



UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Think twice

Literature lessons that matter

Koek, M.

Publication date

2022

[Link to publication](#)

Citation for published version (APA):

Koek, M. (2022). *Think twice: Literature lessons that matter*. [Thesis, externally prepared, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

CHAPTER 2

LITERARY READING AND CRITICAL THINKING

Measuring Students' Critical Literary Understanding in Secondary Education

Abstract. Previous research suggests that literary reading may involve critical thinking. This involvement may facilitate Critical Literary Understanding (CLU), i.e., understanding the literary text in a (re)constructive, de-automatized manner. However, little is known about the cognitive processes this involvement entails. This study aims: 1) to conceptualize CLU, by relating dual process theory to concepts from the domain of literary studies (foregrounding, defamiliarization); 2) to test CLU in an educational context. An instrument to measure CLU was designed. A prospective cohort study was conducted ($N = 271$, Grades 10-12, pre-university education, ages 14-19) at a Dutch secondary school. CLU, critical thinking skills (CTS) and dispositions (CTD) were measured one month after the start of the academic year, CLU was measured again four months later. Results show that students' CLU improved. This improvement is mediated by CTD and moderated by CTS. These results suggest that critical thinking can be engaged in the literature classroom.

Keywords: Literature education, literary reading, critical thinking skills, critical thinking dispositions, dual process theory, literary understanding

1 INTRODUCTION

Literature education may be a perfect context for students to develop critical thinking. At least, that is what the award-winning film *Dead Poets Society* (Haft et al., 1989) vividly suggests to us. In this movie, the main character John Keating utilizes Walt Whitman verses to make his class of adolescent boys “think for themselves”, because, as he puts it, “No matter what anybody tells you, words and ideas can change the world.”

Typical Hollywood romanticism? Not exclusively perhaps. It seems that literary fiction can be instrumental to a wide range of educational objectives. Purves and Beach (1972) reviewed 126 studies on the effects of teaching literature and reported effects on cognitive performance, attitudes, and interest patterns. Forty years later these types of effects are still discerned in educational research (Janssen, 1998; Verboord, 2005; Witte, 2008; Mol & Bus, 2011; Schrijvers et al., 2016), so it seems safe to assume that the literature classroom can be a conducive environment for the impressionable mind of an adolescent. However, little is known about whether lessons in literature do indeed help students to develop their critical thinking.

For long, literary theorists have thought it a hallmark of the literary text that it can enhance readers’ consciousness about the world, by triggering them to think about the familiar as if they encountered it for the first time (Shklovsky, 2017). In addition, recent empirical studies with experienced adult readers indicate that literary reading may foster reflective thinking processes (Hakemulder, 2000; Djikic et al., 2013; Kidd & Castano, 2013). Such processes include slowing down thinking and postponing one’s judgment, both of which are widely considered as prerequisites for critical thinking (Dewey, 1910; Facione, 1990; Ennis, 2011; Facione, 2015).

However, these studies took place outside an educational context. Therefore, we do not know whether their conclusions might also pertain to novice and developing literary readers. Moreover, these studies did not measure how participants interpreted what they read, so we do not know whether there is a relation between critical thinking and meaning making processes in literary reading. There are theoretical studies that address the understanding of literary fiction in the literature classroom and link it explicitly to critical thinking (Rosenblatt, 1995; Faust, 2000; Bean & Moni, 2003), but these studies do not reveal which cognitive processes may constitute the possible interplay between the two. As a result, teachers in literature education who want to engage their students in critical thinking when these students make meaning from literary texts are somewhat in

the dark about what thinking skills and dispositions they should address in their lessons, to achieve that goal.

The aim of the present study is thus twofold: 1) to conceptualize critical thinking in the context of literary understanding; and 2) to assess its development in pre-university students (Grades 10-12), across a four-month timespan.

2 THEORETICAL BACKGROUND

Classroom discussions about literary texts may lead students to question habitual worldviews and stereotypes (Faust, 2000; Bean & Moni, 2003), and reflect on their emotions, in order to understand human nature better (Rosenblatt, 1995). Such possible outcomes of literature education add to the supposition that critical thinking, defined as “[...] reasonable, reflective thinking focused on deciding what to believe or do” (Ennis, 2011, p. 1), is part of understanding literature. Indeed, there are theoretical and empirical indications that the three cognitive processes highlighted in Ennis’ definition of critical thinking – reflection, reason, and focus on a decision – also play a substantial role in reading literary texts.

2.1 Literary reading and reflection as (re)constructive processes

Many current definitions of critical thinking refer to reflection (Lai, 2011), a thinking process conceptualized by John Dewey. According to Dewey, the origin of reflection is “[...] a state of perplexity, hesitation, doubt.”, brought about by the occurrence of a “[...] problem [, that] – no matter how slight and commonplace in character – perplexes and challenges the mind so that it makes belief at all uncertain [...]” (Dewey, 1910, p. 12). This combination of stimulus (the problem) and experience (perplexity) that triggers reflection, can also be discerned in literary reading. Typically, literary readers might experience a sensation of strangeness, that de-familiarizes them (Miall & Kuiken, 1994; Koopman & Hakemulder, 2015; Shklovsky, 2017). Defamiliarization can be induced by an artistic technique that literary scholars refer to as foregrounding (Hakemulder & Van Peer, 2016). Foregrounding theory proposes that literature (and art in general) invites recipients to look at what is overfamiliar to them from a different perspective, with a range of alienation techniques (Shklovsky, 2017) that make some textual features stand out from other, backgrounded features of the text, i.e., themes, patterns and contexts readers are familiar with (Murakovsky, 1964; Dixon et al., 1993; Koopman & Hakemulder, 2015). The reception process allegedly involves a more active involvement, for instance, in overcoming obstructions in comprehension, the interpretation of novel metaphors (Hoorn, 1997), retardation resulting in

deeper, more reflective processing (Hakemulder, 2004; Van Peer et al., 2007; Sopčák, 2011). Therefore, it seems that deviations from familiar, everyday discourse trigger both reflective thinking and literary reading.

In response to the experience of defamiliarization, readers of literary fiction may try to refamiliarize themselves with the text, a process described as “[...] an intra and/or extra textual revision or re-evaluation in order to discern, delimit or develop the novel meanings suggested by the foregrounded passages” (Miall & Kuiken, 1994, p. 394). Refamiliarization may be facilitated by the occurrence of aesthetic feelings, i.e., feelings in response to foregrounding, like being struck by a metaphor (Miall & Kuiken, 2002). These feelings incite readers to re-evaluate their first comprehension of a text, via increasing cognitive engagement (Fialho, 2007) and augmenting “anticipatory reading [that] enables readers to monitor their ongoing response to the text and to shape its significance as new events fall within the scope of the anticipation (or fail to do so)” (Miall & Kuiken, 2002, p. 227). This heightened anticipation and (re-)evaluation can also be understood as a hallmark of reflective thinking. The mind tries to resolve its state of perplexity by “the formation of some tentative plan or project, the entertaining of some theory which will account for the peculiarities in question” (Dewey, 1910, p. 15), constituting “[...] an act of search or investigation directed toward bringing to light further facts which serve to corroborate or nullify the suggested belief” (ibid, p. 12). Thus, we can describe both reflective thinking and literary reading as (re)constructive processes, concerned with building up an enhanced, novel, perspective.

