Effects of intervention on self-efficacy and text quality in elementary school students’ narrative writing

Emily Grennera, Victoria Johanssonb, Joost van de Weijerb, and Birgitta Sahlénac, d

aDepartment of Clinical Sciences, Logopedics, Phoniatrics and Audiology, Lund University, Lund, Sweden; bCentre for Languages and Literature, Lund University, Lund, Sweden

ABSTRACT

Aim: Self-efficacy for writing is an important motivational factor and considered to predict writing performance. Self-efficacy for narrative writing has been sparsely studied, and few studies focus on the effects of writing intervention on self-efficacy. Additionally, there is a lack of validated measures of self-efficacy for elementary school students. In a previous study, we found that a trained panel rated personal narrative text quality higher for girls than for boys, which led to our aim: to investigate boys’ and girls’ self-efficacy for narrative writing before and after an intervention, and to explore associations between self-efficacy and text quality.

Methods: An 18-item self-efficacy scale was developed. Fifty-five fifth-grade students (M 11.2 years, SD 3.7 months) filled out the scale before and after a five-lesson observational learning intervention. Self-efficacy was then related to writing performance as measured by holistic text quality ratings.

Results: The students demonstrated strong self-efficacy, which increased significantly post-intervention. Girls and boys demonstrated similar self-efficacy, despite girls’ higher text quality. There were moderate correlations between self-efficacy and writing performance pre- and post-intervention.

Conclusions: The results support previous findings of strong self-efficacy at this age. The interaction between writing self-efficacy and performance is complex. Young students may not be able to differentiate between self-efficacy, general writing skills, task performance, and self-regulation. Self-efficacy scales should thus be carefully constructed with respect to age, genre, instruction, and to students’ general educational context.

Introduction

Self-efficacy is the belief in one’s own capability to perform successfully within a field. According to social cognitive theory [1], self-efficacy varies widely between domains, in different activities within a domain, and under different task demands. Bandura states that stronger self-efficacy leads to a better performance regardless of skill: a person with stronger beliefs in her or his capabilities, i.e. self-efficacy, will approach a difficult task as an achievable challenge, while a person with weaker self-efficacy will view it as a threat [2,3].

To see an achievable challenge will in turn motivate more effort, while a threatening task leads to less effort and a higher risk of giving up [3]. Self-efficacy has been shown to predict outcome in diverse domains, including sales performance, health, and academic performance [4]. For instance, students with strong self-efficacy participating in math groups of different levels had better performance than group-members with weak self-efficacy [5]. Bandura further states four sources of self-efficacy: (1) enactive experience, i.e. an earlier successful performance within the field or a similar field; (2) vicarious experience, i.e. watching “similar others” perform a comparable task; (3) social persuasion, i.e. others expressing beliefs in or providing adequate feedback on one’s performance; (4) the emotional and physiological state of the person performing the task [3,6].

Research on self-efficacy for writing has mainly followed one of two paths [7]. One path is represented by correlational studies, in which factors that potentially influence self-efficacy have been explored. The other path, which is taken here, includes experimental studies in which the effects of writing intervention on self-efficacy are tested. These studies have explored the influence of various instructional approaches on self-efficacy outcomes. One such approach is the observational learning paradigm [8,9]. The writing instruction in the present study is based on this paradigm, with a design in which students observed peer models working with texts, and thereafter discussed the models’ behavior with their classmates. This method is considered to increase student motivation [10], since it offers students structured opportunities to discuss the observations in groups which induces a comparison of their own writing performance to that of the peer models (i.e. the above-mentioned “vicarious experience”). There is some evidence of
effects on writing performance. For instance, in one study on academic writing, an observational learning intervention increased task knowledge in university students who observed films of dyads of peers writing a literature review. Controls, who did the same exercise with traditional writing instruction did not show a similar positive effect. In parallel, the increased task knowledge led to an increase in self-efficacy in the experimental group, but not in the control group [11].

Self-efficacy development

Young children’s self-efficacy is often strong, in that they rate their capability as high in relation to their performance, and holistic, meaning that they do not analyze what they think they can do in different aspects or components [12]. From the age of 13 or even earlier, students’ beliefs about their performance weaken [12,13], and simultaneously become more diverse across domains. The earlier, more holistic self-efficacy may include a combination of appraisal of their general skills, their performance at a specific task and their perceived effort, but the links between these factors may not be clear to them [14]. Self-efficacy, from the social cognitive point of view, depends on an interaction of cognitive, physical, and social development [2–4,15]. During adolescence, students will begin to experience others’ evaluations of their ability through different forms of feedback, to a greater extent than in younger ages. They will therefore become more analytic and begin to differentiate between different aspects of a certain skill and result in a better correspondence between self-efficacy and actual performance [14]. These associations, sometimes referred to as calibration of self-efficacy or self-efficacy accuracy, are complex. When self-efficacy is much too strong or if it is too weak, learning is impeded. In the first case, it may lead to the student believing she or he masters the task and thereby does not fully engage in it, and in the second case, that she or he focuses too much on basic concepts to be able to move on to complex matters [12,16,17]. For example in writing, a student may focus too much on the spelling of single words, instead of engaging in writing a text with cohesion. In addition, a perfect correspondence between self-efficacy and performance may also be detrimental to development. Judging one’s capability too accurately may block creativity and, as a consequence, limit learning by “trial and error”. When self-efficacy is slightly lower than performance, the performance can still be adequate. However, it may lead to an anxious and perfection-seeking student, worrying that she or he is not sufficiently capable [3,12,18]. The most productive self-efficacy is found in individuals with a slightly stronger belief in their capabilities than their actual performance [3,6].

