DEEPER READING OF POETRY AND PROSE

Effects of a reading process-focused intervention on the text comprehension of 10th grade students in the Netherlands

CORINA BREUKINK[1,2], EWOUT VAN DER KNAAP[2], AND HUUB VAN DEN BERGH[2]

[1]University of Applied Science, Utrecht, the Netherlands; [2]Utrecht University, Utrecht, the Netherlands

Abstract

The main result of this effectiveness study is that a reading program with a focus on students' poetry reading processes, based on observational learning via eye movement modeling examples, can improve students' reading comprehension for different text types. In a pretest-posttest design with an experimental group (ten classes) and a control group (five classes), students' self-efficacy regarding their own reading process and their reading comprehension were measured. Over a six-week period, teachers of Dutch and their students worked with the six experimental lessons, instead of the regular reading program: students observed and evaluated contrasting peer reading processes, reflected on differences with their own reading process, and then they practiced aspects of a deep reading process. The program resulted in significant progress in the reading comprehension of "expository texts" (*ES* = .66), "short stories" (*ES* = .66), and especially "poetry" (*ES* = .81). Furthermore, the self-efficacy test results show that students in the experimental condition experienced significantly more learning effect after the intervention period than those in the control group. Moreover, based on the learning reports, evaluation tasks and interviews, it appears that the participants in the innovative program have become aware of their reading and how they improved their performance.

Keywords: reading processes, observational learning, eye movement modeling examples, teaching-learning conversation, reading comprehension

Breukink, C., Van der Knaap, E., & Van den Bergh, H. (2023). Deeper reading of poetry and prose: Effects of a reading process-focused intervention on the text comprehension of 10th grade students in the Netherlands. L1-Educational Studies in Language and Literature, 23, 1-29. https://doi.org/10.21248/l1esll.2023.23.2.585

Corresponding author: Corina Breukink, Institute for Language Sciences, Trans 10, 3512 JK Utrecht, The Netherlands. Email: c.i.a.breukink@uu.nl

1. INTRODUCTION

Empirical research shows that observational learning can be supportive to learning cognitively complex tasks. For the school subject Dutch, an obligatory subject in Dutch schools, this has mainly been demonstrated for writing and writing processes, where the model can be either a teacher or a peer (see, for example, Bouwer & Koster, 2016; Braaksma et al., 2002; Couzijn, 1999). This intervention study, however, predominantly draws from studies by Couzijn (1995), Keehnen et al. (2015) and Rooijackers (2023), in which learning-by-observation from peers is successfully applied to learning to read different types of expository texts in secondary education. Moreover, Rooijackers (2023) has shown that the use of *eye movement modeling examples* (henceforth abbreviated as EMME) in observational learning in reading lessons enhances the comprehension of expository texts. EMME are dynamic eye-movement video's that represent the natural reading process over time via lines (viewing order) and fixations (where and how long someone looks) (cf. Holmqvist et al., 2015; Kok et al., 2022). Visualizations of reading processes are difficult to observe and interpret, and therefore they are accompanied by students' thinking aloud comments.

1.1 Reading comprehension within the school subject Dutch

In Dutch higher general (the second highest secondary education level in the Netherlands) and pre-university education (the highest secondary education level), reading poetry is not mandatory, even though students in upper higher general and pre-university are expected to be able to understand, interpret and evaluate expository texts and literary texts (Expert Group Learning Trajectories, 2009). We know that students in the Dutch classroom practice applying reading strategies to expository texts and that this does contribute to deeper text comprehension (Okkinga et al., 2018). Nonetheless in Dutch upper secondary education little attention is paid in class to the reading and discussion of literary texts (Janssen, 2009; Schrijvers, 2019). This seems particularly true for poetry (Oberon, 2016). Apart from a legitimization for their poetry teaching, Dutch teachers also need didactic tools to be able to discuss poetry, under the umbrella of "literary texts" (Witte, 2018). Apparently, teaching poetry seems to be a challenge not only in the Dutch classroom (cf. Sigvardsson, 2017).

1.2 Text comprehension and "deep reading"

In his construction-integration model, Kintsch defines "text comprehension" as the construction of a coherent text representation or *situation model* (Kintsch, 1998). This situation-model theory is widely accepted internationally (e.g., Cook & O'Brien, 2019; Zwaan, 2016). Text comprehension is a complex and largely unconscious cognitive process in which a reader must actively engage with a text (Kintsch &

Rawson, 2005). To describe the cognitive process of text comprehension, Kintsch (1998) distinguishes three types of text representation. First, at the *syntactic level*, a reader processes linguistic information (spaces, letters, words, phrases, and sentences) without constructing meaning. Second, at the *text base level*, a reader interconnects micro-textual elements (words, phrases, sentences, lines) and macro-textual elements (paragraphs, stanzas) that are *explicitly* expressed in the text (Kintsch & Rawson, 2005, p. 211). Third, at the *situation model level*, to construct a mental representation of the text, readers need to interconnect text parts, retrieve their prior knowledge, and integrate it into the text base. Moreover, constructing a situation model requires reading efforts and "frequently involves imagery, emotions, and personal experiences" (Kintsch & Rawson, 2005, p. 211). The situation-model theory thus suggests that similar processes underlie the reading of expository, narrative and poetry texts, although there is no consensus on this assumption within the academic discourse. Instead, literary theorists, cognitive psychologists, and empirical literary researchers assume genre-specific forms of reading and reading literacy, such as literary reading and poetic literacy (cf. Bortolussi & Dixon, 2003; Dixon & Bortolussi, 2015; Kleber, 2021; Miall, 2006; Peskin & Hanauer, 2023).

In parallel with "situation model level reading" we use the term "deep reading". "Deep reading" is also described as concentrated, attentive, or repeated reading (e.g., Applebee et al., 2003; Baron, 2021; Wolf, 2018). "Deep reading" is characterized by integrating background knowledge to the new information in the *text base*, making analogies, drawing inferences, and is the opposite of "shallow reading" like skimming and scanning (Baron, 2015; Kintsch & Rawson, 2005; Wolf, 2023). Literary texts require concentrated and repeated reading, both for deep understanding and for aesthetic pleasure. Although the focus in PISA-2018 is solely on measuring fifteen years old comprehension of expository texts, the description of higher-level text processing is consistent with the beforementioned definitions and descriptions of deep text comprehension: "inter-sentence integration, extraction of the central themes and drawing inferences, are critical skills for processing complex or multiple texts for specific purposes. If students fail at performing higher-level text processing functions, it is critical to know whether the failure was due to difficulties in these basic skills in order to provide appropriate support to these students (OECD, 2019, p. 24).

1.3 Findings from empirical literary studies

"What is special about poetry as a linguistic form is the intent to directly pattern sound, form and meaning" (Peskin & Hanauer, 2023, p. 1). Poetry is therefore pre-eminently characterized by linguistic and semantic incoherence, and ambiguity (cf. Van Dijk, 1979). Therefore, reading poetry requires a reader to make considerable cognitive efforts at every level of the comprehension process. For deeper reading and understanding of poetry, this means attention to the form and content of the text, but certainly also to the emotions and experiences that poetry can evoke in a

reader (Eva-Wood, 2004; Peskin & Hanauer, 2023). Thus, a reader must constantly switch between these different forms of understanding (cf. McCarthy & Goldman, 2017). A deep poetry reading experience is thus a combination of cognition and affect (Peskin & Hanauer, 2023, pp. 28-29). In this sense, "deep reading" is not reserved for long texts (OECD, 2019; Education Council & Council for Culture, 2019), but also for literary texts and certainly for poetry.

1.4 Aims and research questions

Learning to read poetry and prose more deeply is central to this experimental reading program. In this intervention study, we investigate to what extent observational learning with EMME is also effective for learning to read short stories and poetry texts (cf. Rooijackers, 2023). Therefore, we focus on the following two research questions:

1. To what extent do 10th graders who participated in the experimental reading program demonstrate better reading comprehension for expository texts, short stories, and poetry than 10th graders in the control group?
2. To what extent do 10th graders express learning effects of the experimental reading program for their own reading process?

