



## UvA-DARE (Digital Academic Repository)

### The story, the self, the other

*Developing insight into human nature in the literature classroom*

Schrijvers, M.S.T.

#### Publication date

2019

#### Document Version

Other version

#### License

Other

[Link to publication](#)

#### Citation for published version (APA):

Schrijvers, M. S. T. (2019). *The story, the self, the other: Developing insight into human nature in the literature classroom*. [Thesis, fully internal, 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, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

## CHAPTER 5

### EFFECTS OF DIALOGIC LITERARY INSTRUCTION ON 10<sup>TH</sup> GRADE STUDENTS' INSIGHT INTO HUMAN NATURE

In this study, we aimed to investigate the effects of dialogic literary instruction on adolescent students' insights into human nature and other transformative reading experiences such as identification, imagery and sympathy for characters. A four-unit intervention centered around short literary stories with various social-moral themes. Students engaged in five types of tasks: 1) tasks to activate prior knowledge about and personal experiences with the theme of a story to be read, 2) internal dialogues with the story, focused on transformative reading experiences, 3) external dialogues in small groups or pairs about the story theme and their reading experiences, 4) external dialogues at the classroom level, and 5) reflection on their learning. Participants were 603 students in grade 10 (13 teachers; 22 classes) of six secondary schools in the Netherlands. In a quasi-experimental pretest-posttest design with switching replications, we assessed students' perceptions of learning via a written learner report task, and their learning outcomes in terms of insight into human nature via a questionnaire and a written story response task. In their learner reports, students repeatedly indicated that the intervention affected their skills to talk about stories, transformative reading experiences, and thinking and opinionating skills. However, questionnaire and story task responses showed no consistent intervention effects on students' insight into human nature. We discuss possible explanations, for example, that the intervention has been too skills-oriented or that students were not fully engaged in what they were reading, as well as implications for practice and future research.

#### 1 INTRODUCTION

I hope that my fear will overwhelm my senses. That I won't be conscious when the airplane crashes. [...] I knew this was going to happen. I knew it. I try to get my wife's attention, but mortal fear has fossilized her. In her eyes, I see that I won't be able to reach her anymore. I cry and I scream that I love her. She doesn't respond. Her gaze is far away. A crucifix peaks from her folded hands. (Koeleman, 2016; translation MS).

The short story *Flight behavior* immediately draws the reader into a frightening situation: facing an inevitable death in a plane crash. We may relive the fear we felt during bad turbulence, consider how we would react in such a situation, or think about why religion offers comfort to some of us when we face our worst fears, as it does for the protagonist's wife, while it does not for others. Like

many short stories, *Flight behavior* offers starting points for considering ideas about human nature: how would we react to a story situation that might happen to us in real life? What kind of human being do our reactions make us? Would others, different from ourselves, react differently? How do we feel about how they think and behave?

Empirical research suggests that one of the merits of reading fictional and literary texts – novels, stories, poetry – is that it offers readers insight into themselves and others (Hakemulder, Fialho, & Bal, 2016; Koopman & Hakemulder, 2015). Fictional worlds have been theorized to be simulations and abstractions of social experience (Mar & Oatley, 2008). Reading fictional and literary texts may thus be a *transformative* experience (Fialho, 2012; 2018), because, in simulating social experiences, it may alter readers' perceptions of themselves and themselves in relation to others. In other words, it may develop their insight into what it means to be human, into human nature.

In the present study, we apply these insights from empirical literary studies to the context of the literature classroom. We investigated whether an intervention in secondary school literature classrooms in the Netherlands could foster students' insight into human nature. We focused on 15-year-old students in 10<sup>th</sup> grade of the higher general secondary education track, which is the second highest track in the Dutch educational system. It prepares for higher vocational education, but not for university studies. In this track, formal literature teaching starts in 10<sup>th</sup> grade. Literary instruction is a subdomain of Dutch language class, like writing and grammar instruction. This particular educational context offers opportunities to develop students' insight into human nature, but it also poses a number of constraints. These will be discussed after we have described how we conceptualize "insight into human nature" and "transformative reading" in this study.

### 1.1 *Conceptualizing Insight into Human Nature*

Philosophers and literary scholars alike have argued that reading fictional and literary texts may affect how we perceive both ourselves and others, or in our terms, impact readers' insight into human nature (e.g., Keen, 2007; Nussbaum, 1995; Zunshine, 2006). It may include – but is not limited to – insight into previously unrecognized personal qualities or shortcomings, insight into self-other relations, understandings of and altered attitudes toward individual others and groups of people, and considerations of difficulties or moral dilemmas that people may face. Empirical studies support the potential of reading fictional and literary texts as a catalyst for developing insight into human nature, a pro-

cess also described as transformative reading (Fialho, 2012; Hakemulder et al., 2016; Koopman & Hakemulder, 2015).

In search of components that underlie transformative reading, Fialho (2018) found that adult readers' responses reflected: vivid imagery of the setting and characters in a story (*imagery*); recognition of aspects of self or others in characters (*identification*); enactment and embodiment of characters' experiences (*experience-taking*); evaluations of characters, positively or negatively (*character evaluation*); feelings of sympathy and compassion for characters (*sympathy*); awareness of particularly striking words, phrases or sentences (*aesthetic awareness*); and new or deeper insights into themselves and others (*self-other insights*). For adult readers, the first six components appeared to precede insight into oneself and others, or what we call in this paper, insight into human nature. Together, the seven components form a preliminary model of transformative reading (Fialho, 2018). Transformative reading, thus, is considered a mode of reading in which insight into human nature comes about.

This model of transformative reading has not yet been validated for adolescent readers. Yet, studies with adolescent participants suggest that they may have similar experiences. Adolescents expressed in interview studies that they, as a result of fiction reading, reflected on who they would or would not like to become (i.e., possible future selves, Richardson & Eccles, 2007), that they better understood other people's experiences and therefore felt connected to others or saw new possibilities for their own lives (Rothbauer, 2011), and that they compared their own lives to story situations and experienced empathetic engagements with characters' feelings (Charlton, Pette, & Burbaum, 2004; Ross, McKechnie, & Rothbauer, 2006). All in all, it appears as if adolescents may also gain insight into human nature as a result of reading.

### 1.2 *Developing Insight into Human Nature in the Literature Classroom*

Research suggests that literary instruction must meet certain preconditions to enable students' development of insight into human nature. First, *engaged reading* may positively affect students' personal, social, and moral development (Ivey & Johnston, 2015; Lysaker, Tonge, Gauson, & Miller, 2011). Because engaged reading is considered to be relational and dialogic, it is assumed to offer opportunities for 'self- and other-construction' (Ivey & Johnston, 2015, p. 301; Rosenblatt, 1938/1983). Indeed, evidence suggests that engaged reading influences readers' social imagination and social behavior (Kaufman & Libby, 2012; Lysaker et al., 2011).

However, whereas Ivey and Johnston (2015) indicated that freedom in choosing reading materials is important for engaged reading, and by extension for personal, social, and moral development, a recent review of empirical intervention studies (Schrijvers, Janssen, Fialho, & Rijlaarsdam, 2018) suggested otherwise. In only one literature classroom intervention that fostered students' insight into human nature, students selected reading materials from a preselected list (Vezzali, Stathi, & Giovannini, 2012); in other interventions, students did not choose their texts themselves, but nonetheless developed insight into human nature. The review suggested that, rather than freedom of choice, *text themes* are important for fostering insight into human nature: fictional texts were used that were thematically relevant for intervention aims, such as texts addressing social relations or moral dilemmas (e.g., Adler & Foster, 1997; Malo-Juvera, 2014; 2016; White, 1995).

