified by students as the focus of this task (34.38 %), or as a motive (62.50 %). Few students (15.63 %) were able to identify a second alternative motive; the negative relationship between two characters.

***Argumentation.*** 1st graders often provided their own invented argumentation to support their answer, instead of using information provided in the task. For the fidget spinner-task, it was found that one student (3.13 %) provided evidence derived from the task to found their conclusion, which was the argument that Tom does not have a fidget spinner anymore. When 1st graders had to use evidence that was provided in the question itself, 21.88 % of the students were able to identify and use this evidence. When conducting reasoning, 9.38 % of the students were able to use provided evidence from the task. For the allergy-task, 78.13 % of the students were able to use evidence from the task to support their conclusion, of whom all but one mentioned that the character became sick of eating the cupcake, and a few students (12.50 %) mentioned that the character told about her allergy. Regarding the harelip-task, when conducting reasoning, and evidence was provided within the question, 31.25 % of the students were able to use evidence. When drawing their conclusion for this task, 28.13 % of the students provided evidence derived from the task with their answer, of whom each mentioned that the characters played together.

Table 10

*Examples of the Line of Reasoning of 1st Grade Students in the Fidget Spinner-Task*

| | What could have happened to the fidget spinner? (*Consider alternatives)* |
|---|---|
| Student 1 | Lost, or uh, yes... someone uh, someone took it and hid it. |
| Student 2 | That someone took it. |
| Student 3 | That it fell in the sand and got broken... |
| | What do you think has happened? And why? (*Create hypotheses*) |
| Student 1 | That it got lost and that is why it is hidden. |
| Student 2 | Uh, it could be that it is broken, that is feel on the ground, or the teacher took it from him, because she saw it... I think he left it on the table. |
| Student 3 | Uh Tom has his... lost his fidget spinner.. because he.. was gone at the second break. |
| | After you have heard all of this, what do you think, happened to the fidget spinner of Tom? (*State own conclusion*) |
| Student 1 | That he left it (the fidget spinner) inside and that she took in during the break. |
| Student 2 | That Kim took it. |
| Student 3 | That it got lost and found by Kim. |

**Grade 3.**

***Line of reasoning.*** Regarding the fidget spinner-task, an increase was found of students suspecting a character, from 15.38 % students who thought of this suspect as focus of the task, to 57.69 % mentioning the suspect when creating hypothesis and drawing conclusions for this task. Other students thought that the fidget spinner got lost and thought of this focus for the task (34.38 %), whereas one student drew this conclusion. Regarding the allergy-task, 92.31 % of the students hypothesized that the main character was allergic for the treat, plus, 53.85 % of the students mentioned an alternative, which was that the main character did not like the treat (an example is provided in Table 11). When drawing a conclusion for this task, all students were convinced of the chocolate allergy of the main character. For the harelip-task, 73.08 % of the students thought that counting out the main character was the focus of this task, whereas 57.69 % of the students mentioned the harelip as focus. When asked for motives of the locking out, 92.31 % of the students identified the harelip as a motive.

***Argumentation.*** Regarding the argumentation, a division was present between 3rd graders, with one part of them using invented arguments, and others using the provided information. For the fidget spinner-task, 38.46 % of the students were able to use evidence of the task to support their hypothesis and 42.31 % regarding the conclusion, of which most of them mentioned either the suspect or the teacher (Table 11 provides an example). When the evidence was provided within the question itself, 34.62 % of the students were able to use this evidence. When conducting reasoning, 11.54 % of the students were able to use provided evidence. In the allergy-task, 92.31 % of the students were able to provide at least one piece of evidence, derived from the task to found their conclusion. Of them, 83.33 % mentioned that the character got sick from eating the cupcake, 29.17 % mentioned that the character told that she was allergic, and 8.33 % mentioned that the character refused to eat the treat. Regarding the harelip-task, when conducting reasoning, and evidence was provided within the question, 23.08 % of the students were able to use one piece of evidence. Each of them mentioned that the main character had a harelip as well, and one also identified Bart as correct in this task. When drawing their conclusion for this task, 53.85 % of the students provided evidence derived from the task with their answer, of whom each student mentioned that the characters played together.

**Grade 6.**

***Line of reasoning.*** Twenty percent of the students immediately suspect a character who stole the fidget spinner, as focus for the fidget spinner-task. When asked to think of alternatives, 56 % of the students mention that the fidget spinner got stolen and 56 % of the students referred to a specific character. When creating a hypothesis, 64 % of the students suspected a specific character whereas an equal amount of students concluded that this character stole the fidget spinner. A few students (16 %) concluded that the fidget spinner got lost. For the allergy-task, 88.46 % of the students started their CT process with the allergy as hypothesis and 57.69 % of the students hypothesised that the character did not like the treat. In the end, all students were convinced of the chocolate allergy. As focus of the

Table 11

*Examples of Students, Recalling and Combining Sources, provided in the Fidget Spinner-Task*

| | |
|---|---|
| | Follow-up question of question assessing skill *State own conclusion*. What caused that you think that? (*Recall & combine sources*) |
| Student 1 | Because the teacher saw her going outside with a fidget spinner. |
| Student 2 | Because Kim had no fidget spinner at first, but then she does... uh... because... the teacher and my classmate saw it. |
| Student 3 | Kim went inside and Kim had the same fidget spinner. |

harelip-task, all but one student (96 %), mentioned the locking out of the main character, whereas 72 % of the students mentioned the harelip as well. When asked to think of a motive for counting out the main character, 84 % of the students identified the harelip and 64 % of the students were able to mention a second alternative motive.

***Argumentation.*** 6$^{th}$ graders often used objective information provided in the task to support their answers. Although 64 % of the students suspected a specific character in their hypotheses for the fidget spinner-task, only 28 % used evidence derived from the task to found their hypothesis. An increased amount of students used evidence from the task to support their conclusion (60 %). An example is provided in Table 12. When students had to use evidence that was provided in the question itself, 36 % of the students were able to identify and use this evidence. When conducting reasoning, 36 % of the students were able to use provided evidence. Regarding the allergy-task, 96 % of the students provided evidence of the task to found their conclusion. Of these students, 87.50 % mentioned that the character got sick after eating the cupcake, 29.17 % mentioned that the character told them about her allergy, and 12.50 % mentioned the character refusing to eat the treat. Regarding the harelip-task, when conducting reasoning and evidence was provided within the question, 36 % of the students were able to use one piece of evidence. Of them, 77.78 % mentioned that the main character had a harelip as well, and 55.56 % identified Bart as correct in this task. When drawing their conclusion for this task, 84 % of the students provided evidence derived from the task with their answer, of whom all but one student, mentioned that the characters played together.