2. METHOD

2.1 Participants

In the intervention study, 15 teachers of Dutch and 15 classes in 10th grade of higher general and pre-university education participated. Ten classes constituted the experimental group (n = 235), while five classes served as the control group (n = 122). Classes were randomly assigned to either the experimental condition or the control condition. Before the start of the experimental classes, teachers in the experimental and control condition were informed through information letters and prepared for their tasks and the implementation of the intervention in a kick-off meeting by one of the researchers. The teachers who implemented the experimental reading program also received a general introduction video and a short instructional video for each lesson, a PowerPoint presentation with embedded EMME, a teacher's guide, and an assignment booklet. The teachers taught the experimental lessons independently to their own classes. Once to twice a week, the first author was present at the school as a point of contact for the teachers who participated in the intervention study. In addition, the researchers were accessible by mail or phone. In four experimental classes, all lessons were observed by the first author and in one class she observed half of the lessons.

2.2 Research design

For this intervention study, we used a quasi-experimental pretest-posttest design with a control group, in which the experimental group participated in the innovative reading program, while students in the control group followed the regular Dutch reading program. All classes in both groups participated in the pre-test and the three posttests (see Table 1). This type of research design allows us to demonstrate the effects of the intervention on students' text comprehension. Furthermore, a pretest and a control group allow us to investigate possible threats to external or population validity (Shadish et al., 2002).

Per school, the pre- and posttests for the groups were administered by their own teacher of Dutch during the same teaching weeks. After the posttest, an at random sample of seventeen 10th graders from the experimental group was subjected to a *stimulated recall* interview on the condition that the students had participated in all six lessons. One boy and one girl were chosen from each class, except in the one case where there was an all-boy class. In three of the ten classes, one student did not show up at the scheduled interview appointment.

2.2.1 Construction of teaching materials

In the eye-tracking study, we collected eye-movement videos from 39 higher general and pre-university students in grades 8, 10, and 12 who performed eight reading comprehension tasks for poetry and prose, without guidance from questions. Immediately after the experiment, during a stimulated recall interview, the students reflected on salient aspects of their own initial reading process for the last two texts, respectively a prose text and a poetry text. Both types of data, the visualizations of students' individual reading processes, and the interview protocols, informed the EMME. In the various EMME, aspects of "shallow" (*textbase level*) and "deep" (*situation model level*) reading processes are peer modelled (see Figure 1). For the application lessons three and six, we designed a variety of assignments aimed at practicing aspects of deep reading, that were observed and discussed earlier in the EMME lessons.

While designing the reading intervention, a build-up in *text complexity* and *task complexity* between lessons was also considered. Assuming that students in grade 10 are more familiar with reading expository texts and narrative prose than poetry, in lessons 1 and 2 students first work with an expository text and a short story. In lesson 3, they make the transition to reading poetry. Furthermore, between the lessons centered on poetry reading processes, the *text order* is dictated by the perceived text complexity that students reported in the text comprehension study and the eye tracking study we conducted earlier. Finally, students observe reading processes first at the micro level (between words/phrases/lines) in the first EMME lessons and only at the macro level (between paragraphs or stanzas, at the level of

the whole text) in the later EMME lessons, to teach students to perform observational tasks.

2.3 Experimental intervention

2.3.1 The experimental reading program

The reading program consists of six 40 to 45-minute lessons. An application lesson is scheduled after every two EMME lessons so that students can also practice what they have learned with other poetry texts. At the end of each lesson, students individually reflect on the development of their own reading process. Finally, in lesson 6 they evaluate the eight texts they have read and discussed during the reading program.

2.3.2 Design principles

Table 2 presents a general description of the design principles of the program and how they were operationalized in learning and teaching activities. *Observational learning* is a form of social learning (Bandura, 1986) that appears to be an effective learning activity in contexts where complex tasks need to be taught. Through observational learning, learners gain insight into the learning process while their cognitive load is reduced (Braaksma et al., 2002). To ensure that students compare their own findings concerning deep reading processes with those of their peers, in both lesson formats, time is systematically planned for *student collaboration* and *teaching-learning conversation* (cf. Applebee et al., 2003; Beach et al., 2021; Janssen, 2009; Schrijvers et al., 2019). Both types of classroom interaction, can be understood as forms of dialogic learning. First, the students read a text individually, and in an EMME-lesson they individually observe and compare the reading behaviors in the EMME (see Table 2). These are designed as for students to learn from their classmates' reading experiences, observations of reading processes and perspectives on text comprehension. Collaborating with a neighbor is safe but can quickly produce consensus. The move to a classroom exchange of findings may then be less challenging, and it may well yield new insights into and conclusions about less and more deep reading behaviors and text comprehension. To support the teacher in this role, the teacher's manuals write out what is seen and heard in the EMME and how the reading processes affect the level of text comprehension (*text base* or *situation model level*). Therefore, the teacher's role in the teaching-learning conversation is crucial.

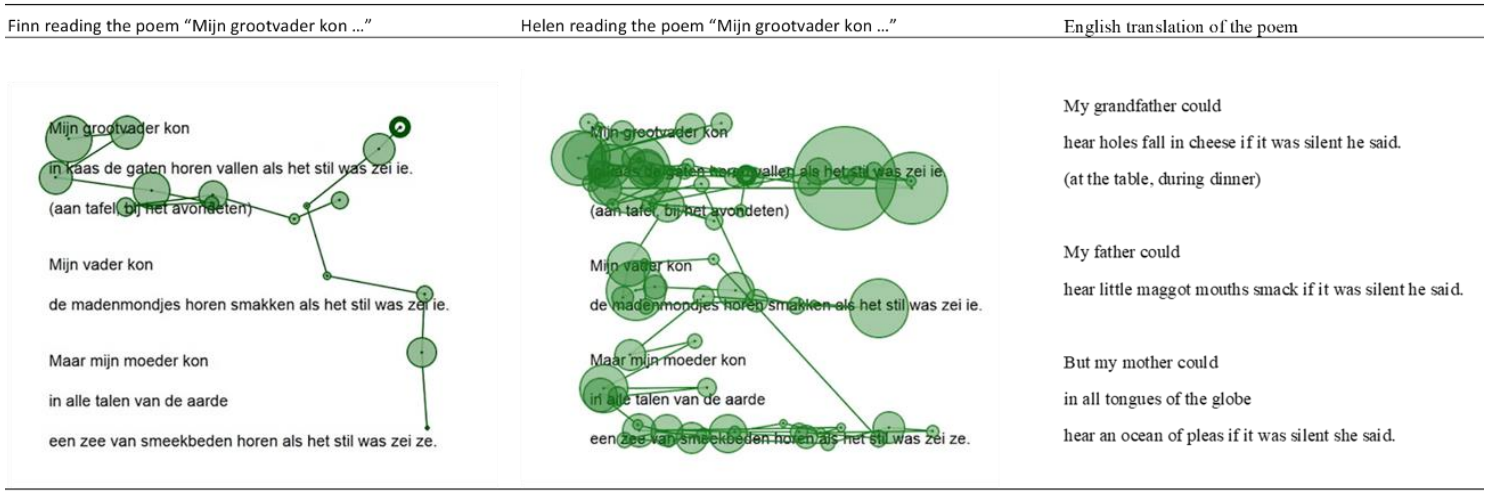
Naturally, not all design principles are covered in every lesson, although the reading process is at the center of both types of lessons.