The importance of dialogue in the literature classroom for fostering insight into human nature, as suggested by Ivey and Johnston (2015), was supported by the recent review study (Schrijvers et al., 2018). In reviewed intervention studies, reading was of a dialogic nature in two ways. First, writing tasks were implemented that prompted students to activate previous personal experiences before reading, to notice their experiences during reading, and/or to write (reflective) responses directly after reading (Eva-Wood, 2004; Halász, 1991; Malo-Juvera, 2014; 2016; White, 1995). As such, they were encouraged to engage in an *internal dialogue* with the text that generated an initial process of meaning-making or, in Rosenblatt's (1938/1983) terms, a transaction between reader and text. Because secondary school students have been found to often read fictional and literary texts superficially, in a uni-dimensional and rather closed way (Andringa, 1995; Earthman, 1992; Janssen, Braaksma, Rijlaarsdam, & Van den Bergh, 2012), learning to actively interact with texts in internal dialogues may help them to construct interpretations, and subsequently, to develop insight into human nature beyond the process of reading.

Second, initial processes of meaning-making were combined with or followed by verbally sharing personal responses related to texts and text themes with peers, in *external dialogues* (e.g., Adler & Foster, 1997; Eva-Wood, 2004; Malo-Juvera, 2014; 2016; White, 1995). Such dialogues were organized in pairs, small groups and/or at the classroom level. Students could explore multiple perspectives on a text and the themes or issues it addresses. Via such explorations, they may come to understand that a variety of interpretations can be valid, as fictional and literary texts are assumed to be, in essence, multi-interpretable. This process may, by extension, elicit alternative or elaborated views, ideas and perceptions of self and others in relation to texts and themes.

Likewise, Schrijvers, Janssen, Fialho, and Rijlaarsdam (2016) showed that if literature teachers reported to allow for more interaction and dialogue in the classroom, their students more often reported to have learned about themselves and others.

### 1.3 *Constraints in Dutch Literature Classrooms*

Although a dialogic approach to literature teaching seems promising for developing students' insight into human nature, implementing such an approach in Dutch literature classrooms is subject to several constraints. These partly determined the design of the intervention in this study.

First, even though teachers and curriculum organizations may strive for students' personal and social development (Dutch Institute for Curriculum Development, 2015; Janssen, 1998; Oberon, 2016), it is not explicitly mentioned as one of the three global objectives that are formulated in the Dutch literature curriculum: in examinations, teachers assess students' literary-historical knowledge, their structural-analytical skills, and their ability to reflect on their literary reading experiences and development (Dutch Institute for Curriculum Development, 2012). Moreover, attending to objectives additional to those mentioned above may be complicated by time constraints. In upper general secondary education (grade 10 and 11), on average only 30 to 40 minutes per week are devoted to literature teaching (Oberon, 2016).

A second constraint pertains to limited choice in text materials for students (Ivey & Johnston, 2015; Lenters, 2006). Up to 9<sup>th</sup> grade, Dutch students mostly read children's and young adult literature, but from 10<sup>th</sup> grade onwards most teachers require students to read increasingly complex literary texts intended for adult readers. Students in 10<sup>th</sup> grade have little experience with such texts, but their freedom to choose otherwise is limited. In adhering to teachers' objective to familiarize students with literary texts of a certain complexity, students' freedom in choosing text materials was also limited in the intervention designed for the present study. Text selection was mainly informed by thematic considerations, as the review study by Schrijvers et al. (2018) suggested that text themes are important in fostering students' insight into human nature. In most Dutch schools, there are no constraints for text choice on the thematic level: it is possible to select literary texts intended for adult readers that address social-moral themes.

Third, a dialogic approach in the literature classroom may be new and unfamiliar for students. This applies to actively attending to their own responses and meaning-making processes in internal dialogues during reading, as well as

to sharing responses with others. In Dutch literature classrooms, responding to literary texts mostly happens *after* reading, individually, and in writing, for instance by writing a book report or review, collecting these reports and other writing tasks in a portfolio, completing a written test on a book (Dirksen, 2007; Oberon, 2016). Responding verbally mostly happens individually as well: students give book presentations in class, or take oral examinations which are often shaped as analytical, reflective conversations that mostly take place between a single student and the teacher, who hardly can be considered an equal conversation partner. All in all, it may be challenging for students to monitor their reading experiences and responses during reading in internal dialogues, as well as to exploratively share these responses with peers in external dialogues.

#### 1.4 *Research Questions*

The aim of the present study was to test the effects of a literature classroom intervention, characterized by attention for transformative reading experiences and by dialogic learning tasks, on students' transformative reading experiences which include their insight into human nature. To yield a first indication of whether the intervention aims were achieved, our first research question concerned students' perceptions of learning; second, we focused on the effects of the intervention in comparison to a control condition:

1. Which learning experiences do 10<sup>th</sup> grade students report at the end of an intervention that is based on a model of transformative reading and that stimulates them to engage in internal and external dialogues with and about short stories with social-moral themes?
2. Does the intervention have a positive effect on students' transformative reading experiences, such as insight into human nature, imagery, identification, and sympathy, in comparison to students who participate in an untreated control condition?

## 2 METHOD

### 2.1 *Research Design*

In a quasi-experimental pretest-posttest design with switching replications, group A (11 classes,  $n = 311$ ) first participated in the intervention and then in an untreated control condition, and group B (11 classes,  $n = 292$ ) vice versa (see Figure 5.1). The strengths of this design lie in the fact that it allows for assessing

the sustainability of potential intervention effects (measurement 3 for group A), as well as their replicability (measurement 3 for group B; see Shadish, Cook & Campbell, 2002).

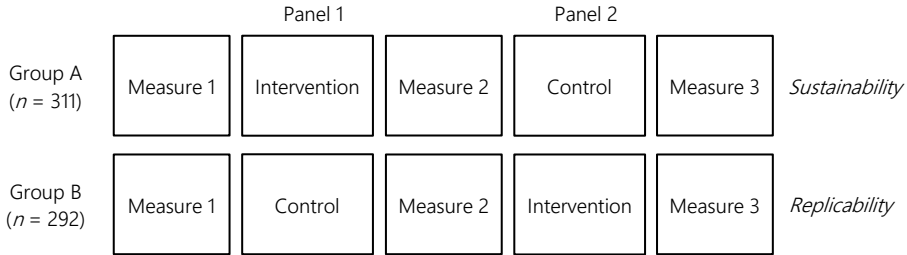


Figure 5.1. Research design.

## 2.2 Participants

We recruited teachers via social media groups for teachers of Dutch and via our network. We found 13 teachers willing to participate, who worked in six different schools across the Netherlands. They participated with either one or two classes. On average, they were 46.2 years old and they had 13.3 years of teaching experience. Two teachers were male. Teachers and their school leaders gave active consent for participation. We randomly assigned teachers to group A or B.

