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].

DESIGNING A LITERATURE CLASSROOM INTERVENTION TO FOSTER 10TH GRADE STUDENTS' INSIGHT INTO HUMAN NATURE

We describe the design of a literature classroom intervention for 15-year-old students in the Netherlands, which aimed to foster their insight into human nature – insight into themselves, fictional others, and real-world others. The design was informed by a theoretical-empirical model of transformative reading, an explorative study in Dutch literature classrooms, and three initial design principles identified in a review of previous intervention studies. We investigated the effects of an iterative design process on two quality indicators – validity and practicality – and on the initial design principles underlying the intervention, to contribute to theory and classroom practice. A first intervention was developed in collaboration with teachers, tested in trial studies, and taught by 13 teachers to 22 classes. In four lessons, students focused on internal and external dialogues with and about short stories addressing various social-moral themes. We assessed validity and practicality by using implementation and evaluation data from teacher logs, time on task observations, students' evaluations, and teachers interviews. Suggestions for improvement were derived from the data (e.g., make learning objectives more explicit, select stories centering around a single social-moral theme). A redesigned intervention was taught by 6 teachers to 6 classes. Teacher and student data supported the validity and practicality of the redesigned intervention and yielded sub-principles for operationalization of the initial design principles. All in all, this study suggested that an iterative design process, grounded in previous empirical research and utilizing both teacher and student data, may result in the design of valid and practical domain-specific interventions.

1   INTRODUCTION

Intervention studies in literature classrooms empirically evaluate whether a particular instructional approach helps students to achieve predetermined objectives, such as improving their interpretative skills (Janssen, Braaksma, & Couzijn, 2009; Levine & Horton, 2013) or rethinking certain social-moral attitudes (Malo-Juvera, 2014; 2016). Researchers usually develop such an instructional approach in an educational design research project, ideally in close collaboration with teachers. Educational design research has been conceptualized as consisting of three phases (Plomp, 2013): the preliminary research phase, in which relevant

literature is reviewed and a theoretical framework is built; the development phase, in which an intervention is developed, improved and refined; and the assessment phase, in which its implementation and effectiveness are evaluated, compared to predetermined specifications. Following these phases increases the probability of designing high-quality interventions.

Researchers have established several quality indicators for interventions: validity, practicality, effectiveness, sustainability and replicability. First, Nieveen (1999) distinguished between content validity, which dictates that the components of an intervention should be based on state-of-the-art knowledge and should be relevant to those using the intervention, and construct validity, which means that all components should be consistently linked. Second, for the intervention to be practical, teachers must consider it to be usable and use it in a way that is generally compatible with the designers' intentions (see O'Donnell, 2008). Third, Nieveen indicated that high-quality interventions should result in the desired outcomes: the intervention should be effective. Rietdijk, Janssen, Van Weijen, Van den Bergh and Rijlaarsdam (2017) pointed to continued use of the developed intervention after the research project (sustainability). Finally, Rijlaarsdam, Janssen, Rietdijk, and Van Weijen (2017) called for more comprehensive descriptions of interventions that strengthen their replicability. Metaphorically speaking, interventions all too frequently remain "black boxes": it is unclear what happened in the classroom, why it happened, and how what happened was developed. This not only threatens the validity and replicability of interventions, but also hampers detailed insights into domain-specific instructional activities that are designed to achieve particular aims (see also Schrijvers, Janssen, Fialho, & Rijlaarsdam, 2018).

With this paper, we aim to respond to the call by Rijlaarsdam et al. (2017) for more comprehensive and replicable intervention descriptions, by describing in detail the iterative design process of an intervention for 10th grade literature classrooms that aims to foster students' insight into human nature – insight into themselves, fictional others, and real-world others. We focus on validity and practicality; effectiveness and sustainability are beyond our current scope, as their evaluation requires different types of studies (e.g., quasi-experimental or longitudinal). As Figure 4.1 shows, the design project included a preliminary research phase, two development phases and two subsequent assessment phases (Plomp, 2013). Both development phases were informed by the preliminary research phase, and development phase 2 was additionally informed by the results of assessment phase 1.

We operated on the micro level of curriculum design (Van den Akker, 2013): the level of the classroom and the instructional materials and strategies used in

it. The intervention was designed for 10th grade of the higher general secondary education track in the Netherlands, which is the second highest track in Dutch secondary education and prepares for higher vocational education but not for university.
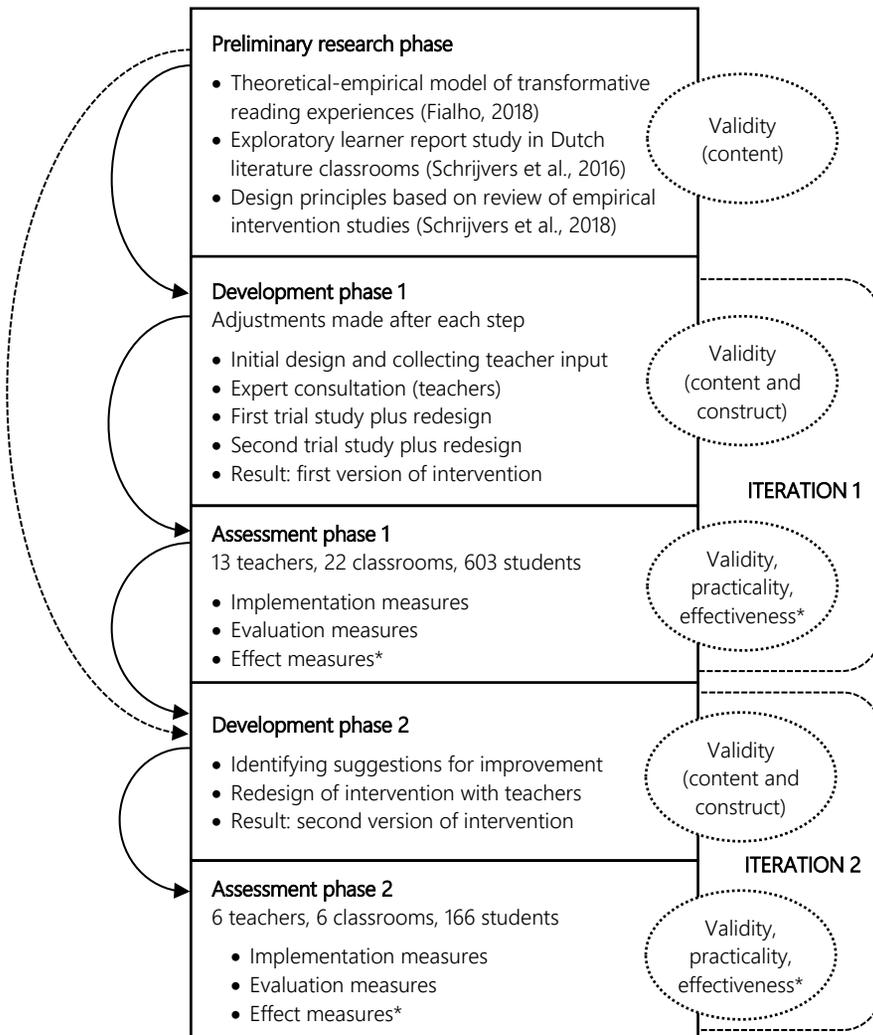
Preliminary research phase
- Theoretical-empirical model of transformative reading experiences (Fialho, 2018)
- Exploratory learner report study in Dutch literature classrooms (Schrijvers et al., 2016)
- Design principles based on review of empirical intervention studies (Schrijvers et al., 2018)

Validity (content)

Development phase 1
Adjustments made after each step
- Initial design and collecting teacher input
- Expert consultation (teachers)
- First trial study plus redesign
- Second trial study plus redesign
- Result: first version of intervention

Validity (content and construct)

ITERATION 1

Assessment phase 1
13 teachers, 22 classrooms, 603 students
- Implementation measures
- Evaluation measures
- Effect measures*

*Validity, practicality, effectiveness**

Development phase 2
- Identifying suggestions for improvement
- Redesign of intervention with teachers
- Result: second version of intervention

Validity (content and construct)

ITERATION 2

Assessment phase 2
6 teachers, 6 classrooms, 166 students
- Implementation measures
- Evaluation measures
- Effect measures*

*Validity, practicality, effectiveness**

Figure 4.1. Overview of the design process (see Chapters 5 and 6).*

In Dutch schools, literature education usually does not have the status of a separate subject. It is a sub-domain within Dutch language classes, like writing and rhetoric. In lieu of regulations, standardized tests and nationwide exams, teachers have much freedom in selecting literary texts and designing tasks and instructions. Students are merely required to work toward three intertwined objectives: acquiring literary-historical knowledge, utilizing structural-analytical skills, and reflecting on their literary reading experiences and development (Dutch Institute for Curriculum Development, 2012). The current design project is related to the third objective, as its focus is on particular literary experiences in relation to "human nature", for example, relating a story theme to the own life, or considering how thoughts, feelings and behaviors of characters in fictional situations represent human responses to similar real-life situations.

Figure 4.1 summarizes the design process. We first outline preliminary research results, including three initial design principles. In the Method section, we describe the design process and instruments used to assess the implementation and evaluation of two subsequent versions of the intervention, which yielded information about their validity and practicality. We then present the two versions of the intervention and their implementation and evaluation outcomes, and indicate how the design process affected the initial design principles.