Measuring self-efficacy

Self-efficacy varies between domains, and must thus be measured in a task-specific way. In his recommendations for construction of self-efficacy scales, Bandura [18] describes how the statements in a self-efficacy scale should tap into the many different skills which are important for the domain of functioning, but not into other skills or general abilities. They should concern self-perceived beliefs in one’s own capability to perform a task, as opposed to constructions targeting intent or comparison to others’ performance. Further, the scale should reflect the genre, but not a specific topic or subgenre, and performance assessment (e.g. text quality) should correspond to the content of the scale. It should not have too few intermediate steps [18]. When the present study was initiated in 2013, validated self-efficacy measures were lacking for the chosen age group, for the narrative genre, and for students in a Swedish school context (for more recent scales measuring self-efficacy in younger students, see Bruning et al. [19,20]). A scale for students in a Dutch context developed by Braaksma et al. [9] was used as a point of departure, since the writing intervention in Braaksma’s study was based on an observational learning paradigm as was the present study. The scale was designed for teenagers writing argumentative texts and we adapted it to fit 10–12-year-olds writing narrative texts within the Swedish curriculum.

Gender differences in writing performance are often reported, e.g. in the OECD assessment PISA (The Program for International Student Assessment) which shows higher performance in girls [21]. Research on gender differences in self-efficacy for writing has shown contradictory results. In two studies by Pajares et al., similar self-efficacy for writing was found between girls and boys, while girls’ performance was assessed as better [22,23]. The students were also asked to compare their own writing ability to the writing ability of other girls and boys. In the first study, on 8–11-year-old students (grades 3–5), both girls and boys on average considered themselves as better writers than classmates of the other gender, but girls did so to a higher degree [22]. In the second study on students aged 11–14 (grades 6–8), girls again considered themselves better writers than boys, while the boys considered themselves poorer writers than girls [23]. Pajares and Valiante concluded that even if self-efficacy is similar for girls and boys, statements in a scale may be judged differently between genders, girls answering more cautiously [23]. Pajares and Valiante further found that gender differences were non-significant when controlling for what they called “gender-stereotypic beliefs” [24]. In a review however, Pajares found that several studies show that girls report stronger self-efficacy than boys during earlier school years, but that the differences even out or reverse later on [25].

Text quality

The Swedish curriculum does not provide set criteria for assessment for the age group 10–12 years old. Instead, a method of holistic text quality ratings based on benchmark texts was chosen for the present study. This method has been used in previous observational learning studies [11] as well as other writing intervention studies, and was tested and validated by Tillema et al. [26]. A previous study has
explored writing performance measured by text quality for the same students as in the present study [27]. This was done through measuring text quality in personal narrative texts on repeated occasions in a waiting control design, before and after the observational learning intervention. The narrative genre was chosen as it is a prerequisite for developing other genres and predicts results in higher education [28]. The students in the age-group of the present study are thus expected to be aware of the structure of a narrative text but still to be developing their abilities for creating such texts [29]. The results showed that the estimated text quality for boys increased from approximately 33 (on a scale from 0 to 100) before the intervention to approximately 40, as a consequence of intervention. For girls, the estimated text quality increased from 45 to 52. Thus, girls had considerably higher text quality than boys, pre- as well as post-intervention, while intervention effects were similar at about seven points on the 0–100 scale.

Cognitive and linguistic abilities are important prerequisites for writing performance, at micro- as well as at macro-level [30]. Measures of working memory capacity, language comprehension, and reading comprehension were collected for our previous study [27], and are presented as demographic data in the present study.

**Aim**

In our previous study, the trained raters found that girls’ texts had higher text quality than boys’ texts [27]. Considering this difference, the aim of the present study is to explore boys’ and girls’ self-efficacy for narrative writing, before and after intervention. Further, considering that intervention effects on text quality were similar for boys and girls, a secondary aim is to explore associations between self-efficacy and text quality.

The study addresses two research questions:

1a. Does self-efficacy for narrative writing change after intervention, and if it does, in what way? 1b. Are there any gender differences?
2. Is self-efficacy related to text quality, and if it is, in what way?

**Methods**

Data regarding the participating students, their results on cognitive and linguistic tasks, the texts and text quality, and intervention were also described in our previous study [27].

**Design and procedure**

Before and after an observational learning intervention, students wrote personal narratives and filled out self-efficacy scales. An overview of the design is found in Figure 1. The students wrote their personal narratives on laptops. A few days or up to a week after the first writing assignment, the students filled out a self-efficacy scale, which is found in Table 1, and were given a working memory test and a language comprehension test. The following week, the intervention (led by researchers) started for one school, replacing the regular Swedish lessons. The other school had regular Swedish lessons with their teachers during this time. The five intervention lessons were given over three weeks. In the second school, intervention started after nine weeks (week 10 in Figure 1). After the intervention period, all students wrote another personal narrative. A few days or up to a week later, they filled out the self-efficacy scale for the

Table 1. Self-efficacy statements.

| Self-efficacy statements                                                                 |
|----------------------------------------------------------------------------------------|
| 1. I can quickly write a text on the computer.                                          |
| 2. I can find all the letters on the keyboard.                                         |
| 3. I can write a text without spelling errors.                                        |
| 4. I can use periods, question marks and exclamation marks in the right places.       |
| 5. I can use commas in the right places.                                                |
| 6. I can partition my text into paragraphs.                                            |
| 7. I can write a heading that fits with the content.                                   |
| 8. I can write a story in such a way that the reader understands where the story took place and who was in it. |
| 9. I can divide my text into beginning, middle, and end.                               |
| 10. I can write a story with an unexpected ending.                                     |
| 11. I can write a story in such a way that the reader understands what happened.       |
| 12. I can begin the story in an exciting way, so that the reader wants to continue reading. |
| 13. I can change one or several words in my text if I am not satisfied with what I had written initially. |
| 14. I can write both long and short sentences.                                         |
| 15. I can decide which parts of my story need to be explained to the reader.           |
| 16. I can read through my text and correct spelling mistakes.                         |
| 17. I can read through my text and make changes to improve it.                        |
| 18. I can write a text which is one page long.                                         |

![Figure 1](Image)  
*Figure 1. Overview of data collection and intervention. Narrative writing assignments pre- and post-intervention (Text), self-efficacy scales pre- and post-intervention (Self-efficacy), tests of working memory and language comprehension pre-intervention (tests), intervention, regular lessons, and test of reading comprehension (test).*
second time and participated in a reading comprehension task. For practical reasons, working memory and language comprehension tasks were administered pre-intervention, and the reading comprehension task post-intervention. The whole data collection, including intervention, took 13 weeks. Teachers were explicitly asked not to work on written narratives with their students during this time.