Table 12

*Examples of a Student's Argumentation in the Fidget Spinner-Task*

| | |
|---|---|
| | What do you think has happened? And why? (*Create hypotheses*) |
| Student | I think the other girl might took it, because she has the same. |
| | Follow-up question of question assessing skill *State own conclusion*. What caused that you think that? (*Recall & combine sources*) |
| Student | That Kim took it, because Kim also says 'Yes I got a fidget spinner last weekend', but that is.. I guess that is not true, because I think she just lies really badly. |

*Note.* The assessed skill of each question is add in italic font. The students uses different pieces of evidence to answer each of the questions.

## Discussion

Current available information regarding CT among elementary school students is retrieved from studies investigating aspects of CT, instead of overall CT ability. This study aimed to enhance understanding of the overall CT ability of elementary school students. Therefore, the results of this study are a starting point for investigation of CT ability among elementary school students. To operationalize CT, CT was conceptualized for elementary school students, including a model that displayed the CT process. Based on the proposed framework and model, an assessment tool (the CTAEE) was developed that assessed the CT performance of elementary school students. The data generated from the CTAEE outcomes, provided insight in the growth of CT ability, as well as insight in the CT process of the students. The same data was also used to confirm the CT model as proposed in this study. The CTAEE was completed by 83 students, divided over three grades, respectively 1$^{st}$, 3$^{rd}$, and 6$^{th}$ grade.

The main outcome of this study, is that the results of the CTAEE showed significant differences between overall CT ability between each of the assessed grades. For each underlying construct and skill, a significant difference was found between 1$^{st}$ and 6$^{th}$ grade. It was also found that grade and CT ability are related to each other. Moreover, the pace in which students' CT ability of constructs and skills grew, differed. Most of the 1$^{st}$ graders show signs of CT constructs and constituent skills when reasoning, whereas 6$^{th}$ graders seem to possess a wide variety of CT skills.

### Growth of Critical Thinking Ability during Elementary Education

Comparing the CT ability of 1$^{st}$ and 6$^{th}$ grade students, it appeared that some of the CT skills showed a great improvement, whereas others showed small improvements. It was observed that the skills for which 1$^{st}$ graders achieved high scores, often were skills that showed small improvements during elementary education. A possible explanation for this difference is the ceiling effect, which means that a high score at the starting point (respectively 1$^{st}$ grade), leaves room for a small chance of development (Salkind, 2010). The identified growth in CT performance is in line with findings of scholars that investigated concepts similar to CT, such as metacognition (Kuhn, 1999).

Conducting deductions seems to be a difficult skill for all students when looking at the ability scores of students. Piaget's theory could explain this observation. According to this theory, children only learn conducting deductions after the age of twelve. The ability scores of other reasoning skills were compared to other scores, high. In combination with the extreme low score of students regarding conducting deductions, this could explain the lower internal consistency of the reasoning construct.

A possible explanation for high ability levels of CT skills, could be that these are related to other skills students learn during elementary education (e.g., theory of mind and language understanding). It would be interesting and relevant to study the development of these skills combined with CT ability, to identify possible causal relationships. For example, reading comprehension starts in 2$^{nd}$ grade in

Dutch elementary education, which shows overlap with CT (e.g., identify the focus of an article). If reading comprehension correlates with the development of CT, it might be more convenient to integrate CT in existing reading comprehension instruction. Moreover, the CTAEE was read out to the students, to overcome influence of reading skills. However, differences in students' ability in performing listening skills might have influenced the outcomes of the CTAEE. Lastly, student characteristics might influence CT performance as well (e.g., intelligence, gender, and socioeconomic state).

**Differences in Critical Thinking Process; Line of Reasoning and Argumentation**

In-depth differences between CT ability of elementary school students are derived from the answers of the students. These differences were found within the used line of reasoning and provided argumentation. Regarding the fidget spinner-task, approximately halve of the 1$^{st}$ graders hypothesised that the fidget spinner got lost, whereas 6$^{th}$ graders mostly hypothesised and concluded that their classmate stole the fidget spinner (including recalling the name of the suspect). Regarding this task, 3$^{rd}$ graders were more divided in their hypotheses. The lines of reasoning of the grades regarding the other tasks (themed allergy and harelip) were quite similar. It was revealed that 1$^{st}$ graders more often drew conclusions irrelevant to the task, compared to 3$^{rd}$ and 6$^{th}$ graders. Moreover, the irrelevant conclusions drawn by the students had different characteristics (Table 13 provides examples of irrelevant conclusions). Comparing the answers students provided regarding the skill *consider alternatives* revealed similar findings (Table 14 provides an example). Moreover, 1$^{st}$ and 3$^{rd}$ graders mention different aspects than 6$^{th}$ graders when they had to judge sources, such as personal reasons.

In general, it was found that, especially 1$^{st}$ graders were less able to judge the provided information compared to 6$^{th}$ graders. Moreover, 1$^{st}$ graders seemed to invent their own arguments and reasons to support their answers. In this matter, 3$^{rd}$ graders were divided with some of them inventing arguments and others using provided information. Both findings concerning 1$^{st}$ grade students could relate to each other and could mean that 1$^{st}$ graders had to invent their own arguments because they were unable to judge provided information that could be used as argumentation.

Another example contributing to this finding, is that 1$^{st}$ and 3$^{rd}$ graders tend to conduct reasoning within their own mental representations, which they show by creating hypotheses that are in line with their own world. It seems that it is difficult for these students to adapt their mental representations or to make use of information that is new to them, and perhaps unfamiliar. Findings concerning the use of provided information are consistent with findings of Lai (2011), who stated that younger children were less able to identify and make use of provided information, especially if this information was new to them.

Table 13

*Examples of Illogical Conclusions Students provided in the Fidget Spinner-case*

| | Question: After you have heard all of this, what do you think, happened to the fidget spinner of Tom? |
|---|---|
| Grade 1 | Tom got sad suddenly, and that is why he has no fidget spinner anymore. |
| Grade 3 | Uh, that he gave it away. |
| Grade 6 | *Is thinking* ... uh Tom dropped it, no left it in the classroom, Kim took it and the next day she no fidget spinner anymore, and gave it back to Tom. |

*Note.* This question assessed skill *state own conclusion*, as part of construct *draw conclusions*. Each of these answers receives the skill-code *illogical conclusion* and argument-code *others*, because all of them are not in line with the provided storyline.