The critical quality of this (re)constructive process may depend on how long a conclusion is suspended. According to Dewey, the mind, when in a state of perplexity, needs time to suggest another conceptual scheme that resolves the strangeness induced by the facts at hand. Accepting the first suggestion that comes to mind is typical of “uncritical thinking” (Dewey, 1910, p. 15). This implies that reflective thinking is cognitively taxing because it comprises a prolonged state of doubt and estrangement (Dewey, 1910). Therefore:

“[...] the most important factor in the training of good mental habits consists in acquiring the attitude of suspended conclusion [...]” (Dewey, 1910, p. 15).

Empirical studies of literary reading show literary effects that mirror Dewey’s suspended conclusion. For instance, reading literary fiction prolongs reading time (Miall & Kuiken, 1994) and induces a “suspension of urgency and permanence” that facilitates a short-time decrease of cognitive closure, i.e., the desire to arrive at conclusions quickly (Djikic et al., 2013, p. 153). These effects may be induced through the sole notion that one is reading a literary narrative (Zwaan, 1994),

and/or can be seen as a direct effect of foregrounding, that is said to de-automatize and slow down the reading process (Murakovsky, 1964). It is important to note, however, that the time span of this slowdown seems related to the reader's experience in reading literary texts (Djicic et al., 2013). When readers encounter foregrounding and defamiliarization while lacking the skills and/or experience to engage with it, they may deem the text as too difficult, and may "want to spend the least effort and do not try to look for more efficient [refamiliarization] strategies" (Fialho, 2007, p. 120). Thus, both reflective thinking and literary reading involve suspended conclusion, but both processes need to be trained for this suspension to be sustained.

To sum up, characteristics of reflective processes can be discerned in reading to understand literary texts.

2.2 Literary reading and reasoning as de-automatized processes

Another similarity between critical thinking and literary reading concerns both the pitfalls and the possibilities of human reasoning. These are probably best described by psychological, empirically founded, dual process theories (Kahneman & Frederick, 2002; Kahneman, 2011; Evans & Stanovich, 2013). According to these theories, there are two distinct systems at work in the mind, each with different cognitive processes to take care of. System 1 processes are automatic and enable us to judge situations quickly and deliver a fast response, making it possible for us to smoothly interact with the world around us, without having to deploy much cognitive energy. System 2 processes are de-automatized and facilitate so called cognitive decoupling: the ability to separate an imaginary representation of the world from our primary representation of the world as it is (Stanovich & Toplak, 2012), making it possible for humans to think hypothetically, a prerequisite for critical thinking (West et al., 2008; Facione, 2015). Due to its minimal cognitive load and subsequent fast responses, System 1 steers our daily life and is always processing in the background (Kahneman, 2011). This implies that, by default, humans are cognitive misers (Fiske & Taylor, 1984), i.e., they are prone to focus on easy to evaluate attributes of the world around them, even if the more difficult may yield more accurate understanding of that world (Kahneman & Frederick, 2002; Frederick, 2005; Kahneman, 2011). Therefore, System 1 processes need to be inhibited before System 2's de-automatized processes can override them (West et al., 2008; Stanovich & Toplak, 2012; Evans & Stanovich, 2013).

It is likely that dual process theories also apply to literary reading. Because literary texts contain both foregrounded and backgrounded elements (cf. Jacobs,

2015), literary reading may comprise different mental processes. Typically, foregrounded elements may incite System 2 processing. Refamiliarization, for instance, involves comparing a first reading of a text with alternative readings (Fialho, 2007), a process of cognitive decoupling. Furthermore, most literary effects are likely to be emergent, i.e., in need of study and reflection to evolve over time (cf. Dixon et al., 1993), which means that effects of literary reading may depend on sustained de-automatized processing. In contrast with foregrounded elements, backgrounded elements, such as story context and/or imagery a reader is accustomed to (Dixon et al., 1993), activate “familiar cognitive schemata, situation models, and affective responses [...] which correlate with a fluent reading mode” (Jacobs, 2015, p. 7). As such, backgrounded text elements may reinforce System 1 processing in literary reading. An example of this reinforcement is the reception of the Robert Frost (1920) poem *The Road Not Taken*. To most readers this poem depicts individuality and freedom of choice, which may be due to the vivid and familiar imagery in its final lines (Orr, 2015): “I took the road less travelled by - And that has made all the difference”. Other elements of the poem that may problematize this popular interpretation, such as its title, rarely pop up in the meaning making process, even to the extent that the poem is often mistakenly referred to as *The Road Less Travelled By* (Orr, 2015). This example illustrates that – in line with dual process theory – de-automatized literary understanding may depend on the extent to which literary readers inhibit System 1 processes.

Frederick (2005) and West et al. (2012) found that the tendency to inhibit System 1 processing correlates with certain thinking dispositions, mainly the Need for Cognition, defined as the willingness to engage in cognitive tasks, and Actively Open-minded Thinking, which refers to the willingness to think about acts and beliefs outside one’s own frame of reference. Therefore, if dual process theories apply to literary reading, one would expect that readers who possess these critical thinking dispositions might be more inclined to inhibit familiar schemata and to look actively for novel perspectives hidden in the story. So far, we do not know if such an effect exists, but there are indications that reader characteristics do play a role in making meaning from literary texts. Schraw and Bruning (1996) distinguished two epistemic beliefs held by readers: transmission beliefs, that refer to the idea that there is only one correct author’s message to be reconstructed by the reader, and transactional beliefs, that constitute the position that readers should interact with the author and the text to discern a possible meaning out of manifold. They found that only transactional beliefs correlated positively with the number of critical responses towards a text, and with the willingness to consider multiple, ambiguous meanings. Readers with

transactional beliefs seem “more likely to engage a text, rather than merely process it” (Schraw & Bruning, 1996, p. 301).

Recent empirical studies indicate that there might also be a converse relation between literary reading and critical thinking dispositions: literary reading may support the maturation of such dispositions. Djikic et al. (2013) found that literary reading decreases the need for cognitive closure; Kidd and Castano (2013) indicated that reading literary texts supports the development of a cognitive Theory of Mind, i.e., the ability to understand the thoughts of others; and Hakemulder’s (2000) results indicate that reading literature may contribute to the development of moral reflection. Other empirical research has shown that more experienced readers of literature tend to read in a more open manner, with an eye for multiple interpretations, while less experienced readers tend to focus primarily on story events and view the literary text from one perspective only (Earthman, 1992; Andringa, 1995; Janssen et al., 2012). Thus, literary reading might serve as a training module for critical thinking dispositions to develop and, as such, may foster System 1 inhibition.

To sum up, the degree in which literary reading may result in de-automatized literary understanding may depend on the extent to which System 1 processes in literary reading are inhibited. At the same time, reading to understand literary texts might help to stimulate this inhibition. Critical thinking dispositions might play a key role in both these respects.