In an *observation lesson*, students compare and evaluate reading processes of peers, and compare them to their own reading process. In an observation lesson,

Table 1. Research design with a pretest, an intervention period, and a three-tiered post-measurement, combined with a post-then-pretest (: PTP) and stimulated recall interviews (: SRIs)

Samples	Pretest	Intervention period (6 weeks)	Posttests			PTP	SRIs
	(expository text)		Posttest 1 (expository text)	Posttest 2 (short story)	Posttest 3 (poetry)		
Experimental group (*n* = 10 classes)	*n* = 252	Experimental lessons (*n* = 262)	*n* = 231	*n* = 231	*n* = 231	*n* = 223	*n* = 17
Control group (*n* = 5 classes)	*n* = 130	Regular lessons (*n* = 130)	*n* = 123	*n* = 123	*n* = 123	*n* = 79	

Figure 1. Screen Shots from two EMME in which Finn (left image) and Helen (right image), respectively, Read "My Grandfather Could …" for the First Time (Lesson 5)



students perform a reading task and an observation task consecutively. During an observation task, students watch the reading process of two peer-models reading, while thinking aloud. After the observation task, that is, in the peer talks and during teaching-learning conversations, students evaluate which peer model is the better reader and why. Then students compare and evaluate their findings about others' reading processes successively in pairs and during the teaching-learning conversation. During the teaching-learning conversation, it is explicitly the teacher's role to draw clear conclusions about (aspects of) deep reading and deep text comprehension together with the students. In lesson 5 for example, students first read the poem, “My grandfather could …” (see Figure 1). The poem is built on repetitions, parallels and a contradiction and it contains metaphors. Associating and connecting emotions to what is in the text are essential for a deeper understanding of this text. Therefore, lesson 5 focuses on associating and linking emotions to the text base to model deep comprehension. Students first read the text individually and actively by highlighting salient parts of the text while reading and then check off in a scheme which mood they associate the three individual stanzas with. Then, students observe the peer models “Helen” and “Finn”, who are also reading the poem for the first time, meanwhile reflecting aloud on their reading process and situation modelling. Afterwards, students compare and evaluate the modeled reading processes and contrast them with their own reading process. In the teaching-learning conversation, they subsequently discuss and motivate which learner model reached deep comprehension of the text. Teachers are supported by sample answers in the teacher's manuals.

An *application lesson* has a similar structure: from independently preparing or performing a reading task to interactive learning in the form of peer interaction and a teaching-learning conversation led by the teacher. These lessons focus on the development of students' *own* reading processes and text comprehension. Students perform various (preparatory) reading tasks individually or with a classmate, which trigger application and practice with insights from the observation lessons.

Table 2. Design principles, related to (Aspects of) the Reading Process and Type of Didactics of "Deeper learning to read poetry and prose", translated to Learning and Teaching Activities

Design principles			
Focus	*Types of didactics*	*Learning activities*	*Teaching activities*
1. Characteristics of reading processes, e.g., skim reading; attentive reading; reading order; (*rereading*); identifying comprehension problems; asking questions; recognizing text structure (e.g., *repetitions, parallels, paradoxes, contrasts, enjambments*); interconnecting text parts; relating to personal emotions and experiences (*associating*); concluding (e.g., *filling in gaps, thinking about ambiguity, metaphors*)	a. Observational learning b. Collaborative learning c. Teaching-learning conversation	I. Observing and analyzing of (aspects of) reading processes in the EMME: How do learners read and to what extent do they understand a text (fragment)? How do you observe this? How is this reflected in their thinking-aloud comments? II. After an observation task: Discuss observations with a classmate, draw conclusions and argue III. After an observation task: Sharing observations, listening to each other and the teacher	i. Presenting of the EME ii. Monitoring students' task performance iii. After an observation task: Identifying and discussing student findings; drawing conclusions about (aspects of) reading processes in the EMME

Design principles			
Focus	*Types of didactics*	*Learning activities*	*Teaching activities*
2. Contrast between (a) reading processes of student models and (b) the observed reading processes and one's own reading process	a. Observational learning	IV. After an observation task: Compare and contrast your own reading process with (aspect of) the reading processes of the student models, and ask for instance what you do differently from Finn or Helen when you experience comprehension problems with this text (fragment)?	i. Presenting of the EME
	b. Collaborative learning	V. After an observation task: Compare and evaluate observations and draw conclusions: which student model reads and understands a text (fragment) more deeply? How do you observe this? What do you infer from the think-aloud commentary?	ii. Monitoring students' task performance
	c. Teaching-learning conversation	VI. After an observation task: Evaluate observations and draw conclusions, listening to each other and the teacher: which student model reads and understands a text (fragment) more deeply? How do you recognize this?	iii. After an observation task: Identifying and discussing student findings; drawing conclusions about (aspects of) reading processes in the EMME

Design principles			
Focus	*Types of didactics*	*Learning activities*	*Teaching activities*
3. Contrast between (a) reading processes of student models and (b) the observed reading processes and one's own reading process	a. Observational learning	VII. Before and after a reading task: Activating prior knowledge; applying and practicing the aspects of deep reading discussed in class, such as making connections between sections of text; asking comprehension and essential questions about text parts or a text; detecting text structure; reflecting on one's own reading process	ii. Monitoring students' task performance
	b. Collaborative learning	VIII. During and after a reading task: Applying and practicing the aspects of deep reading discussed in class	ii. Monitoring students' task performance
	c. Teaching-learning conversation	IX. After a reading task: Evaluate one's own reading process and how it can be deepened: e.g., How do you read now? What would you do differently from now on?	iv. After an reading task: Identifying and discussing student findings; drawing conclusions about (aspects of) reading processes in the EMME
4. Evaluating texts	d. Practice	X. After all reading tasks: Compare texts in relation to perceived difficulty and personal rating: How difficult do you find this text?	ii. Monitoring students' task performance

2.4 Instruments to measure the effects of the experimental reading program

Pretest and posttest. To evaluate the effectiveness of the intervention, several tests were developed in the form of cloze tasks. In four texts, words were omitted in strategic places (*rational deletion strategy*), prompting readers to think about connections between text parts, i.e., text comprehension at the macro level (Kleijn, 2018). Two cloze tasks were created for two expository texts, one for a short story, and one for a poem.

For the pretest, we employed a 30-item cloze task on an expository text, which has previously been tested and used in another study (Rooijackers, 2023). To identify possible differences in text comprehension of the text types expository texts, narrative prose and poetry, we used three cloze tasks for the post-measurement: one for "expository texts", one for "short stories", and one for "poetry", each consisting of 20 items. The posttest "expository text" has also previously been tested (Rooijackers, 2023). To increase the feasibility of the cloze test "poetry", a list of the 20 missing words plus three additional words were added, from which students could select the best fitting word. One point was scored for each fitting answer; zero points for each non-fitting answer.

Post-then-pretest. To also assess how students in both conditions self-assessed changes in their reading process after the experimental or regular reading instruction, we developed a retrospective post-then-pretest. This test is a well-established and approachable way to collect participants' self-perceived changes in knowledge, awareness, skills, confidence, attitudes, or behavior (Hill & Betz, 2005). Students in both conditions were able to indicate on a five-point scale (1: "totally disagree" and 5: "totally agree") the extent to which they read after the lessons compared to otherwise. This data was collected immediately after the posttest. The test consists of 20 statements covering characteristic aspects of reading processes for poetry and prose, all of which are explicitly addressed in the lesson series (see Table 3). We categorized these statements into four dimensions, namely "reading process", "identifying comprehension problems", "making connections within the text", and "making connections with prior knowledge". The "reading process" dimension includes four statements about "reading more carefully" and "reading back through the text more often". The dimension "interconnecting text parts" consists of six statements related to linking text parts at micro level and at macro level. The four statements within the dimension "identifying comprehension problems" relate to awareness of comprehension problems and posing comprehension or essential questions while reading. Asking questions yourself while reading a text encourages reflection on a text and deeper processing of a text (Janssen, 2009). Finally, the dimension "making connections with prior knowledge" consists of six statements about linking one's own experiences, emotions, and prior knowledge to the text base.

Learning reports. At the end of each lesson, students completed a short learner report in their assignment booklet, for which the reflection question varied depending on the focus point in the lesson. To formulate the learning outcome of the lesson, students could use sentence stems (cf. Levine, 2019; Schrijvers, 2019), for example, "Because of this lesson, I now know how to ..." (lesson 2). The students in the experimental condition completed a total of 1264 learning reports. An average of 91.9% learning reports completed per lesson was handed in. The relatively low response rate for lesson 3 corresponds to the average teacher experience that there was relatively little time for conducting lesson 3.

