



UvA-DARE (Digital Academic Repository)

Creative writing: Thinking beyond the standard text

Teaching high school students to write original texts

ten Peze, A.A.

Publication date

2024

[Link to publication](#)

Citation for published version (APA):

ten Peze, A. A. (2024). *Creative writing: Thinking beyond the standard text: Teaching high school students to write original texts*. [Thesis, externally prepared, Universiteit van Amsterdam].

General rights

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

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

CHAPTER 5
CREATIVE WRITING IN SECONDARY EDUCATION
A Meta-Analysis of Intervention Studies

Abstract. The aim of this meta-analysis was to gain insight into effective interventions for improving students' creative writing in order to formulate design principles for effective creative writing instruction. We examined the extent to which overall effectiveness varied by study variables, research methodology and assessment method, treatment (focus and content), and instructional approach for both writing quality and writing process. Thirteen studies met our inclusion criteria and were reviewed. The overall effect sizes were .88 for writing quality and .68 for writing process, confirming that instruction helps improve both students' creative writing products and processes. None of the included variables related to study characteristics, research methodology, and assessment method contributed to the effect sizes of the interventions for either writing quality or writing process. Results indicated that the quality of creative writing did improve when instruction focused on the writing process and when fictional texts (excerpts) were used. Furthermore, the effect of a focus on the writing process is consistent with previous meta-analyses showing that writing process instruction has a positive effect on writing quality; our meta-analysis indicated that this is also explicitly true for creative writing quality. If the goal is to influence the writing process, then instruction should focus on the writing process or a combination of the writing and creative processes. Our findings can be used to design follow-up creative writing interventions or as guidelines for teachers and curriculum developers who want to develop creative writing materials for secondary schools. In addition, we recommend that more research be conducted on the effects of creative writing instruction in secondary schools. Finally, we make recommendations for a clear and insightful method of reporting, as the quality of reporting was often suboptimal.

Keywords: creative writing, text quality, writing processes, education, intervention, instruction

1 INTRODUCTION

'Teaching students how to write creatively helps teach them how to approach life in a creative way.' – Robert J. Sternberg

'Teaching is a creative profession.' – Sir Ken Robinson

Interest in creativity has grown tremendously over the past fifteen years. Between 2010 and 2023, the number of hits for "creativity" in Google Scholar increased from 205,000 to 4,170,000. When it comes to creativity in education, Hernández-Torrano & Ibrayeva (2020) reported that interest in creativity increased significantly in the last decades. They proposed that this is partly due to the current understanding that creativity contributes positively to school performance (Gajda et al., 2016). In addition, they argue that the rapidly changing world requires creative skills that appear to be teachable (Chan & Yuen, 2014; Davies et al., 2013; OECD, 2019). A clear indication that creativity is on the education policy agenda is that the OECD has included creativity as one of five constructs

in its new Learning Framework 2030 (OECD, 2018), alongside critical thinking, responsibility, resilience, and collaboration.

In this study, we focus on creative writing in secondary education. Creative writing is of fundamental importance (Vincent-Lacrin et al., 2019). Yet it is under-researched in education and also rarely taught: '...many people become creative writers not by virtue of their education, but in spite of it.' (Sternberg, 2009, p. xvi).

Indeed, curriculum analyses indicated that there is much to be gained from teaching creativity (Halpern, 2010; Sternberg, 2010). Some countries, such as Singapore and Korea (Vincent-Lacrin, 2013), have made efforts to promote creative teaching practices, but many countries still pay very little attention to creativity. This variation was confirmed in a study by Wyse & Ferrari (2015) on the position of creativity in the national curricula of the 27 EU member states and the United Kingdom. They reported that all curricula included the word "creativity". However, the relative density varied widely, from 0.04 per 100 words in the Netherlands to 1.78 in Northern Ireland. They also found that creativity was relatively under-represented in language arts curricula, particularly in writing and reading.

This lack of attention to creativity in the curriculum seems to be a global phenomenon (Sternberg, 2010; Wyse & Ferrari, 2015). In addition, Patston et al. who examined creativity in the curricula in twelve countries around the world concluded that '... there were many times when the general statements reinforced the idea that creativity is a desirable attribute and a key skill to be developed. The translation of these general statements into the disciplines, however, was inconsistent and seemed ad hoc' (Patston et al., 2021, p. 223). Their study also found that teachers received little support for embedding creativity promotion in their subject areas, and that policymakers are apparently unaware of the importance of such support for teachers.

This lack of inclusion of creative writing education, policy consciousness, and teacher support in the Dutch secondary curriculum prompted us to conduct a meta-analysis of studies on creative writing, to inform researchers, teachers, curriculum specialists, and instructional designers about the extent to which creative writing education can be effective, and which instructional approaches and learning content might be crucial. Because the need for creativity in education is global, as is the curriculum gap, the outcomes of this meta-analysis could be informative for other curriculum regions as well.