2.3 Literary reading and focusing on a decision: underlying thinking skills

The last element highlighted in Ennis’ definition of critical thinking is focusing on a decision. It may seem far-fetched to look for similarities with literary reading here, because decision-making is not a part of reading to understand literary texts (cf. Djikic et al., 2013). However, in critical thinking theory a set of specific thinking skills (Facione, 1990; Ennis, 2011) constitutes the decision-making process. We argue that three of these thinking skills – clarification, assessment, and inference – are related to reading strategies that experienced readers of literary texts are likely to apply.

First, clarification skills are needed to establish focus (Ennis, 2011). Relevant questions and arguments first need to be identified before one can attempt to use them in a decision-making process. Clarification is aided by categorization, i.e., the skill to apprehend and use appropriate categories (Facione, 1990), and decoding significance, i.e., the skill to detect and describe underlying functions and motives in communication, such as in language, signs and symbols (Facione,

1990, p. 14). Finding significance behind the surface code, guided by the skill to discern relevant categories, might be typical for (expert) literary readers. They seem to possess “[...] metalinguistic awareness [...] a rich knowledge of literary devices, conventions and common themes [...]” (McCarthy, 2015, p. 109), that enables them to find significance in “discourse aspects of a narrative (e.g., point of view, tone, diction, ‘style’) [...]” (Vipond & Hunt, 1984, p. 271) as opposed to finding significance only in the aspects that keep the story going, such as plot, setting and character (Vipond & Hunt, 1984). Such readers have a point-driven attitude, i.e., they assume 1) that there is a point to the text that is not literally expressed in the text, and 2) that seemingly disparate elements of the text might contribute to establishing that point (Vipond & Hunt, 1984; Miall & Kuiken, 2002; McCarthy, 2015). Thus, both critical thinking and (point-driven) literary reading seem to address skills for identifying the relevant pieces of information out of which meaning is to be inferred.

Second, in critical thinking one needs to assess whether the identified information is reliable before one can make inferences from it (Ennis, 2011). An important sub-skill is judging the credibility of a source, the skill to determine whether the information presented may be colored by the source (Ennis, 2011). This skill may also be of special importance in literary reading because of the implications of point of view in understanding a story. By default, readers are inclined to trust narrator information about the narrative world (Claassen, 2012). To become aware of the undertones in a character’s point of view, readers first have to perceive the text as written by an intentional being who has a particular point to make; a perception of the text that enables them to distinguish authors from narrators (Vipond & Hunt, 1984). Based on this distinction, readers may come to view a character’s thoughts and actions from a meta-standpoint, which may help them to discern ironies (Vipond & Hunt, 1984) and judge whether and how the (un)reliability of the narrator may influence their own position towards the text and the author (Iser, 1978). Thus, both critical thinkers and (point-driven) literary readers tend to assess the reliability of the information they are presented with before making inferences.

Third, critical thinkers need to derive their own conclusion about what to believe or do from the data at hand (Ennis, 2011). Sub-skills that play a role here are 1) deductive reasoning, which is the ability to reach a conclusion by reasoning from the general to the specific, and 2) inductive reasoning, i.e., the ability to reach a conclusion by reasoning from the specific to the general (Ennis, 2011). In reading, the extent to which inferencing takes place seems related to a reader’s goal (Van den Broek et al., 2001, p. 1082). This goal subsequently affects a reader’s “standards of coherence that act as criteria for comprehension [and] in

turn, dictate the inferential activities in which the readers engage at each point during reading" (Van den Broek et al., 2001, p. 1082). This suggests readers might address higher level inferences in point-driven literary reading than in other modes of reading, because in point-driven reading their reading goal sprouts from recognizing the text as an artefact (Vipond & Hunt, 1984). Point-driven readers "[...] recognize intentionality behind, or impute it to, the text, [so] they realize that what is presented in the text [...] cannot be taken at face value but instead must be interpreted in light of the fact that the characters are mere creations of an intentional being" (Vipond & Hunt, 1984, p. 272). In paying attention to the world outside the text as well as the textual world, these literature readers are likely to have a high standard of coherence (McCarthy, 2015, p. 116). Such coherence "requires the integration of multiple parts of the texts as well the activation of prior knowledge into the reader's representation" (McCarthy, 2015, p. 116). Therefore, these readers may use inductive reasoning skills, for instance when trying to infer meaning from relations between characters, and deductive skills, for instance when trying to infer meaning from a metaphor.

To sum up, critical thinkers need to develop clarification, assessment, and inference skills (amongst other critical thinking skills) to arrive at the best possible decisions. Experienced readers of literature may apply these skills when they aim to make meaning from what they are reading.

In conclusion, critical thinking and reading to make meaning from literary texts may involve similar cognitive processes. These processes seem to gain vigor from training. Therefore, the literature classroom might provide a suitable context to foster critical thinking. At the same time, critical thinking may provide the necessary skills and dispositions that help students to evolve their understanding of literature into *Critical Literary Understanding*, hereby defined as de-automatized, (re)constructive meaning making in response to literary texts.

3 RESEARCH QUESTIONS AND HYPOTHESES

Our concept of Critical Literary Understanding (CLU) includes the assumption that it can be discerned in an educational context. One way to test this assumption is to assess CLU's growth within an educational context. Therefore, our first question is: Do upper secondary pre-university students show growth in Critical Literary Understanding after attending literature education for four months? If there is no growth, that would indicate that CLU is not viable as an educational concept. However, we hypothesize that after four months of literature education, which constitutes half a school year, students will show growth in CLU. After all,

our concept of CLU implies that making meaning from literary texts may improve Critical Literary Understanding.

We also assume that critical thinking skills and dispositions each might play a specific role in the development of Critical Literary Understanding. This assumption leads to the second question: Is growth in Critical Literary Understanding mediated by critical thinking dispositions and moderated by critical thinking skills? Since we proposed – in line with dual process theories – that CLU (perceived as a System 2 process) depends on the inhibition of System 1 processes, we hypothesize that critical thinking dispositions will mediate growth in CLU. After all, 1) mediation variables help explain how specific growth is achieved (Hayes, 2013, p. 7), and 2) critical thinking dispositions are correlated with System 1 inhibition. Also, we expect that critical thinking skills will moderate growth in CLU. Moderation variables help explain when/with whom specific growth is more likely (Hayes, 2013, p. 8). We assume that critical thinking skills constitute the mindware (Stanovich & Toplak, 2012; Heijltjes, 2014) people need to focus on and to sustain System 1 inhibition. If that is indeed the case, students with better critical thinking skills are more likely to improve their Critical Literary Understanding over time.

In addition, we formulate a third, conditional, hypothesis. If critical thinking skills moderate and critical thinking dispositions mediate growth in CLU, then we expect that these moderation and mediation variables will explain more of that growth per grade. After all, students with more years of education are more likely to apply their critical thinking skills and dispositions when they make meaning from literary texts, because learning to rely on critical thinking skills and dispositions in an educational context takes time, effort and confidence (Van Gelder, 2005; Heijltjes, 2014).

4 METHOD

4.1 Design

We conducted a prospective cohort study with two measurement occasions: one in October, six weeks after the beginning of the school year (T1) and one in February, halfway through the school year (T2). At T1 we assessed students' critical thinking skills (CTS), critical thinking dispositions (CTD) and Critical Literary Understanding (CLU). At T2 we assessed CLU again, to evaluate if any changes had occurred and, if so, where and how. To reduce test recall we did not report or evaluate test results at T1 and we split the original items we measured at T1 into two distinct records. We then complemented each of the two records with new

items. In our final analysis of T1 and T2 we only considered the scores on the items that were presented at both measurements.