Table 14

*Examples of Alternatives Students provided in the Allergy-case*

| | Question: Why would Inge not eat the cupcake? |
|---|---|
| Grade 1 | Uh, allergic.... or does not like it... uh uh, I don't know more things.. |
| Grade 3 | Because she is allergic to chocolate, or she does not like chocolate. |
| Grade 6 | Because she does not like it, or because it contains something she is allergic to. |

*Note.* This question assessed skill *consider alternatives*, as part of construct *analyse the problem*. Each of these answers receives the skill-code *≥2 relevant alternatives* and argument-codes *allergic* and *does not like it*. Most of the students provided the same two alternatives in this case.

**Implications for Theory and Practice**

CT literature concerning elementary school students was lacking, as was a conceptualization of it. In this study, a framework and CT process were proposed, based on existing frameworks concerning adults. The proposed model was found responsible for a third to halve of the variance in CT performance of consecutive phases. In future research, causal relationships between the phases could be investigated. The current executed analysis regarding the CT model confirms the idea of CT being a process among elementary school students as well.

The CTAEE was developed specifically to fit the purpose of this study. It seems that the CTAEE is a suitable tool to measure CT ability of elementary school students. It was constructed in such a way that influences due to lack of background knowledge or reading skills were minimized as much as possible. Moreover, the questions and cases were developed in such a way that it was understandable for 1st graders, but not too simple for 6th graders. The script of the CTAEE included additional explanatory materials for the allergy- and harelip-task. Although these materials were added to overcome lack of background knowledge, it could still have influenced the scores of students that made use of these materials. To reveal the actual influence of background knowledge in the CTAEE, a study using an experimental design could be conducted where different groups of students have to complete only one of the three tasks included in the CTAEE. Comparing the results of such a study reveals differences in the tasks. Moreover, existing studies regarding CT development among elementary school students often investigated CT in scientific contexts, such as physics or mathematics. It could be that the social context is easier to understand for children compared to scientific contexts. Hence, it would be useful to conduct a similar research, using tasks with different contexts, to identify the influence of the context in which students conduct CT.

This study suggests improvement of CT ability during elementary education. However, this study used a cross-sectional design, which suggests development of CT. In future, a longitudinal research design could be used to confirm the findings of this study and to enhance understanding of the precise moments of development of skills. Results confirm development of CT skills within the students themselves. Such findings can be used by instructional designers to develop instruction that is levelled to the CT development of elementary school students.

**Conclusion**

This study is a first exploration into the development of CT (as measured by the increase in CT performance) during elementary education and contributes to literature and practice. The results provide information and insights in the growth of CT performance among elementary school students. An assessment tool was developed for this study, which made use of simple social contexts and was based on our proposed framework and model. It was found that even 1st graders already show signs of most of the CT constructs and constituent skills, and 6th graders were almost competent for most of them. Not all of the assessed constructs and skills developed in an equal pace. These findings show differences in CT ability between students of different grades. Results can be used for the development of effective CT instruction for elementary education. Studies further investigating the development of CT among elementary school students, including other variables and longitudinal research, is important for future research and contribute to the development of effective CT instruction as well.

**References**


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

Boudah, D. J. (2011). *Conducting Educational Research: Guide to Completing a Major Project.* Thousand Oaks: SAGE Publications.

Browne, M. N., Freeman, K. E., & Williamson, C. L. (2000). The importance of critical thinking for student use of the Internet. *College Student Journal*, *34*, 391-398. Retrieved from https://go.galegroup.com/ps/i.do?p=AONE&sw=w&u=googlescholar&v=2.1&it=r&id=GALE%7CA66760560&sid=googleScholar&asid=b38e6586

Christensen, L. B., Burke Johnson, R., & Turner, L. A. (2014). *Research methods, design, and analysis.* New York: Pearson.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6,* 284-290. doi:10.1037/1040-3590.6.4.284

Cottrell, S. (2011). *Critical Thinking Skills: Developing effective analysis and argument*. New York: Palgrave Macmillan.

D'Angelo, E. (1971). *The Teaching of Critical Thinking.* Amsterdam: B.R. Grüner.

Dede, C. (2010). Comparing frameworks for 21st century skills. In J. Bellanca & R. Brandt (Eds.), *21st century skills: Rethinking how students learn* (pp. 51-76). Bloomington: Solution Tree Press.

Dierking, L. D. & Falk, J. H. (2016). 2020 Vision: Envisioning a new generation of STEM learning research. *Cultural Studies of Science Education, 11*, 1-10. doi:10.1007/s11422-015-9713-5

Doherty, J. J., Hansen, M. A., & Kaya, K. K. (1999). Teaching information skills in the information age: The need for critical thinking. *Library Philosophy and Practice, 1*, 1-10. Retrieved from https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1017&context=libphilprac

Duron, R., Limbach, B., & Waugh, W. (2006). Critical thinking framework for any discipline. *International Journal of Teaching and Learning in Higher Education*, *17*, 160-166. Retrieved from http://www.isetl.org/ijtlhe/pdf/IJTLHE55.pdf

Ennis, R. H. (1962). A concept of critical thinking. *Harvard Educational Review*, *32*, 81-111. Abstract retrieved from http://psycnet.apa.org

Ennis, R. H. (1991). Critical thinking: A streamlined conception. *Teaching Philosophy, 14*, 5-24. doi:10.5840/teachphil19911412

Ennis, R. H. (1993). Critical thinking assessment. *Theory Into Practice, 32*, 179-186. doi:10.1080/00405849309543594

Ennis, R. H. (2011). *The Nature of Critical Thinking: An Outline of Critical Thinking Dispositions and Abilities.* Retrieved from https://education.illinois.edu/docs/default-source/faculty-documents/robert-ennis/thenatureofcriticalthinking_51711_000.pdf?sfvrsn=7bb51288_2

Ennis, R. H. (2015). Critical thinking: A streamlined conception. In M. Davies & R. Barnett (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 31-47). New York: Palgrave Macmillan.

Expertisecentrum Nederlands (2010). *Begrijpend Lezen* [Reading Comprehension]. Retrieved from http://www.leerlijnentaal.nl/page/70/begrijpend-lezen.html

Facione, P. A. (1990). *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. Research Findings and Recommendations.* Retrieved from ERIC database. (ED315423)

Facione, P. A. (2015). *Critical thinking: What it is and why it counts.* Retrieved from https://www.nyack.edu/files/CT_What_Why_2013.pdf

Florea, N. M. & Hurjui, E. (2015). Critical thinking in elementary school children. *Procedia-Social and Behavioral Sciences*, *180*, 565-572. doi:10.1016/j.sbspro.2015.02.161

Ford, C. L. & Yore, L. D. (2011). Toward convergence of critical thinking, metacognition, and reflection: illustrations from natural and social sciences, teacher education, and classroom practice. In A. Zohar & Y. J. Dori (Eds.), *Metacognition in Science Education: Trends in current research* (pp. 251–271). doi:10.1007/978-94-007-2132-6_11

Foundation for Critical Thinking (2007). *Defining Critical Thinking.* Retrieved from http://www.criticalthinking.org/aboutCT/define_critical_thinking.cfm.