The full data collection and intervention took place in the classroom. All students who typically attended the classes were included, as this study was meant to test the intervention in a natural school environment. All data collection and intervention was carried out by the first author and a research assistant.

Participants
All students came from two schools in areas of similar socio-economic status. With 35 students in fifth grade in one school and 44 in the other, there was a total of 79 students. Informed consent was collected from 59 students and their parents. Following the advice from the regional ethics board (EPN dnr. 2013/270), no students were a priori excluded from participation in the intervention or data collection. The inclusion criteria were, aside from written consent: taking part in at least four of the five intervention lessons, adequate Swedish listening comprehension and speech production skills, and being regularly present for whole-class activities. Four students, two from each school, did not meet these criteria and were excluded from the dataset. Thus, 55 students (30 girls and 25 boys) remained, 32 from one school and 23 from the other. Their age was 10:9 to 11:9 years (M 11:2 years, SD 3.7 months) and five students had other first languages in addition to Swedish.

One student was not present for writing the pre-intervention personal narrative, two students’ post-intervention narratives were lost due to software issues, and one student was not present for filling out the post-intervention self-efficacy scale. Additionally, two students did not complete the self-efficacy scales. Thus, the total number of students in the analyses varies between 49 and 55.

Cognitive and linguistic tasks
The students performed according to age norms and results were similar across the two schools and between boys and girls on norm-referenced or standardized tests of working memory [31], language comprehension [32], and reading comprehension [33]. As reported in our previous study [27], the students’ mean result was 31.2 (SD 6.5) on the working memory task Lilla Duvan, ranging from 14 to the maximum 36. The norms for the fifth grade are 31.7 (SD 5.8) [31]. The mean was 16.5 (SD 2.3) on the language comprehension task TROG-2, ranging from 5 to 19 (maximum score is 20). Scores 15–19 represent percentiles 23–73 according to age norms [32]. In the reading comprehension task SL40, students read sentences and chose the corresponding pictures. The mean result was 37.5 (SD 4.1), ranging from 15 to the maximum 40. Scores 37–39 equal percentiles 25–75 [33].

Self-efficacy scale
The self-efficacy scale was adapted from the scale by Braaksma et al. [9] to fit the narrative genre as well as the goals for writing stated in the curriculum for the subject of Swedish [34] for the age group. There were 18 statements which are listed in Table 1. Six of them were identical or very similar (statements 1, 3, 6, 9, 12, and 18) to the original scale. The 18 statements reflected aspects of form as well as content, of writing processes as well as the finished text. For example, statement 2, “I can find all the letters on the keyboard” merely concerned low-level writing processes, while statement 9, “I can divide my text into beginning, middle, and end” concerned form (the structure of a narrative) and content (relevant content in the beginning, middle, and end of a story), and to some degree also writing processes (how to go about writing a story with these elements). The students filled out the self-efficacy scale once before and once after the intervention. The instructions were read aloud to the class and written in the booklet containing the self-efficacy scale:

Imagine the following scenario: In school, you get a writing assignment where you are to write a narrative story about something you have experienced. For example, it could be writing a story about the most exciting time you had during the summer vacation. It should be written so that somebody in your class could understand it, and the text should be about one page long. You are not going to write this text yourself, but please answer some questions about what writing such an assignment would be like. Answer each question by marking the horizontal line beneath each question with a vertical line. The further to the right your mark is, the more you agree with the statement.

Beneath each statement was a 100-mm, visual-analogue scale [35]. The VAS scale was marked with the phrases “not at all” and “yes, completely” below the left and right endpoints. One of the researchers explained how students should mark the scales according to their beliefs, with illustrations on the whiteboard. This included demonstrating that the students should put a mark in the middle of the scale if they believed their ability to be average. The marks on the scales were measured resulting in possible values ranging from 0 to 100 for each of the 18 statements.

Personal narrative texts
Personal narratives were written before and after the intervention (Figure 1). The students wrote on laptops with a basic word-processing interface, in the classroom. Students were asked to write a personal narrative about “one time you were saved from a jam you had got into, or when you saved somebody else from a jam” (the first text) and about “one time when you were afraid” (the second text). The topics of the narratives have been used and found suitable for the age group [29,36]. Students were not given feedback after the writing tasks.
**Text quality**

Text quality was assessed by raters who were trained by assessing comparable texts. Four benchmark texts written by age peers were given to the raters, six university students. These texts were given a holistic score, and rated 25, 40, 50, and 95 on a 0–100 scale. The scores of the benchmark texts were motivated by short, written summaries, describing aspects of content, structure, genre, organization, grammar, spelling, punctuation, and text length. Following this, the raters were practicing in the method of holistic scoring through the rating of another six texts in the same way. In the final data set, each text was rated by three or in some cases two raters. Their inter-rater reliability, calculated on the averages of ratings of the whole rated dataset (including about 150 texts written by students who were not included in the present study) was high (Cronbach’s alpha = 0.90). The raters and texts were distributed at random with the restrictions that each text was rated three times, texts written by one student would be rated by different raters, and each rater would rate texts from pre- and post-intervention. Raters were not aware of intervention or that one student had produced more than one text.

**Intervention**

The five-lesson intervention was developed within the observational learning paradigm, a method relying on vicarious experience and structured reflection. The intervention is thoroughly described in Grenner et al. [27]. Each lesson had a different theme, which in design and content was based on the Swedish curriculum [34] and writing development for the age group: (1) the reader’s perception – what does the reader find important in a story? (2) structure – different ways to start a story; in what order should the events unfold? (3) conclusion – how to finish a story; (4) editing someone else’s text; (5) editing during writing – what changes do writers make while they write? The lessons were structured around short video clips of 12 different peers. The students were not given any information about the peers’ (low, average, or advanced) they were watching. Thus, instruction was implicit.