4.2 Setting

We collected data at one Dutch secondary school. This school has implemented a curriculum in literary arts education that includes several pedagogical elements that might foster students' Critical Literary Understanding, by stimulating them to 1) inhibit System 1 processing, 2) (re)construct meaning, and 3) engage in point-driven reading. First, students are stimulated to inhibit System 1 processing. They are systematically encouraged to discern deviations from familiar discourse in the texts they read. Poetical background and literary concepts (theme, symbols, motives, point of view, characters, time, and space) are not presented as a self-contained body of knowledge, but rather as metalinguistic focus points, to help the students notice and reflect on peculiarities in the texts. Second, students are actively stimulated to (re)construct meaning. At the start of each school year, they are asked to define a given topic, for instance 'freedom', and discuss three theorems on that topic in small group discussions. At the end of the school year, they are asked to evaluate their definition and opinions, with reference to the literature portfolio they created during the school year, in which they communicated their meaning making choices with regard to five literary novels. With each of these novels, the meaning making process is also centered around (re)construction: students have to construct literary mind maps in small reading groups, starting with their first reaction towards the text and culminating in the formulation of a shared meaning (Koek, 2010). Students also (re)construct meaning in lessons about historic poetical background: these lessons focus on ideas about life that students have to reflect on, for instance, by depicting the current refugee crisis through the eyes of a naturalistic writer (for example, Emile Zola). Third, the curriculum invites students to focus on point-driven reading. The first literary concept that students are introduced to is 'theme', which is presented to them as the outcome of a meaning making process in which, ideally, all elements are connected.

Students spent three hours per week on literature education: 90 minutes at school and 90 minutes outside school. This implies that in the four months between our two measurements students had spent approximately 48 hours on literature education, during which they read, discussed, and reported on three contemporary novels. Of these novels one was read by all students in a grade (Grade 10: *Tschick*, by Wolfgang Herrndorf; Grade 11: *Extremely Loud and Incredibly Close*, by Jonathan Safran Foer; Grade 12: *Kafka on the Shore*, by Haruki

Murakami) and two were chosen by the students themselves. Besides reading contemporary novels, students learned and discussed literary concepts (Grade 10), medieval and renaissance literature (Grade 11) and early 20th century literature (Grade 12). The school permitted students to read all foreign literary novels and stories in a Dutch translation.

The choice to sample participants from this school was deliberate. Its students follow a clearly outlined curriculum in literary arts, so we expected less variation between teachers' practices than at other Dutch schools, where pedagogical approaches can differ substantively between literature teachers (Janssen, 1998; Verboord, 2005). Moreover, since we wanted to examine whether CLU is discernible within an educational context, it seemed opportune to use a context in which we could actually expect to discern it.

4.3 Participants

We recruited students in Grades 10, 11 and 12. Participants were 14-19 years old ($M = 16.4$, $SD = 1.1$, 52% female), and attended pre-university level education, which is the highest track of secondary education in the Netherlands. 45% of them followed a curriculum in the Humanities or Social Sciences, 55% in the Physical Sciences, in addition to the compulsory subjects Dutch, English, Social Studies, Literature, and Physical Education. A sample of the students ($N = 234$, all grades evenly represented) completed the Author Recognition Test (Schrijvers et al., 2016) and their scores ($M = 8.7$, $SD = 4.8$, out of a maximum score of 40) indicated that, on average, participants were not experienced literary readers. Using a six-point Likert scale, ranging from not in the least to greatly, the students also reported on their interest in sports ($M = 5.2$, $SD = 1$), going out ($M = 5$, $SD = 1.1$), reading ($M = 3.6$, $SD = 1.4$), gaming ($M = 4$, $SD = 1.5$), studying ($M = 2.6$, $SD = 1$) and debating ($M = 3.7$, $SD = 1.2$). These means indicate that our participants had a moderate interest in intellectual activities. Participants' informed consent was obtained via an e-mail about the study's goal and procedure.

4.4 Instruments

Critical Literary Understanding (CLU). The Critical Thinking in a Literary Context (CTLC) instrument was designed to measure CLU on the basis of a format for critical thinking assessment, recommended by Ennis (1993), Halpern (1997) and Heijltjes (2014), among others. In this format each question consists of two test-items: a multiple choice item followed by an item in which participants are asked

to justify their choice. As a result, both the choice itself and the reasons that led to the choice can be assessed, which makes it possible to assess to what degree participants' answers were de-automatized and (re)constructive. This format thereby facilitates a comprehensive assessment of the decision-making process and allows for differences in participants' background and their interpretation of the questions (cf. Ennis, 1993). Critical thinking tests that follow this format (for instance, Halpern, 1997) typically use short logical problems for the questions. Instead, in the CTLC all questions are concerned with making meaning from literary texts. The choice of text fragments and formulation of test-items were based on six design principles we derived from our theoretical framework (Table 2.1).

There was also a practical requirement concerning the length of the fragments. To give all students the opportunity to think about their answers thoroughly and still be able to finish the test within the timeframe of one lesson period (45 minutes), we decided that reading all the literary fragments should take them 10 minutes at most. Since the average reading speed of adult readers (as measured over 17 languages) was found to be 184 ($SD = 29$) words per minute (Trauzettel-Klosinski & Dietz, 2012), we set the maximum number of words for all the literary fragments combined at 1500.

A first version of the CTLC consisted of ten literary fragments, each with 1 to 8 items, with 22 items in total. This version was reviewed by three expert educational researchers and piloted for comprehensibility (Grade 12, $N = 25$). Based on the feedback, we dropped two literary fragments that did not seem to concur with Principle 2 of the design and replaced them with one other fragment. Six items were rewritten because they produced answers that were hard to code. Lastly, we added one item. Thereafter, the CTLC consisted of nine literary texts, each with one to eight items, with 23 items in total. This version was piloted (Grade 11, $N = 28$). The reliability proved to be satisfactory (Cronbach's Alpha .78), so this became the version we used in our study. Table 2.2 shows which literary texts were included and describes the items that refer to these texts. Two full example items from the CTLC can be found in Appendix A. To code students' responses to the open questions of the CTLC, we designed a rubric with levels of critical understanding (Table 2.3).

Table 2.1. *Design principles, rationales and operationalization of the Critical Thinking in Literary Context test (CTLCT)*

Principle	Rationale	Operationalization
The test-items must ask participants to make a decision in relation to the particular text fragment.	Decision making establishes focus (Ennis, 2011) and is a cognitive process in which the degree of System 1-inhibition is often clearly discernible (Kahneman, 2011; Stanovich, 2011).	We formulated each item as a choice. For instance, on what occasion would you read this poem aloud? What would be the best title for this story?
Style and/or content of the literary fragments must allow both a fast, automatized response as well as a de-automatized response.	In decision making-processes, System 1-inhibition is best discerned when an automatized response is strongly provoked by the text (Frederick, 2005).	We chose fragments and poems by acclaimed, prize-winning literary authors that, on the surface, described situations, topics, and characters concretely, allowing for easy recognizability.
The test-items should not contain any references to literary theory.	Such references might trigger mere reproduction of learned knowledge, which would be System 1 processing (West et al., 2008).	We formulated our questions without any reference to literary concepts, with the exception of 'poem', 'story', 'title' and 'character', because we needed those four terms to be able to refer to the texts in a question.