1.1   *Outcomes of Preliminary Research Phase*

The preliminary research phase consisted of the development of a theoretical-empirical model of transformative reading (Fialho, 2012; 2018), an exploratory learner report study in Dutch literature classrooms (Schrijvers, Janssen, Fialho, & Rijlaarsdam, 2016), and a review study that resulted in a set of design principles upon which the intervention construct was based (Schrijvers et al., 2018).

Theoretical-empirical model of transformative reading. One of the acclaimed merits of reading fictional and literary texts is that, via processes of empathy and reflection, it offers readers insights into who they are, how they position themselves in the world, and how they see themselves in relation to other human beings (for overviews, see Koopman & Hakemulder, 2015; Hakemulder, Fialho & Bal, 2016). In short, it may offer readers "insight into human nature". This mode of reading has been conceptualized as a 'transformative experience', because it may alter readers' perceptions of themselves and themselves in relation to others (Fialho, 2012).

In a theoretical-empirical model, Fialho (2018) distinguished two outcomes of transformative reading – insights into oneself and into others – and identi-

fied six underlying components. Adult readers who talked in Fialho's phenomenological interview studies about reading experiences that had a transformative impact on them, indicated that they vividly imagined the setting and characters in a story (*imagery*), recognized something of themselves or others in characters (*identification*), enacted and embodied the experiences of a character (*experience-taking*), evaluated characters positively or negatively (*character evaluation*), felt sympathy and compassion for characters (*sympathy*), and noticed which words, phrases or sentences were particularly striking to them (*aesthetic awareness*). For adult readers, these particular experiences preceded new or deeper insights into themselves and others (*self-other insights*).

Whereas the transformative reading model (Fialho, 2012; 2018) has not yet been validated for adolescent readers, studies suggested that similar experiences may occur in them, even though researchers used other terms than "transformative reading" or "insights into self and others", or worked with non-literary texts. For example, adolescents were found to consider their possible future selves as a result of fiction reading: they reflected on who they would (not) like to become (Richardson & Eccles, 2007). In addition, they were found to compare their own lives to story situations and to experience empathetic engagements with characters' feelings (Charlton, Pette, & Burbaum, 2004), as well as to regard fiction reading as a way of understanding others' experiences, which made them feel connected to others or offered them new options for their own lives (Rothbauer, 2011). Therefore, it seems likely that adolescents may as well engage in a "transformative" mode of reading.

Exploratory learner report study. In an exploratory, descriptive study in which we asked students to complete a written learner report, we found that learning about self and others was among the learning outcomes in Dutch upper secondary literature classrooms (Schrijvers et al., 2016). In their learner reports, students reported that literature education, for instance, made them learn about their own and other people's personalities, relations and behavior, consider their future selves, and identify life lessons in literary texts. Such experiences occurred more frequently if students' teachers reported that they allowed for more student autonomy and interaction in the classroom.

Literature education may thus foster students' insight into human nature, departing from themes and issues raised in texts. This potential learning outcome appears to be valued by curricular organizations and teachers in the Netherlands. For example, a team of Dutch language and literature teachers, who are working on an intended curriculum reform, suggested that literary reading may familiarize students with other worlds, contributes to moral devel-

opment, and helps them to think about people's choices, about themselves, others, and the world (Curriculum.nu, 2018a). Moreover, teachers reported to consider fostering students' personal growth or personal development as an important aim of literature teaching (Janssen, 1998; Oberon, 2016). However, little is known about *how* literature education might foster these learning outcomes.

State-of-the-art: design principles. A review of intervention studies shed light on instructional approaches that may foster students' insight into human nature in the literature classroom (Schrijvers et al., 2018). Studies were included in this review if they aimed at fostering some form of insight into self, fictional others, or real-world others, were conducted in regular, first-language secondary education classrooms, and used (quasi-)experimental research designs. Thirteen studies were included. Nine studies provided empirical support for fostering students' insight into human nature (e.g., Adler & Foster, 1997; Malo-Juvera, 2014; 2016; Vezzali, Stathi, & Giovannini, 2012; White, 1995), from which three design principles were derived.

The first principle suggested that fictional texts should be read that are thematically relevant for the aim of an intervention. If fictional texts address relevant social situations, readers may consider how they would position themselves in those situations and how they would impact themselves and others (see Mar & Oatley, 2008). As researchers seldomly considered whether the fictional texts used would be "literary" texts, the review remained inconclusive as for how to conceptualize "literature" and whether it might successfully foster students' insight into human nature.

The second principle suggested to design writing tasks, related to texts and themes, that stimulate students to activate previous personal experiences before reading, to notice and annotate responses during reading, and/or to write down responses or reflections directly after reading. These tasks stimulate students to engage in an *internal dialogue* with the text (see Janssen et al., 2009). Such dialogues may generate transactional processes of meaning-making (Rosenblatt, 1938/1983), because students may become aware of the responses evoked by texts and of how these responses are related to the outer-textual world.

The third principle suggested to design exploratory dialogic activities that stimulate students to verbally share personal experiences related to texts and text themes. Responses and reflections noticed in internal dialogues, thus, may be shared in *external dialogues*, which allows for exploring multiple perspectives on a text and the issues it addresses. This may alter students' opinions,

views and perceptions of themselves and others, or offer them new ones. Such dialogues may take place in pairs, small groups, or the classroom.

In short, the preliminary research phase yielded a central premise for the intervention design: to optimize the probability of fostering their insight into human nature, students should be invited to engage in internal and external dialogues with and about texts and to focus on transformative reading experiences such as imagery, identification and sympathy.

1.2 *Aims and Research Question*

Designing an intervention can be compared to designing a questionnaire. Both processes start by reviewing relevant literature, conducting exploratory work, developing a theorical model, and, in case of an intervention, identifying design principles. In a first iteration, an inital construct is designed: an intervention prototype, or a set of questionnaire items. In practice tests, data are collected to assess validity and practicality (of interventions) or reliability (of questionnaires). If these are unsatisfactory, a second iteration will follow: the construct is adjusted and tested again. Iterations continue until a valid, practical intervention or reliable questionnaire is constructed. Just like questionnaire design may contribute to theory about the measured variable, intervention design may add to the state-of-the-art by further specifying the initial design principles in terms of their operationalization in practice.

The primary aim of this study was, therefore, to design a literature classroom intervention for 10th grade students, in which the model of transformative reading, the text selection principle, the internal dialogue principle, and the external dialogue principle were integrated. Because validity – both at the content and construct level – and practicality are considered to be amongst the indicators of the quality of an intervention (Nieveen, 1999; O'Donnell, 2008), the study was guided by the following research question:

> To what extent is the intervention that is designed a valid and practical instructional approach for upper secondary literature classrooms, according to both students and teachers?

An additional aim of the study was to contribute to the state-of-the-art by reflecting on how the initial design principles are affected by their operationalization in the classroom.

2   METHOD

2.1 *Design Process and Participants*

The two iterations in the design process both consisted of several steps (see Figure 4.1), which we describe below. For both assessment phases, we also describe the sample of teachers and students involved, as we collected implementation and evaluation data from them to assess validity and practicality.

2.1.1   *Development phase 1*

Initial design. As the research team, we collected teachers' input via email, asking for practical suggestions, for instance, concerning the number of units, as well as for examples of suitable short stories and tasks. This led to the decision to use short stories that could be read in about 10-15 minutes, to ensure that reading could be accompanied by both internal and external dialogue activities within a single unit. Teachers also indicated that devoting four lessons (of 50 minutes, the conventional length at Dutch secondary schools) to the project would be preferable, in view of their regular teaching program. We therefore decided that each unit should fit within a single lesson. We further discussed how transformative reading experiences could be integrated in the units and designed a prototype of the intervention.

Expert consultation. The prototype was discussed in a meeting with four experienced teachers. This resulted in two major changes, concerning text choice and external dialogues. First, we originally considered offering students freedom of choice in reading materials, because this might increase their engagement (e.g., Lenters, 2006). However, the research literature is ambiguous at this point: in only one out of the nine intervention studies on which the text choice principle was based, students chose their reading material (from a preselected list; Vezzali et al., 2012). Moreover, the consulted teachers suggested that if students would read the same stories, this would enable "deeper" talk about story themes and related insights into themselves and others. Second, the teachers suggested to end the intervention by a teacher-led classroom dialogue in which students would discuss what they had learned, to establish closure and to talk about how students could apply what they learned in future literature lessons. We adjusted the units according to their suggestions.

First trial study. One female teacher taught each of the four units separately to four different classes. The first author observed these lessons. Overall, the teacher felt that teaching the units was feasible and that students were en-

gaged in the tasks. To further operationalize the external dialogue principle, she suggested to include, in unit 1, a task with "questioning cards" to guide students in using follow-up questions. As the first author also observed that students struggled with follow-up questions, a task was designed accordingly. Related to the text choice principle, the teacher questioned one story: she felt its theme was not familiar enough for students. We decided to assess the appropriateness of this story again in a second trial study. Some other minor adjustments were made, mainly to materials and organization (e.g., bundling tasks in a booklet; ways to compose small groups of students).