Gelman, S. A. & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*, 183–209. doi:10.1016/0010-0277(86)90034-X

Halpern, D. F. (1993). Assessing the effectiveness of critical-thinking instruction. *The Journal of General Education*, *42*, 238-254. doi:10.1353/jge.2001.0024

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

Halpern, D. F. (2003). *Thought & Knowledge.* New Jersey: Lawrence Erlbaum Associates.

Harvey, L., Moon, S., Geall, V., & Bower, R. (1997). *Graduates' Work: Organisational Change and Students' Attributes.* Retrieved from ERIC database. (ED410404)

Heyman, G. D. & Legare, C. H. (2005). Children's evaluation of sources of information about traits. *Developmental Psychology, 41*, 636–647. doi:10.1037/0012-1649.41.4.636

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). New York: Routledge.

Kettler, T. (2014). Critical thinking skills among elementary school students: Comparing identified gifted and general education student performance. *Gifted Child Quarterly*, *58*, 127-136. doi: 10.1177/0016986214522508

Kivunja, C. (2014). Do you want your students to be job-ready with 21st century skills? Change pedagogies: A pedagogical paradigm shift from Vygotskyian social constructivism to critical thinking, problem solving and Siemens' digital connectivism. *International Journal of Higher Education, 3*, 81-91. doi:10.5430/ijhe.v3n3p81

Ku, K. Y. L. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity, 4*, 70–76. Retrieved from ERIC database. (EJ833393)

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

Lai, E. R. (2011). Critical thinking: A literature review. Retrieved from https://images.pearsonassessments.com/images/tmrs/CriticalThinkingReviewFINAL.pdf

Lipman, M. (1987). Critical thinking: What can it be? *Analytical Teaching, 8*, 5-12. Retrieved from ERIC database. (EJ376244)

Massa, S. (2014). The development of critical thinking in elementary school: The role of teachers' beliefs. *Procedia - Social and Behavioral Sciences*, *141*, 387-392. doi:10.1016/j.sbspro.2014.05.068

Mayer, R. E. (2001). *Multimedia Learning*. New York: Cambridge University Press.

National Research Council (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington: The National Academies Press.

Olson, D., & Astington, J. (1993). Thinking about thinking: Learning how to take statements and hold beliefs. *Educational Psychologist, 28*, 7-23. doi:10.1207/s15326985ep2801_2

Partnership for 21st Century Skills (2009). *P21 Framework Definitions.* Retrieved from ERIC database. (ED519462)

Paul, R. W. (1984). Critical thinking: fundamental to education for a free society. *Educational Leadership*, *42*, 4-14. Retrieved from ERIC database. (EJ306669)

Paul, R. W. (1985). Bloom's taxonomy and critical thinking instruction. *Educational Leadership*, *42*, 36-39. Retrieved from ERIC database. (EJ319813)

Paul, R. W. & Elder, L. (2005). *Critical Thinking Competency Standards.* Retrieved from http://www.criticalthinking.org/resources/PDF/CT-competencies%202005.pdf

Perner, J. & Wimmer, H. (1985). "John thinks that Mary thinks that…" attribution of second-order beliefs by 5-to 10-year-old children. *Journal of Experimental Child Psychology*, *39*, 437-471. doi: 10.1016/0022-0965(85)90051-7

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

Saavedra, A. R. & Opfer, V. D. (2012). Learning 21st-century skills requires 21st-century teaching. *Kappan, 94*, 8-13. doi:10.1177/003172171209400203

Salkind, N. J. (2010). *Encyclopedia of research design*. doi: 10.4135/9781412961288

Shultz, T., & Mendelson, R. (1975). The use of covariation as a principle of causal analysis. *Child Development, 46*, 394-399. doi:10.2307/1128133

SLO, 2017. *Concept-leerlijnen voor 21e eeuwse vaardigheden: kijkwijzers voor het volgen van ontwikkeling* [Concept-educational goals for 21st century skills: visualized guides to follow development]. Retrieved from http://downloads.slo.nl/Documenten/SLO%20(2017)%2021e%20 eeuwse%20vaardigheden%20kijwijzers%20en%20leerlijnen%20concept.pdf

Swartz, R. J. & Parks, S. (1994). *Infusing the Teaching of Critical and Creative Thinking into Content Instruction: A Lesson Design Handbook for the Elementary Grades*. Retrieved from ERIC database. (ED407059)

Thayer-Bacon, B. J. (2000). *Transforming Critical Thinking: Thinking Constructively.* New York: Teachers College Press.

Timperley, H. (2008). *Teacher professional learning and development.* Retrieved from http://iaoed. org/downloads/EdPractices_18.pdf

Voogt, J. & Roblin, N. P. (2010). *21st Century Skills: Discussienota* [21st Century Skills: Discussion paper]. Retrieved from http://development.todosmedia.com/klassetheater/wp-content/uploads/2015/ 04/discussie-nota- 21_st_century_skills-.pdf

Washburn, P. (2010). *The Vocabulary of Critical Thinking.* Oxford: Oxford University Press.

Willingham, D. T. (2007). Critical thinking: Why is it so hard to teach? *American Educator,* 8-19. doi:10.3200/AEPR.109.4.21-32

**Appendix A.** Calculated Test-scores MANOVA Test with Bonferroni Corrections, Constructs and Constituent Skills

Table 15

*Calculated MANOVA with Bonferonni Corrections for Construct Analyse the problem, and Constituent Skills, Differences Between Grades*

| | *F* | *df* | *p* |
|---|---|---|---|
| Analyse problems | 58.72 | 2 | **<.001** |
| Grade 1 + Grade 3 | | (1,56) | **<.001** |
| Grade 1 + Grade 6 | | (1,57) | **<.001** |
| Grade 3 + Grade 6 | | (1,51) | **<.001** |
| Identify the focus | 61.73 | 2 | **<.001** |
| Grade 1 + Grade 3 | | (1,56) | **<.001** |
| Grade 1 + Grade 6 | | (1,57) | **<.001** |
| Grade 3 + Grade 6 | | (1,51) | **<.001** |
| Consider alternatives | 8.73 | 2 | **<.001** |
| Grade 1 + Grade 3 | | (1,56) | .060 |
| Grade 1 + Grade 6 | | (1,57) | **<.001** |
| Grade 3 + Grade 6 | | (1,51) | .072 |
| Create hypotheses | 10.50 | 2 | **<.001** |
| Grade 1 + Grade 3 | | (1,56) | **.001** |
| Grade 1 + Grade 6 | | (1,57) | **<.001** |
| Grade 3 + Grade 6 | | (1,51) | .500 |
| Identify gaps in problems | 39.95 | 2 | **<.001** |
| Grade 1 + Grade 3 | | (1,56) | **<.001** |
| Grade 1 + Grade 6 | | (1,57) | **<.001** |
| Grade 3 + Grade 6 | | (1,51) | **<.001** |

*Note.* Bold p-values are significant p <.05.