Initially, 23 classes with 634 students participated in the study. Students' parents received an informed consent letter and could object to their child's participation. None of them withheld their consent. Class size ranged from 13 to 33 students ( $M = 27.4$ ,  $SD = 4.2$ ). In one class – of a teacher who initially participated with three classes – data collection procedures were not executed according to plan, which led to a large amount of missing data at the second measurement. Therefore, we removed this class from the data. Thus, 22 classes remained, with 603 students. Students in group A and B did not differ in terms of gender (50.2% female in group A; 56% in group B;  $\chi^2(1) = 2.06$ ,  $p = .151$ ), age ( $M = 15.9$  years in group A;  $M = 15.8$  in group B;  $t(575) = 1.36$ ,  $p = .176$ ), grade for literature ( $M = 6.2$  for group A on a grading scale of 1-10;  $M = 6.3$  for group B;  $t(531) = 1.19$ ,  $p = .235$ ), and familiarity with fiction as measured by an Author Recognition Test (see p. 135;  $M = 4.9$  for group A;  $M = 5.1$  for group B;  $t(537) = .82$ ,  $p = .411$ ).

### 2.3 *Intervention*

Taking into account theory and previous empirical research on developing insight into human nature in the literature classroom, transformative reading, and dialogic literary instruction, as well as constraints that Dutch literature classrooms are subject to, we developed a four-unit intervention, titled Transformative Dialogic Literature Teaching, or TDLT for short. Teachers of Dutch language and literature were consulted in the design process. This process, as well as the validity and practicality of the intervention, are described in another paper (see Chapter 4).

During the intervention, students received four units of 50 minutes of instruction, over a period of two to four weeks in April and May 2017, depending on the schedules of the participating teachers. TDLT was restricted to four units due to the time constraints as described in Section 1.3 of this chapter; teachers who were consulted in the design process also expressed their concerns about implementing an intervention of a longer duration.

The overall objective of TDLT was to help students to identify connections between short literary stories and themselves (i.e., their personalities, the way they are, the way they think) and their views on the social world (i.e., how other people are, behave and think). TDLT consisted of one preparatory unit and three reading-and-dialogue units. Appendix C provides a more detailed description of the four units and the stories that were read during these units.

#### 2.3.1 *Preparatory unit*

The purpose of this unit was to familiarize students with guidelines for small-group dialogues. These external dialogues were a major part of the subsequent units, but we assumed students were not by definition familiar with engaging in them. Students were therefore introduced to a strategy for conducting small-group dialogues about stories and reading experiences, via explicit instruction, modeling, and practice (e.g., McKeown, Beck, & Blake, 2009; Merrill, 2002; Guthrie & Wigfield, 2000). Students first were asked to consider what characterizes good and less good dialogues and were then given five guidelines that were used throughout the intervention: Listen carefully, Ask follow-up questions, Postpone a first judgment, Equally distribute speaking time, and Deepen the content of the talk (in Dutch, these formed an acronym that translates as “fluent”; students were encouraged to engage in a “fluent dialogue”). Next, students observed their teacher who modeled how to ask follow-up questions. Subsequently, students discussed a reading-related topic in small groups, that is, which considerations they took into account when choosing a book to read.

They used cards with follow-up questions to practice using such questions. Finally, experiences that were discussed in each group were shared in class.

### 2.3.2 *Reading-and-dialogue units*

In each of the three subsequent units, students read a short literary story (see Table 4.2, p. 103). The stories centered around various social-moral themes like family and parent-child relationships, mistreatment of children, religion, and prejudice and racism. All stories could be read in 7-15 minutes. The dialogic approach in these units was operationalized by including five types of learning activities, in a fixed order, that were implemented after a short teacher-led introduction. Thus, in each unit, the same basic structure was applied:

1. Activating prior knowledge about and personal experiences with a story theme, prior to reading, to prepare for internal dialogues. For example, before reading the story *She was everywhere* in unit 2, students were asked to write down their thoughts about how someone might react when a relationship ends (time: 2-3 minutes).
2. Internal dialogue with a story: noticing responses during reading and individual reflection on the reading experience directly after reading; in line with the theoretical-empirical model of transformative reading, attention focused in particular on experiences like imagery, identification and sympathy. For example, in unit 2, students were asked to focus on the responses a story evoked in them during reading. Directly after reading, they were asked to indicate to which extent they had noticed experiences such as imagery, identification and sympathy. Thereby, they determined what kind of reading experience was prominent to them, to prepare for external dialogues (time: 10-15 minutes).
3. External dialogue in small groups or pairs: completing tasks that stimulated sharing personal responses, for example, by comparing and contrasting. For instance, in unit 2, students who had indicated that they felt sympathy for a character were asked to compare the moments in the story when each of them experienced this, and to share in their group what they thought and felt at those moments. Next, they were asked to brainstorm about what kind of help would be of avail to the protagonist. As a third step, they reached a conclusion about what kind of help they would offer the protagonist, by talking about issues like: how feasible would the ideas be? What would be best for the protagonist? How would you take action? How would the protagonist respond? Students were asked to take notes of dialogic tasks; in this case, for instance, one student in the group would

write down the ideas that emerged during the brainstorm (time: 15-20 minutes).

4. External dialogue at the classroom level, to explore an additional layer of perspectives. In small-group dialogues, students had shared and compared their responses to a story. In the final learning activity, they were asked to share their group's outcomes with the rest of the class. For example, in unit 2, this enabled students to realize that they were entitled to deepen a particular reading experience (e.g., imagery), but that other groups of students had explored other kinds of experiences (e.g., identification, sympathy). We assumed that this second layer of exploration would help students to consider alternative or elaborated self-other perceptions in relation to texts and themes (time: 5-10 minutes).
5. Short written reflection: students reflected individually on their learning in the unit. For example, in unit 1, students were asked to indicate what went well during the small-group dialogue and what could be improved; in unit 3, they reflected on whether their views on a particular issue (as activated before reading) had changed after reading and talking about a story.

### *2.3.3 Role of the teacher*

Teachers were asked, in preparatory face-to-face meetings with the first author as well as in a written teacher manual, to minimize their interference during small-group dialogues, to allow for a safe space in which students would be free to express and share their ideas, thoughts, experiences, and emotions. When observing a need for guidance or feedback, teachers were asked to provide this in a non-directive way, for instance, by prompting questions instead of explaining what a story is about.

## *2.4 Instruments*

We collected data on four types of variables: students' backgrounds, their perceptions of learning, their learning outcomes, and implementation fidelity of TDLT.

### *2.4.1 Background variables*

*Demographics.* Students indicated their gender, age, exam subjects (educational profile), and most recent grade for Dutch and literature. If no separate grade had been received for literature, we asked students to estimate their grade.

*Author Recognition Test.* We assessed students' familiarity with fiction prior to the intervention with an adapted version of the Author Recognition Test (originally developed by Stanovich & West, 1989). An adapted version of the original test was used, which has been validated for Dutch-speaking teenagers (Koek, Janssen, Hakemulder, & Rijlaarsdam, 2016; Schrijvers et al., 2016). The ART has predictive validity for real-world reading, while avoiding socially desirable answering to questions about reading frequency and motivation (Mar, Oatley, Hirsh, dela Paz & Peterson, 2006; Rain & Mar, 2014). The test presents a list of contemporary and canonical authors' names as well as foils. Participants encircle what they believe to be real authors' names. Its instruction read: "Encircle those names which you know for sure are authors' names. Some of these people are not authors, so do not guess." The number of correctly recognized names minus the wrongly encircled names indicates one's familiarity with fiction. We created three versions, varying the order of the listed names.

The ART was administered as pretest only. It was completed by 539 students (response rate 89.4%). Scores could range from -40 (all foils but no real names encircled) to 40 (all real names and no foils encircled); students scored on average 5.0 ( $SD = 3.3$ ), with a range from -7 to 21. Their familiarity with fiction, thus, seemed to be rather low.