The teacher also commented on a first version of the teacher guideline she received. She felt it was clear and manageable, but suggested to include information about the background, goals and structure of the units, as well as concrete suggestions for teachers' roles and their interaction with students. We cross-checked her suggestions with the second trial teacher.

Second trial study. Another female teacher, at another school, taught the units to a single class, in four lessons over the course of two weeks. Again, the first author observed. The trial resulted in two adjustments following up on the previous ones. The story we specifically attended to was replaced, as the second teacher also felt it was unclear for students. Further, we designed an observation task in unit 1, in which teachers would demonstrate using follow-up questions during a dialogue about reading experiences, as the teacher felt that using the questioning cards did not suffice. Additionally, she confirmed the suggestions for the teacher guideline, which was revised accordingly.

All in all, consulting experts and conducting trials led to adjusting the initial design in operationalizing the text choice and external dialogue principle. The trial teachers were satisfied with the operationalization of the internal dialogue principle.

2.1.2   Assessment phase 1

Assessment focused on implementation and evaluation of the intervention, which we call Transformative Dialogic Literature Teaching, version 1 (for short: TDLT-1). We assessed content validity by searching evaluation data for endorsements and skepticism regarding the intervention and its relevance. Construct validity was assessed by focusing on data concerning the coherence and clarity of the units. Finally, we assessed practicality by analyzing whether units were implemented as intended, and by focusing on data that concerned feasibility. We collected data of teachers and students. Whereas validity and practi-

cality of interventions has previously been conceptualized mainly from teachers' perspectives – for example, when assessment is conceptualized as '[evaluating] whether target users can work with the intervention and are willing to apply it in their teaching' (Plomp, 2013, p. 30; italics added) – we regarded students as equally important stakeholders, who can provide valuable information for a redesign.

The phase started with individual preparatory meetings with thirteen teachers from six schools across the Netherlands, for a walk-through of the teacher guideline and the units. They then taught TDLT-1 to either one or two classes, in two to four weeks. The teachers had 13.3 years of teaching experience; two of them were male. Student participants (N = 603, from 22 classes) were on average 15.9 years old; 52.5% was female. Their parents received a passive consent letter and could object to their child's participation; none of them withheld their consent.

2.1.3   Development phase 2

By analyzing data from assessment phase 1, we identified suggestions for improvement. These were discussed during a meeting of the first author and three teachers who had been involved in the first assessment phase, resulting in adjustments of the operationalization of the initial design principles. For example, a suggestion for improvement was that students needed more time to learn to engage in internal and external dialogues; therefore, it was decided that the redesigned intervention, called TDLT-2, would consist of six units. The first author adjusted the teaching and learning materials and returned them to the teachers for final comments, which only led to some improvements in formulation.

2.1.4   Assessment phase 2

Six teachers taught TDLT-2 to a single class: those involved in development phase 2, as well as three new teachers. They had on average 18.7 years of teaching experience; all were females. The new teachers participated in a workshop led by the first author, which consisted of: (a) information about the theoretical background of the intervention, (b) a walk-through of the material, (c) an exercise to put themselves in the role of students when writing down first responses to a story, (d) practicing to give feedback on students' dialogues, and (e) time for questions. One of the teachers involved in the redesign was present and discussed, for example, student talk she had observed and challenges she faced in giving feedback on dialogues.

Again, we collected implementation and evaluation data from teachers and students. TDLT-2 was taught to 166 students of six classes in four schools. They were on average 15.5 years old; 49.2% was female. Parents were again asked for passive consent. After the second assessment phase, we reevaluated validity and practicality, after which a third design iteration was not deemed necessary.

2.2 *Instruments and Data-Analysis*

In design research, triangulation of data sources and data collection methods, as well as empirical testing of the practicality of interventions is important for data interpretation (McKenney, Nieveen, & Van den Akker, 2006). Therefore, we used five instruments to collect data of teachers and students: (a) teacher logs, (b) time on task observations, (c) teacher interviews, (d) student evaluation forms, and (e) student evaluation tasks. Table 4.1 shows the indicators of validity and practicality (appreciation and relevance, structure and coherence, clarity and feasibility) in which the instruments provided insight, whether they were used to assess TDLT-1, TDLT-2, or both, and how many responses and observations were collected. Below, we provide further details of instruments and analysis.

Teacher logs. Teachers completed one online log per unit, which consisted of several phases; for each phase, teachers indicated whether it was fully, partly or not completed. If they did not complete a phase, they indicated the reason (e.g., "not enough time", "forgot about it") and were asked to add elaborations. We also asked teachers to evaluate fully or partly completed phases. They indicated on 5-point scales (letters A-D refer to items in Table 4.1):
1.   How interested and engaged students seemed to be (A);
2.   How clear the phase seemed for students (B);
3.   How attainable it was to teach the phase (C);
4.   How much was order and discipline there was in the classroom (D).
Here, too, we asked them to elaborate. We analyzed how many phases were fully, partly, and not completed and how units were evaluated. Furthermore, teachers' elaborations informed the interview guidelines.

Table 4.1. Overview of instruments and derived quality indicators

Instrument	Quality indicators	TDLT	Response rate or N observations
Teacher logs	*Practicality:* - Percentages fully, partly, not completed phases - Evaluation of feasibility of teaching (C) - Evaluation of maintaining order in class (D) *Content validity:* appreciation and relevance for students (A) *Construct validity:* clarity for students (B)	1 2	96.6% for 21 phases from 4 units, in 22 classes 94.5% for 31 phases from 6 units, in 6 classes
Time on task observations	*Practicality:* percentages time on task (overall, and for intended activities)	1 2	N = 1690 in 22 visits; 1 unit in each class N = 877 in 12 visits; 2 units in each class
Teacher interviews	*Practicality, content and construct validity:* indications of appreciation and relevance, structure and coherence, clarity and feasibility	1	100%
Student evaluation form	*Content validity:* - Overall appreciation and relevance - Appreciation and relevance of external dialogues - Appreciation and relevance of internal dialogues - Appreciation of stories read - Sense of safety as aspect of appreciation *Construct validity:* - Overall clarity and comprehensibility - Clarity and comprehensibility of dialogue tasks *Additional support for indicators:* strengths, suggestions for improvement	1	85.4%
Student evaluation task	*Content validity:* relevance of intervention elements, story appreciation *Construct validity:* coherence between intervention elements and objectives *Practicality:* story difficulty	2	90.4% (story items) 70.5% (intervention elements)

Time on task observations. To assess the proportion of available learning time that students were engaged in intervention tasks, we conducted time on task observations (Karweit, 1984). Behavior was coded "on task" if students 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, writing in the workbook, talking to the teacher, or asking questions. We defined "off task" behavior as obvious non-learning behavior, when students were not working on the given task but were, for instance, being disruptive, talking about something irrelevant, or looking at cell phones.

We formulated two additional coding instructions: if students seemed disengaged or were waiting for a next task without being disruptive, we coded this as "on task", because they might still be listening or thinking. If we could not get a clear view on a student (e.g., if another student moved into the line of sight) we used "unclear". In addition to task behavior, we coded for each observation moment the learning activity the teacher intended at that moment: teacher-led activity (explanation, instruction); individual student task; reading and/or listening to a (read-aloud) story; dialogic pair or small-group activity; whole-class activity (presenting, discussing); or unclear.

During a classroom visit, we randomly selected six students – or, in case of group work, one student from each group – and observed them in multiple rounds. We observed the first student for twenty seconds, coded task behavior and intended activity, observed the same student for another twenty seconds, again coded task behavior and intended activity, and observed the next student. After observing all six students twice for twenty seconds, there was a one-minute break, after which the second round started, which continued until the end of the unit. If a student left the classroom, we chose a substitute. All units were observed at least once. We analyzed which proportion of observed time students were on and off task, exploring differences between units and activities.

Teacher interviews. The first author held semi-structured interviews with the teachers involved in teaching TDLT-1. To stimulate recall, they were encouraged to browse through materials (e.g., workbook, teacher guideline) whenever needed. Interviews were analyzed directly from notes and audio recordings. We focused in particular on experiences shared by multiple teachers. The interview consisted of three parts:
1. General evaluation of main intervention components: Teachers were asked to tell about a moment in TDLT-1 that went particularly well and one that went not as well. Next, stories, internal and external dialogue tasks, the

teacher's own role, achievement of intervention goals, teacher guidelines and the preparatory meeting were discussed.

2. Teacher-specific questions based on teacher logs. Questions were asked about the teacher's elaborations in the logs. For example, one teacher noted: "I need a lot of words to ask a student: 'Do you mean to say...?' Then he or she replies 'Yes, exactly,' and I switch to another student. I will take that more into account in other lessons." In the interview, she was asked to elaborate: why did she feel this was important, and how was she taking it into account in other units?

3. Tips for new teachers. Teachers were asked about tips and tricks for new teachers who would teach TDLT-1. Finally, they were asked if there was anything else left to discuss and they received a gift card as a token of appreciation.