**Analyses**

Pearson’s bivariate correlations were calculated between each statement and the self-efficacy mean pre- and post-intervention. Means and standard deviations for the self-efficacy statements and the self-efficacy mean pre- and post-intervention were calculated. Repeated measures ANOVA were calculated to test intervention effects of each self-efficacy statement, including interactions with gender. Each self-efficacy statement and self-efficacy mean pre-intervention was correlated to text quality pre-intervention by using Pearson’s bivariate correlations. The same was done for these measures post-intervention. The alpha level was set at .05 for all analyses.

**Results**

**Internal consistency within the self-efficacy scale**

To explore the internal consistency of the self-efficacy scale, each statement (Table 1) pre-intervention was correlated to the self-efficacy mean pre-intervention. The same procedure was followed for the statements post-intervention. The results are presented in Table 2 (pre-intervention) and Table 3 (post-intervention). The correlations between the individual statements and the self-efficacy mean varied between $r = .463$ and $r = .858$. All correlations were statistically significant indicating strong internal consistency, and suggesting that the students had consistent beliefs in their ability, across the

**Table 2.** Pearson bivariate correlations between self-efficacy statements before intervention and mean self-efficacy before intervention, text quality before intervention, mean self-efficacy after intervention, and text quality after intervention.

| Statement | Mean SE pre | Text quality pre | Mean SE post | Text quality post |
|-----------|-------------|------------------|--------------|------------------|
| 1         | .624**      | .323*            | .531**       | .085             |
| 2         | .601**      | .353**           | .555**       | .245             |
| 3         | .491**      | .256             | .510**       | .258             |
| 4         | .595**      | .188             | .530**       | .353*            |
| 5         | .614**      | .062             | .490**       | .343*            |
| 6         | .706**      | .222             | .519**       | .320*            |
| 7         | .521**      | .155             | .437**       | .266             |
| 8         | .806**      | .328*            | .653**       | .161             |
| 9         | .791**      | .314*            | .501**       | .252             |
| 10        | .677**      | .251             | .458**       | .108             |
| 11        | .658**      | .379**           | .630**       | .264             |
| 12        | .616**      | .396**           | .525**       | .220             |
| 13        | .727**      | .362**           | .548**       | .237             |
| 14        | .676**      | .192             | .478**       | .109             |
| 15        | .820**      | .386**           | .683**       | .227             |
| 16        | .604**      | .145             | .573**       | .163             |
| 17        | .795**      | .247             | .672**       | .266             |
| 18        | .716**      | .328*            | .589**       | .170             |
| Mean      | .394**      | .816**           | .327*        |                  |

Significant correlations $p < .05$.

**Table 3.** Pearson bivariate correlations between self-efficacy statements after intervention and mean self-efficacy before intervention, text quality before intervention, mean self-efficacy after intervention, and text quality after intervention.

| Statement | SE mean pre | Text quality pre | SE mean post | Text quality post |
|-----------|-------------|------------------|--------------|------------------|
| 1         | .640**      | .319*            | .565**       | .178             |
| 2         | .507**      | .166             | .603**       | .086             |
| 3         | .557**      | .151             | .633**       | .260             |
| 4         | .480**      | .270             | .658**       | .368**           |
| 5         | .682**      | .291*            | .653**       | .251             |
| 6         | .597**      | .337*            | .651**       | .296*            |
| 7         | .490**      | .314*            | .612**       | .345*            |
| 8         | .693**      | .340*            | .790**       | .170             |
| 9         | .569**      | .130             | .592**       | .046             |
| 10        | .387**      | .322*            | .572**       | .130             |
| 11        | .640**      | .256             | .733**       | .098             |
| 12        | .497**      | .313*            | .637**       | .134             |
| 13        | .327**      | .050             | .580**       | .161             |
| 14        | .676**      | .168             | .742**       | .026             |
| 15        | .612**      | .012             | .755**       | .111             |
| 16        | .482**      | .144             | .703**       | .186             |
| 17        | .325**      | .080             | .589**       | .259             |
| 18        | .433**      | .261             | .581**       | .246             |
| Mean      | .816*       | .394**           | .322*        |                  |

*Significant correlations $p < .05$.  
**Significant correlations $p < .01$.  

different statements. The variability between the various self-efficacy statements is illustrated in Figure 2.

**Self-efficacy increased post-intervention**

Mean values for the 18 self-efficacy statements varied between 60.8 and 88.0 pre-intervention and between 63.2 and 92.8 post-intervention on the scale with possible values from 0 to 100. All mean values post-intervention were higher than those pre-intervention. Pre-intervention values were $M = 76.2, SD = 14.5$ for the whole group of students (boys and girls), and post-intervention, the self-efficacy mean had increased to $M = 81.7, SD = 13.5$ for the whole group of students. Means and standard deviations for each statement and for the self-efficacy mean for boys and girls are found in Table 4. Some students had considerably lower self-efficacy mean values than the rest of the group, which is illustrated in Figure 3.

For each self-efficacy statement as well as for self-efficacy mean, a repeated measures ANOVA was run with intervention and gender as the independent variables, and change in self-efficacy (from pre- to post-intervention) as the dependent variable. The results are presented in Table 4. The self-efficacy mean increased significantly from pre- to post-intervention, $F(1, 51) = 22.423, p < .000, \eta^2_p = .305$. There was no significant effect of gender and no significant interaction between gender and intervention. The increase in self-efficacy was significant in 13 of the 18 statements (numbers 1, 2, 6, 9, 10, 12, 13, 14, 15, 16, 17, 18, and 19) but not in the remaining five statements. For the individual statements, there were no main effects of gender, nor interactions with gender. This means that boys’ and girls’ self-efficacy increased similarly.