Table 16

*Calculated MANOVA with Bonferonni Corrections for Construct Identify & Judge Sources, and Constituent Skills, Differences Between Grades*

| | *F* | *df* | *p* |
|---|---|---|---|
| Gather reliable information | 22.70 | 2 | **<.001** |
| Grade 1 + Grade 3 | | (1,56) | **<.001** |
| Grade 1 + Grade 6 | | (1,57) | **<.001** |
| Grade 3 + Grade 6 | | (1,51) | **.027** |
| Identify motives | 7.37 | 2 | **<.001** |
| Grade 1 + Grade 3 | | (1,56) | **.035** |
| Grade 1 + Grade 6 | | (1,57) | **<.001** |
| Grade 3 + Grade 6 | | (1,51) | .244 |
| Identify opinions | 6.52 | 2 | **.001** |
| Grade 1 + Grade 3 | | (1,56) | .143 |
| Grade 1 + Grade 6 | | (1,57) | **.001** |
| Grade 3 + Grade 6 | | (1,51) | .103 |
| Identify others' point of views | 5.22 | 2 | **.004** |
| Grade 1 + Grade 3 | | (1,56) | .500 |
| Grade 1 + Grade 6 | | (1,57) | **.004** |
| Grade 3 + Grade 6 | | (1,51) | **.037** |
| Judge the reliability of information | 8.61 | 2 | **<.001** |
| Grade 1 + Grade 3 | | (1,56) | **.002** |
| Grade 1 + Grade 6 | | (1,57) | **<.001** |
| Grade 3 + Grade 6 | | (1,51) | .500 |
| Judge the reliability of observations | 11.25 | 2 | **<.001** |
| Grade 1 + Grade 3 | | (1,56) | **.001** |
| Grade 1 + Grade 6 | | (1,57) | **<.001** |
| Grade 3 + Grade 6 | | (1,51) | .500 |

*Note.* Bold p-values are significant p <.05.

Table 17

*Calculated MANOVA with Bonferonni Corrections for Construct Conduct Reasoning, and Constituent Skills, Differences Between Grades*

| | *F* | *df* | *p* |
|---|---|---|---|
| Conduct reasoning | 11.76 | 2 | **<.001** |
| Grade 1 + Grade 3 | | (1,56) | .051 |
| Grade 1 + Grade 6 | | (1,57) | **<.001** |
| Grade 3 + Grade 6 | | (1,51) | **.018** |
| Conduct deductions | 5.27 | 2 | **.004** |
| Grade 1 + Grade 3 | | (1,56) | .500 |
| Grade 1 + Grade 6 | | (1,57) | **.005** |
| Grade 3 + Grade 6 | | (1,51) | **.018** |
| Conduct inductions | 8.13 | 2 | **<.001** |
| Grade 1 + Grade 3 | | (1,56) | **.027** |
| Grade 1 + Grade 6 | | (1,57) | **<.001** |
| Grade 3 + Grade 6 | | (1,51) | .210 |
| Establish causal relationships | 4.94 | 2 | **.005** |
| Grade 1 + Grade 3 | | (1,56) | .135 |
| Grade 1 + Grade 6 | | (1,57) | **.004** |
| Grade 3 + Grade 6 | | (1,51) | .270 |

*Note.* Bold p-values are significant p <.05.

Table 18

*Calculated MANOVA with Bonferonni Corrections for Construct Draw conclusions, and Constituent Skills, Differences Between Grades*

| | *F* | *df* | *p* |
|---|---|---|---|
| Draw conclusions | 23.67 | 2 | **<.001** |
| Grade 1 + Grade 3 | | (1,56) | **<.001** |
| Grade 1 + Grade 6 | | (1,57) | **<.001** |
| Grade 3 + Grade 6 | | (1,51) | **.014** |
| Recall & combine sources | 22.61 | 2 | **<.001** |
| Grade 1 + Grade 3 | | (1,56) | **<.001** |
| Grade 1 + Grade 6 | | (1,57) | **<.001** |
| Grade 3 + Grade 6 | | (1,51) | **.020** |

*Note.* Bold p-values are significant p <.05.

**Appendix B.** Calculated Results Paired Comparison Tests of Skills, per Grade

Table 19

*Calculated Test Statistics Paired Comparisons, as Measured by Paired Samples t-tests*

| Skills | Grade 1 | Grade 3 | Grade 6 |
|---|---|---|---|
| Skill 1+2 | $t$ -10.54, $p$ <.001 | $t$ -4.34, $p$ <.001 | |
| Skill 1+3 | $t$ -5.40, $p$ <.001 | $t$ -3.58, $p$.001 | |
| Skill 1+4 | | $t$ -2.67, $p$.013 | |
| Skill 1+5 | $t$ -6.66, $p$ <.001 | $t$ -4.47, $p$ <.001 | |
| Skill 1+6 | $t$ -9.08, $p$ <.001 | $t$ -3.48, $p$.002 | $t$ -2.09, $p$.047 |
| Skill 1+7 | $t$ -4.53, $p$ <.001 | | |
| Skill 1+8 | $t$ -4.70, $p$ <.001 | $t$ -2.67, $p$.013 | |
| Skill 1+9 | $t$ -6.84, $p$ <.001 | $t$ -5.73, $p$ <.001 | |
| Skill 1+10 | $t$ -4.44, $p$ <.001 | | $t$ 3.92, $p$.001 |
| Skill 1+11 | $t$ -9.21, $p$ <.001 | $t$ -5.51, $p$ <.001 | $t$ -2.45, $p$.022 |
| Skill 1+12 | $t$ -10.39, $p$ <.001 | $t$ -4.87, $p$ <.001 | |
| Skill 1+13 | $t$ -4.88, $p$ <.001 | $t$ -3.19, $p$.004 | |
| Skill 2+3 | $t$ 2.16, $p$.039 | | |
| Skill 2+4 | $t$ 5.44, $p$ <.001 | | |
| Skill 2+5 | | | |
| Skill 2+6 | | | $t$ -2.47, $p$.021 |
| Skill 2+7 | $t$ 2.70, $p$.011 | $t$ 3.94, $p$.001 | |
| Skill 2+8 | $t$ 2.98, $p$.006 | | |
| Skill 2+9 | | $t$ -2.14, $p$.042 | |
| Skill 2+10 | $t$ 4.36, $p$ <.001 | $t$ -5.78, $p$ <.001 | $t$ 4.21, $p$ <.001 |
| Skill 2+11 | | $t$ -3.07, $p$.005 | $t$ -3.70, $p$.001 |
| Skill 2+12 | | $t$ -2.18, $p$.039 | $t$ -2.69, $p$.013 |
| Skill 2+13 | $t$ 3.32, $p$.002 | | |
| Skill 3+4 | $t$ 2.86, $p$.008 | | $t$ -2.38, $p$.026 |
| Skill 3+5 | | | |
| Skill 3+6 | $t$ -2.87, $p$.007 | | $t$ -4.10, $p$ <.001 |
| Skill 3+7 | | $t$ 3.80, $p$.001 | |
| Skill 3+8 | | | |
| Skill 3+9 | | | $t$ -2.62, $p$.015 |
| Skill 3+10 | | $t$ 5.44, $p$ <.001 | $t$ 3.09, $p$.005 |
| Skill 3+11 | $t$ -3.22, $p$.003 | $t$ -2.61, $p$.015 | $t$ -3.38, $p$.003 |