<p>The test should contain a variety of literary genres and topics.</p>	<p>This variety increases the chance that students discern foregrounded textual features (Dixon et. al., 1993) and experience de-familiarization (Shklovsky, 2017), that may trigger reflective thinking (Dewey, 1910).</p>	<p>We used poetry and prose and were careful to incorporate different styles (dialogue heavy, description heavy, high on imagery, low on imagery) and topics (love, science, insecurity, life goals, nature, education) in the material.</p>
<p>Items have to facilitate both a quick response and reflection.</p>	<p>Hereby the items enable assessment of whether reflection on the choice made is (re)constructive (Dewey, 1910; Fialho, 2007; Miall & Kuiken, 1994) and sustained (Dewey, 1910; Ennis, 2011).</p>	<p>In formulating the multiple choice items we took care to ensure that the choices we presented to the participants included one that was easy to link with the text's face-value, as well as choices that would probably be hard to link with the text at first reading, or even might seem to be in contrast with it.</p>
<p>Items have to focus on the thinking skills Clarification, Judging credibility and Inference (deductive and inductive).</p>	<p>These are thinking skills we presume to constitute the mindware needed for sustaining System 1 inhibition in literary reading (Stanovich & Toplak, 2012; Vipond & Hunt, 1984).</p>	<p>We chose to formulate three types of questions: 1) questions that asked participants about what they found striking in a text, or which character they thought stood out in a dialogue (clarification) 2) questions about whether a character's account of a given situation justifies a proposed action (judging credibility) 3) questions about the meaning of the whole text, such as how a fragment/poem with one word or the title left out should be completed again (inference).</p>

Table 2.2. *The Critical Thinking in Literary Context test (CTLC):
Overview of literary sources, genres, word count, type of item, thinking skills and tasks*

Question	Literary text	Genre	Word-count	Item	Type	Thinking skill	Task
1	<i>When I heard the learn'd astronomer</i> [Toen ik de astronoom zo hoorde], Whitman (1900c)	Poem	98	1	MC	Clarification	Decide which lines in the poem are important. Justify your choice.
				2	Open		
				3	MC	Induction	Decide which occasion is suitable for a reading of this poem. Justify your choice.
				4	Open		
2	<i>My sister is a monk sea</i> [Ik ben de sterkste], Frascella (2009)	Novel (fragment)	436	5	MC	Judging credibility	Decide on responsibility of the narrator. Justify your choice.
				6	Open		
3	<i>A game of thrones</i> , Martin (1996)	Syllogism	12	7 – 9	MC	Deduction	Deduce meaning from metaphor.

4	<p><i>Turkish Delight</i> [Turks fruit] Walkers (1969), <i>Interview with Spinvis</i>: Hoorntje & Doesborgh (2012) <i>Still lucky</i> [Nog een geluk dat], De Coninck (1975) <i>Turkish Delight</i> [Turks fruit] Walkers (1969),</p>	<p>Novel (fragment) Interview Poem Novel (fragment)</p>	<p>215 83 83 215</p>	<p>10</p>	<p>Open</p>	<p>Induction</p>	<p>Justify which two fragments taken together would offer a broad perspective on love sickness.</p>
5	<p><i>Autumn</i> [Herfst], Lucebert (2002a)</p>	<p>Poem</p>	<p>9</p>	<p>12 13</p>	<p>Open Open</p>	<p>Induction</p>	<p>Decide whether mothers generally are a factor in their children's relationship problems. Justify your choice. Justify which word fits this poem the best.</p>
6	<p><i>Writer</i> [Schrijver], Sijnders (2009)</p>	<p>Short story (full)</p>	<p>95</p>	<p>14-21 22</p>	<p>MC MC</p>	<p>Clarification Induction</p>	<p>Decide on importance of a sentence for understanding this story Justify which title fits this story the best.</p>
7	<p><i>Uncle Wiggly in Connecticut</i> [Oom Wiggly in Connecticut], Salinger (1954b)</p>	<p>Short story (fragment)</p>	<p>349</p>	<p>23</p>	<p>Open</p>	<p>Clarification</p>	<p>Justify which character behaves most strikingly in this dialogue.</p>

Table 2.3. Rubric with levels of critical understanding in open answers of the Critical Thinking in Literary Context test (CTLIC), with example answers per skill and level

Skill	Non-critical (1 pt) *	Beginning critical literary under-standing (2 pt)	Developing Critical standing (3 pt)	Advanced Critical standing (4 pt)
Clarification	Takes text at face-value. Does not discern peculiarities.	Shows general awareness of peculiarities, a sense that 'something' is or could be strange. Does not elaborate on this.	Describes peculiarities at discourse level. Directly points at what is happening/said/done in the text, also in an inferential way, like noticing something is getting more positive/negative.	Frames peculiarities: describes what is happening/said/done and why these happenings are peculiar.
Example answers (item 23)	They just have a conversation. Nothing peculiar about it.	The mother is a hideous person.	The mother is constantly lecturing her child. This stands out for me.	The mother acts strange. First, she introduces a subject in the conversation (her child's boyfriend), but from thereon she does not seem to be listening anymore.
Credibility of a source	Takes the text for granted; does not question credibility.	Distances from the text by acknowledging the	Acknowledges point of view and (possible) questionable credibility.	Acknowledges point of view and elaborates on why it is credible or not in this specific case.

Example answers (item 9)	He says so in the last sentence.	text as an 'incomplete' source. It might be the case here, but you cannot say that it is always the case.	I don't think the I is very trustworthy. You only see his side of the story.	The way the I describes his ex and her mother shows his emotion, anger. It is therefore unlikely that he is telling the whole story. What was his role in the breakup?
Inference	Does not infer. Direct (emotional) response, focus on what is immediately understood.	Makes inferences, but guessingly (It could be that....).	Makes inferences, but leaves out textual elements, like explaining only one side of a difference.	Makes inferences by integrating multiple elements in the text, and/or outside the textual world, including those that are ambiguous.
Example answers (item 4)	It is about stars and astronomers.	It could be that the I is dying at the end of the poem, going to the stars.	In the last part of the poem the I is alone, and tranquil again. Just like you'd sometimes want to be at a funeral.	There is a difference between the first few lines and the last. The I-figure does not like the cold facts of science, and rather wants to be in nature, wondering.

*answers left blank were awarded 0 points

This rubric was then validated by three trained and experienced literature teachers in three rounds, for each item, under supervision of the first author. In every round the teachers each coded the work of five randomly selected students and compared and evaluated their decisions afterwards, thereby expanding the rubric with examples of answers for each category and for each item. This resulted in a coding scheme. Appendix B presents excerpts from the coding scheme.

Next, the coding scheme was tested for reliability by having 10% of the CTLC (only open items, $N = 50$) coded independently by two raters. The total scores of both raters were strongly correlated (T1: $r = .97, p < .001$; T2: $r = .96, p < .001$), indicating a satisfactory inter-rater reliability. Next, the remaining 90% of the CTLC was divided between the two raters for further coding.