#### 2.4.2 *Perceptions of learning: learner report task*

To measure students' perceptions of learning, we asked them, at the end of the last TDLT unit, to write in their workbooks what they had learned from the project. This task was inspired by the open-ended learner report used by Schrijvers et al. (2016), which was originally developed by De Groot (1980a; 1980b) and has been validated in studies in literature education (Janssen, 1998; Janssen & Rijlaarsdam, 1996) and arts education (Van der Kamp, 1980). Students were given writing prompts such as "I learned that..." and "I now know that I (do not)...", and were asked to write down two responses.

#### 2.4.3 *Learning outcomes*

*Transformative Reading Experiences Questionnaire (TREQ).* First, to assess students' learning outcomes, including insight into human nature, we administered the newly developed TREQ. Items were not story-specific but referred to reading experiences in general (i.e., "When I read stories, I tend to experience X"). A similar approach has been taken, for example, in studies in which partici-

pants completed Miall and Kuiken's (1995) Literary Response Questionnaire (Eva-Wood, 2004; Fialho, Zyngier, & Miall, 2011).

The TREQ included eight scales, each indicating a transformative reading component (Fialho, 2018; see Table 5.1): 1) imagery, 2) identification, 3) experience-taking, 4) character evaluation, 5) sympathy, 6) aesthetic awareness, and self-other insights which was split into: 7) self-insight, and 8) insight into others. The instruction read that the questionnaire was about fictional stories (i.e., short stories and books or novels; not expository books or articles) and explained how items should be scored, using a 5-point scale (*completely disagree – completely agree*).

An initial version of the questionnaire was tested in a pilot study (57 items,  $N = 198$ ). After analyzing internal consistency, we shortened the questionnaire by removing items from scales if a smaller item set maintained sufficient internal consistency. To ensure that students would not complete three identical measures, we created three versions by varying the items in the scales. For example, if an internally consistent scale contained items A-B-C-D, three-item combinations were selected (A-B-C, B-C-D, and A-C-D) if these were internally consistent as well. We were able to apply this procedure to four scales; the remaining four scales were identical in the three versions (see Table 5.1). We ordered items differently in the versions to reduce remembrance effects as a threat to internal validity. Depending on the version, three or four items were formulated negatively.

Table 5.1 includes an example item of each scale and shows that internal consistency for all scales was sufficient across versions and measurements. Therefore, we calculated scale scores. For each measurement point, correlations between imagery on the one hand and aesthetic awareness, self-insights and insights into others on the other hand were rather low (around .20), whereas correlations between self-insights and insight into others were high (around .75; see Appendix E, p. 281).

*Table 5.1. Transformative Reading Experiences Questionnaire: subscales, number of items, example items, and internal consistency per version (A, B, C) and measurement occasion (1, 2, 3)*

Subscale	Items	Example item	Measure	Internal consistency in Cronbach's $\alpha$ ( <i>n</i> students)		
				A	B	C
Imagery	3 <sup>a</sup>	When reading, I clearly picture in my mind the setting where the story takes place.	1	.75 (188)	.69 (183)	.83 (168)
			2	.77 (166)	.73 (175)	.81 (173)
			3	.78 (174)	.75 (161)	.76 (186)
Identification	4 <sup>a</sup>	When I read stories, I recognize something of myself in the protagonist or other characters.	1	.80 (188)	.83 (182)	.69 (168)
			2	.84 (166)	.87 (179)	.88 (174)
			3	.84 (172)	.86 (161)	.82 (186)
Experience-taking	4 <sup>b</sup>	While reading stories, I see, think and feel what a character sees, thinks and feels.	1	.79 (186)	.77 (180)	.75 (166)
			2	.76 (166)	.77 (179)	.77 (173)
			3	.73 (172)	.81 (161)	.75 (183)
Character evaluation	3 <sup>a</sup>	When I read stories, I notice that I evaluate characters positively or negatively.	1	.74 (188)	.74 (183)	.76 (168)
			2	.81 (166)	.75 (179)	.85 (172)
			3	.80 (172)	.83 (161)	.82 (185)
Sympathy	3 <sup>b</sup>	If something bad happens to the protagonist in stories, that doesn't do much to me.	1	.86 (187)	.86 (181)	.76 (169)
			2	.80 (167)	.84 (178)	.80 (173)
			3	.78 (173)	.82 (159)	.76 (186)
Aesthetic awareness	3 <sup>a</sup>	When I read stories, particular words, sentences or passages really stand out for me.	1	.80 (188)	.74 (183)	.73 (167)
			2	.79 (166)	.76 (177)	.83 (173)
			3	.82 (174)	.84 (161)	.80 (185)
Self-insight	5 <sup>b</sup>	I have the idea that I understand myself better because of story reading.	1	.82 (186)	.78 (183)	.77 (167)
			2	.81 (164)	.83 (177)	.87 (172)
			3	.84 (172)	.81 (163)	.82 (186)
Insight into others	4 <sup>b</sup>	Story reading offers me more insight into how other people are.	1	.83 (187)	.84 (182)	.78 (168)
			2	.83 (165)	.83 (177)	.87 (172)
			3	.85 (172)	.85 (161)	.83 (185)

*Note.* For missing values at one measurement point, data was excluded listwise. <sup>a</sup> In all three versions of the questionnaire, items were identical. <sup>b</sup> Items varied per version.

*Story response task.* Second, we collected responses evoked during reading a particular story. Students were asked to read a story or excerpt from a novel and to write their first responses next to it during reading, but they were also allowed to add responses after finishing the story. We provided examples of responses: “something the story reminds you of”, “you find strange or unclear”, “that evokes a feeling in you”, and so forth. Next, students wrote in response to three open-ended questions about 1) what “message” the story conveyed to them, 2) what they expected would happen to the characters and how they relate to each other, and 3) what was, according to them, the value or significance of story reading, for which they could list as little or as many values as came to mind.

To warrant generalizability and avoid remembrance effects, we used three texts at each measurement, all featuring characters who struggled to communicate in a traumatic situation: 1) the opening excerpt of the young adult novel *Birk* by Jaap Robben (2014), in which a boy seems to struggle to tell his mother that he saw his father drown; 2) a short story titled *Something needs to happen* by Maartje Wortel (2015), in which a couple attempts to communicate after the – suggested – suicide of their son; 3) the opening excerpt of the novel *The asylum seeker* by Arnon Grunberg (2003), in which a man’s girlfriend is incurably ill, announces she wants to marry an asylum seeker, and asks her boyfriend to be her witness at the wedding.

#### 2.4.4 Implementation fidelity

To monitor the implementation fidelity of TDLT (O’Donnell, 2008), we used three data sources: teacher logs, time on task observations and students’ workbooks.

*Teacher logs.* To assess the extent to which units were completed as intended, teachers completed logs online after each TDLT unit. For each phase as described in the teaching plans, they indicated whether it was fully, partly or not completed. If their answer was “fully” or “partly”, they evaluated on 5-point scales how feasible it was to teach the phase, how clear the phase was for students, how much order there was in the classroom, and how interested and engaged students were. If they did not complete a phase, they indicated the reason for this by selecting options such as “I did not have enough time” or “I forgot about it”. The response rate was high (96.6% of administered logs). In total, we collected 446 indications of the completeness of a phase (16 missing).