Student evaluation form. Students evaluated TDLT-1 by filling in a form. All items were evaluated on 5-point agreement scales, unless indicated otherwise. First, students scored evaluation words that followed the phrase "I found the units...", for instance "fun", "useful", "clear" and "confusing" (10 items). Negative items were recoded. Principal components analysis with Varimax rotation (KMO = .87; Bartlett's test p < .001) revealed two components which accounted for 57% of the total variance:

- Overall appreciation and relevance (eigenvalue 4.27, 42.7% of variance), including words like 'useful', 'fun' and 'boring';
- Overall clarity and comprehensibility (eigenvalue 1.45, 14.5% of variance), including 'clear', 'difficult' and 'confusing'.

Next, students completed 12 items on how meaningful, difficult (items recoded) and enjoyable they found internal and external dialogues. Principal components analysis with Varimax rotation (KMO = .74; Bartlett's test p < .001) revealed three components, which together accounted for 55% of the variance:

- Appreciation and relevance of external dialogues (eigenvalue 3.48, 29.0% of variance), indicating how meaningful and enjoyable small-group and whole-class dialogues were, such as "Talking in small groups about stories was meaningful".
- Clarity and comprehensibility of dialogue tasks (eigenvalue 2.08, 17.3% of variance) containing the recoded difficulty items, such as "Talking in small groups about stories was difficult".
- Appreciation and relevance of internal dialogues (eigenvalue 1.09, 9.1% of variance; content validity indicator), indicating how meaningful and enjoyable internal dialogue tasks were.

Students also indicated their appreciation of the stories they read (3 items), by assigning each story a grade (1-10). When asked to indicate which story they read in unit 3 and to evaluate it, only 396 of the 515 students who completed the form did so (i.e., almost 25% left blank which story they read). Apparently, many students could not remember which one they read in the third unit.

Further, we assessed students' sense of safety as an aspect of appreciation. Four items formed an internally consistent scale (Cronbach's α = .71). Finally, we asked students to list strengths and suggestions for improvement. They listed 683 strengths and 528 suggestions. Of the latter, 36 (6.8%) concerned the research in which TDLT-1 was embedded, such as comments about effect measures (e.g., "boring", "repetitive"). Ten responses (1.9%) referred to how students themselves could improve (e.g., "Pay more attention in class"). We decided to leave out these responses, after which 482 suggestions for improvement remained. We analyzed responses inductively to see if they would support other validity and practicality indicators.

Student evaluation task. In assessment phase 2, to assess story appreciation and difficulty, we asked students to evaluate these aspects for the story they read for the final task of TDLT-2, which they had chosen from four options. As two of these stories were also used in TDLT-1, this was deemed to give a valid impression of story appreciation and difficulty.

In addition, students evaluated the relevance of various intervention elements and the coherence between these elements and four main intervention objectives. They indicated whether or not elements (e.g., dialogue guidelines, teachers' explanations) were helpful for achieving progress in learning to 1) notice responses during reading, 2) deal with incomprehension during reading, 3) gain insights into reading experiences, and 4) actively contribute to dialogues about stories and reading experiences. We analyzed the frequency of binary responses (helpful or not) across elements and objectives. The 117 students who responded could all have indicated that each element was helpful for each objective, thus assigning "helpful" 32 times. In practice, they were more selective and indicated 1,784 times that an element had been helpful (on average 15 indications per student).

3   RESULTS

3.1 *First Development Phase*

This phase resulted in TDLT-1, a four-unit intervention. Its overall objective was, as described in the teacher guideline, "to help students identify connections between short literary stories and themselves (i.e., their personalities, the way they are, the way they think) and their view on the social world (i.e., how other people are, behave and think)". TDLT-1 consisted of one preparatory unit, aimed at understanding and applying strategies for engaging in and deepening external dialogues about stories. The subsequent units were "reading-and-dialogue" units. The goal of each unit was presented in students' workbooks. The workbook further contained all instructions, tasks, and space to write answers and notes during group dialogues. Stories were bundled in a separate booklet. We also designed PowerPoint slides with instructions. The units of TDLT-1 are described in closer detail in Appendix C; here, we indicate how we operationalized the initial design principles.

3.1.1   *Operationalization of text choice principle*

In line with this principle – thematically relevant fictional texts should be selected – we used texts addressing peculiar, presumptuous or painful social interactions between characters (see Table 4.2 for descriptions). In response to the teacher input collected during the initial design, we selected both canonical short stories often used in 10th grade (e.g., *A plate with spaghetti* and *Blood*) and more recently published stories (e.g., *She was everywhere* and *Flight behavior*). Their thematic appropriateness was confirmed via expert consultation and trial studies. Students read a story in units 2, 3, and 4. For unit 3, they selected one story: in unit 1, they read several sentences and a description of each story and indicated which two they preferred. The teacher decided which of these a student would read, to group students for the external dialogue task in unit 3.

3.1.2   *Operationalization of the internal dialogue principle*

To operationalize the second principle – design writing tasks, related to texts and themes, that stimulate students to activate previous personal experiences before reading, notice their experiences during reading, and/or reflect on evoked experiences directly after reading – we used pre-reading tasks, reading instructions and individual reflection tasks that focused on story themes and

transformative reading experiences. As an example of a pre-reading task, students wrote in unit 4 about their ideas of an afterlife, to prepare for reading about someone in a plane crash, who has, to his own disbelief, a religious experience. Personal views were thus activated in a safe and time-efficient way. An example of a reading instruction was: "Try to pay close attention to your own responses while reading the story: which thoughts, ideas and feelings does it evoke in you? What in the story stands out to you?" Finally, as an example of reflection directly after reading, students were asked to indicate which part of the story stood out most to them (aesthetic awareness) and the extent to which the story evoked, for example, imagery and sympathy.

Table 4.2. Stories used in TDLT-1

Story	Author	Description	Unit
1. *Ze was overal* (She was everywhere)	Ed van Eeden	Confused and suspicious man keeps thinking about ex-girlfriend; he ends up on the roof of a library, while people in the street below are staring at him.	2
2. *Een bord met spaghetti* (A plate with spaghetti)	Adriaan van Dis	Man in restaurant thinks a black man stole his plate; he passive-aggressively confronts him, but finds out he misjudged.	3*
3. *De biefstuk van het zoete water* (The freshwater steak)	Hans Dorresteijn	Boy is nervous during a fishing trip, because his father always physically punishes him several days after he has misbehaved.	3*
4. *Het recht* (The right)	Annelies Verbeke	Man distrusts his black cleaning lady; he tries to trick her into stealing, but she gets the better of him.	3*
5. *Bloed* (Blood)	Gerard Reve	Child is physically abused by guardian and takes revenge by causing him to take a deathly fall.	3*
6. *Merkwaardig verhaal* (Curious story)	Elke Geurts	Girl is phoned by her grandmother, who instructs her to act as if she has died; awkward conversation about the cause of grandmother's made-up death.	3*
7. *Vluchtgedrag* (Flight behaviour)	Bertram Koeleman	Man experiences a plane crash and, in the final moments of his life, is confronted with how religion may play a role in such circumstances.	4

Note. * In unit 3, students read one of these five stories.

3.1.3   *Operationalization of the external dialogue principle*

Expert consultation and trial studies already indicated that the third principle – design exploratory dialogic activities that stimulate students to verbally share personal experiences related to texts and text themes – might be difficult to operationalize: teachers felt that students needed guidance for talking in small groups or pairs, not led by the teacher. We indicate below how the external dialogue principle was operationalized in the preparatory unit dedicated to dialogue guidelines, as well as in subsequent reading-and-dialogue units.

Preparatory unit. Students first considered what characterizes good and less good dialogues and were provided with five guidelines: Listen carefully, Ask follow-up questions, Postpone a first judgment, Distribute speaking time equally, 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"). After observing their teacher modeling how to ask follow-up questions, they talked in small groups about a reading-related topic (i.e., what is important when choosing a book to read), using cards with follow-up questions, for instance, "Can you give an example?" or "Who has a different opinion?". Small-group experiences were then shared in class and students reflected individually, in their workbooks, on strengths and points for improvement of their dialogues.

Reading-and-dialogue units. Whenever students started an external dialogue task, teachers reminded them of the guidelines. In unit 2, students had indicated which reading experience was most prominent and were grouped accordingly: for example, students who strongly pictured the story in their minds (imagery) formed a group. Each group completed a dialogue task to deepen their prominent reading experience. For instance, the imagery group talked about what the images in their minds looked like, explored the atmosphere the images evoked, discussed and selected photographs that represented the atmosphere, and talked about which five words from the story fit the chosen photographs best. In unit 3, the dialogue started with deciding on the most important story moments, by drawing a story board. Students then talked about which life lesson they derived from the story. In unit 4, the dialogic task was a speed date: students talked in the first round about how they would feel and react if they were the story characters; in the second, about which associations the story and its characters evoked; in the third, about what happened at the end and how they thought it connected to the theme of religion, which was prominent in the story.

External dialogues were also operationalized at the classroom level at the end of each unit. For instance, in unit 2, group representatives explained what their group had talked about, while other students indicated in their workbooks whether what they heard made them reconsider their own small-group task. In unit 4, students first individually wrote down what they learned from the project, shared this in pairs, and exchanged it in a teacher-led classroom dialogue.

3.2 *First Assessment Phase*

In examining the validity and practicality of TDLT-1, we first address the intervention as a whole, followed by selected stories, internal dialogue tasks, and external dialogue tasks, in line with the design principles. Quantitative results from teacher logs, time on task observations, and evaluation forms are substantiated with results from teacher interviews (see Table 4.1, p. 98) and strengths and suggestions for improvement listed by students (the latter are summarized in Figure 4.2 and 4.3).