Critical thinking skills (CTS). To assess CTS, we chose the Cornell Critical Thinking Test, level X (CCTT; Ennis et al., 2005) for two reasons. First, it has proven to be a reliable instrument in educational research for the age group 12-18 (Ennis et al., 2005). Second, the assignments represent four distinct thinking skills, of which three – induction, deduction, and judging credibility of a source – correspond to our concept of CLU. The CCTT tells a story about a group of explorers who land on an alien planet and who then have to find out what happened to an earlier group. In 70 multiple choice questions participants had to determine which of two statements presented was more likely to be true, or credible, or if both were equally credible, given the information they were presented with.

Critical thinking dispositions (CTD). We compiled a Critical Thinking Dispositions Questionnaire (CTDQ), consisting of three scales: 1) The Need For Cognition (NFC) scale, that assesses the disposition to examine what is not immediately understood (Cacioppo & Petty, 1982, 2) the Actively Open-minded Thinking (AOT) scale, that assesses the disposition to think about what lies beyond one's frame of reference (West et al., 2012) and 3) the Intolerance for Ambiguity (IFA) scale (Eysenck, 1954), that assesses the disposition to avoid ambiguous thinking. In a recent empirical study, the first two scales were found to correlate with System 1 inhibition (West et al., 2012), and in our conceptualization of CLU we proposed that these dispositions are addressed when trying to understand literary fiction. The third scale was empirically associated with the need for cognitive closure (Webster & Kruglanski, 1994), a need we proposed could be decreased by reading to understand literary fiction (see also Djikic et al., 2013). Using these three scales we presented participants with a total of 68 statements about their thinking dispositions. For each they were asked, on a 6-point Likert scale, to indicate to what degree they thought the statements applied to them. We

translated the CCTT and the CTDQ to Dutch and the translations were checked by a native speaker of English, who provided feedback on the translation.

4.5 Procedures

All tests were administered within two school weeks, in the classroom, by the regular teachers of the particular school subject that was on the schedule. In the written introduction to the tests, we told students that these tests were part of an investigation into how people think when they read literary fiction, and that their answers would not be graded. Teachers who administered the tests were instructed by e-mail to follow the same standardized protocol they always do when administering a school exam.

Student responses to the tests that were clearly not serious or out of context (for instance 'La, la, la', 'D.J. Fredje is the best', or all left blank) were considered as not valid and were removed from the data set. Table 2.4 shows the number of valid responses acquired with each measurement, with the final data set, i.e., the number of participants who completed all four measurements, presented in the last row.

Table 2.4. Number of participants at the first (T1) and second (T2) measurement

Measurement	Grade 10	Grade 11	Grade 12	Total
Critical Literary Understanding T1	141	155	165	461
Critical Literary Understanding T2	122	115	105	342
Critical Thinking Skills	165	162	157	484
Critical Thinking Dispositions	178	167	157	502
Final data set	108	80	83	271

Table 2.4 shows that the number of valid responses decreased at T2. This was due to three factors: 1) one teacher forgot to administer the test in her classes; 2) a relatively large number of students was absent from classes that week due to health reasons or preparations for exam presentations (Grade 12), and 3) our decision to split the measurement of T2 over two sessions records influenced the *N* substantially. Only students who were present at both records could be participants in the study.

General linear modelling showed that the group of students that participated in the study and the group that dropped out differed significantly: Wilks' Lambda

= .97, $F(3, 375) = 3.4$, $p = .02$. Univariate analysis showed that the groups differed significantly only in critical thinking skills, $F(1) = 10.25$, $p = .001$, with the dropped-out group scoring lower ($M = 46.1$; $SD = 10.9$) than the participants ($M = 49.5$; $SD = 8.5$).

In the end, 15 classes were represented in the study, five in each Grade (10–12). The number of students who completed the CTLC at both T1 and T2 varied between 8 to 24 per class, with a mean of 16.

4.6 Reliability

Table 2.5 shows that the internal reliability of the instruments ranged from acceptable to high, except for the Intolerance for Ambiguity scale (IFA) and the subscales of the CTLC. For the subscales this finding corresponds with a problem that scholars often encounter when they want to discern sub-skills while assessing critical thinking: these skills are often too interdependent of each other to assess as separate thinking skills (Ennis et al., 2005).

Table 2.5. Reliability of instruments (Cronbach's Alpha) at T1 and T2

Instrument	Variable	T1			T2		
		N	N	α	N	N	α
			items		items		
Critical Thinking in Literary Context (CTLC)	Critical Literary Understanding (CLU)	461	23	.76	343	23	.71
	Clarification (CLA)	461	11	.60	343	11	.40
	Credibility (CRE)	461	4	.54	343	4	.53
	Inference (INF)	461	8	.63	343	8	.62
Cornell Critical Thinking Test Level X (CCTT)	Critical thinking skills (CTS)	424	70	.86			
Critical Thinking Dispositions Questionnaire (CTDQ)	Critical thinking dispositions (CTD)	502	68	.85			
	Need for cognition (NFC)	502	18	.86			
	Actively Open-minded Thinking (AOT)	502	41	.78			
	Intolerance for Ambiguity (IFA)	502	9	.54			

We decided to leave the IFA scale and the subscales of the CTLC out of our further analyses, and to involve the CCTT, the total score of the CTLC, the total score of the CTDQ, and the scores of the CTDQ-subscales AOT and NFC.

All three instruments were intended to measure criticalness, so we expected positive correlations between the scores they produced. All correlations were indeed significant ($r = .22 - r = .46, p < .01$), indicating construct validity.

4.7 Analysis

We analyzed our data in seven steps. First, to see whether growth in CLU can be discerned in an educational context, we applied general linear model analyses (repeated measures) to determine the effect of time on CLU, for the total sample and for each Grade. Second, to see whether students with more literature education were better at CLU, we then applied analysis of variance, with post-hoc analysis (Scheffé), to discern whether the mean scores of the CTLC differed significantly between grades, at T1 and T2. Third, to see whether the CTLC score at T1 was a predictor of the score at T2, we applied linear regression analysis. Fourth, to see whether this prediction would explain more variance at T2 with critical thinking dispositions (CTD) as mediator variable and critical thinking skills (CTS) as moderator variable, we applied Hayes' SPSS macro PROCESS (Hayes, 2013). Fifth, to check whether the hypothesized role of CTD and CTS explained most variance at T2, we compared our model with four alternative models, each representing either CTD or CTS as moderator or as mediator. Sixth, to check the respective explanatory power of the CTD-subscales Actively Open-minded Thinking (AOT) and Need for Cognition (NFC) we compared their separate effect for the model that explained the most variance.

Finally, we investigated whether the explanatory power of the model we discerned in steps 3-6 was stronger in higher grades by repeating our analysis, this time per grade. In doing so, we focused exclusively on the explanatory power of the model and refrained from a detailed mediator/moderator analysis per grade because of the statistical pitfalls of such analyses for subgroups (Hayes, 2013). We used an alpha level of .05 for all statistical tests.