---

**Table 4.** Means and standard deviations of self-efficacy statements and self-efficacy mean; effects of intervention, gender effects, and interaction effects between intervention and gender.

|                | Pre-int. Boys | Post-int. Boys | Pre-int. Girls | Post-int. Girls | Intervention effect | Gender effect | Intervention × gender |
|----------------|---------------|----------------|----------------|-----------------|---------------------|--------------|-----------------------|
|                | $N$ | $M$ | $SD$ | $M$ | $SD$ | $M$ | $SD$ | $M$ | $SD$ | $F(1, 51)$ | $p$ | $\eta^2_p$ | $F(1, 51)$ | $p$ | $\eta^2_p$ | $F(5, 250)$ | $p$ | $\eta^2_p$ |
| 1              | 53  | 65.0| 23.9| 74.6| 18.1| 69.0| 23.9| 73.8| 24.3| 7.240 | .010 | .124 | .076 | .784 | .369 |
| 2              | 52  | 81.7| 22.0| 90.2| 14.1| 77.6| 25.3| 82.1| 22.3| 7.725 | .008 | .134 | 1.208 | .277 | .397 |
| 3              | 53  | 67.0| 18.9| 70.6| 24.3| 55.6| 27.3| 57.1| 30.9| 0.898 | .348 |        |
| 4              | 53  | 83.5| 19.3| 86.7| 23.8| 81.6| 20.8| 87.3| 15.1| 4.010 | .051 | .073 | 0.017 | .896 | .580 |
| 5              | 53  | 70.2| 23.1| 75.8| 21.0| 68.7| 25.8| 79.7| 25.0| 8.582 | .005 | .144 | 0.040 | .843 | .531 |
| 6              | 53  | 75.4| 24.6| 75.6| 27.3| 75.6| 20.7| 82.1| 19.6| 1.119 | .295 |        |
| 7              | 53  | 84.2| 15.1| 88.6| 16.9| 88.5| 14.3| 87.8| 16.0| 0.872 | .355 |        |
| 8              | 53  | 78.2| 20.0| 82.1| 18.2| 76.9| 18.4| 82.1| 19.1| 7.231 | .010 | .124 | 0.018 | .894 | .689 |
| 9              | 53  | 73.0| 20.5| 85.3| 15.9| 75.7| 22.8| 80.8| 23.9| 6.886 | .011 | .119 | 0.036 | .850 | .285 |
| 10             | 51  | 82.4| 20.9| 89.2| 12.1| 80.9| 20.4| 84.6| 19.6| 3.944 | .053 | .074 | 0.449 | .506 | .551 |
| 11             | 51  | 78.5| 17.7| 85.2| 14.6| 84.4| 17.8| 85.7| 15.6| 4.245 | .045 | .080 | 0.615 | .437 | .200 |
| 12             | 51  | 67.1| 23.5| 81.2| 17.9| 77.9| 24.9| 83.2| 18.7| 11.541 | .001 | .191 | 1.422 | .239 | .134 |
| 13             | 51  | 73.0| 24.0| 91.9| 8.7| 77.2| 22.0| 82.3| 24.1| 9.858 | .003 | .167 | 0.342 | .561 |        |
| 14             | 51  | 89.4| 15.2| 94.6| 7.0| 87.0| 16.7| 91.6| 12.9| 6.405 | .015 | .116 | 0.625 | .433 | .863 |
| 15             | 49  | 79.9| 19.5| 80.2| 16.2| 68.9| 24.8| 72.0| 27.8| 5.652 | .022 | .070 | 0.721 | .400 | .245 |
| 16             | 49  | 74.5| 21.2| 82.0| 17.8| 63.8| 23.3| 69.7| 33.3| 6.076 | .017 | .114 | 0.239 | .131 | .785 |
| 17             | 49  | 77.7| 20.6| 84.3| 16.7| 76.4| 25.0| 83.4| 24.2| 4.760 | .034 | .092 | 0.040 | .842 | .948 |
| 18             | 49  | 86.3| 19.9| 90.8| 14.8| 86.9| 17.0| 92.0| 12.7| 4.650 | .036 | .091 | 0.042 | .838 | .692 |
| Mean           | 53  | 76.0| 13.9| 82.5| 12.5| 76.4| 15.2| 81.0| 14.5| 22.423 | <.000 | .305 | 0.018 | .893 | .416 |

Values that were missing either pre- or post-intervention were excluded from the analysis. Statistically significant effects are indicated by bold text. Numbers 1–18 in the left column indicate the self-efficacy statements, and mean is the self-efficacy mean.
**Self-efficacy and text quality**

Correlations were calculated between the self-efficacy mean pre- and post-intervention and the text quality pre- and post-intervention. Results are found in Table 2 (pre-intervention) and Table 3 (post-intervention). Self-efficacy mean pre-intervention had a strong, significant correlation to self-efficacy mean post-intervention ($r = .816, p = .000$). Self-efficacy pre-intervention and text quality pre-intervention had a significant, moderate correlation ($r = .394, p = .003$), and the same was found for self-efficacy and text quality post-intervention ($r = .322, p = .021$). Thus, self-efficacy and text quality had an association, but not as pronounced as the association of self-efficacy pre- and post-intervention. Text quality as measured by trained raters was considerably lower than students’ perception of their skills as shown by their self-efficacy [2]. This means that the students showed a general “overestimation” of their capability compared to their actual performance as measured by text quality.

Correlations between self-efficacy statements (Table 1) and text quality were calculated. Pre-intervention, nine of the 18 self-efficacy statements were moderately, statistically significantly correlated with text quality (Table 2). These nine statements were all included in the 13 statements in which self-efficacy increased significantly (Table 4). Post-intervention, statements 4 (I can use periods, question marks and exclamation marks in the right places), 6 (I can partition my text into paragraphs), and 7 (I can write a heading that fits with the content) had a statistically significant correlation with text quality (Table 4). None of the three was significantly correlated with text quality pre-intervention. When there is little variation in the values, correlations will not be significant.

**Summary of results**

The students demonstrated strong self-efficacy pre- as well as post-intervention, however with considerable individual variability. The self-efficacy mean increased significantly after intervention, even though the effects of the intervention on performance, i.e. the increased text quality, were mild, as our previous study showed [27]. Scores on 13 out of the 18 self-efficacy statements increased significantly.

There were no interactions between gender and intervention effects for individual statements or for the self-efficacy mean. Although girls’ texts were assessed as having higher text quality [27], boys and girls demonstrated similar self-efficacy. At group level, there were moderate, statistically significant correlations between the self-efficacy mean and text quality, pre- as well as post-intervention. There were statistically significant correlations between half of the self-efficacy statements and text quality pre-intervention, but only three statistically significant correlations post-intervention.