3.2.1   *Intervention as a whole*

Content validity. In their logs, teachers generally agreed students were interested and engaged in TDLT-1 (M = 3.8, SD = .8). Students themselves neither highly valued nor highly disliked it (Overall appreciation and relevance: M = 3.0; SD = .8): 47.6% responded negatively (M < 3.0), and 45.4% responded positively (M > 3.0; for 7%, M = 3.0). Students agreed to feel safe during the intervention (M = 3.9, SD = .7).

Students' mixed responses were reflected in strengths and suggestions they listed. Of the strengths, 23.7% referred to valuable, relevant learning outcomes. 7.5% concerned general strengths (TDLT-1 was "important" or "fun"). A small share (2.3%) concerned a safe social atmosphere in class. Of the suggestions for improvement, 21.7% concerned TDLT-1 not being fun or engaging. Its purpose and relevance were sometimes questioned as well (6%). Teachers agreed in interviews that students needed clearer goals and more insight into steps to work toward those goals. One teacher suggested that a rubric for "noticing reading experiences and talking about it" might be helpful.
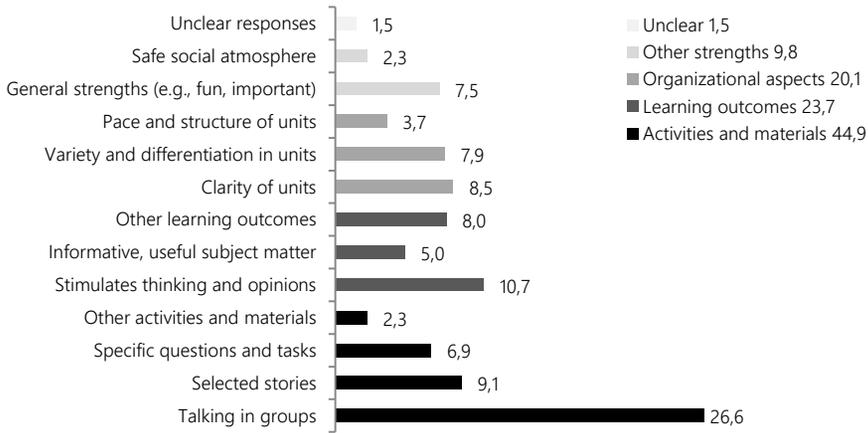
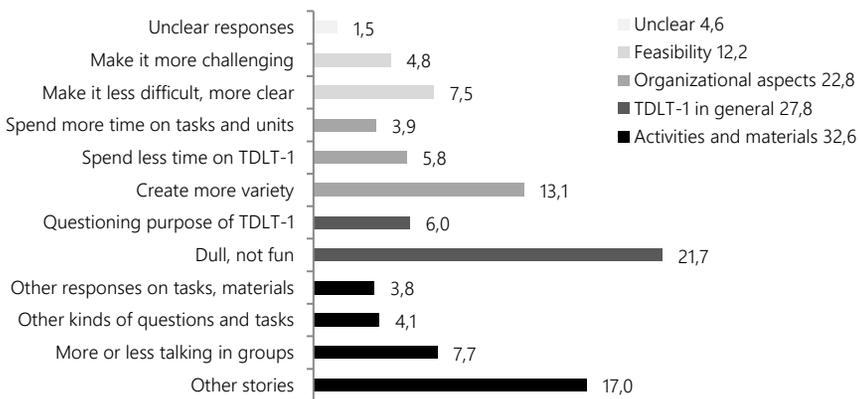Figure 4.2. Categories of strengths (% of 683 strengths).

Figure 4.3. Categories of suggestions for improvement (% of 482 suggestions).

Construct validity. In their logs, teachers indicated the units were clear for students (M = 4.1, SD = .5). Students also were neutral to positive about Overall clarity and comprehensibility (M = 3.6, SD = .8). In addition, 8.5% of the strengths they listed concerned the clarity of the units, and 3.7% their pace and structure. As 13.1% of the suggestions for improvement concerned calls for more variety, we hypothesized that TDLT-1 might even have been too coherently structured.

Practicality. Teacher logs showed that 88% of the phases was fully completed as intended; 9.4% was partly completed, and 2.7% was not completed. For

partly completed phases, teachers mostly noted they had spent somewhat less time on them than planned. In general, organizing various dialogic activities around a short story in a single 50-minute unit seemed manageable. Teachers agre/ed that teaching the phases was attainable (M = 4.2, SD = .5) and that there was order and discipline in the classroom (M = 4.0, SD = .6).

Time on task findings, however, showed different results. Students were on task in 72.5% of the observed time and off task in 24.9% of the time (in 2.5%, task behavior was "unclear"). The on task percentage was below the standard of 80% that has been suggested in studies on effective teaching (Kauchak & Eggen, 1993; Muijs & Reynolds, 2010). In addition, it varied significantly among units (χ^2(6) = 15.09, p = .020), which was due to unit 3 (67.7% versus 73.1%, 73.9%, 74.2%). The same was true for variance among teachers (χ^2(24) = 124.21, p < .001): only two of them scored above the 80% norm (range 54.1% – 84.8%). Similarly, on task percentages varied significantly among activities (χ^2(8) = 56.54, p < .001), as will be illustrated in relation to the design principles.

For teachers, feasibility seemed related to the intended active pedagogy of TDLT-1 (e.g., short phases in a high pace, group work, organizing short presentations). In interviews, teachers either said to appreciate it ("it's something else", "it challenges students", "you can organize so much more in a lesson than I expected"), or suggested it was "too much fuss" or "strange for students, who are too passive for this". Not for all teachers, thus, an intervention that required such a strict organization was feasible. Although they all appreciated the introductory meeting, some suggested to organize an active workshop to become familiar with the intervention.

Students considered feasibility a point for improvement (12.2% of all suggestions). They felt that units were either not feasible, for example, "It was hard to complete everything in detail in the time we were given", or not challenging enough, for instance, "I'd like to go deeper into the discussions but the tasks didn't really allow for it". In addition, in 3.9% of the cases they suggested to devote more time to the project.

3.2.2 Selected stories

Content validity. Students rated story appreciation on a scale of 1 to 10. Two stories were read by all students were rated on average 6.3 (*She was everywhere*, SD = 1.3) and 6.7 (*Flight behavior*, SD = 1.3). Stories in unit 3 were evaluated by fewer students; mean scores varied between 6.1 and 6.9, with standard deviations similar to the ones above. Overall, although standard deviations indicated considerable variety, students evaluated the stories neutrally to posi-

tively. However, 17% of the suggestions for improvement concerned stories ("Select better stories, these were vague"), against 9.1% of the strengths ("Nice stories").

Teachers evaluated the stories positively, in particular *Flight behavior*. For example, one teacher said the story was "really imaginable for [students] and concrete enough to talk about, they can really put themselves in the [plane crash] situation". Some teachers had reservations about *She was everywhere*: on the one hand, it was well-chosen because of its recognizable setting and comprehensible style, but it was also characterized as "intangible" and "too open". Yet, when asked if they would replace a story, none of the teachers indicated that they would.

Construct validity. In student data, we found no responses about the coherence of the stories. Three teachers commented on it during interviews. Even though they appreciated most stories, they suggested to select stories centering around a single theme. One teacher said: "I think that stories centered around a theme help students to identify relations between stories, themselves and how others think. It prompts them to consider a theme in multiple units. Something like 'injustice', maybe."

Practicality. In the interviews, most teachers indicated the stories were practical in terms of reading time: only occasionally, students could not finish their story in time. Some teachers read stories aloud, so that students finished and started the next task simultaneously; others had their students read individually, to fully focus on the responses stories evoked. Several teachers mentioned that, in practical terms, they preferred short stories over excerpts from novels, to offer students "the sense of a beginning and end", as one teacher said. The on task percentage for story reading was high (91%).

3.2.3   Internal dialogues

Content validity. For Appreciation and relevance of internal dialogues, students' mean score was 2.7 (*SD* = .7). On average, thus, they did not value internal dialogue tasks highly. In strengths and suggestions for improvement, they did not refer to internal dialogues. We suspect that internal dialogues were not prominent enough for students to reflect upon them and, potentially, value them. Several teachers endorsed this: they said that the purpose of internal dialogues could be made more explicit, as these were "strange" for students. For instance, one teacher said: "What students find strange is that these writing tasks are very open. For them it feels [...] as if anything goes."

Apart from this suggestion, teachers evaluated pre-reading activation tasks, tasks to notice responses, and reflection tasks after reading positively. For example, one of them said: "Thinking about a theme prior to reading is valuable and safe. We should do it more often." Teachers also recalled that, despite the purpose perhaps not being entirely clear, students were generally engaged in the tasks, as this response illustrates: "My students completed the tasks rather seriously and felt they were heard and taken seriously."

Construct validity. Students evaluated Clarity and comprehensibility of dialogue tasks negatively to neutrally (M = 2.8; SD = 1.0; applies to internal and external dialogues). Most teachers felt the tasks were clear and well-structured. For instance, one of them said: "The tasks had a clear buildup. I think that helped my students to focus." In terms of overall coherence, some teachers noticed that internal dialogue tasks created starting points for external dialogues, which were then more profound than they usually observed. However, others said this happened "not as much as you would expect".