5 RESULTS

5.1 Descriptive results

Table 2.6 presents the mean scores and standard deviations at T1 and T2 of the participants that were present at both measurements of the CTLC. From this

table it can be understood that for all grades students' mean scores were around 60% of the maximum score of 62. No student reached a perfect score; the scores ranged between 27% and 85% of the maximum.

Table 2.6. Descriptive results Critical Thinking in a Literary Context test (CTLC) at T1 and T2

	Grade 10 (N = 116)		Grade 11 (N = 96)		Grade 12 (N = 96)		Total (N = 308)	
	M	SD	M	SD	M	SD	M	SD
T1*	37,4	4,1	35,9	6,3	40,4	4,6	37,9	5,3
T2*	38,9	3,9	39,4	5	41,4	4,7	39,8	4,6

*max = 62

Table 2.7 represents means, standard deviations, and minimum/maximum scores for the Critical Thinking Dispositions Questionnaire (CTDQ) and Cornell Critical Thinking Test Level X (CCTT) of the participants that were present at all measurements of all instruments (the final dataset).

Table 2.7. Descriptive results Cornell Critical Thinking Test (CCTT) and Critical Thinking Dispositions Questionnaire (CTDQ)

Test	Grade 10		Grade 11		Grade 12		Total	
	M	SD	M	SD	M	SD	M	SD
CCTT *	48,9	7,2	47,8	10,4	51,5	7,8	49,4	8,6
CTDQ (tot)**	3,9	.33	3,9	.32	4	.30	3,9	.32
CTDQ (AOT)	4,1	.33	4,2	.30	4,2	.33	4,2	.32
CTDQ (NFC)	3,9	.58	4	.66	4	.58	3,9	.60

*Mean of total score (Min = 0, Max = 70); ** mean scores (6-point Likert scale)

From table 2.7 it can be inferred that on average our participants scored around 70% of the maximum score at both the CCTT and the CTDQ. Means and standard deviations of the CCTT were on par with those obtained from measurements at American high schools and freshman colleges (N = 22, Ennis et al., 2005).

These results indicate that our participants showed a level of critical thinking skills and dispositions common for their age group and level of education.

5.2 Growth in Critical Literary Understanding (CLU)

Table 2.6 shows that on average students improved their score on the CTLC at T2, indicating growth in CLU. For the total sample this effect of time was significant: Wilks' Lambda = .86, $F(1,307) = 48,3, p < .001, \eta = .14$, also for each separate Grade: Grade 10, Wilks' Lambda = .87, $F(1, 115) = 16,91, p < .001, \eta = .13$; Grade 11: Wilks' Lambda = .75, $F(1, 95) = 31,7, p < .001, \eta = .25$; Grade 12: Wilks' Lambda = .95, $F(1,95) = 4,7, p = .032, \eta = .05$.

Analysis of variance showed that Grades differed significantly in mean score at T1, $F(2, 305) = 20,04, p < .001, \eta = .12$, and T2, $F(2,305) = 8,55, p < .001, \eta = .05$. Post-Hoc analysis revealed that at T1 and T2 Grade 12 differed significantly from Grades 10 ($p < .001$) and 11 (T1: $p < .001$; T2: $p = .009$).

5.3 Mediating and moderating variables

Table 2.8. Standardized regression coefficients (β) to explain variance in Critical Literary Understanding at T2

Model	Direct and (indirect) effect X-Y	Mediator (M)		Moderator (V)	Ex- plained variance (R ²)	F-ratio (Df)
		X – M	M – Y			
0	.50** (----)	-----	-----	-----	.25	$F(1, 269) = 90,5^{**}$
1 MED=CTD MOD=CTS	.40** (.06)	.24**	.26**	.14**	.37	$F(4, 266) = 38,9^{**}$
2 MED=CTD	.43** (.07)	.24**	.29**	-----	.33	$F(2, 268) = 66,3^{**}$
3 MOD=CTD	.43** (----)	-----	-----	.01	.33	$F(3, 267) = 44^{**}$
4 MED=CTS	.46** (.05)	.31**	.15*	-----	.27	$F(2, 268) = 50,1^{**}$
5 MOD=CTS	.45** (----)	-----	-----	.15**	.31	$F(3, 267) = 39,1^{**}$

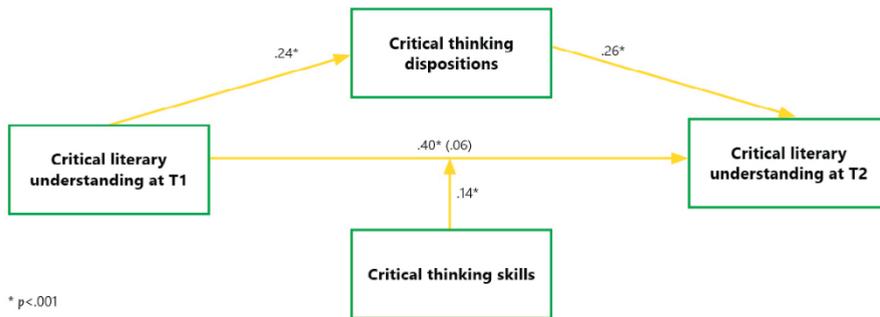
* $p < .05$; ** $p < .001$. X stands for CLU at T1 (predictor), Y for CLU at T2 (predicted), M for mediation variable and V for moderation variable.

CLU at T1 was a predictor of CLU at T2 ($R^2 = .25, p < .001$). To test mediating and moderating variables, we compared this explained variance in CLU at T2 with that of our proposed explanatory model: critical thinking dispositions (CTD) as mediator, critical thinking skills (CTS) as moderator. We also checked whether other models explained more variance (Table 2.8).

As Table 2.8 shows, more variance was explained in models in which CTD and/or CTS was a mediator variable and in models in which CTS was a moderator variable. Model 1 explained most variance (37%).

Mediation. Figure 2.1 depicts model 1. Growth in Critical Literary Understanding (CLU) was mediated by critical thinking dispositions (CTD): the standardized indirect effect of CTD was .06, which was significant at the 95% confidence interval (.03-.11).

Figure 2.1 Model of mediation and moderation effects (standardized β) in development of Critical Literary Understanding (CLU)



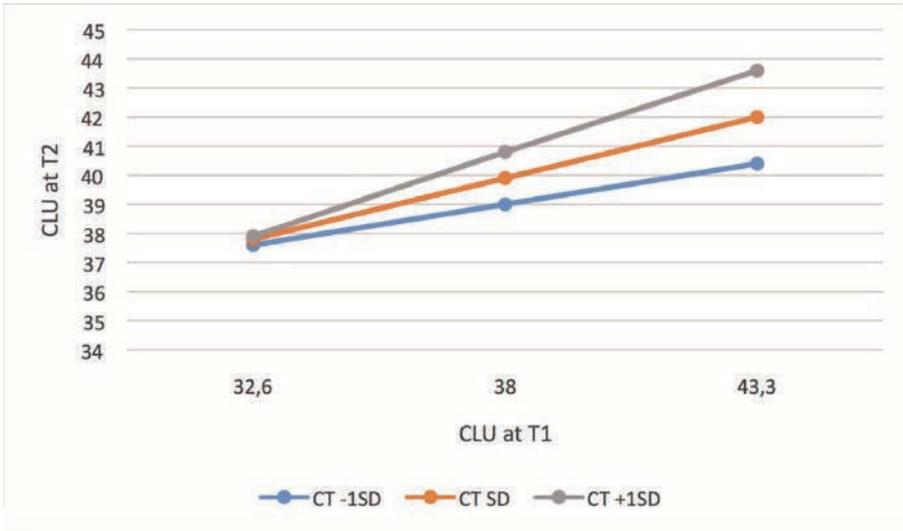
Taken separately, Actively Open-minded Thinking (AOT) and Need for Cognition (NFC) both had a statistically significant mediation effect ($p < .001$), with AOT producing more indirect effect (.06, significant at the 95% confidence interval, .03-.10) than NFC (.04, significant at the 95% confidence interval, .02-.08).