| Skills | Grade 1 | Grade 3 | Grade 6 |
|---|---|---|---|
| Skill 3+12 | *t* -3.40, *p*.002 | | *t* -3.36, *p*.003 |
| Skill 3+13 | | | |
| Skill 4+5 | *t* -4.53, *p* <.001 | | |
| Skill 4+6 | *t* -5.48, *p* <.001 | | |
| Skill 4+7 | *t* -2.90, *p*.007 | | *t* 2.52, *p* .019 |
| Skill 4+8 | *t* -3.02, *p*.005 | | *t* 2.68, *p*.013 |
| Skill 4+9 | *t* -4.59, *p* <.001 | *t* -3.25, *p*.003 | |
| Skill 4+10 | | *t* 3.49, *p*.002 | *t* 4.47, *p* <.001 |
| Skill 4+11 | *t* -6.68, *p* <.001 | *t* -3.49, *p*.002 | |
| Skill 4+12 | *t* -8.78, *p* <.001 | *t* -2.78, *p*.010 | |
| Skill 4+13 | *t* -3.48, *p*.002 | | |
| Skill 5+6 | *t* -2.25, *p*.032 | | |
| Skill 5+7 | | *t* 2.86, *p*.008 | |
| Skill 5+8 | | | |
| Skill 5+9 | | *t* -2.60, *p*.015 | |
| Skill 5+10 | *t* 3.22, *p*.003 | *t* 5.28, *p* <.001 | *t* 3.17, *p*.004 |
| Skill 5+11 | *t* -3.09, *p*.004 | *t* -2.13, *p*.043 | *t* -2.40, *p*.024 |
| Skill 5+12 | *t* -3.48, *p*.002 | | *t* -2.19, *p*.038 |
| Skill 5+13 | *t* 2.40, *p*.023 | | |
| Skill 6+7 | *t* 3.31, *p*.002 | *t* 3.11, *p*.005 | *t* 2.98, *p* .007 |
| Skill 6+8 | *t* 3.39, *p*.002 | | *t* 3.85, *p*.001 |
| Skill 6+9 | | | |
| Skill 6+10 | *t* 4.63, *p* <.001 | *t* 5.21, *p* <.001 | *t* 6.35, *p* <.001 |
| Skill 6+11 | | | |
| Skill 6+12 | | | |
| Skill 6+13 | *t* 3.47, *p*.002 | | |
| Skill 7+8 | | *t* -2.41, *p*.024 | |
| Skill 7+9 | *t* -2.08, *p*.046 | *t* -4.35, *p* <.001 | *t* -2.87, *p*.008 |
| Skill 7+10 | | | |
| Skill 7+11 | *t* -4.46, *p* <.001 | *t* -5.62, *p* <.001 | *t* -3.48, *p*.002 |
| Skill 7+12 | *t* -5.26, *p* <.001 | *t* -4.36, *p* <.001 | *t* -3.07, *p*.005 |
| Skill 7+13 | | *t* -3.43, *p*.002 | |
| Skill 8+9 | *t* -2.30, *p*.028 | | *t* -3.36, *p*.003 |
| Skill 8+10 | | *t* 3.98, *p*.001 | |
| Skill 8+11 | *t* -3.74, *p*.001 | *t* -2.67, *p*.013 | *t* -3.92, *p*.001 |

| Skills | Grade 1 | Grade 3 | Grade 6 |
|---|---|---|---|
| Skill 8+12 | *t* -4.09, *p* <.001 | *t* -2.48, *p*.020 | *t* -4.04, *p* <.001 |
| Skill 8+13 | | | *t* -2.61, *p*.015 |
| Skill 9+10 | *t* 3.51, *p*.001 | *t* 7.86, *p* <.001 | *t* 5.10, *p* <.001 |
| Skill 9+11 | | | |
| Skill 9+12 | *t* -2.18, *p*.037 | | |
| Skill 9+13 | *t* 2.95, *p*.006 | | |
| Skill 10+11 | *t* -7.04, *p* <.001 | *t* -9.38, *p* <.001 | *t* -6.82, *p* <.001 |
| Skill 10+12 | *t* -7.16, *p* <.001 | *t* -5.43, *p* <.001 | *t* -5.53, *p* <.001 |
| Skill 10+13 | | *t* -5.53, *p* <.001 | *t* -4.03, *p* <.001 |
| Skill 11+12 | | | |
| Skill 11+13 | *t* 5.35, *p* <.001 | *t* 3.61, *p*.001 | |
| Skill 12+13 | *t* 7.00, *p* <.001 | *t* 2.38, *p* .025 | |