Some teachers observed that students struggled to notice responses during reading and suggested they might benefit from teacher think-alouds, to model that all sorts of responses may be evoked during reading. Another suggestion was to teach strategies for when students encounter difficulties in stories. Several teachers noticed that students tended to simply ignore difficulties and continue reading, which potentially hampered their thinking about story themes in relation to themselves and others. Teachers felt that a "step-by-step strategy to deal with difficulties" could be coherently connected to attention for reading responses. As one teacher said: "Things like 'Huh?' or 'I don't get it' are genuine responses that occur frequently among 10th graders, but they often just don't know what to do when it happens".

Practicality. Teachers mostly felt that the internal dialogue tasks were well-organized and were practical to work with. For example, a teacher said: "This went really well. The students understood the steps and had enough time to complete the tasks." Teachers further indicated that annotating or highlighting the text when noticing a response and guided reflection after reading (e.g., scoring statements about imagery and sympathy) worked well for students, "because it narrows down what students are asked to reflect upon", as another teacher said. The on task percentage for individual student activities was 77%.

3.2.4   External dialogues

Content validity. Like the internal dialogues, students did not value external dialogues highly (Appreciation and relevance of external dialogues: M = 2.8; SD = .8), even though "talking in groups" was most frequently mentioned as a strength of the project (26.6% of all strengths). Of the suggestions for improvement, only 7.7% concerned group talk activities.

Teachers appreciated that external dialogue tasks invited for talking about stories, which they considered a valuable activity in itself. They were most enthusiastic about the speed dates in unit 4. Evaluations of other tasks were mixed. Some teachers were positively surprised, as this response illustrates: "There was some real improvement in my class. I didn't expect it, but students were actually engaged in practicing to talk about reading experiences. Even now, after the project, they sometimes refer to personal experiences." However, the relevance of external dialogues was not always obvious for students, as another teacher indicated: "It was quite difficult to get them talking. You know, they did do it, they came up with new ideas, but they just find it weird to share personal experiences." Other teachers observed a lack of follow-up questions and general disinterest in literature. "The dialogues were... well, mediocre," one teacher said. "Students are too easily satisfied. They don't ask follow-up questions, it just doesn't interest them. The talk remains artificial." Teachers also indicated that students were "quickly done talking" or "completed tasks superficially", even though there were always groups that engaged in more extensive, serious dialogues. Teachers suggested to offer students a set of follow-up questions they could use anytime and to clarify why talking about literature can be beneficial.

External dialogues also took place at the classroom level. Teachers noticed these dialogues were not always necessary for students, who felt "they had already discussed a topic," as one teacher said. Students might as well reflect individually on dialogues, "by scoring themselves and their group", as another teacher proposed.

Construct validity. As indicated, students evaluated Clarity and comprehensibility of dialogue tasks negatively to neutrally (M = 2.8; SD = 1.0). In terms of coherence, teachers mostly felt that the guidelines, introduced in unit 1, helped students to engage in dialogues, although "it doesn't come naturally, I constantly reminded them", as one teacher said.

Practicality. Concerns were raised about the practicality of external dialogue tasks. Although only 2.7% of all unit phases was not completed, this mostly

occurred toward the end of a lesson, when a classroom dialogue would take place. Mostly, teachers indicated there was not enough time left. Three of them suggested that whole-class dialogues might be shifted to a next lesson. Units, thus, could potentially cross the boundaries of 50-minute lessons. This suggested that teachers felt the coherence among units could be released somewhat. During whole-class activities, students were on task in 74.4% of the time.

In contrast, students were on task in only 65.5% of the time devoted to dialogues in pairs or small groups. Several teachers said that students needed more time to get used to talking about reading experiences, which they expected to result in less superficial talk and more on task behavior during small-group dialogues. Finally, several teachers explained how they prompted students to deepen a dialogue whenever they fell silent or went off task, for example: "I asked something like: 'Okay, I hear your conclusion. Could you now talk about whether others could have alternative opinions?' Then they explored other perspectives as well." They suggested to include a list of such questions and prompts in the teacher guideline, to offer new teachers more guidance in scaffolding students' dialogues.

3.3 *Second Development Phase*

The second development phase resulted in a six-unit intervention (TDLT-2). Like TDLT-1, it started with a preparatory unit in which the main objective was to understand and apply strategies for engaging in and deepening dialogues about stories and reading experiences. The five subsequent units were reading-and-dialogue units. To relevantly connect TDLT-2 to the regular curriculum, we designed a final individual writing task in which students were asked to apply what they learned. Teachers provided feedback and included the task in students' curricular literature portfolios. TDLT-2 is described in closer detail in Appendix D; here, we present specific reoperationalizations of the initial design principles.

3.3.1   *Reoperationalization of text choice principle*

Although all stories in TDLT-1 were considered to be thematically relevant for its aim, teachers suggested that more thematic coherence would help students to identify relations between stories and to consider more deeply how a story theme might be connected to themselves and their perceptions of others. In following the suggestion to select stories centering around "justice and injustice", we maintained some stories in TDLT-2 but substituted others (see Table

4.3). The story from TDLT-1 that teachers and students appreciated most, *Flight behavior*, did not fit this theme, but was given another role: we recorded two short videos of students modeling a good and bad example of a dialogue about this story, which were observed and discussed in the preparatory unit.

Table 4.3. Stories used in TDLT-2

Story	Author	Description	Unit
1. Excerpt from *Flight behavior*	Bertram Koeleman	See Table 4.2.	1
2. *Dood (Death)*	Martin Bril	A girl realizes her love interest has died because of senseless violence, and thinks about all that will never happen anymore.	2 & 5*
3. *Blood*	Gerard Reve	See Table 4.2.	3 & 5*
4. *Volgens de regels (Following the rules)*	Mirjam Bonting	A father has always set strict rules for his daughter; when he falls during mountaineering, she follows the rules and leaves him behind.	4 & 5*
5. *Van geluk spreken (Count oneself lucky)*	Marga Minco	A woman meets an acquaintance after World War II; they have a painful conversation about who survived the war and who didn't.	6
6. *The freshwater steak*	Hans Dorresteijn	See Table 4.2.	6**
7. *Een najaarsdag (An autumn day)*	Thomas Heerma van Voss	A guard feels compassion for a summer camp host on death row, who murdered children; after the execution, the guard tells his son a bedtime story about a friendly summer camp host, but his son falls asleep before the end.	6**
8. *The right*	Annelies Verbeke	See Table 4.2.	6**
9. *Hoela (Hula)*	Cees Nooteboom	From behind a window, a boy at a birthday party watches his little nephew drown in a garden pond, without doing anything to help.	6**

Note. * In unit 5, students were asked to compare these three stories. ** In unit 6, students chose one of these four stories for the final task and read it at home.

3.3.2   Reoperationalization of internal dialogue principle

Overall, teachers evaluated internal dialogue tasks in TDLT-1 positively: they were clear, well-structured, and practical to work with. A relatively high on task percentage endorsed this. However, students were not too appreciative of the tasks and hardly commented on them in the qualitative data. As suggested by

various teachers, it may be important to explicate for students why and how they learn to engage in internal dialogues, as they were not used to this approach.

For TDLT-2, we therefore designed a rubric with learning objectives, among which a) learning to notice responses during reading, b) learning to deal with incomprehension while reading, and c) gaining insights into reading experiences, for which three levels were described (see Appendix D, p. 280). Students indicated their starting level in unit 1 and evaluated their progress at the end of the intervention.

In addition, we postponed explicit attention for transformative reading experiences until unit 3. We included a unit focused on evaluative responses (e.g., "fun", "unexpected", "boring", "challenging"; unit 2), as students were more familiar with such responses than with responses related to transformative reading: evaluating fictional texts is usually attended to in lower grades of secondary education.

Whereas students read stories in unit 1 and 2, attention for internal dialogues remained implicit. Unit 3, then, was dedicated to noticing responses during reading. After discussing explicitly why this could be beneficial for students, teachers explicated personal responses by thinking aloud during reading, while being observed by students. Only after these explicit preparations, students were asked to focus on their own responses and to reflect on transformative reading experiences. Moreover, dealing-with-difficulties strategies (i.e., stop reading for a moment, write a question mark, think about possible meanings, ask the teacher or classmates for help) were introduced and summarized on a 'first aid card', which was available to students at all times.

3.3.3   Reoperationalization of external dialogue principle

Whereas students considered it a strength of TDLT-1 that they could talk in groups during literature class, they also indicated it was not always clear why this was relevant and how they could do it well. Teachers indicated that students asked too few follow-up questions and were quickly done talking. The low on task percentage substantiated their impression. Teachers felt their students needed more time to practice and get used to external dialogues, as well as more feedback on and scaffolding of dialogues. They also questioned the need for whole-class dialogues at the end of each unit in TDLT-1.