Student self-evaluation task. In lesson 6, students were prompted to fill in a chart about the extent to which, after the lesson series, they felt more competent to *analyse*, *motivate*, *evaluate*, and *reflect* on their own reading process (*self-regulation*). They were then asked to write down the extent to which they thought they read differently or better after the lessons. 88.2% of students present in lesson 6 completed this evaluation task; only a very small number (0.76%) of students did not complete the task, while 10.3% were absent.

(Stimulated recall) interviews. Two students from each experimental class (n = 10) were randomly chosen to be interviewed about their experiences with and evaluation related to the essential components of the reading program, such as the EMME, the observation and reflection tasks and the teaching-learning conversations, and the experienced learning effects (see Appendix A). To refresh their memory, we gave students their own completed assignment booklets for review. A total of seventeen interviews were conducted; three students turned out to be absent at the scheduled time. Each interview lasted about 30 minutes.

2.5 Data analysis

Pretest and posttest. With three raters we scored all items according a previously discussed and pretested answer model, in which the allowed answers were made explicit. To determine the effect of the intervention, a multilevel model was constructed, in which the mean test score was estimated for each condition for the three text types, in addition to the variance between classes and students.

Assignment booklets. Three raters independently scored the assignment booklets ("1": assignment made, "0": assignment not made) in Excel. Next, the frequency distribution was calculated of the number of assignments made in SPSS.

Post-then-pretest. To assess whether the mean item scores in the experimental and control condition were significantly different, we performed an independent samples t-test in SPSS.

The first author coded the *learner reports*, the *evaluation tasks,* and the *student interviews* in Atlas.ti, using a thematic analysis (TA) approach (Braun & Clarke, 2006; Friese, 2019). In Atlas.ti, all student statements were systematically coded in relation to the design principles. In each sample (learner reports, evaluation tasks, and student interviews), the codes were eventually clustered into three code groups: (1)

Characteristics of reading processes, (2) Contrasting reading processes a. between peer models and b. between the observed reading processes and one's own reading process, and (3) Development of one's own reading process.

3. RESULTS

In this section, we will sequentially answer the research questions 1 and 2, concerning 10th grade students' reading comprehension and reading process. The first question relates to students' reading comprehension for expository prose, short stories, and poetry; the second question addresses students' reflections on and evaluations of their own changed reading process.

3.1 Effects of the experimental intervention on reading comprehension

Reliability estimates of pre- and posttests. A total of four cloze tests (the pre-test "expository text" and the posttests "expository text", "short story", and "poetry") were administered. The reliability estimates of the different tests are relatively high and ranges from *GLB* = .79 ("short story") to *GLB* = .86 ("poetry"). Thus, the test items are internally highly consistent. Analysis also shows that the reading comprehension scores are more or less normally distributed (i.e., skewness varies from -.4 to .2, and the kurtosis varies from -2. to .4). Only for the comprehension of poetry there are relative many students with low scores (skewness = -1.73) and the distribution is a bit too peaked (kurtosis = 2.10). However, this can be interpreted as a relative minor deviation from normality.

To answer research question 1, we ran a multilevel analysis in which students are nested within classes. Table 4 shows that the mean score on the pretest is almost the same in the experimental condition (M = 15.88) and the control condition (M = 15.47); this difference is not significant (χ^2 (1) = .21, p = .65). On the posttests, the differences between the two conditions are clearly larger. For example, for expository texts, the experimental group (M = 7.03) scored significantly higher on average than the control group (M = 4.78).

Table 4. Mean Test Scores and their Standard Errors (se) in the Intervention Group and Control Group in addition to the Standard Deviations (SD) between Classes and Students at both Measurement Occasions

		Posttests		
	Pre-test (se)	Expository text (se)	Short story (se)	Poetry (se)
Experimental group	15.88 (.50)	7.03 (.42)	8.65 (.25)	18.32 (.28)
Control group	15.47 (.73)	4.78 (.57)	6.73 (.32)	15.78 (.36)
SD_class	1.60	1.20	0.40	0.60
SD_students	3.90	3.20	2.90	3.10

To test the interaction between measurement moment and condition a test for planned contrasts is used (Goldstein, 1987). Results show that for all three types of posttests the interaction between condition and occasion reached significance: “expository texts: (χ^2 (1) = 9.32, p < .01); short stories: (χ^2 (1) = 3.91, p = .048) and “poetry” (χ^2 (1) = 8.51, p < .01). The effect size was also calculated. As a measure of effect size, we used the d_{pcc2} coefficient; this coefficient expresses differences between an experimental and a control group, considering differences already present on the pretest (Morris, 2008). The effect sizes are medium to high: “expository texts: (*ES* = .66); short stories: (*ES* = 66) and “poetry” (*ES* = .81). Thus, for all three text types, students' text comprehension improved in the experimental condition.

3.2 Engagement of students during the experimental lessons

As an indication of students' effort in the experimental condition, we calculated the frequency distribution of the mean number of assignments made. It turns out that 77.9% of these students made 75% or more of the assignments in their booklets. Thus, students' average effort was high during the experimental lessons. However, for none of the three dependent variables (“expository text”, “short story”, and “poetry”) the relationship between students’ effort during the lessons and performance on the post-measurement was found to be significant. Thus, we cannot show that student effort affects performance. This unexpected lack of an effect is possibly related to the high effort of the students who participated in the experimental lessons.

3.3 Self-assessment of changes in reading skills

Changes in reading comprehension were evaluated with three cloze tests. Moreover, students of both groups were asked to indicate the extent to which their reading had changed after the intervention period. Table 3 summarizes the means and variances of this retrospective self-assessment test for each item. The *t* test is relatively robust to nonnormal distributions (see e.g., Sawilowsky & Blair, 1992). We used Cohen’s d as an indication of the effect sizes. In two classes, teachers did not administer this post-then-pre-test.

The results show that students in the experimental condition (*n* = 222) scored all 20 items on a five-point Likert scale not only more positively on average than students in the control condition (*n* = 79), but also above average.

We discuss the results successively based on the following four dimensions: “Reading process” (Items 1 to 4); “Interconnecting text parts” (Items 5 to 10); “Connecting with prior knowledge” (Items 11 to 16), and “Identifying comprehension problems” (Items 17 to 20). The dimensions “Reading process” and “Identifying comprehension problems” represent the (deep) reading process (Janssen, 2009), while the dimensions “Interconnecting text parts” and “Connecting

with prior knowledge, emotions, experiences" are characteristic aspects of deep reading.

Reading process. There is a significant difference for the reading process between the conditions "reading more carefully" and "reading back into the text". For each item, students in the experimental group scored higher on average for poetry and for prose than students in the control group (t (299) = 7.34, $p \leq .00$).

Interconnecting text parts. The students in the experimental condition scored higher than the students in the control group for each item here too (t (299) = 2.14, $p \leq .033$). All mean difference scores are significant, except for item 7.

Connecting with prior knowledge, emotions, experiences. For both items "I more often connect what is in a prose or poetry text to my prior knowledge", the differences between experimental and control group were significant, but not for the items related to "Connecting to personal experiences and emotions". An explanation could be that for the students these items coincided with "Connecting to prior knowledge" (t (299) = 2.49, $p \leq .013$).

Signaling comprehension problems. For the items (13) and (14), students in the experimental condition scored significantly higher than students in the control group (t (299) = 2.07, $p \leq .040$). For the items (17) and (18), the students in the experimental condition did score higher than the students in the control condition, but these differences are not significant.

Conclusion: students in the experimental condition scored all items more positively than students in the control condition, of which 65% of the difference scores were significant. Moreover, these self-reported differences are in line with the improvements in student performance for text comprehension for poetry and prose on the pre- and posttests.