Moderation. Critical thinking skills (CTS) was a moderator variable for the outcome at T2 (see Figure 2.1). The effect of CTS increased with higher values of CLU at T1 (see Figure 2.2). So, students with higher CTS at T1 were more likely to improve their CLU over time.

The Johnson-Neyman technique showed that the moderating effect of CTS was significant from a CCTT score of 34.7 up to 64, which concerned 92.6% of the sample.

Overall, our results show that, for the total sample, critical thinking skills moderated, and critical thinking dispositions mediated growth in Critical Literary Understanding.

Figure 2.2 Simple slopes of growth in CLU over time for three levels of critical thinking skills: one SD below the mean, the mean, and one SD above the mean



5.4 Explained variance per Grade

The explanatory power of the model was stronger in higher grades, indicating that the part that CTD and CTS play in the development of CLU is larger with students who had more years of literature education (see table 2.9).

Table 2.9. Explained variance (R^2) in CLU at T2, per grade

Grade	Explained variance	Significance of overall model
10	.33	$F(4,103) = 12,7, p < .001$
11	.43	$F(4,75) = 13,9, p < .001$
12	.44	$F(4,78) = 15,5, p < .001$

6 DISCUSSION

Do students in secondary literature education engage in critical thinking when they read to understand literary texts? We proposed that they do, on account of multiple similarities between cognitive processes in critical thinking and reading to understand literary fiction. To make it possible to detect students' critical thinking in the context of literary education, we conceptualized Critical Literary Understanding, defined it as de-automatized, (re)constructive meaning making in response to literary texts, and designed and tested an instrument to measure it.

We hypothesized that students would improve their Critical Literary Understanding in four months, during which they followed, among other subjects, a curriculum of literary arts that explicitly addresses their critical thinking skills and dispositions. On account of dual process theory and philosophical conceptualizations of critical thinking, we further hypothesized that critical thinking skills would moderate growth in Critical Literary Understanding and critical dispositions would mediate it.

The results of the present study confirm both hypotheses. Our findings suggest that critical literary understanding can be improved in an educational context, that this improvement can be partly explained by critical thinking dispositions that might have been stimulated, and that it is partly conditioned by students' critical thinking skills.

Finally, we checked whether the role critical thinking skills and dispositions play in making meaning from literary texts increases with years of literature education. Students in higher grades have been subjected to more assignments that may have invoked their critical thinking skills and dispositions than students in lower grades. They therefore may have become increasingly inclined to address their skills and dispositions when they read to make meaning from literary texts, compared to lower grade students. The results of the present study also confirm this hypothesis.

6.1 Limitations

Our study does not show whether the average growth in CLU that we found is the result of literature education. Maturation or other school subjects that our participants attended during the four months of our study could also explain growth in CLU. However, the aim of the present study was not to investigate a direct effect of literature education on CLU, but to empirically test the concept of CLU. Because that concept implies that the literature classroom potentially is a suitable environment for CLU to evolve, we chose a clearly outlined and

comprehensive curriculum of literary arts as a context for our study. This enabled falsification of our concept: finding no growth in CLU in this particular context would have indicated that CLU is not a viable educational concept after all.

One could argue that our results are not generalizable. We conducted our measurements at just one school, with a particular curriculum in literary arts, without including a control group. We do not know whether growth in critical literary understanding can be discerned in other curricula, at other schools. However, our aim was to test that literary understanding may involve addressing critical thinking skills and dispositions, within an educational context. Therefore, we chose a school where we could expect to discern it, and we do not claim generalizability.

We reported considerable data reduction at the second measurement (T2). However, this data reduction does not invalidate our results. The group of students that participated in the study and the group that dropped out differed significantly only in critical thinking skills. This implies that if data reduction had not occurred, our results would have shown more growth in critical literary understanding and a stronger moderation effect of critical thinking skills, which would have strengthened the relations we reported in this paper.

The average increase in test scores that we found could be perceived as a test-retest effect. However, we argue that we did measure growth. First, we took technical measures to lessen the test-retest effect: at T2 we mixed the original items of the CTLC with new, similar items and text fragments. We did not code nor analyze these new items, but they did serve to change the appearance of the test at T2. Second, the questions in the CTLC were reflective, not reproductive. We did not give any feedback on the results, so if students did recognize a question, they had no way of knowing the quality of their previous answer. Third, our moderation/mediation analysis implies growth between the measurements because nothing would have been mediated or moderated if there were no growth.

6.2 Deliveries and future studies

Our study's main delivery is the Critical Thinking in Literary Context test (CTLC). Following the construction parameters and the internal reliability, this instrument presents a first step in measuring Critical Literary Understanding. Still, in subsequent studies further validation should be undertaken, for instance by comparing scores of the CTLC with scores on instruments that assess other literary reading variables, like the Literary Response Questionnaire (LRQ) (Miall & Kuiken, 1995). Three of its subscales - Leisure Escape, Story-Driven Reading, and

Rejection of Literary Values – seem to refer to modes of reading that do not address cognitive decoupling (the main feature of System 2), while three other subscales - Insight, Imagery Vividness, and Concern with Author might refer to modes of reading that do. Therefore, we would expect no or negative correlations between the CTLC and the former subscales of the LRQ, and positive correlations with the latter.

A second delivery is our conceptualization of Critical Literary Understanding. In integrating literary understanding with dual process theory, we distinguished an automatized, impulse-driven level of literary understanding from a de-automatized, (re)constructive level of meaning making in response to literary texts. This concept and its underlying theory are corroborated by our results and are open to testing in subsequent studies.

A next step in investigating the potential of literature education in fostering critical thinking would be to study which pedagogical interventions could stimulate growth in Critical Literary Understanding. Our model presents an answer to the question how, via critical thinking dispositions, and under what condition, a level of specific critical thinking skills, such a growth comes about. This suggests that future studies could focus on investigating pedagogical interventions that enlarge critical thinking skills and/or stimulate critical thinking dispositions in the literature classroom.

6.3 Conclusion

The main conclusion of this study is that the literature classroom is potentially a suitable environment for “[...] reasonable, reflective thinking focused on deciding what to believe or do” (Ennis, 2011, p.1). In providing an empirically tested concept of Critical Literary Understanding and an instrument to assess it, this study may inspire subsequent investigations into how teachers of literature can implement critical thinking in their lessons. In light of the marginalization of literature education (Fialho et al., 2011) and the small role of critical thinking in current educational practice (Elder, 2015), research into this potential of literature education becomes all the more urgent, maybe even critical.