First, as indicated, we extended TDLT-2 from four to six units. This enabled a buildup in how challenging and unfamiliar external dialogues were. In unit 1, dialogues concerned famous quotes about literature and reading (e.g., "We

read to know we're not alone", William Nicholson). Students identified what a quote meant, evaluated if they agreed with it, and explained their opinion. In unit 2 they talked about their opinions about a story, explained them by referring to literary devices in the text (e.g., flashbacks, gaps), and presented conclusions on a poster. These units prepared students for subsequent ones, in which they were asked to identify, evaluate and explain, for instance, connections between story themes, themselves and others. For example, in unit 4 they read part of a story, were instructed to imagine to be in the protagonist's position, and talked from that point of view about possible story ends.

In TDLT-2, more time was available for small-group dialogues because whole-class dialogues were no longer implemented in all lessons: because units crossed the boundaries of lessons, these dialogues were sometimes shifted to a next lesson. In other occasions, whole-class dialogues were replaced by individual reflections, for example, when students indicated how well they applied dialogue guidelines and wrote down how they might improve in a next unit. Finally, more 'incubation time' was available because TDLT-2 taught in four to eight weeks, rather than two to four weeks.

Furthermore, teachers explicated why and how students should engage in external dialogues, in line with reoperationalizing the internal dialogue principle. First, the rubric contained the objective "contributing actively to dialogues about stories and reading experiences". Second, teachers discussed in unit 3 the purpose and relevance of sharing reading experiences. Moreover, dialogue guidelines were summarized on the first aid card and were thus available to students at all times. The card also included examples of follow-up questions and topics to bring up in dialogues. Finally, we used peer modeling videos in unit 1, to show examples of what dialogues about reading experiences should (not) be like.

Finally, the teacher guideline was adapted. We implemented more feedback moments (e.g., when students presented opinions and support on a poster in unit 2, teachers wrote comments and questions on post-its attached to it). We also added specific instructions for providing students with process-oriented feedback, including a list of example responses and questions teachers could use to prompt students to deepen a dialogue. During the workshop for new teachers, such feedback situations were imagined, discussed and rehearsed.

3.4 *Second Assessment Phase*

We examine the validity and practicality of TDLT-2. If applicable, we indicate whether significant differences compared to TDLT-1 were found.

3.4.1 Intervention as a whole

Content and construct validity. Teachers indicated in their logs that students were generally interested and engaged in the intervention (M = 3.7, SD = .8) and that it was clear for them (M = 4.0, SD = .7). Scores did not differ significantly from TDLT-1.

Practicality. Teachers indicated that of all phases, 77.6% was fully completed as intended, 14.4% was partly completed, and 8.0% was not completed. These percentages differed significantly from TDLT-1 ($\chi^2(2)$ = 12.43, p = .002), in which more phases were fully completed (88%) and fewer were partly and not completed (9.4% and 2.7%). As in TDLT-1, teachers had spent somewhat less time than planned on partly completed phases. Non-completed phases were due to time constraints or deliberate decisions (e.g., one teacher felt her students were too agitated and distracted at the end of the day to discuss life lessons). In general, however, teaching was deemed attainable (M = 4.1, SD = .7) and there was order and discipline in the classrooms (M = 3.9, SD = .8). Scores did not differ significantly from TDLT-1.

Students were on task in 85.2% of the observed time, and off task in 14.7% of the time. The on task percentage was well above the standard of 80% (Kauchak & Eggen, 1993; Muijs & Reynolds, 2010) and was significantly higher than in TDLT-1 (85.2% versus 72.5%, $\chi^2(1)$ = 39.36, p < .001). On task percentages varied significantly across activities ($\chi^2(8)$ = 17.44, p = .026), as illustrated in subsequent sections. In contrast with TDLT-1, percentages did not vary significantly across units, nor across teachers. All teachers seemed to be equally able to engage their students in efficiently spent learning time.

3.4.2 Selected stories

Content and construct validity. For the four stories they could choose from for the final intervention task, students' mean appreciation score was 3.5 (SD = .8), ranging from 2.9 for *Hula* to 3.8 for *An autumn day*. Students thus evaluated the stories neutrally to positively. Even though we used a different scale than in assessing TDLT-1, story appreciation was rather alike. In terms of relevance, students indicated repeatedly that simply reading these stories helped them to achieve their goals (15.2% of all indications; see Table 4.4). We collected no data that indicated the construct validity of selected stories, as we only adjusted the central theme of the intervention.

Table 4.4. Distribution of TDLT-2 elements perceived as helpful (N = 1,784)

Intervention element	%	% per objective*			
		1	2	3	4
1. Observing peer modeling videos: good and less good dialogue about reading experiences	5.8	2.0	0.9	0.8	2.1
2. Dialogue guidelines on first aid card, e.g., listening carefully, asking follow-up questions	9.5	2.5	1.6	1.8	<u>3.6</u>
3. Other suggestions on card, e.g., strategies to deal with incomprehension, follow-up questions	9.6	2.4	2.1	2.3	2.9
4. Reading the selected stories	15.2	3.6	<u>4.1</u>	<u>4.1</u>	3.4
5. Observing the teacher modeling to notice responses during reading (teacher think-aloud)	14.5	<u>4.9</u>	3.1	<u>3.4</u>	3.1
6. Focus on noticing own responses through reading instructions (e.g., annotating)	13.7	<u>4.5</u>	3.1	<u>3.4</u>	2.7
7. External dialogue tasks in pairs, groups or class	14.7	3.1	<u>3.4</u>	3.1	<u>5.1</u>
8. Teacher's explanations about noticing responses, sharing responses, and other topics	16.9	<u>4.0</u>	<u>4.3</u>	<u>4.1</u>	<u>4.4</u>
Totals	100	27.0	22.6	23.0	27.4

Note. * Elements helpful for objectives: 1) noticing responses while reading; 2) dealing with incomprehension while reading; 3) gaining insights in reading experiences; 4) contributing actively to dialogues about stories and reading experiences (see Appendix D). For each objective, the three elements most frequently perceived as helpful are underlined.

Practicality. Students indicated that the story they read for their final intervention task was not too difficult (M = 2.5; SD = 1.1). The mean value for three of the four stories was below the scale mean of 3; only *Hula* was considered more difficult (M = 3.5, SD = 1.1). As an additional practicality indicator, the on task percentage for story reading was 84.6%, which was only slightly lower than in TDLT-1 (91%).

3.4.3   Internal dialogues

Content and construct validity. Students indicated that learning to notice responses relatively often was helpful for them (13.7% of all indications; see Table 4.4). Thus, at least for part of the students, it was relevant to engage in internal dialogues with stories. However, suggestions on the first aid card, which included those for dealing with difficulties, were somewhat less often considered helpful (9.6%). In contrast, teachers' explanations about, amongst other topics, literary reading and noticing responses were most often deemed helpful (16.9%). The same was true for the teacher modeling how to notice responses

(14.5%). As Table 4.4 indicates, the latter was most often considered helpful for learning to notice responses, which suggested a coherent connection between the reoperationalization of the design principle and the objective.

Practicality. During individual internal dialogue activities, students were on task in 86.4% of the time, which suggested the activities were completed as intended.

3.4.4   External dialogues

Content and construct validity. Students indicated that external dialogue tasks relatively often helped them to achieve progress on intervention objectives (14.7% of all indications; see Table 4.4). In particular – and not surprisingly – they felt these tasks helped them to learn to actively contribute to dialogues about stories and reading experiences. This indicated a coherent connection between the reoperationalization of the design principle and the objective. Support for external dialogues was less often seen as helpful (peer modeling videos: 5.8%; dialogue guidelines: 9.5%). Teachers' explanations, which also concerned how to engage in small-group dialogues, were most helpful to students, as indicated above (16.9%). Clearly, the role of the teacher in the intervention cannot be underestimated.

Practicality. The on task percentage for whole-class activities (76%), that included classroom dialogues, was lower than for the other activities. However, for dialogues in pairs or small groups, it was clearly higher (90.1%). This was in stark contrast with TDLT-1, where the on task percentage for such activities was only 65.5%.

3.5 *From Initial Design Principles to Subprinciples for Operationalization*

The design process described in this paper started from three rather broad design principles. Here, we complement them with subprinciples that represent how the initial principles were eventually operationalized. By specifying the initial principles, we offer practical suggestions for designing literary instruction that focuses on transformative reading and gaining insight into human nature:
1. Fictional texts should be selected that are thematically relevant for the intervention aim.
    a. These texts should coherently center around a single relevant theme. The purpose of thematic coherence (here: justice and injustice) was to help students to identify relations between stories, themselves and

their perceptions of others. Thereby, rather than introducing a new theme in each unit, students could build on outcomes of previous dialogues.

b. For 10th grade students, texts can be selected that teachers consider to be *literary texts*. We used stories that teachers considered to be appropriate in terms of literary quality, for which on task percentages during reading were high. Whereas previous studies were indecisive as to the role of literary reading in fostering students' insight into human nature (Schrijvers et al., 2018), the current study indicated at least that 10th grade students and their teachers considered reading literary texts to be appropriate and helpful for the objective "gaining insight into reading experiences", even though its effects on students' insight into human nature requires further analysis.