3.4 Characteristics of reading processes

Learning reports. Since the lessons show a constructive level of difficulty and vary in terms of focus, text type, and learning activities, it was expected that over the course of the reading program, students would become increasingly proficient in reflecting on features of reading processes and in formulating their learning outcomes. This proved to be the case: from lesson 1, most students formulated their learning outcomes as features of reading processes generally, "Najim" among them:

> As a result of these learning activities, I now know how to understand a text. The assignment was quite nice, and I was able to keep my attention because there are some videos to accompany it. This kept me thinking. [quote 1, lesson 1]

The learning reports for the observation lessons 4 and 5 are relatively more comprehensive. An important new insight students gained through the observation tasks is that *reading processes differ between readers and texts*. Furthermore, in the learning reports for the EMME lessons, students frequently note that they have learned how to "interconnect between text parts" by "rereading" and "connecting

the text to one's own experiences" can support a deeper understanding of a complex text. Students also frequently mention that they have learned that asking questions about a text can help for deep reading and text comprehension. "Reading back" and "making connections between text parts" and "relating to personal experiences" (associating) are three important aspects of deep reading. What is striking is that students reportedly seldom learned through a lesson that "relating to personal emotions" can help with deep text comprehension.

Student interviews. Through the EMME in the intervention, many students were introduced to the concept of "reading process" for the first time. Some students had to get used to observing via short, information-dense EMME, but they appreciated this new learning activity as interesting and instructive. They also reported that observational learning through EMME had had a positive effect on their engagement in the lesson, as they identified with the peer models. These experiences are congruent with observations by the teachers and the first author that students were full of attention when the teacher presented an EMME. "Klara", for example, said that in lesson 5 she learned through an observation task to improve your understanding of a poem or text in general if you connect personal experiences to the text base while reading, i.e., construct a situation model:

> With this video, I did like that Helen started linking everything to personal experiences, because actually ... I hadn't thought about it, well, that you can just link your own personal experiences to texts or poems. And yes, I also liked to hear what exactly Helen then thought of that whole text ["My grandfather could ..."], because it was also explained very well that way. [quote 2, lesson 5]

Table 3. Mean (M) and Standard Deviation (SD) per Item in the Post-then-pre-test, for the Experimental Group (n = 224) and the Control Group (n = 79) on a five-point Likert Scale (1: totally disagree; 5: totally agree; ES: Effect Size, Cohen's d)

Dimensions and statements		Experimental group		Control group		
		M	*SD*	*M*	*SD*	*ES*
A.	*Reading process*					
1.	I reread prose texts more often	3.68*	1.07	3.00	1.34	0.60
2.	I reread poetry texts more often	3.63*	1.11	2.97	1.39	0.56
3.	I read prose texts more carefully	3.43*	1.08	2.89	1.19	0.49
4.	I read poetry texts more carefully	3.55*	1.12	2.80	1.30	0.64
B.	*Interconnecting text parts*					
5.	When I read prose texts, I interconnect sentences more often	3.40*	1.14	3.06	1.30	0.29
6.	When I read poetry texts, I interconnect lines more often	3.41*	1.17	3.05	1.29	0.30
7.	I interconnect paragraphs more often when reading prose texts	3.32	1.18	3.05	1.30	0.22
8.	I interconnect stanza's more often when reading poetry texts	3.35*	1.16	2.63	1.33	0.60
9.	I interconnect words and phrases more often when I read prose texts	3.32*	1.18	2.89	1.33	0.35
10.	I interconnect words and phrases more often when I read poetry texts	3.35*	1.17	2.90	1.37	0.37
C.	*Connecting textbase to prior knowledge*					
11.	I more often connect what is in a prose text to my prior knowledge	3.37*	1.34	2.91	1.58	0.33
12.	I more often connect what is in a poetry text to my prior knowledge	3.38*	1.35	2.81	1.54	0.41
13.	I more often connect what is in a prose text to my experiences	2.96	1.42	2.59	1.62	0.25
14.	I more often connect what is in a poetry text to my experiences	2.94	1.43	2.57	1.53	0.25
15.	I more often connect what is in a prose text to my emotions	2.57	1.43	2.44	1.56	0.09
16.	I more often connect what is in a poetry text to my emotions	2.69	1.46	2.41	1.54	0.19
D.	*Signalling comprehension problems*					
17.	While reading a prose text, I more often ask comprehension questions	3.13*	1.24	2.78	1.39	0.27
18.	While reading a poetry text, I more often ask comprehension questions	3.17*	1.21	2.70	1.41	0.37
19.	In prose texts, I more often highlight a difficult part of the text	2.73	1.42	2.62	1.48	0.08
20.	In poetry texts, I more often highlight a difficult part of the text	2.60	1.36	2.47	1.38	0.10

*$p < .05$

3.5 Contrasting reading processes

3.5.1 Contrasts between student models

Learning reports. Learning reports 1, 2, 4, and 5 show that most students can distinguish the reading processes of a strong and a weak reader well. Incidentally, this should also follow from the teaching-learning discussion, which draws conclusions about characteristics of deep and weak reading processes. The students indicate that the weak readers in the EMME: (a) read the text only scanning (skipping words); (b) are only aware of comprehension problems at word level (micro-level); (c) do not understand how a verse or sentence runs syntactically; (d) read back because they no longer understand what they are reading; (e) understand the text only locally or literally. According to the students, strong readers in the EMME, on the other hand, demonstrate characteristics of deep reading and deep text comprehension: first, they read carefully and attentively; second, they are aware of comprehension problems at the paragraph or stanza level and at text level; third, they read backwards in the text to be able to interconnect text parts; fourth, they actively look for text structure at the macro-level of a text; fifth, they relate what they read in the text to their personal experiences or feelings.

Student interviews. Almost all the students interviewed indicated that they were able to observe a clear contrast between the reading processes in the different EMME pairs. This made it easier to conclude during the teaching-learning interview that one of the two learner models was the better reader, evidenced, for example, by a statement from “Nabeel”:

> Yes, because you saw for example, I think Syb was the one who went through it much faster and with Nadine, she went through it much slower. Then you could also see the effect of whether you understood it or not, because they [Syb and Nadine] all gave comments during those videos. We said [during the teaching-learning conversation], "Because she reads quietly and goes through it well, she will understand the text better than Syb who goes through it quickly." Actually, everyone agreed with that. Yes. [quote 3, lesson 2]

However, a few students reported that after a few EMME lessons that they had the impression that the different reading processes only amounted to “careful reading” or “skim reading”.

3.5.2 Contrasts between reading processes of student models and students

Learning reports. The learning reports accompanying the observation lessons show that most students are becoming increasingly aware of their own reading process and how they could change or improve it, by comparing it with the different modeled reading processes. “Giovanni”, for example, recognizes his own reading process in that of a sloppy reader (Bas) and takes an example from a careful reader (Imme):

> Through these reading activities, I now know how others look at a text compared to myself. Others read more quietly and skip fewer words. I also now know that I do need to read more quietly to better understand texts. [quote 4, lesson 1]

The learning question for observation lesson 4 reads, “What would you do differently from now on while reading an ambiguous text?” “Lars”, for example, responded: “This lesson I learned that while reading an ambiguous text [the poem “Window in the sky”] I should look more carefully at the sentence structure." [quote 5, lesson 4]

Student interviews. The student interviews show that these students were also able to identify well with the reading process of either student model. In addition, they were able to articulate well in what ways they had read and understood a text differently from the student models and what they had learned from it. In an EMME in lesson 5, prototypical strong reader Helen articulates the negative feelings and associations she experiences while reading stanza 2 of “My grandfather could ...” and the notion “maggot mouths” in particular. She therefore colors the tone in stanza 2 negatively. “Emilie” said during the interview that this EMME gave her the new insight that associating can also help her understand poetry more deeply:

> And so, I hadn't thought about it that way, so when I was shown that, then I thought ”Ooh yeah, of course that's how you can think about it”, because I just didn't think anything at all with “maggot mouths”. I [first] just thought of how weird that that's in there. But that is indeed also something negative when you have maggots. So yes, I found that very interesting at the time. I learned a little bit again how to go deeper into poetry, I thought. [quote 6, lesson 5]

3.6 Developing one’s own reading process

Learning reports. Awareness of the concept of “reading process” and of one’s own reading process obviously precedes the development of that reading process. Especially the learning activities “observing”, “comparing”, and “evaluating” other students’ reading processes appear to have made students aware of the complexity of texts, but also of text structure (e.g., contrasts, repetitions); of their own reading process and of different ways in which they can change or deepen it . Students regularly report that the EMME help them to better read and understand texts, and poetry in particular. The learning reports also show that carrying out the active reading tasks that recur in each lesson, combined with classroom interaction on student findings, have certainly also contributed to raising awareness of comprehension problems in ambiguous and incoherent texts and of how to read and understand them deeply. Students successively mentioned “asking questions” (>156 comments); “reading back into the text” (119 comments); “thinking about ambiguity” (79 comments); “becoming aware of one's own reading process” (69 comments), and “reading attentively” (59 comments) as most helpful in being able to understand a complex text more deeply and as the main learning outcome of the lesson series.