**Appendix C.** Distribution Tables, Shares and Frequencies, Constructs & Skills, per Grade.

Table 20a

*Construct: Analyse problems*

| Score | Grade 1 | | | Grade 3 | | | Grade 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | | | | | | | | | |
| .07 (1) | 2 | 6.3 | 6.3 | | | | | | |
| .13 (2) | | | 6.3 | | | | | | |
| .20 (3) | 3 | 9.4 | 15.6 | | | | | | |
| .27 (4) | 2 | 6.3 | 21.9 | | | | | | |
| .33 (5) | 7 | 21.9 | 43.8 | | | | | | |
| .40 (6) | 5 | 15.6 | 59.4 | 5 | 19.2 | 19.2 | | | |
| .47 (7) | 8 | 25.0 | 84.4 | 3 | 11.5 | 30.8 | | | |
| .53 (8) | 2 | 6.3 | 90.6 | 5 | 19.2 | 50.0 | | | |
| .60 (9) | 2 | 6.3 | 96.9 | | | 50.0 | 2 | 8.0 | 8.0 |
| .67 (10) | 1 | 3.1 | 100 | 6 | 23.1 | 73.1 | 5 | 20.0 | 28.0 |
| .73 (11) | | | | 3 | 11.5 | 84.6 | 3 | 12.0 | 40.0 |
| .80 (12) | | | | 2 | 7.7 | 92.3 | 6 | 24.0 | 64.0 |
| .87 (13) | | | | | | 92.3 | 7 | 28.0 | 92.0 |
| .93 (14) | | | | 2 | 7.7 | 100 | 1 | 4.0 | 96.0 |
| 1.00 (15) | | | | | | | 1 | 4.0 | 100 |

Table 20b

*Skill: Identify the Focus*

| Score | Grade 1 | | | Grade 3 | | | Grade 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | 21 | 65.6 | 65.6 | 3 | 11.5 | 11.5 | | | |
| .33 (1) | 10 | 31.3 | 96.9 | 13 | 50.0 | 61.5 | 3 | 12.0 | 12.0 |
| .67 (2) | 1 | 3.1 | 100 | 9 | 34.6 | 96.2 | 11 | 44.0 | 56.0 |
| 1.00 (3) | | | | 1 | 3.8 | 100 | 11 | 44.0 | 100 |

Table 20c

*Skill: Consider Alternatives*

| Score | Grade 1 | | | Grade 3 | | | Grade 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | 1 | 3.1 | 3.1 | | | | | | |
| .17 (1) | 1 | 3.1 | 6.3 | | | | | | |
| .33 (2) | 4 | 12.5 | 18.8 | 2 | 7.7 | 7.7 | | | |
| .50 (3) | 10 | 31.3 | 50.0 | 9 | 34.6 | 42.3 | 4 | 16.0 | 16.0 |
| .67 (4) | 12 | 37.5 | 87.5 | 5 | 19.2 | 61.5 | 5 | 20.0 | 36.0 |
| .83 (5) | 3 | 9.4 | 96.9 | 7 | 26.9 | 88.5 | 12 | 48.0 | 84.0 |
| 1.00 (6) | 1 | 3.1 | 100 | 3 | 11.5 | 100 | 4 | 16.0 | 100 |

Table 20d

*Skill: Create Hypotheses*

| Score | Grade 1 | | | Grade 3 | | | Grade 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | 8 | 25.0 | 25.0 | | | | | | |
| .33 (1) | 9 | 28.1 | 53.1 | 6 | 23.1 | 23.1 | 2 | 8.0 | 8.0 |
| .67 (2) | 12 | 37.5 | 90.6 | 13 | 50.0 | 73.1 | 16 | 64.0 | 72.0 |
| 1.00 (3) | 3 | 9.4 | 100 | 7 | 26.9 | 100 | 7 | 28.0 | 100 |

Table 20e

*Skill: Identify Gaps in problems*

| Score | Grade 1 | | | Grade 3 | | | Grade 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | 15 | 46.9 | 46.9 | 2 | 7.7 | 7.7 | | | |
| .33 (1) | 13 | 40.6 | 87.5 | 9 | 34.6 | 42.3 | 2 | 8.0 | 8.0 |
| .67 (2) | 3 | 9.4 | 96.9 | 9 | 34.6 | 76.9 | 7 | 28.0 | 36.0 |
| 1.00 (3) | 1 | 3.1 | 100 | 6 | 23.1 | 100 | 16 | 64.0 | 100 |

Table 20f

*Construct: Gather reliable information*

| Score | Grade 1 | | | Grade 3 | | | Grade 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | 1 | 3.1 | 3.1 | | | | | | |
| .07 (1) | | | 3.1 | | | | | | |
| .13 (2) | | | 3.1 | | | | | | |
| .20 (3) | 2 | 6.3 | 9.4 | | | | | | |
| .27 (4) | 4 | 12.5 | 21.9 | | | | | | |
| .33 (5) | 2 | 6.3 | 28.1 | | | | | | |
| .40 (6) | 3 | 9.4 | 37.5 | | | | | | |
| .47 (7) | 5 | 15.6 | 53.1 | 3 | 11.5 | 11.5 | | | |
| .53 (8) | 5 | 15.6 | 68.8 | 5 | 19.2 | 30.5 | 1 | 4.0 | 4.0 |
| .60 (9) | 4 | 12.5 | 81.3 | 6 | 23.1 | 53.8 | 4 | 16.0 | 20.0 |
| .67 (10) | 1 | 3.1 | 84.4 | 2 | 7.7 | 61.5 | 3 | 12.0 | 32.0 |
| .73 (11) | 2 | 6.3 | 90.6 | 2 | 7.7 | 69.2 | 3 | 12.0 | 44.0 |
| .80 (12) | 2 | 6.3 | 96.9 | 5 | 19.2 | 88.5 | 5 | 20.0 | 64.0 |
| .87 (13) | | | 96.9 | 3 | 11.5 | 100 | 6 | 24.0 | 88.0 |
| .93 (14) | 1 | 3.1 | 100 | | | | 3 | 12.0 | 100 |
| 1.00 (15) | | | | | | | | | |

Table 20g

*Skill: Identify Motives*

| Score | Grade 1 | | | Grade 3 | | | Grade 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | 3 | 9.4 | 9.4 | 1 | 3.8 | 3.8 | | | |
| .33 (1) | 14 | 43.8 | 53.1 | 6 | 23.1 | 26.9 | 5 | 20.0 | 20.0 |
| .67 (2) | 12 | 37.5 | 90.6 | 12 | 46.2 | 73.1 | 8 | 32.0 | 52.0 |
| 1.00 (3) | 3 | 9.4 | 100 | 7 | 26.9 | 100 | 12 | 48.0 | 100 |

Table 20h

*Skill: Identify Opinions*

| Score | Grade 1 | | | Grade 3 | | | Grade 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | 3 | 9.4 | 9.4 | 2 | 7.7 | 7.7 | | | |
| .33 (1) | 8 | 25.0 | 34.4 | 3 | 11.5 | 19.2 | 2 | 8.0 | 8.0 |
| .67 (2) | 12 | 37.5 | 71.9 | 8 | 30.8 | 50.0 | 4 | 16.0 | 24.0 |
| 1.00 (3) | 9 | 28.1 | 100 | 13 | 50.0 | 100 | 19 | 76.0 | 100 |