2. Writing tasks should be designed that stimulate students to activate previous personal experiences before reading, to notice and annotate responses during reading, and/or to write down or reflect on responses directly after reading, to stimulate an internal dialogue between reader and text as preparation for external dialogues.

a. Teachers should explicitly explain the purpose and importance of internal dialogue tasks. Students appeared to need guidance in determining the relevance of engaging in internal dialogues with a text. In TDLT-2, we found that they appreciated their teacher's explanations about this topic as well as other topics.

b. Students should observe the teacher modeling how to notice responses while reading. Students in 10th grade are not necessarily used to paying attention to the responses evoked by a literary text. The redesigned intervention showed that students found it helpful to observe their teachers, who modeled noticing responses during reading.

c. Students should be taught strategies for dealing with difficulties during reading a story, which they may consult at any time via specific support tools. Incomprehension is a rather common response that students may notice while reading. Attending to strategies for dealing with difficulties stimulates students to not simply ignore them, but to work toward finding solutions for them. A support card or other tool that summarizes such strategies may remind students of applying them.

d. Internal dialogue tasks should first focus on more familiar responses (i.e., opinions about a story) before turning to transformative reading experiences. We implemented a buildup in the intervention: students

first engaged in tasks they recognized from lower grades of secondary education (i.e., thinking about their overall opinion about a story) and then gradually moved to less familiar internal dialogues in which they were asked to attend to responses related to transformative reading.

3. Students should engage in exploratory dialogic activities that stimulate them to verbally share their personal experiences related to texts and text themes.

   a. Teachers should explicitly explain the purpose and importance of external dialogue tasks. Similar to internal dialogues, students were given guidance in establishing the relevance of external dialogues. Teachers thus attended to the importance of talking about multiple reading experiences and to the characteristics of a good dialogue.

   b. Students should be taught guidelines and other suggestions for external dialogues. The first aid card in the intervention included dialogue guidelines and other suggestions for deepening the talk about a story (e.g., examples of open-ended questions to be asked, suggestions for what to tell about the own reading experience). Students may be familiarized with such guidelines in a first, preparatory unit.

   c. Teachers should be given guidelines for providing students with feedback and guiding their dialogic processes. For teachers who are new to dialogic approaches in the literature classroom, it may be challenging to assist students without being too involved in the content of their dialogue. Therefore, teachers were given prompts and questions to stimulate small-group dialogues (e.g., "What (else) does this story make you think about?", "Why might anyone have different opinions?", "I hear you struggle with [x]. Could we think about how to solve that issue?")

   d. Students should be given enough time to learn to engage in external dialogues about stories. Four units, designed for four 50-minutes lessons, appeared to be too few for students to get used to talking about stories and reading experiences; in TDLT-2, we opted for six units. Ideally, however, attention for external dialogues and preparation for them by internal dialogues should not be confined to several units or lessons, but may be interwoven in the regular literature curriculum in secondary schools.

4   DISCUSSION

This study has resulted in a literature classroom intervention that aims to foster 10th grade students' insight into human nature. The intervention (TDLT-2) is characterized by an emphasis on learning to notice responses in internal dialogues that are related to transformative reading, and share these in external dialogues. For example, students are asked to reflect on and talk about whether they picture the setting and characters in a story in their minds, or feel sympathy and compassion for characters.

   In adult readers, such experiences have been shown to result in insights into themselves and others (Fialho, 2018). The present study has paved the way for investigating whether these effects also occur in adolescents in the context of the literature classroom, as it resulted in a valid and practical intervention that can function as a dependent variable in future effect studies. In addition, the study has offered concrete suggestions for how the three initial design principles are to be operationalized in teaching practice. In what follows, we reflect on the validity and practicality of the eventual intervention (TDLT-2) and evaluate the design process, before turning to the conclusions.

4.1 *Validity and Practicality of the Intervention*

The study confirmed, first, the validity of the selected stories. The main outcome of the iterative design process was to use stories that centered around a single relevant theme in TDLT-2, to increase coherence and offer students multiple opportunities to reflect on the theme and consider how it connects to themselves and the lives of others. Teachers felt the stories were appropriate in terms of difficulty and literary quality. Students moderately appreciated them and indicated they were relevant for achieving progress on the intervention objectives. In terms of practicality, teachers preferred short stories over excerpts from novels and felt that organizing various learning activities before and after the story was manageable. For students, the stories were not too difficult to read.

   The main adjustment in redesigning internal dialogues was to discuss their purpose and the strategies to establish them more explicitly with students. In TDLT-2, students gained insights into different kinds of responses and reading experiences via, for example, the rubric used in unit 1, teachers' explanations, and observing their teacher who modeled reading with purposeful attention for the responses a story evoked. Overall, students felt that the attention for noticing responses evoked by a text and the teacher modeling what this looked like were helpful, relevant intervention elements.

The main challenge in the design process was to establish external dialogues in which students would share responses and reading experiences in such a way that it would stimulate their insight into human nature. Even though students felt that the opportunity to talk in groups was a strength of TDLT-1, they questioned the purpose of external dialogues, showed remarkably low on task percentages during small-group dialogues, and demonstrated artificial and superficial talk. Teachers ascribed this to students not being used to talking about stories, let alone about transformative reading experiences. They also felt there was not enough time in TDLT-1 for students to learn how to engage in external dialogues and considered it a challenge to guide student groups toward more profound dialogues.

To remedy these validity and practicality concerns, we extended the intervention from four to six units and let go of the strict structure of pre-reading task, noticing responses while reading, individual reflection, small-group dialogue, and classroom dialogue in a single 50-minute lesson; in particular, whole-class dialogues were shifted to next lessons or substituted by individual reflections on small-group dialogues. More attention was paid to the potential benefits of sharing reading experiences, as teachers explicitly discussed this with students, and to how students could achieve progress in contributing actively to external dialogues (rubric, support tools, teacher feedback). In addition, teachers were provided with more extensive guidelines for providing students with feedback.

After these adjustments, the on task percentage during small-group dialogues in TDLT-2 was considerably higher than in TDLT-1, and students indicated that teachers' explanations and dialogic tasks were helpful for learning to contribute actively to dialogues. As a whole, TDLT-2 was generally used as it was intended, as teacher logs and students' overall on task percentage indicated. All in all, the iterative design process resulted in a valid and practical intervention that may be implemented in literature classrooms to potentially foster students' insight into human nature.

4.2 *Evaluation of the Design Process*

The design process has several strengths, but is also subject to limitations. First of all, only a few teachers were able to commit to the project for a longer period of time. This led us to recruiting new teachers for the various steps and phases of the design process, which is both a strength and a limitation. On the one hand, the number of teachers involved yielded a large variety of experiences and perspectives on TDLT, for example, during expert consultation, trial

studies and interviews. On the other hand, some teachers mentioned that it was challenging to teach the intervention precisely because they had not been involved in developing it.

The instruments used in the assessment phases yielded valuable information regarding the validity and practicality of TDLT. We recommend to consider both teachers and students as stakeholders of an intervention, and to collect evaluation and implementation data of both groups. In the present study, teacher interviews and student evaluation forms and tasks were complementary and thereby offered a nuanced picture of validity and practicality. Moreover, digital teacher logs and time on task observations proved to be practical and time-efficient instruments for collecting implementation data. The high response rates for teacher logs indicated that it was feasible for teachers to complete them after each unit. Moreover, in the first assessment phase, they provided useful input for teacher interviews.

A potential criticism on the design process and our conclusions about validity and practicality is that they strongly depend on comparisons of data collected in the first and the second assessment phases. We purposefully compared teacher log data and time on task results of both TDLT versions, but these findings must be considered cautiously as the groups of teachers and students may not have been fully comparable.

Moreover, the instruments used in the assessment phases of the design process tell us little about how students talked about stories and reading experiences. Therefore, a process-oriented perspective on student dialogues and analysis of how students would express considerations of themselves and others in literary dialogues may be important to further develop the dialogic approach presented in this paper. While qualitative studies into dialogic reading discussions are available (Janssen & Pieper, 2009), to date they have not focused on students' insights into themselves and others that potentially emerge during such dialogues.

Finally, we assessed not only the validity and practicality of TDLT-1 and 2, but also their effectiveness, on which we will report in future publications. To this end, the units were embedded in quasi-experimental studies with strict research designs and data collection procedures. This may have influenced teachers, who were sometimes concerned about the time frame, and students, who may have felt to be involved in "just" a research project which was not part of their regular curriculum. Even though TDLT was taught by the students' own teachers and was embedded as optimally as possible in the curriculum, its ecological validity may have been affected. In future design projects, researchers may choose to focus in a first assessment phase solely on implementation

and evaluation. Once a valid and practical intervention has been established, a subsequent assessment phase can take the form of an effect study with a more rigorous research design.

4.3 *Conclusions*

All in all, this study has shown that iterative cycles of development and assessment may result in the design of a valid and practical domain-specific intervention. Triangulation of data sources explicated a wide range of experiences of teachers and students, from which valuable information for redesigning the intervention was derived. This, in turn, resulted in a set of design principles in which the initial ones were complemented by subprinciples for operationalization in the classroom. Our next step is to assess the effects of both versions of TDLT, to shed further light on their quality and to contribute to ongoing research on transformative reading by expanding its scope to adolescents in the literature classroom. In the meantime, we hope that this paper offers an example of how a comprehensive description of intervention development may be conceptualized, in order to open the metaphorical "black box" and to enhance the validity and replicability of domain-specific intervention research.