**Discussion**

In the present study, our aim was to explore self-efficacy for narrative writing in girls and boys, before and after an observational learning intervention. We also aimed to explore associations between self-efficacy and text quality in boys and girls, as our previous study on the same students showed higher text quality in girls’ texts than boys’ texts [27]. The students wrote personal narratives and filled out a self-efficacy scale before and after a short observational learning intervention. Our first research question was whether and in what way self-efficacy for narrative writing changes after intervention. Most students displayed strong self-efficacy already before intervention, and the self-efficacy ratings increased after intervention. There were, however, large individual variations. To sum up, at group level self-efficacy was strong, especially in relation to the text quality ratings made by trained raters. Strong self-efficacy in this age group is consistent with earlier research [19,25]. Social cognitive theory postulates several sources of self-efficacy [1,2,18]. Our intervention design embraces several of these sources. Observing video clips of 12 different peers offered the students many examples of how peers work with narrative writing (“vicarious experience”). Thus, the students had access to a richer and wider range of skills than using film clips of only an expert writer, which has been tried in e.g. a study by van de Weijer et al. [37] or showing only two contrasting film peers, e.g. as in the study by Braaksma et al. [9], in which the video clips showed only two students at a time working with texts. During the intervention in the present study, one important component consisted of time for...
structured reflection, where the participating students were encouraged to talk about the filmed peer models’ oral or written contributions. This was done in small groups and shared with the whole class and can be considered as a form of “social persuasion”. Although students were not receiving direct feedback on their written texts, they were given “enactive experience” when they wrote them. The second part of the first research question was whether there were gender differences in self-efficacy ratings. No such differences were found.

Our second research question was whether, and in what way, self-efficacy was associated with text quality. The self-efficacy results were compared to intervention effects on text quality from our previous study [27] on the same students. There was a statistically significant, moderate correlation between self-efficacy mean and text quality before intervention. Post-intervention, the correlation was less pronounced, but still moderate. The importance of self-efficacy for writing performance has been discussed by several researchers, and evidence remains inconclusive. For example, in a recent study by Graham et al. [38] of older secondary school students, it was shown that writing attitudes and self-efficacy accounted for statistically significant and unique variance in essay-writing after a range of other variables were controlled for, i.e. gender, eligibility for free lunches, reading self-efficacy, and first language. These findings indicate that motivational factors such as self-efficacy and attitudes towards writing must be taken into account in research on students’ writing and writing development. Nine self-efficacy statements had statistically significant correlations with text quality pre-intervention. Post-intervention, three statements had statistically significant correlations with text quality. These concerned punctuation (4), partitioning one’s text into paragraphs (6), and writing headings (7). The three statements did not have statistically significant correlations with text quality pre-intervention. Interestingly, few students actually wrote headings for their narrative texts and very few divided their texts into paragraphs. If there is a lack of distribution along a scale, correlations will not be significant, which may explain that there are only three statistically significant correlations post-intervention, as at least statements 4 and 7 had very high values. Another possible explanation may be that the students got insights into their own (and peers’) competence in these aspects of writing from the intervention, and were able to rate their self-efficacy more accurately post-intervention, though they did not, for some reason, demonstrate these particular aspects of writing in their texts. Our results show that students had strong beliefs about their writing capabilities which is in accordance with previous studies [9,11,19].

Previous research also indicates that an “overestimation” of skills, i.e. stronger self-efficacy than (raters’ assessment of) writing performance, can be expected in this age group [7,14]. In the present study, no students’ self-efficacy mean decreased after intervention. Instead, the self-efficacy mean and the score on most self-efficacy statements increased significantly. We found no significant differences in self-efficacy between girls and boys, neither before nor after intervention. However, as our earlier study showed that girls’ text quality was higher [27], the girls did not overestimate their capability in relation to performance (text quality) to the same degree that the boys did. The argument that girls may be more cautious when reporting self-efficacy [23] may thus apply to the results in this study.

There are some drawbacks in the design that may explain why no student had decreased their self-efficacy score. For logistic and practical reasons, the same two researchers distributed self-efficacy scales, administered the narrative writing tasks and performed the intervention lessons. This may have influenced students trying to be compliant towards the researchers by stating increased self-efficacy. The increase in self-efficacy due to study participation per se, i.e. so-called Hawthorne effects [39], could also explain the increase. Just performing the intervention could make students feel more confident in their abilities. However, the text quality as measured by trained raters not aware of intervention or that each student wrote more than one text, also increased significantly (around seven points on the 0–100 scale) [27].

The validity of the scale merits some further considerations. One issue is internal consistency between statements in the self-efficacy scale. In our study, the correlations between statements in the self-efficacy scale and the self-efficacy mean were strong pre- as well as post-intervention, indicating that the students’ self-efficacy for narrative writing was quite holistic. The mean values of the 18 self-efficacy statements were higher at post-intervention than pre-intervention (with similar standard deviations). Some of the self-efficacy statements were related to specific aspects of narrative writing (e.g. statement 2, “I can find all the letters on the keyboard”). In other statements, there was an overlap between form, content, and organization or between process and product aspects of narrative writing. As an example, statement 9, “I can divide my story into beginning, middle, and end”, may concern writing processes more in this age group than in older adolescents, as they still have not mastered lower-level writing processes fully [40,41]. Thus, we cannot conclude with certainty that self-efficacy for specific aspects of writing increased as an effect of the intervention. This may, however, be further explored by relating each self-efficacy statement to the other statements.