*Time on task observations.* We used time on task observations as an indicator of the proportion of available learning time that students were engaged in the tasks assigned to them (e.g., Karweit, 1984). In each class, either the first author or one of six trained research assistants observed one of the four TDLT units. Observations were scheduled in consultation with teachers. In seven classes, we observed unit 1; in five classes, unit 2; in four classes, unit 3; and in six classes, unit 4.

Students' behavior was coded "on task" if they worked on the given task and did what was asked of them, for example, listening to the teacher or a peer, talking about a task, reading a story, talking to the teacher, or asking questions. It was coded "off task" when students were not working on given tasks or subject matter, but were, for example, looking at cell phones, talking about something else than their task, or being disruptive. If a student's task behavior could not be determined, for example, when another student got into the line of sight, we coded it as "unclear".

We selected six students in the classroom, evenly distributed over the seating arrangement (one toward the front on the right, one toward the back on the right, etcetera). If possible, we selected three boys and three girls, except for classes where gender was unevenly distributed. These six students were observed in multiple rounds of six minutes. The focus was on a single student for one minute, during which this student was observed twice for twenty seconds. This allowed for sufficient time for coding and taking notes. After six minutes (one minute per student), there was a one-minute break, after which the next round started. This process continued until the end of the lesson. In total, we collected 1690 units of observations.

*Students' workbooks.* We randomly selected workbooks of one third of the students ( $n = 198$ ; nine from each class) and checked to which extent tasks were completed. We selected 15 out of 19 tasks to be eligible for coding: four in unit 1, 2, and 3, and three in unit 4. Not eligible were dialogic tasks that did not involve any writing, nor tasks in which one group member took notes, because, if left blank, we could not determine if a student had not completed the task or was not the one appointed to take notes.

We coded the tasks as completed, partly completed, or not completed. If a student had been absent during a unit and thus left blank all tasks in that unit, we treated these cases as missing data, because students' absence told us little about the implementation of the units. In contrast, assessing single tasks in comparison to other tasks in a particular unit or the intervention as a whole might point to implementation issues such as the difficulty of the task or the

time frame in which it was to be completed. Therefore, we deemed it necessary to make a distinction between truly noncompleted tasks and tasks left blank because a student did not attend the unit at all. Of 2,970 observations (15 tasks for 198 students), 138 were missing due to students' absence, which left 2,832 tasks to assess.

### *2.5 Procedures*

The first author met with each teacher before the study started, to familiarize teachers with the materials, structure of the units, and theoretical-empirical background of TDLT. Each teacher then scheduled the units within a maximum of four weeks: some teachers distributed the units over four weeks, whereas others needed to teach them in two weeks, depending on school schedules (excursions, test weeks). TDLT was preceded or followed by an untreated control period.

Trained research assistants or the first author administered the TREQ and story response task three times during regular hours of Dutch class: first, shortly before the start of panel 1 in the research design; second, in between panel 1 and 2; and third, directly after panel 2. Students completed the measures individually on paper. When students completed the learner report task, which was in their workbook, no representative of the research team was present.

### *2.6 Control Condition*

We opted for an untreated control condition in which no literary instruction took place, in view of the time constraints that teachers faced. Consequently, teachers followed their regular program for Dutch language class. They could teach any language domain (e.g., writing skills, grammar), except for literature. To monitor the control condition, teachers completed two digital logs during this period. In total, 40 logs were collected (response rate 90.1%). On average, 3.5 lessons of 50 minutes were taught in the control period. Mostly, reading skills of nonfictional, expository texts (32% of all indications) and rhetoric and argumentation (e.g., formulating opinions, substantiating them with arguments, recognizing sophisms; 22%) were taught. The majority of teachers indicated that students read literary novels at home and completed a task subsequently, which is conventional in Dutch upper secondary schools. In the preparatory meetings, teachers indicated that these tasks were administered several months before the study started (usually in September or October 2016). The tasks, thus, were not adapted to TDLT aims and continued as usual.

## 2.7 Data Analysis

### 2.7.1 Perceptions of learning

We randomly selected learner report tasks of 198 students (nine from each class). As the task did not apply to the control condition, these data are of a descriptive nature. In total, 420 responses were included; occasionally, students wrote down more responses than the two they were asked for. We coded responses inductively, looking for those that occurred more than once. We distinguished ten categories, as shown in Table 5.2. The first author coded all learning experiences. A trained research assistant coded 10% of the data (42 learning experiences). Agreement was 83.4%,  $\kappa = .78$ .

*Table 5.2. Coding scheme for students' perceptions of learning*

Code	Description	Example
Human nature	Learning about one's own or others' personality, life, ways of thinking, behavior, and views on the world	"I learned that people's reactions in certain situations are very different from mine."
Transformative reading experiences (other)	Learning that stories may evoke imagery, identification, experience-taking, evaluations of characters, sympathy, and aesthetic awareness	"I learned to put myself in a character's experience, that became more important to me during reading."
Talking about stories	Learning to talk to peers about stories and reading experiences, apply dialogue guidelines, listen carefully to others, ask follow-up questions, keep a dialogue going	"I learned how you can have a good dialogue because of the dialogue guidelines."
Noticing responses during reading	Learning to focus on emerging thoughts and feelings while reading, to write down responses while reading	"I discovered that I feel quite a lot of responses during reading a story."
Multiple interpretations and responses	Learning that peers express different responses evoked by the same story, that others interpret stories differently, that multiple opinions about stories exist	"You can understand stories in various ways, as became clear from the different opinions in class."
In-depth processing of stories	Learning to consider a topic profoundly, to discuss "deeper layers" in stories, to think "deeply" about a story theme	"I learned to think more in-depth about stories and to consider the deeper layers."
Thinking skills and forming opinions	Learning to formulate an opinion about stories or topics, support ideas by referring to story characteristics, postpone a judgment, reconsider a first impression	"I learned how to formulate my responses and my opinion about a story."

Code	Description	Example
Literary reading and interpretation skills	Learning how to interpret stories, finding out which stories are (not) appreciated, evaluating reading in general	"I learned how to interpret stories."
No learning, negative comments	When students wrote to have learned nothing from TDLT or wrote down a negative remark	"I didn't learn anything. It wasn't interesting."
Unclear	Incomprehensible responses	"Supah."

### 2.7.2 Learning outcomes

*Questionnaire data.* Descriptive statistics of TREQ scale scores for both groups across measurements are included in Appendix E (see p. 284). Analysis of these data consisted of three steps. First, we investigated whether TREQ scales might represent underlying factors. Via principal components analysis with Varimax rotation for all measurements, we extracted two factors (see Table 5.3). Factor 1, *Insights*, contained self-insights, insights into others, and aesthetic awareness as exclusive factor loadings. In addition, identification loaded high on this component. Factor 2, *Imagery*, contained imagery as an exclusive factor loading, but experience-taking also yielded high loadings on this factor. Factors scores were used for subsequent analyses.