Students’ estimates of self-regulation regarding their reading process. On being competent in *analysing* their own reading process, 73.3% of the students reported

they feel more competent to read a text deeply. Moreover, 11.7% reported that they also know which aspects of deep reading they apply or do not apply themselves. Furthermore, with respect to *motivating*, we notice that 59.0% of students perceive that they can clearly explain to what extent their own reading process is effective and how they could possibly improve or change their reading process. In contrast, only 1.3% of students believe they can mostly not explain the extent to which their own reading process exhibits characteristics of deep reading. Moreover, 90.5% of students perceive themselves as competent after the lessons to *evaluate* the quality of their own reading (from moderate to good competence), while only 9.5% of students think they cannot. Finally, 41.6% of students indicated that they *reflect* regularly or systematically on the quality of their own reading process after the lesson series, while 23.8% of students indicated that they rarely reflect on the quality of their own reading process. All in all, the gain is that students reflect on the quality of their own reading process after six lessons. A further 34.6% doubt the quality of their own reading process: learners indicate that they need examples of reading processes in order to improve their own reading process.

Experienced development of own reading processes. Most students write that they read differently after participating in the reading program. Many students also explicitly state that they are more aware of their own reading process. This dual response implies that most students reflect on the development of their reading process after the lesson series. In students' formulations, the distinction between what they have learnt or will do differently is not always clear. Students mainly articulate their changed reading process as reading texts more attentively, more deeply, and more carefully. In addition, many students say that the series of lessons made them aware of their own reading process and its quality. Furthermore, students frequently mention that they apply certain aspects of a deep reading process more often, namely ”asking questions” and “interconnecting text parts”. “Inaya”, for instance, reflects on the positive effects of the lesson series on her reading process and acknowledges how differently she reads now and what this means for deep text comprehension:

> Before, I always read the text only looking for the answer and not to understand the whole text. Now that I have had these lessons, I have learned that when you read a text it is also important to empathize with it, because that way you understand it better. [quote 7]

Students also mention, but less frequently, that they more often deploy characteristic aspects of a deep reading process, such as ”recognizing text structure”'; “thinking about ambiguity”, and “relating to personal experiences and emotions”. However, a small number of students (*n* = 27) indicated that they do not henceforth read texts differently, such as “Jelle”: “[I'm not changing] anything, on the [school reading] test we are not going to read poetic texts”. [quote 8]

Experienced competence in relation to their own reading process. When asked what they can do better in relation to their reading process after the lesson series, many students (*n* = 55) answer very convincingly: above all, they perceive that they

can read and understand texts *in different ways more deeply*, such as “Linde”: ”I can understand deeper meanings in a text and talk about them”. [quote 9] In addition, some students mention that they feel they can better understand, evaluate, and reflect on their own reading process, such as “Björk”:

> I am better at preparing essential questions, because before the lesson series, I didn't even know what an essential question was. I can [also] reflect better on myself: I know better what I am doing right and wrong. [quote 10]

Moreover, six students explicitly state that they can read and understand “poetry” or “poems” better than before the experimental lesson series. In conclusion, by far most students say they experienced benefits from participating in the experimental lesson series. The extent to which students estimate that they have mastered aspects of a deep reading process (Table 3) corresponds to a large extent to their own formulations of the development of their reading process.

Stimulated recall interviews. The retrospective stimulated recall interviews also revealed how learners differed in the extent to which they reflected on the perceived learning outcomes of the lesson series and individual lessons and the way in which they formulated them. These differences also seem to be related to learner conceptions of what ”learning” is, despite the sentence stems offered in the learning reports. For the individual learning activities and types of didactics, most students interviewed were able to explain well how these were helpful for deeper reading and comprehension of texts.

Transfer. Some students noted that they apply what they have learned from the EMME lessons in subsequent lessons and while reading new texts. “Klara” for instance was able to identify with the strong reader and indicated that this helped her to think back to the observed reading process of the strong reader while reading a new difficult text.

> Yes, well, I could kind of empathize with the person who read the text best. I could apply this after the lesson. If I didn't understand a text or a poem, I could think back to the previous lessons of "But how can I understand this text better?" and “What did that girl or what did that boy do again to understand the text better?". [quote 11]

In Lesson 6, students also appear to have applied aspects of deep reading, such as 'asking yourself questions while reading' on the short story “Pleasure”, among them “Alice”:

> Yes, I started thinking about this too, and then I thought, “yes, that man, he can have it carefree or the hare, he's just carefree.” And yes, then you start thinking about that further, like: “so why would the hare suddenly start running, say, out of nowhere?” “Is there something somewhere, then, that he senses?” “Or maybe the man is a hunter, that he is standing there like that in the kitchen?” Then you start thinking about that again. But I didn't write all that much down [in the assignment booklet] but I think I did write down a bit. [quote 12, assignment booklet lesson 6]

Also, in the assignment booklets in which they worked, the majority of the students did a fine job of reflecting on how they read differently compared to the peer models or how they would do so in the future.

4. CONCLUSION AND DISCUSSION

4.1 Conclusion

The convincingly positive learning effects of the reading intervention are twofold. First, using a pretest-posttest design, we showed that the reading comprehension of the 10th grade students who participated in the experimental lesson series improved, also compared to the students who received regular Dutch reading lessons during the intervention period (RQ1). This is true for the text types “expository texts” (*ES* = .66), “short stories” (*ES* = .66) and “poetry” (*ES* = .81). These results are congruent with students' high commitment as revealed by the assignment booklets. Second, we explored via learning reports, evaluations tasks, a post-then-pre-test, and stimulated recall interviews the extent to which students in the experimental condition experienced changes in their reading process due to the experimental reading program (RQ2). Students in the control condition were also surveyed about changes in their reading process through a post-then-pre-test. Students in the experimental condition rated changes in their reading process on the four dimensions “reading process”, “identifying comprehension problems”, “interconnecting texts parts”, “connecting with prior knowledge” above average and higher than students in the control condition.

In short, the quantitative and qualitative student data from this study indicate that a reading program for poetry and prose that focuses on observing and discussing students' reading processes and uses contrasting EMME is an effective way to deepen students' reading skills and text comprehension in upper secondary higher general and pre-university for both expository and non-expository texts. Our findings therefore not only validate previous positive effects of observational learning to read for expository texts (Couzijn, 1995, 1999; Rooijackers, 2023), but also stress that this didactic technique enhances reading skills and text comprehension of short stories and especially poetry.

For the vast majority, if not all, students, participation in the experimental intervention was a first introduction to reading poetry. It was clear from the student interviews that reading and discussing poetry in class were new learning activities in the classes studied, as opposed to reading expository texts. Against this context, it is understandable that students' average rating for poetry in the experimental condition was relatively low.

We investigated how to improve students' reading skills and text comprehension in 10th grade students for poetry and prose. A potential question is whether this observational-learning-reading approach with EMME could also be feasible for other school types, such as vocational education and primary education, and other grade levels, e.g., higher general and pre-university lower secondary education. Given the positive effects of this experimental intervention on reading skills and text comprehension, this is certainly worth investigating.