The difficulty in deciding the level of task-specificity has been addressed repeatedly [9,11,18]. The level of specificity which was chosen for the scale in this study aligned with Bandura’s recommendations [18]. Further, the scale should concern belief in one’s capability (“I can…”), rather than e.g. one’s intent or judgment of performance [18]. Our measure of self-efficacy for narrative writing for 10–12-year-olds was adapted from Braaksm et al. [9]. Their scale was targeting self-efficacy for argumentative writing in teenagers, but had the same number of statements, which were similar in their specificity for the genre. Five statements which were independent of genre were similar or identical, e.g. “I can write a text which is one page long” (statement 18) and “I can partition my text into paragraphs” (statement 6). The genre- and age-specific statements were adapted to suit the narrative genre and the age group of the present study,
basing the content on the Swedish curriculum. For example, the statement “I can present the arguments and subordinate arguments in a structured way in my text” from the original scale was adapted to “I can divide my text into beginning, middle and end” (statement 9). Thus, it was not our intention to make the statements more specific for the task at hand, as such scales may end up in a pure self-assessment scale. The students in the present study had not been made explicitly aware of all the skills mentioned in the statements during intervention, and they were not told which peers showed stronger or poorer writing abilities.

Despite using a different scale than previous researchers, our results corroborate earlier findings reporting a somewhat poor calibration, i.e. accuracy of self-efficacy, in relation to writing performance [9,11,19]. A self-efficacy for writing scale (“SEWS”) for middle and high school students by Bruning et al. comprises statements representing three levels of self-efficacy: ideation (capability to generate ideas), convention (capability to express ideas with language-related tools), and self-regulation (capability to manage one’s behavior and writing decisions during writing) [19,20]. This scale was also used by De Smedt et al. [16]. SEWS was not available when the present study was planned and piloted in 2013, but still, it might not have been a better choice of scale for the current study. A third of the statements in SEWS regard self-regulation, which may require metacognitive skills not yet developed in students of our age group. In our scale, we wanted to avoid tapping meta-cognitive skills such as self-assessment or self-regulation. However, the question remains whether we succeeded. It is a challenging task to construct valid and reliable self-efficacy scales in younger students. This can offer an explanation to why children in this age span perhaps have difficulties in differentiating their skills [12,19]. If they consider their abilities in a holistic way, this may be a reason why it is difficult to isolate clear findings in this study. It is not straightforward to disentangle whether students report on their general capability to write good narratives (i.e. self-efficacy) or if they report their self-regulation capacity or their recent performance on a particular narrative, when they are responding to statements in self-efficacy scales.

In conclusion, the students in the present study had strong self-efficacy for narrative writing, which increased after intervention. This supports previous findings. The students may still have a holistic view of their capabilities, not separating their self-efficacy from related factors such as assessment of performance, self-regulation, or general writing skills. Self-efficacy had significant, moderate correlations separating their self-efficacy from related factors such as self-assessment or self-regulation. However, the question remains whether we succeeded. It is a challenging task to construct valid and reliable self-efficacy scales in younger students. This can offer an explanation to why children in this age span perhaps have difficulties in differentiating their skills [12,19]. If they consider their abilities in a holistic way, this may be a reason why it is difficult to isolate clear findings in this study. It is not straightforward to disentangle whether students report on their general capability to write good narratives (i.e. self-efficacy) or if they report their self-regulation capacity or their recent performance on a particular narrative, when they are responding to statements in self-efficacy scales.

Acknowledgments
We thank Martine Braaksma for fruitful discussions on observational learning intervention and self-efficacy scales. We also thank the Linnaeus environment Cognition, Communication and Learning (CCL) at Lund University and Lund University Humanities Lab where the ‘film peers’ for the observational learning intervention were filmed. Lastly, we thank the anonymous reviewers, whose comments have greatly improved this manuscript.

Disclosure statement
The authors report no conflict of interest.

Funding
The present work was financially supported by The Marcus and Amalia Wallenberg foundation [2012.0038].

Notes on contributors
Emily Grenner is a speech-language therapist and PhD student at the Department of Clinical Sciences, Logopedics, Phoniatrics and Audiology, Lund University, Sweden. Her PhD studies concern processes and products of narrative writing in school children with normal hearing and school children with hearing loss.

Victoria Johansson is an Associate Professor in Linguistics at the Centre for Languages and Literature and Deputy Director at the Lund University Humanities Lab, Lund University, Sweden. Her research focus is on language development through the lifespan, primarily cognitive aspects of writing.

Joost van de Weijer is an Associate Professor in Linguistics at the Centre for Languages and Literature at Lund University. His research focuses on the perception of foreign-accented speech. He is also affiliated as methodologist with the Lund University Humanities Lab, where he mainly works with the analysis of experimental data.

Birgitta Sahlén is a Professor in Speech Pathology at the Department of Clinical Sciences, Logopedics, Phoniatrics and Audiology, Lund University, Sweden. Her research focus is on cognition and communication in children with language disorder and/or hearing loss. Comprehension, listening effort, motivation, and learning in relation to children’s cognitive capacity. She is currently heading comprehensive intervention projects aiming at improving children’s narrative writing by observational learning, and teachers’ communicative techniques for language learning interaction in the classroom.

ORCID
Emily Grenner http://orcid.org/0000-0002-6818-1093
Victoria Johansson http://orcid.org/0000-0002-0684-6491
Joost van de Weijer http://orcid.org/0000-0001-9843-3143
Birgitta Sahlén http://orcid.org/0000-0002-8468-0546

References
[1] Bandura A. Social foundations of thought and action: a social cognitive theory. Englewood Cliffs (NJ): Prentice-Hall; 1986.
[2] Bandura A. Social learning theory. Englewood Cliffs (NJ): Prentice-Hall; 1977.

[3] Bandura A. Self-efficacy: the exercise of control. Basingstoke: W. H. Freeman; 1997.

[4] Schunk DH. Self-efficacy and academic motivation. Educ Psychol. 1991;26:207–231.

[5] Coutinho SA, Neuman G. A model of metacognition, achievement goal orientation, learning style and self-efficacy. Learn Environ Res. 2008;11:131–151.

[6] Bandura A. Perceived self-efficacy in cognitive development and functioning. Educ Psychol. 1993;28:117–148.

[7] Bruning R, Kauffman DF. Self-efficacy beliefs and motivation in writing development. In: MacArthur C, Graham S, Fitzgerald J. Handbook of writing research. New York: Guilford Publications; 2016.