Table 5.3. Factor loadings of PCA for TREQ scales, per measurement

Scale	Measurement 1		Measurement 2		Measurement 3	
	Factor 1 <i>Insights</i>	Factor 2 <i>Imagery</i>	Factor 1 <i>Insights</i>	Factor 2 <i>Imagery</i>	Factor 1 <i>Insights</i>	Factor 2 <i>Imagery</i>
Imagery		.85		.85		.87
Identification	.68	.56	.68	.58	.77	.44
Experience-taking	.34	.75	.31	.79	.30	.81
Character evaluation	.39	.58	.37	.63	.41	.58
Sympathy	.56	.52	.51	.64	.52	.64
Aesthetic awareness	.69		.70		.72	
Self-insight	.88		.91		.90	
Insight into others	.87		.88		.86	
Eigenvalue	4.30	1.08	4.63	1.06	4.42	1.21
Variance explained (%)	53.8	13.4	57.9	13.2	55.3	15.2

*Note.* Factor loadings < .30 not displayed. Sampling adequacy and sphericity assumptions were met: KMO values .87, .87 and .84; Bartlett's test of sphericity all  $p$ 's < .001.

Second, we identified potential outliers. We first inspected boxplots for both factors at the three measurement points and identified 17 outliers. As a next step, we applied case wise diagnostics, using regression analyses and graphs to plot students' scores of measurement 1 against 2, and measurement 2 against 3, on both components, using group (A or B) as a factor in the analyses. We looked for cases deviating more than 2.5 standard deviations from the regression line and found 32 outliers, including the 17 cases we already detected using boxplots: 20 in group A and 12 in group B (5.3% of total  $N$ ). We created a filter variable to exclude the outliers from the data.

Third, we treated the remaining data ( $n = 571$ ) as those of two repeated quasi-experiments: we first performed analyses for panel 1, in which group A ( $n = 291$ ) represented the experimental condition and group B ( $n = 280$ ) the control condition, and then repeated the analyses for the second panel in which the order was switched. We took the hierarchical data structure into account: as measurements were nested within students, we applied mixed models analyses with Student as random factor. For both panels and factors (*Insights* and *Imager*), we tested four models. We started with a basic null model in which only the intercept was included (Model 1). In Model 2, we added the effect of pretest as a fixed factor to test to which extent posttest scores can be explained by pretest scores (measurement 1 as pretest in the first panel, and measurement 2 as pretest in the second panel). In Model 3, we added condition as a fixed factor to test whether the average posttest scores differed between the experimental and control condition. Finally, in Model 4, we added the interaction effect of condition and pretest, to test if effects of condition on posttest scores were dependent of pretest scores.

*Story task data.* At each measurement, 198 story tasks were selected (nine from each class, three of each version). When exploring the data inductively, we noticed differences in the extensiveness of responses written next to the stories, as well as in what kind of responses they were, for which we developed two coding schemes. Similarly, we developed coding schemes for the questions about Message, Prediction, and Values of stories. Incomprehensible responses were coded as "unclear" and were not included in further analyses. A trained assistant coded 10% of the selected tasks to assess inter-rater reliability. Appendix E (see p. 282) presents all coding schemes and reliability values.

Next, we treated story task data as those of two repeated quasi-experiments. At the posttest of panel 1, group A ( $n = 99$ ) represented the experimental condition and group B ( $n = 99$ ) the control condition. At the posttest of panel 2, conditions had been switched. Descriptive statistics for story task data

are included in Appendix E (see p. 284). For scale-coded variables (Extensiveness, Message, and Prediction), we used mixed-models analyses with individual students as random factor, testing four models: 1) a basic null model; 2) a model in which we added the effect of pretest as a fixed factor; 3) a model in which we added condition as a fixed factor; and 4) a model in which we added a pretest\*condition interaction effect.

For Story responses, we first analyzed *self-related* and *socially-related* responses (see p. 282). As students could have written down multiple of these responses, we initially focused on whether they had done so at all (1 for "yes, at least once", 0 for "no", for both response types). We used loglinear analyses to test for significant three-way interaction effects of pretest, posttest and condition, which would indicate that relations between pretest and posttest are different for both conditions. If so, we would perform additional analyses on the number of responses written down by the students. If an intervention effect would occur, we would further specify the analysis for additional codes assigned to these responses. For Values of stories, we applied loglinear analysis to test for significant three-way interaction effects on the six Insights categories (see p. 283).

### 3 RESULTS

#### 3.1 *Implementation Fidelity*

Teacher logs indicated that no units were skipped in any of the classes. Overall, units were completed as intended: 88% of the phases was fully completed, 9.4% partly, and only 2.7% was not completed. Teachers evaluated feasibility ( $M = 4.2$ ,  $SD = .5$ ), clarity for students ( $M = 4.2$ ,  $SD = .5$ ), order in the classroom ( $M = 4.0$ ,  $SD = .6$ ), and students' interest and engagement ( $M = 3.8$ ,  $SD = .8$ ) overall positively. Phases that were not completed mostly occurred toward the end of a unit. Teachers indicated ten times there was too little time left, once that there had been a misunderstanding of the teacher guidelines, and once that a phase simply had been forgotten.

Students were on task in 72.5% of the observed time, and off task in 24.9% of the time (2.5% coded "unclear"). There was a large variability between classes ( $\chi^2(42) = 155.02$ ,  $p < .001$ ), with on task percentages ranging from 54.1 to 88.6. The threshold for effective teaching lies around 80% time on task (e.g., Kauchak & Eggen, 1993; Muijs & Reynolds, 2010). In only five classes, the percentage was above 80%, which was cause for concern.

Of 2,832 workbook tasks, 76.7% was fully completed, 10.1% was partly completed, and 13.2% was not completed. Two individual reflection tasks, at the

end of units 2 and 3, were remarkably often not completed: the task in unit 2 was left blank by 90% of the students, and the task in unit 3 by 54%. In light of teachers' logs, we suspect there was too little time left at the end of these units for completing the tasks.

All in all, implementation fidelity seemed to be satisfactory in terms of the number of units and unit phases that were completed, as well as the tasks in students' workbooks. However, task-oriented behavior among students was not up to standard. The difference among classes suggests that TDLT has been implemented effectively in some, but not in all classes.

### 3.2 *Students' Perceptions of Learning*

Students' responses about what they felt they learned from TDLT offered a first indication of whether the intervention aims were achieved. Figure 5.2 shows the percentages of responses in each category.

The largest share of learning experiences (28.1%) concerned talking about stories. Examples included: "I learned how you can have a good dialogue because of the dialogue guidelines", "I learned to talk about stories, you never really talk to anyone about a book or story, but now we did and I think that's good", and "I always thought I was not good at talking about literature. But after the lessons, I know I can do it."



Figure 5.2. *Distribution of learning experiences, in percentages of N = 420 responses.*

Next, 15% of the learning experiences concerned transformative reading experiences other than insight into human nature, such as: "I learned how to put

myself in the experience of a character and that has become more important to me during reading", "I found out that I can recognize things in stories, and then I compare it to real life, so I can better put myself in the story situation", and "I learned that I can really identify with most characters." Third, 12.6% concerned thinking skills and forming opinions, for instance: "I discovered how a good dialogue can make you think differently and reconsider a story", "I learned to postpone my first judgments", and "I learned how to formulate my responses and my opinion about a story." Responses in other categories occurred less frequently.

### 3.3 Learning Outcomes: TREQ Data

Results of the fit and comparison for the four models, for Insight and Imagery in panel 1 and 2, are shown in Table 5.4. All Models 2 fit the data significantly better than Models 1, which indicated that students' average posttest scores were partly explained by their pretest scores. Effects of condition and interaction effects were not consistent (see Table 5.5 for estimates).