Table 20i

*Skill: Identify Others' Point of View*

| Score | Grade 1 | | | Grade 3 | | | Grade 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | 10 | 31.3 | 31.3 | 5 | 19.2 | 19.2 | 2 | 8.0 | 8.0 |
| .33 (1) | 10 | 31.3 | 62.5 | 9 | 34.6 | 53.8 | 6 | 24.0 | 32.0 |
| .67 (2) | 9 | 28.1 | 90.6 | 10 | 38.5 | 92.3 | 8 | 32.0 | 64.0 |
| 1.00 (3) | 3 | 9.4 | 100 | 2 | 7.7 | 100 | 9 | 36.0 | 100 |

Table 20j

*Skill: Judge the Reliability of Information*

| Score | Grade 1 | | | Grade 3 | | | Grade 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | 8 | 25.0 | 25.0 | 2 | 7.7 | 7.7 | | | |
| .33 (1) | 13 | 40.6 | 65.6 | 4 | 15.4 | 23.1 | 7 | 28.0 | 28.0 |
| .67 (2) | 9 | 28.1 | 93.8 | 15 | 57.7 | 80.8 | 12 | 48.0 | 76.0 |
| 1.00 (3) | 2 | 6.3 | 100 | 5 | 19.2 | 100 | 6 | 24.0 | 100 |

Table 20k

*Skill: Judge the Reliability of Observations*

| Score | Grade 1 | | | Grade 3 | | | Grade 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | 5 | 15.6 | 15.6 | 2 | 7.7 | 7.7 | | | |
| .33 (1) | 10 | 31.3 | 46.9 | 1 | 3.8 | 11.5 | 1 | 4.0 | 4.0 |
| .67 (2) | 10 | 31.3 | 78.1 | 7 | 26.9 | 38.5 | 8 | 32.0 | 36.0 |
| 1.00 (3) | 7 | 21.9 | 100 | 16 | 61.5 | 100 | 16 | 64.0 | 100 |

Table 20l

*Construct: Conduct Reasoning*

| Score | | Grade 1 | | | Grade 3 | | | Grade 6 | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | 1 | 3.1 | 3.1 | | | | | | |
| .11 (1) | 2 | 6.3 | 9.4 | | | | | | |
| .22 (2) | | | 9.4 | | | | | | |
| .33 (3) | 2 | 6.3 | 15.6 | 2 | 7.7 | 7.7 | | | |
| .44 (4) | 7 | 21.9 | 37.5 | 2 | 7.7 | 15.4 | | | |
| .56 (5) | 7 | 21.9 | 59.4 | 5 | 19.2 | 34.6 | 1 | 4.0 | 4.0 |
| .67 (6) | 7 | 21.9 | 81.3 | 7 | 26.9 | 61.5 | 7 | 28.0 | 32.0 |
| .78 (7) | 2 | 6.3 | 87.5 | 9 | 34.6 | 96.2 | 11 | 44.0 | 76.0 |
| .89 (8) | 4 | 12.5 | 100 | 1 | 3.8 | 100 | 4 | 16.0 | 92.0 |
| 1.00 (9) | | | | | | | 2 | 8.00 | 100 |

Table 20m

*Skill: Conduct Deductions*

| Score | | Grade 1 | | | Grade 3 | | | Grade 6 | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | 7 | 21.9 | 21.9 | 6 | 23.1 | 23.1 | 1 | 4.0 | 4.0 |
| .33 (1) | 19 | 59.4 | 81.3 | 13 | 50.0 | 73.1 | 13 | 52.0 | 56.0 |
| .67 (2) | 6 | 18.8 | 100 | 7 | 26.9 | 100 | 7 | 28.0 | 84.0 |
| 1.00 (3) | | | | | | | 4 | 16.0 | 100 |

Table 20n

*Skill: Conduct Inductions*

| Score | | Grade 1 | | | Grade 3 | | | Grade 6 | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | 3 | 9.4 | 9.4 | | | | | | |
| .33 (1) | 5 | 15.6 | 25.0 | 1 | 3.8 | 3.8 | | | |
| .67 (2) | 14 | 43.8 | 68.8 | 13 | 50.0 | 53.8 | 7 | 28.0 | 28.0 |
| 1.00 (3) | 10 | 31.3 | 100 | 12 | 46.2 | 100 | 18 | 72.0 | 100 |

Table 20o

*Skill: Establish Causal Relationships*

| Score | | Grade 1 | | | Grade 3 | | | Grade 6 | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | 2 | 6.3 | 6.3 | 1 | 3.8 | 3.8 | | | |
| .33 (1) | 5 | 15.6 | 21.9 | 4 | 15.4 | 19.2 | | | |
| .67 (2) | 15 | 46.9 | 68.8 | 5 | 19.2 | 38.5 | 8 | 32.0 | 32.0 |
| 1.00 (3) | 10 | 31.3 | 100 | 16 | 61.5 | 100 | 17 | 68.0 | 100 |

Table 20p

*Construct: Draw conclusions*

| Score | Grade 1 | | | Grade 3 | | | Grade 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | 3 | 9.4 | 9.4 | | | | | | |
| .20 (1) | 4 | 12.5 | 21.9 | 1 | 3.8 | 3.8 | | | |
| .40 (2) | 12 | 37.5 | 59.4 | 3 | 11.5 | 15.4 | | | |
| .60 (3) | 6 | 18.8 | 78.1 | 10 | 38.5 | 53.8 | 4 | 16.0 | 16.0 |
| .80 (4) | 6 | 18.8 | 96.9 | 6 | 23.1 | 76.9 | 9 | 36.0 | 52.0 |
| 1.00 (5) | 1 | 3.1 | 100 | 6 | 23.1 | 100 | 12 | 48.0 | 100 |

Table 20q

*Skill: Recall & Combine Information*

| Score | Grade 1 | | | Grade 3 | | | Grade 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| (n points) | n | % | Cum.% | n | % | Cum.% | n | % | Cum.% |
| .00 (0) | 6 | 18.8 | 18.8 | | | | | | |
| .33 (1) | 16 | 50.0 | 68.8 | 9 | 34.6 | 34.6 | 1 | 4.0 | 4.0 |
| .67 (2) | 9 | 28.1 | 96.9 | 10 | 38.5 | 73.1 | 12 | 48.0 | 52.0 |
| 1.00 (3) | 1 | 3.1 | 100 | 7 | 26.9 | 100 | 12 | 48.0 | 100 |