4.2 Limitations

Of course, the encouraging results also should be screened for limitations of the study. First, even though the sample size has ample power due to the participation of 15 classes in the intervention study, the sample still is relatively small, and classes have been randomly assigned to conditions, not students. Second, ”reading comprehension” is a highly complex construct, especially if we extend it to the reading and comprehension of short stories and poetry at situation model level and beyond. Although we examined reading comprehension for three text types, we note that the focus was on poetry texts: in the lesson series, students read an expository text, two short stories and five poems. Third, we measured text comprehension only with cloze tasks. Although cloze tests seem to be adequate measures of reading comprehension (e.g., Kleijn, 2018), we do not know whether the effects can be generalized to other measures of reading comprehension. Fourth, although learning-by-observation of peer models is the most essential and innovative component in the lesson series, it is plausible that the combination with other components contributed to the extent to which the reading intervention could be effective. Observational learning is a type of instruction that obviously enables and stimulates (self-) reflection and (self-)evaluation of reading processes or students’ metacognitive and self-regulation skills (see also Couzijn, 1995, 1999). Using collaborative learning and the teaching-learning conversation, the reading intervention meets forms of interaction that have proven to strengthen the learning process (Applebee et al., 2003; Van Leeuwen & Janssen, 2019). However, these types of instruction serve the main topic of the classroom discussions: reading processes of students. In contrast, the classroom discussions in regular Dutch reading lessons, is about the correct answer to text-dependent questions, as formulated in the teaching methods or test assignments (e.g., Rooijackers, 2023).

Finally, we note that self-reporting by students in learner reports and interviews may have some degree of social desirability in them. This effectiveness study is part

of an intervention study. Against this context, we sought to compensate as best we could for possible shortcomings in self-reporting through triangulation of method and perspective: the results on the posttests, the teacher observations, and the classroom observations by the first author support the students' reporting.

4.3 Implications for education

The findings from this intervention study have implications for the school subject Dutch and as such, for all other subjects in which reading is at the core. First of all, a focus on observing, analyzing, comparing, evaluating and discussing characteristics of students' reading processes turns out to be effective for learning to read and understand texts deeply, that is, at situation model level. This applies not only to expository texts, but also to short stories and poetry. The result that students experience and make progress after a series of only six lessons, four of which are EMME lessons, is very encouraging. The EMME play a crucial role in this: for the EMME, most students were full of attention, which contrasts with the lack of student motivation for regular reading and literature lessons. The assignment booklets and learning reports also tell that students could usually identify well with the reading process of either student model (cf. Braaksma et al., 2002).

Variety of text types. For learning to read deeply, a variety of text types proved effective, although there was a strong emphasis on reading poetry in this lesson series. Thus, reading poetry and observing student models' poetry reading processes have positive effects on reading comprehension and text comprehension for "expository texts" and "short stories". To encourage the engagement of 10th grade students in reading instruction, a variety of assignments and text types proved effective. The students interviewed indicated that the lesson series was quite different from how they usually worked on reading skills in a reading lesson Dutch and in language classes in general: practicing and testing with expository texts with questions. For these effects alone, poetry may be given a permanent place in the secondary education curriculum.

Interaction. Students, to achieve learning and deep understanding of complex reading texts, also need the classroom exchange about their observations, reading experiences, and interpretations, so that they take away other perspectives on their own findings and a deeper understanding of a read text. This proved especially necessary for the short stories and poetry. From the student's point of view, it is apparent how important the teacher's role is as a facilitator of the learning process.

It seems highly recommendable to have students read texts from paper as well and do assignments on paper, next to reading from electronic devices (Honma et al., 2022; Mol & Bus, 2011; Wolf, 2018). *Reading from paper is needed to propel deep reading*. Students reported that they liked this in alternation with laptop work in class and that they took more from texts they read from paper.

ACKNOWLEDGEMENTS

We sincerely thank the students and teachers who participated in this study for their time and efforts.

DECLARATION OF CONFLICTING INTERESTS

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

This research was funded by the Netherlands Organization for Scientific Research (NWO) under Grant 023-011-077.

REFERENCES

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal*, *40*(3), 685–730. https://doi.org/10.3102/00028312040003685

Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology, 4*(3), 359–373. https://doi.org/10.1521/jscp.1986.4.3.359

Baron, N. S. (2015). *Words onscreen: The fate of reading in a digital world*. Oxford University Press.

Baron, N. S. (2021). *How we read now: Strategic choices for print, screen, and audio*. Oxford University Press.

Bortolussi, M., & Dixon, P. (2003). *Psychonarratology: Foundations for the empirical study of literary response*. Cambridge University Press.

Bouwer, R., & Koster, M. (2016). *Bringing writing research into the classroom: the effectiveness of Tekster, a newly developed writing program for elementary students*. Thesis Utrecht University.

Braaksma, M., Rijlaarsdam, G., & Van den Bergh, H. (2002). Observational learning and the Effects of model-observer similarity. *Journal of Educational Psychology*, *94*(2), 405–415. https://doi.org/10.1037/0022-0663.94.2.405

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77−101. https://10.1191/1478088706qp063oa

Couzijn, M. J. (1995). *Observation of writing and reading activities: Effects on learning and transfer.* Thesis University of Amsterdam.

Couzijn, M. (1999). Learning to write by observation of writing and reading processes: Effects on learning and transfer. *Learning and Instruction, 9*(2), 109−142. https://10.1016/S0959-4752(98)00040-1

Dixon, P., & Bortolussi, M. (2015). Fluctuations in literary reading: The neglected dimension of time. In L. Zunshine (Red.), *The Oxford handbook of cognitive literary studies* (pp. 541–566). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199978069.013.0027

Education Council & Council for Culture. (2019). *Lees! Een oproep tot een leesoffensief* [Read! A call for a reading offensive]. Raad voor Cultuur/Onderwijsraad [*Education Council/Council for Culture*].

Eva-Wood, A. L. (2004). How think-and-feel-aloud instruction influences poetry readers. *Discourse Processes*, *38*(2), 173–192. https://doi.org/10.1207/s15326950dp3802_2

Expert Group Learning Trajectories. (2009). Referentiekader taal en rekenen: De referentieniveaus [Reference framework language and arithmetic: The referential levels]. Retrieved from the Ministry of Education, Culture and Science: http://www.taalenrekenen.nl/downloads/referentiekader-taal-en-rekenen-referentieniveaus.pdf

Friese, S. (2019). *Qualitative data analysis with ATLAS.TI*. Third edition. Sage.

Goldstein, H. (1987). *Multilevel models in educational and social research*. Charles Griffin.

Hill, L. G., & Betz, D. L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation*, *26*(4), 501–517. https://doi.org/10.1177/1098214005281356

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van den Weijer, J. (2015). *Eye-tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Honma, M., Masaoka, Y., Iizuka, N. et al. (2022). Reading on a smartphone affects sigh generation, brain activity, and comprehension. *Scientific Reports – Nature*, *12*, 1589. https://doi-org/10.1038/s41598-022-05605-0

Insingel, M. (2015). Lof der vergissing. [Praise of error.] In L. Bernaerts, In lussen gevat: het oeuvre van Mark Insingel, 55. *Ons erfdeel*, *58(*2), 46−57. Stichting Ons Erfdeel.

Jansma, E. (2000). Raam in de lucht [Window in the sky]. In E. Jansma, *Dakruiters: gedichten* [Roof-turrets: Poems], (p. 25). De Arbeiderspers.

Janssen, T. (2009). Learning to read literature in dialogue: Reader questions as a tool for learning to interpret short stories [Literatuur leren lezen in dialoog: lezersvragen als hulpmiddel bij het leren interpreteren van korte verhalen]. Amsterdam University Press.

Janssen, T. (2012). Flexibility in reading literature: Differences between good and poor adolescent readers. *Scientific Study of Literature, 2*(1), 83−107. https://doi.org/10.1075/ssol.2.1.05jan

Janssen, T., Braaksma, M., & Couzijn, M. (2009). Self-questioning in the literature classroom: Effects on students' interpretation and appreciation of short stories. *L1- Educational Studies in Language and Literature*, *9*(1), 91–116. https://doi.org/10.17239/L1ESLL-2009.09.01.05

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kintsch, W., & Rawson, K.A. (2005). Comprehension. In Snowling, M.J., & Hulme, C. (Eds.), *The science of reading: A handbook* (pp. 209−226). Blackwell Publishing Ltd.