*Panel 1.* A main effect of condition on Insights in panel 1 occurred: mean posttest scores of TDLT students differed significantly from mean posttest scores of control students. Contrary to our expectations, the regression coefficient for the control condition was larger than for the TDLT condition (pretest  $\beta = .744$ ; control condition added  $\beta = .148$ ). Thus, at the posttest, control students scored significantly higher than TDLT students. The pretest\*condition interaction for Insights was not significant. For Imagery, only Model 2 had a better fit than Model 1: there was no effect of condition, nor a pretest\*condition interaction effect.

*Panel 2.* Table 5.4 shows there was no main effect of condition on Insights in panel 2. However, adding a pretest\*condition interaction effect significantly improved the model. Parameter estimates (see Table 5.5) indicated that for TDLT students, posttest scores were more strongly dependent on the pretest than for control students: in the TDLT condition, the effect of condition was larger for those students who had relatively high pretest scores ( $\beta = .117$ ; Cohen's  $d = .10$ ). For Imagery in panel 2, there was a main effect of condition (see Table 5.4), in favor of TDLT (pretest  $\beta = .756$ , experimental condition added  $\beta = .124$ ; see Table 5.11). Thus, at the posttest in panel 2, TDLT students scored significantly higher than control students ( $d = .10$ ). The interaction effect of pretest\*condition was not significant.

Table 5.4. Fit and comparison of models 1-4 for Insights and Imagery in panel 1 and 2

Panel and variable	Model	-2LL	$N_{\text{pars}}$	Models	Comparison $\chi^2$	$df$	$p$
Panel 1 <i>Insights</i>	1. null	1252.30	3				
	2. + pretest	854.02	4	2 vs 1	398.28	1	< .001
	3. + condition	847.77	5	3 vs 2	6.25	1	.012 <sup>a</sup>
	4. + pretest*condition	844.42	6	4 vs 3	3.45	1	.067
Panel 1 <i>Imagery</i>	1. null	1239.13	3				
	2. + pretest	868.79	4	2 vs 1	370.34	1	< .001
	3. + condition	866.27	5	3 vs 2	2.52	1	.112
	4. + pretest*condition	864.37	6	4 vs 3	1.90	1	.168
Panel 2 <i>Insights</i>	1. null	1216.67	3				
	2. + pretest	794.87	4	2 vs 1	421.80	1	< .001
	3. + condition	792.12	5	3 vs 2	2.75	1	.097
	4. + pretest*condition	787.99	6	4 vs 2	6.88	1	.001 <sup>b</sup>
				4 vs 3	4.13	1	.042 <sup>b</sup>
Panel 2 <i>Imagery</i>	1. null	1201.84	3				
	2. + pretest	810.09	4	2 vs 1	391.75	1	< .001
	3. + condition	805.50	5	3 vs 2	4.59	1	.032 <sup>a</sup>
	4. + pretest*condition	804.10	6	4 vs 3	1.40	1	.237

Note. <sup>a</sup> Main effect of condition. <sup>b</sup> Pretest\*condition interaction effect.

Table 5.5. Parameter estimates for main and interaction effects on Insights and Imagery

Panel, variable, model	Parameter	B	SE	$df$	$t$	$p$
Panel 1 <i>Insights</i> Model 3	Intercept	-.070	.041	448	-1.71	.088
	Pretest	.744	.029	448	25.25	< .000
	Control condition	.148	.059	448	2.51	.012
Panel 2 <i>Insights</i> Model 4	Intercept	-.015	.040	439	-.39	.698
	Pretest	.709	.040	439	17.54	< .001
	Intervention*pretest	.117	.057	439	2.04	.042
Panel 2 <i>Imagery</i> Model 3	Intercept	-.044	.040	439	-1.08	.280
	Pretest	.756	.030	439	25.15	< .000
	Intervention condition	.124	.058	439	2.15	.032

*Conclusions.* In short, the questionnaire data yielded incoherent results. We found a significant interaction effect for Insights and a main effect of condition for Imagery in panel 2, both in favor of the experimental condition, but these outcomes were not consistent with panel 1. Moreover, effect sizes were small. To look for additional explanations, we systematically added background variables to the models for both panels: gender, educational track, grade for Dutch, estimated grade for literature, and familiarity with fiction. None of the variables significantly improved the models in either panel; therefore, they could not further explain the results.

### 3.4 Learning Outcomes: Story Task Data

*Panel 1.* A main effect of condition on Extensiveness occurred in panel 1 (see Table 5.6), but in favor of the control condition (pretest  $\beta = .256$ ; control condition added  $\beta = .349$ ; see Table 5.7). Thus, taking pretest into account, control students responded more extensively to stories at the posttest than TDLT students. For Self-related and Socially-related story responses, loglinear analyses revealed no significant three-way interaction effects, indicating that relations between pretest and posttest had not been affected by TDLT.

Table 5.6. Fit and comparison of models 1-4 for Extensiveness in panel 1

Model	-2LL	$N_{\text{pars}}$	Models	Comparison		$p$
				$\chi^2$	$df$	
1. null	546.25	3				
2. + pretest	527.94	4	2 vs 1	18.31	1	< .001
3. + condition	520.86	5	3 vs 2	7.08	1	.008*
4. + pretest*condition	520.56	6	4 vs 3	.30	1	.584

Note. \* Main effect of condition.

Table 5.7. Parameter estimates under model 3 on Extensiveness in panel 1

Parameter	$\beta$	$SE$	$df$	$t$	$p$
Intercept	-.070	.041	448	-1.71	.088
Pretest	.744	.029	448	25.25	< .000
Control condition	.148	.059	448	2.51	.012

For Message and Prediction, mixed-models analyses indicated no main effects of condition, nor significant pretest\*condition interactions. For Values of stories, loglinear analysis revealed one significant three-way interaction effect, for the number of students who reported 'gaining insights in life and the world' as a value ( $\chi^2(1) = 5.81, p = .016$ ). For this variable, thus, the relation of pretest and posttest differed significantly between conditions. Yet, two separate chi-square tests showed that this difference occurred for control students ( $\chi^2(1) = 9.40, p = .002$ ), but not for TDLT students. The odds ratio indicated that control students were 2.2 times more likely to mention this value than TDLT students.

*Panel 2.* For all scale variables, mixed-models analyses indicated no main effects of condition, nor significant pretest\*condition interaction effects. Likewise, for categorically coded variables, loglinear analyses revealed no significant three-way interactions, indicating that the relations between pretest and posttest had not been affected by the intervention.

## 4 DISCUSSION

We set out to investigate whether a dialogically oriented literature classroom intervention that was based upon a model of transformative reading would foster students' transformative reading experiences – most importantly, their insight into human nature. Our main conclusion is that, whereas students reported relevant perceptions of learning, we were not able to demonstrate any effects of the intervention on their insight into human nature, nor on other transformative reading components. Students seemed to perceive the intervention as a project that taught them how to engage in dialogues about stories, whereas this was only partly what the intervention aimed for. In this section, we contextualize our findings in an attempt to explain them. We address the limitations of this study and consider next steps for future intervention studies in the literature classroom.

### 4.1 *Explaining the Absence of Effects*

In this study, we set up a quasi-experimental pretest-posttest design with switching replications. Ideally, intervention effects would be demonstrated at the posttest of panel 1, which would sustain for group A and be replicated for group B at the posttest of panel 2. However, the main effects or interaction effects we found were either in favor of the control condition or, when in favor of TDLT, negligibly small.