[8] Rijlaarsdam G, Braaksma M, Couzijn M, et al. Observation of peers in learning to write: practice and research. J Write Res. 2008:1:53–83.

[9] Braaksma MAH, Rijlaarsdam G, van den Bergh H. Observational learning and the effects of model-observer similarity. J Educ Psychol. 2002;94:405–415.

[10] Raedts M, Rijlaarsdam G, van Waes L, et al. Observational learning through video-based models: impact on students’ accuracy of self-efficacy beliefs, task knowledge and writing performances. In: Hidi S, Boscolo P, editors. Writing and motivation. Oxford: Elsevier; 2007: 219–238.

[11] Raedts M, Steendam E, Grez L, et al. The effect of different types of video modelling on undergraduate students’ motivation and learning in an academic writing course. J Writ Res 2017;8: 399–415.

[12] Muenks K, Wigfield A, Eccles JS. I can do this! The development of achievement motivation. San Diego (CA): Academic Press; 2002. p. 309–332.

[13] Schunk DH, Meece JL. Self-efficacy development in adolescence. In: Pajares F, Urdan T, editors. Self-efficacy beliefs of adolescents. Greenwich (CT): Information Age Publishing; 2006. p. 71–96.

[14] De Smidt F, Merchie E, Barendse M, et al. Cognitive and motivational challenges in writing: studying the relation with writing performance across students’ gender and achievement level. Read Res Q. 2018;53:249–272.

[15] Klassen R. A question of calibration: a review of the self-efficacy beliefs of students with learning disabilities. Learn Disabil Quart. 2002:25:88–102.

[16] Bandura A. Guide for constructing self-efficacy scales. In: Pajares F, Urdan T, editors. Self-efficacy beliefs of adolescents. Greenwich (CT): Information Age Publishing; 2006. p. 307–337.

[17] Bruning R, Dempsey M, Kauffman DF, et al. Examining dimensions of self-efficacy for writing. J Educ Psychol. 2013;105:25–38.

[18] Bruning R, Dempsey M, Kauffman DF, et al. Self-efficacy for writing scale. Database record retrieved from PsycTESTS; 2013.

[19] OECD. PISA 2015 results. (Volume I), Excellence and Equity in Education, PISA. Paris: OECD Publishing; 2016. doi:10.1787/9789264266490-en

[20] Pajares F, Miller MD, Johnson MJ. Gender differences in writing self-beliefs of elementary school students. J Educ Psychol. 1999:91:50–61.

[21] Pajares F, Valiante G. Grade level and gender differences in the writing self-beliefs of middle school students. Contemp Educ Psychol. 1999;24:390–405.

[22] Pajares F, Valiante G. Gender differences in writing motivation and achievement of middle school students: a function of gender orientation? Contemp Educ Psychol. 2001;26:366–381.

[23] Pajares F. Self-efficacy beliefs, motivation, and achievement in writing: a review of the literature. Read Writ Quart. 2003;19: 139–158.

[24] Tillema M, van den Bergh H, Rijlaarsdam G, et al. Quantifying the quality difference between L1 and L2 essays: a rating procedure with bilingual raters and L1 and L2 benchmark essays. Lang Test. 2013;30:71–97.

[25] Grenner E, Åkerlund V, Askér-Årnason L, et al. Improving narrative writing skills through observational learning: a study of Swedish 5th-grade students. Educ Rev. 2018. doi:10.1080/10775462.2018.1536035

[26] Feagans L, Appelbaum MI. Validation of language subtypes in learning disabled children. J Educ Psychol. 1986;78:358–364.

[27] Johansson V. Developmental aspects of text production in writing and speech [dissertation]. Lund: Department of Linguistics and Phonetics, Centre for Languages and Literature, Lund University; 2009.

[28] McCutchen D. From novice to expert: implications of language skills and writing-relevant knowledge for memory during the development of writing. J Writ Res. 2011;3:51–68.

[29] Wolff U. Lilla DUVAN. Stockholm: Hogrefe Psychologiörflaget AB; 2010.

[30] Bishop D. Test for reception of grammar. version 2. Swedish version. Stockholm: Pearson Education; 2009.

[31] Magnusson E, Naucler K. SL40: läsförståelse: meningar [SL40: reading comprehension: sentences]. Lund: Studentlitteratur AB; 2010.

[32] Skolverket. Curriculum for the compulsory school, preschool class and school-age education 2011. Stockholm: Norstejd juridik; 2018.

[33] Wewers ME, Lowe NK. A critical review of visual analogue scales in the measurement of clinical phenomena. Res Nurs Health. 1990;13:227–236.

[34] Strömqvist S. Discourse flow and linguistic information structuring: explorations in speech and writing (Gothenburg papers in theoretical linguistics: 78). Gothenburg: Göteborg University; 1996.

[35] van de Weijer J, Åkerlund V, Johansson V, et al. Writing intervention in university students with normal hearing and in those with hearing impairment: can observational learning improve argumentative text writing? Logoped Phoniatr Voc. 2018;44:1–9.

[36] Graham S, Daley SG, Atkens AA, et al. Do writing motivational beliefs predict middle school students’ writing performance? J Res Read. 2018;41:642–656.

[37] Schwartz D, Fischhoff B, Krishnamurti T, et al. The Hawthorne effect and energy awareness. Proc Natl Acad Sci USA. 2013;38:15242–15246.

[38] Bourdin B, Fayol M. Even in adults, written production is still gender and achievement goal oriented. J Writ Res. 2018;53:249–272.

[39] Chanquoy L. Revision processes. In: Nystrand M, editors. The SAGE handbook of writing development: explorations in speech and writing (Gothenburg papers in theoretical linguistics: 78). Gothenburg: Hogrefe Psychologiörflaget AB; 2010.

[40] Nystrand M, editors. The SAGE handbook of writing development: explorations in speech and writing (Gothenburg papers in theoretical linguistics: 78). Gothenburg: Hogrefe Psychologiörflaget AB; 2010.

[41] Skolverket. Curriculum for the compulsory school, preschool class and school-age education 2011. Stockholm: Norstedts juridik; 2018.