Kleber, C. (2021). *Inferenzbildung bei literarischem Textverstehen: Theoretische Grundlagen und empirische Erforschung*. Cuvillier Verlag. Thesis Pädagogische Hochschule Heidelberg.

Kleijn, S. (2018). *Clozing in on readability: How linguistic features affect and predict text comprehension and on-line processing*. Thesis Utrecht University.

Kok, E., Hormann, O., Rou, J., Van Saase, E., Van der Schaaf, M., Kester, L., & Van Gog, T. (2022). Reviewing performance: Showing eye-tracking data as feedback to improve performance monitoring in a complex visual task. *Journal of Computer Assisted Learning*, 38(4), 1087–1101. https://doi-org/10.1111/jcal.12666

Kuipers, F. (2014). Mijn grootvader kon … [My grandfather could …]. In F. Kuipers, *Molwerk: gedichten* [Molework: Poems], (p. 34). Atlas Contact.

Levine, S. (2019). Using everyday language to support students in constructing thematic interpretations. *The Journal of the Learning Sciences*, *28*(1), 1–31. https://doi.org/10.1080/10508406.2018.1485023

McCarthy, K. L., & Goldman, S. R. (2017). Constructing interpretive inferences about literary text: The role of domain-specific knowledge. *Learning and Instruction*, *60*, 245–251. https://doi.org/10.1016/j.learninstruc.2017.12.004

Miall, D. S. (2006). *Literary reading: Empirical & theoretical studies*. Peter Lang.

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological bulletin*, *137*(2), 267–296. https://doi.org/10.1037/a0021890

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364−386. https://doi.org/10.1177/1094428106291059

NRC Next (2016, 12 January). Voornemens: vaag uitstel beteugelt de snacktrek [Intentions: Vague postponement restrains craving for snacks]. Retrieved from https://advance- lexis-com.proxy.library.uu.nl/api/document?collection=news &id=urn: contentItem:5HV1-JKJ1-DYRY-N4WM-00000-00&context=1516831.

Oberon (2016). *Lees- en literatuuronderwijs in havo/vwo* [Reading and literature teaching in higher general and pre-university secondary education]. Stichting Lezen [Dutch Reading Foundation].

OECD. (2019). *PISA 2018: Assessment and Analytical Framework*. PISA, OECD. https://doi-org/10.1787/b25efab8-en

Okkinga, M., Van Steensel, R., Van Gelderen, A., Van Schooten, E., Sleegers, P., & Arends, L. R. (2018). Effectiveness of reading-strategy interventions in whole classrooms: A meta-analysis. *Educational Psychology Review*, *30*(4), 1215–239. https://doi.org/10.1007/s10648-018-9445-7

Perquin, E.N. (2017). Guard duty. In E.N. Perquin, *The hunger in plain view: Selected poems* (p. 32). Translated by D. Colmer. White Pine Press.

Perquin, E.N. (2008). Op wacht [Guard duty]. *Het Liegend Konijn*, *6*(1), 138. Retrieved from https://www.dbnl.org/tekst/_lie005200801_01/_lie005200801_01_0089.php

Peskin, J. (1998). Constructing meaning when reading poetry: An expert-novice study. *Cognition & Instruction, 16*(3), 235–263. https://doi.org/10.1207/s1532690xci1603_1

Peskin, J. (2007). The genre of poetry: Secondary school students' conventional expectations and interpretive operations. *English in Education*, *41*(3), 20–36. https://doi.org/10.1111/j.1754-8845.2007.tb01162.x

Peskin, J., & Hanauer, D. I. (2023). *A life with poetry: The development of poetic literacy.* John Benjamins.

Rooijackers, P. (2023). *Oog voor diep begrip: Onderzoek naar het tekstbegrip van vwo-leerlingen* [Focusing on deep reading: Research on reading comprehension of Dutch pre-university students]. Cito. Thesis Utrecht University. https://dspace.library.uu.nl/handle/1874/425157

Sawilowsky, S. S., & Blair, R. C. (1992). A more realistic look at the robustness and type II Error properties of the t test to departures from population normality. *Psychological Bulletin*, *111*(2), 352–360. https://doi.org/10.1037/0033-2909.111.2.352

Schrijvers, M. S. T. (2019). *The story, the self, the other: Developing insight into human nature in the literature classroom*. Thesis University of Amsterdam.

Schrijvers, M., Janssen, T., Fialho, O., & Rijlaarsdam, G. (2019). Gaining insight into human nature: A review of literature classroom intervention studies. *Review of Educational Research*, *89*(1), 3–45. https://doi.org/10.3102/0034654318812914

Sigvardsson, A. (2017). Teaching poetry reading in secondary education: Findings from a systematic literature review. *Scandinavian Journal of Educational Research*, *61*(5), 584–599. https://doi.org/10.1080/00313831.2016.1172503

Van den Broek, P., & Kendeou, P. (2022). Reading comprehension I: Discourse. In Snowling, M. J., Hulme, C., & Nation, K. (Eds.), *The science of reading: Handbook* (pp. 239−260) (Second, Ser. Wiley Blackwell handbooks of developmental psychology). Wiley-Blackwell. https://doi-org/10.1002/9781119705116.ch11

Van Dijk, T. A. (1979). Cognitive processing of literary discourse. *Poetics Today, 1*(1-2), 143−159. https://doi.org/10.2307/1772045

Van Essen, R. (2015). Het meisje van de garderobe. [The cloakroom girl.] In *Nederland leest: de mooiste korte verhalen voorafgegaan door Utrecht leest* [The Netherlands reads: The most beautiful short stories preceded by Utrecht reads], (pp. 216–217). Selected by A. L. Snijders. CPNB.

Witte, T. (2018). De kunst van het onderwijzen [The art of teaching]. *Nederlandse letterkunde*, *23*(3), 359–383. https://doi.org/10.5117/nedlet2018.3.007.witt

Wolf, M. (2018). *Reader, come home: The reading brain in a digital world*. HarperCollins.

Wolf, M. (2023, 21 February). "Success in Circuit lies": How do we cultivate deep reading processes in a digital age? *ABC Religion & Ethics*. Retrieved from https://www.abc.net.au/religion/maryanne-wolf-cultivating-deep-reading-in-a-digital-age/102001224

Zeitz, C. M. (1994). Expert-novice differences in memory, abstraction, and reasoning in the domain of literature. *Cognition and Instruction 12*(4), 277–312. https://doi.org/10.1207/s1532690xci1204_1

Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review*, *23*(4), 1028–1034. https://doi.org/10.3758/s13423-015-0864-x

APPENDIX A

INTERVIEW GUIDELINE STUDENTS

Student experiences with the "Deep Learning to Read Prose and Poetry" lesson series, by lesson, by assignment, by working form based on their own assignment booklets.

A. Questions about each lesson, related to essential didactic components:
   1. How did you experience this lesson?
   2. How would you evaluate this lesson?
   3. How did you evaluate the text(s)?
   4. How did you experience the eye movement videos and the peer models?
   5. To what extent were the assignments feasible and practicable?
   6. To what extent did you appreciate the collaborative learning tasks?
   7. To what extent did you appreciate the teaching-learning conversations?
   8. To what extent did you contribute to the teaching-learning conversations?

B. Questions about the entire lesson series, related to acceptability and experienced learning effects:
   9. To what extent is this lesson series similar to regular Dutch reading lessons? What is different?
   10. How did you feel about having to write down answers in an assignment book?
   11. To what extent did you contribute to the teaching-learning conversations?
   12. How did you feel about writing learning reports at the end of each lesson?
   13. To what extent did the reading program help deepen your reading process, do you think?
   14. What would you have liked to see different?
   15. Do you have any other comments?