Effects may be absent if an intervention is not implemented as intended in its original design (O'Donnell, 2008). Yet, implementation data – teacher logs, time on task observations, and completed workbook tasks – indicated that TDLT was, overall, implemented as intended, even though the overall percentage of observed time that students were on task (72.5%) was below the 80% threshold for effective teaching and learning. Furthermore, our findings cannot be ascribed to insufficient sample sizes, nor to insufficient reliability of the questionnaire or the coding systems used to analyze the story task data. Finally, as we took pretest scores and the hierarchical data structure into account, these aspects cannot explain the absence of effects. However, other limitations, in terms of intervention design and methodology, may function as explanations for our findings.

#### *4.1.1 Explanations based on intervention design*

First, TDLT may have been too skills-oriented. Students felt that they learned what they were most explicitly trained to do: engaging in external dialogues about short stories. Theoretically, we expected that sharing responses and considering multiple perspectives on relevant story themes would result in insight into human nature, but attention for gaining such insights may have remained too implicit for this group of students. Indeed, when assessing the practicality of TDLT, students as well as teachers indicated that its ultimate purpose was not clear enough and might have been made more explicit (see Chapter 4; TDLT-1).

Second, we are unsure whether transformative reading (Fialho, 2018) may at all occur in 10<sup>th</sup> grade students, who can be considered novice readers of literary texts. Students reported some transformative reading experiences in the learner report task, but they did so rather infrequently and seldomly reported about insight into human nature. Perhaps, TDLT would have been more effective if it were implemented in grade 11, the final year of the higher general secondary education track. Moreover, Dutch secondary school students do not often read for enjoyment, have a neutral to negative attitude toward fiction, and their motivation for literary reading decreases in higher grades of secondary school (OECD, 2010; Van Schooten, 2005). Their relatively low ART scores suggest that the participants in this study were no exception. As a consequence, students may not have had sufficient experience with literary story reading to reflect profoundly on their own reading experiences and their insight into human nature. This was confirmed by several of their teachers, who indicated that the focus on noticing transformative reading experiences was

completely new to their students and that students sometimes felt confused or even frustrated when being asked to reflect upon what a literary text evoked in them.

Furthermore, as a potential consequence of their limited freedom of text choice, students may not have been fully engaged in reading. As outlined in the Introduction of this chapter, it was challenging to bridge the gap between teachers' and curricular demands on the one hand, and students' autonomy that may foster engaged reading on the other hand. Moreover, the research literature is ambiguous: studies have suggested that self-selected reading may foster students' insight into human nature (e.g., Ivey & Johnston, 2015; Richardson & Eccles, 2007; Rothbauer, 2011), but similar results have been found in a review of intervention studies in which students were hardly given any autonomy in selecting reading material (Schrijvers et al., 2018). The role of freedom of choice in reading materials for developing insight into human nature, thus, is yet to be further illuminated.

In addition, the number of stories or units may have been too few. Time constraints allowed for only four intervention units of 50 minutes; in three of these, a story was read. Reading more stories or interacting with a story more in-depth may result in different learning outcomes. Moreover, as we opted for variety and differentiation (i.e., using stories with various themes), there was no accumulated reflection on one particular social-moral theme. Selecting stories around a single social-moral theme might offer students the opportunity to consider more extensively, in multiple units, how such a theme affects and relates to themselves and other human beings.

Finally, the outcomes of our study raise the intriguing question why students scored higher on various posttests when they had just participated in the untreated control condition, compared to those who had received the intervention. The control condition was unrelated to the intervention: no instructional attention was paid to literary texts and literary responses. Why attention for reading skills of expository texts and rhetoric and argumentation have resulted in higher posttest scores remains unclear. As effects in favor of the control condition occurred not only in panel 2 but also in panel 1, they cannot be explained as 'delayed' intervention effects: on the contrary, raw TREQ scores (see Appendix E, p. 284) indicated that intervention students' mean scores mostly decreased from pretest to posttest. We can only speculate why this happened: perhaps students experienced reluctance, boredom or frustration, even though their perceptions of learning did not support this assumption.

### 4.1.2 *Methodological explanations*

One might argue that potential reluctance, boredom or frustration may also have applied to repeated administration of measurements in a short period of time. Even though a research design with switching replications is, theoretically, a strong design to assess intervention effects (Shadish, Cook, & Campbell, 2002), in this study it involved three measurement points in at most eight weeks. However, if repetition would have negatively affected students' posttest scores, we would have expected to see a similar effect in the control conditions of both panels.

Specifically for story task data, an explanation for the absence of effects might be found in the role of students' writing abilities. We collected no information about their writing skills, for example, in the form of their most recent results on writing tests. If writing is difficult for students, think-aloud tasks (e.g., as applied by Janssen, Braaksma, & Couzijn, 2009) may be a more valid way of gaining insights into students' responses to stories.

### 4.2 *Implications for Practice and Research*

This study was the first to empirically investigate, in the specific context of 10<sup>th</sup> grade of higher general secondary education in the Netherlands, whether dialogically oriented literature lessons would foster students' insight into human nature. Because we did not find considerable effects of condition, implications for practice can only be based on descriptive learner report data. This study suggests that strategy instruction helps students to engage in external dialogues about stories and reading experiences. In particular, students found practical guidelines for dialogues helpful, as illustrated by a large share of learning experiences such as: "I learned how to engage in a good dialogue, due to the dialogue guidelines (FLUENT)" and "I learned 'FLUENT', so how to follow a good conversation. And how to go more into depth". Research into the particular needs of students in the higher general education track in the Netherlands confirms that they prefer clear structures and guidelines in their learning materials: they feel the need to know what to do and how to do it (Vermaas & Van der Linden, 2007). This may also imply that the ultimate purpose of the intervention – developing insight into human nature – remained too implicit for this population of students.

Therefore, our first step in future research is to assess whether redesigning the intervention would result in the outcomes that could not be demonstrated in this paper. The redesign should operationalize several points for improvement as discussed here, among which putting more emphasis on the purpose

of eliciting transformative reading experiences and developing insight into human nature, a longer duration that allows for this group of relatively novice readers to read and interact with a larger selection of stories, and centering those stories and interactions around a single social-moral theme.

In addition, as our measures relied on individual self-reports and could therefore not capture external dialogues among students, in future studies researchers might choose to administer dialogue tasks as pretest and posttest. Analysis of observations might demonstrate changes in how students refer in their group talk to transformative reading experiences, including insight into human nature. Finally, such studies may also add to existing knowledge about transformative reading by conducting interview and think-aloud studies with adolescent participants. This would offer valuable insights into their reading and reflection processes, which may offer relevant input for the design of interventions like the one described here.

#### 4.3 *Conclusion*

The current literature classroom intervention study demonstrated that students, as a result of only four 50-minute units, reported that they learned to talk about their responses to literary stories, and, to a lesser extent, to notice transformative reading experiences and to develop their thinking and opinionating skills. However, the study could not demonstrate any consistent increase of students' transformative reading experiences, including their insight into human nature. This was true for their reflections on their reading experiences in general, as well as for story-specific responses as expressed in writing during and directly after reading.

Nonetheless, our study provides a number of suggestions for how future interventions might successfully foster students' insight into human nature. After all, the absence of statistical effects does not necessarily mean that such outcomes cannot be achieved in the literature classroom, as is suggested by a response written by a student in the workbook: "During the lessons, I learned that the meaning of life for other people is much more complex than I thought it was." Responses such as this one suggest that it is worthwhile, for teachers and researchers alike, to pursue the systematic design and empirical testing of literature classroom interventions that aim to foster students' insights into who they are, how they relate to others, and how they position themselves in life and in the world around them.