In the Netherlands, writing instruction at the secondary level focuses mostly on argumentative and expository writing (Van Burg, 2010; Van Gelderen, 2010). Creative writing has become virtually absent from the curriculum since it was removed from the Dutch national examination program in 1998 (Stuurgroep

Profiel Tweede Fase [Steering Group Second Phase Profile], 1995), which focused on functional and academic literacy. Since then, the writing of poetry and fiction has been largely confined to primary and lower secondary education (Van Burg, 2010; Van Gelderen, 2010). At the same time, students between the ages of 12 and 18 reported writing creatively in their free time, including all kinds of texts (LKCA, 2017). They also reported a significant difference between this self-initiated writing and school writing tasks (LKCA, 2017).

Moreover, adolescence seems to be the most sensitive period of life for improving creative thinking. Two studies support this assumption. Kleibeuker et al. (2016) investigated whether training could improve creative idea generation in both adolescents (13-16 years) and adults (23-20 years). Adolescents performed as well as adults before the training in terms of idea flow and flexibility, and even better in terms of originality. Almost more remarkable was that adolescents continued to improve their originality up to two weeks after training, while adults did not. Stevenson et al. (2014) confirmed these findings in a similar study in the verbal domain. Adolescents improved more in terms of originality than adults. The authors concluded that 'Given the strong associations between creative ideation, prefrontal cortex, and cognitive control functionality (e.g., Dietrich, 2004; Groborz & Necka, 2003; Keating, 2004), adolescence provides a favorable time window for progression in creative ideation' (2018, p. 13). Furthermore, with age, brain networks become increasingly specialized and integrated, at the expense of cognitive plasticity in adults compared to adolescents (Jolles & Crone, 2012; Stevenson et al., 2014). Given that the adolescent brain is both capable of flexible thinking and receptive to training in creative ideation, it seems a promising idea to train students in creative writing at this age.

1.1 *Defining creative thinking & writing*

There is some agreement in research on the definition of creativity, but there are also differences in emphasis. All definitions assume two important characteristics: originality and task appropriateness. Sternberg & Lubart define creativity as '...the ability to produce work that is both novel and appropriate' at the same time (1999, p. 3). Thus, a creative product must be both new or original, and appropriate to the task, the intended audience, and the context. The OECD defined creative thinking for 15-year-old students as: 'the competence to engage productively in the generation, evaluation and improvement of ideas, that can result in original and effective solutions, advances in knowledge and impactful expressions of imagination' (OECD, 2019, p. 8).

Several process models within creativity research are based on the interaction between generative and explorative processes (e.g., Finke et al., 1992; Mumford et al., 1991). In the cyclical Geneplore model, proposed by Finke et al. (1992), there is a continuous interaction between these two processes. The generative process generates ideas that the explorative process evaluates and elaborates, which in turn may trigger the generative process. Csikszentmihalyi (1996) and Doyle (1998) confirmed this cyclical process in their interview studies with professional novelists and poets. They concluded that it is this interaction that characterizes the creative writing process: on the one hand, the more unconscious associative processes that generate ideas, and on the other hand, the critical examination of those generated ideas. When these two processes are in optimal balance, the writer achieves a state of flow (Csikszentmihalyi, 1996, p. 264). Doyle added another dimension to the cyclical model. What the authors she interviewed also experienced is that they moved back and forth between a "fiction-world" and a "writingrealm" (1998, p. 36). This allows the writer to empathize with the characters in the story. Thus, although all forms of writing require some degree of creative thinking, writing fiction, such as poems or stories, creates additional creative space because the writer creates and relates to a fictional world and fictional characters.

1.2 *Aims and research questions*

If we want to encourage, develop, and improve students' creative writing skills, we need to design effective instruction. To do so, we need to define the key components of creative writing instruction. To date, however, there is no review of research that examines the effects of instruction on high school students' creative writing performance. In addition, in the present study we aim to report the intervention effects on processes, as improvement of creativity and originality may be observed in writing or thinking processes. Meta-analyses are available on effective writing instruction (Graham & Perin, 2007; Graham & Sandmel, 2011), effective creative learning environments (Davies et al., 2013; Lasky & Yoon, 2020), and creative pedagogies (Cremin & Chappell, 2021, Karwowski et al., 2022; Liu & Chang, 2017). More recently, Graham et al. (2023) found a small positive effect ($ES = 0.27$) for teaching critical/creative thinking skills for writing. However, of the 357 studies they included, only six studies focused on these skills, and even fewer of those focused exclusively on creative thinking. Given the lack of research on creative writing in secondary education, we decided to conduct a meta-analysis as well. We did not limit our review to text quality, but also looked at processes: both writing processes and creative thinking processes.

The aim of the current study was to gain insight into effective interventions for creative writing instruction and propose design principles based on the outcomes, so that teachers can develop effective creative writing lessons in secondary schools. As Sir Ken Robinson (Robinson & Aronica, 2015) states, it is teachers who can have the greatest impact, but they must be equipped with the right knowledge and insights to do so.

To enable the design of effective creative writing instruction, we need to define instructional design principles for effective creative writing instruction. Design principles are parameters that indicate the relationship between a desired learning outcome, such as writing quality, and an instructional activity, such as divergent thinking. They are formulated as 'if/then' statements. Following Van den Akker (1999) and Reigeluth (1999), Van Ockenburg et al. (2019) formulated them as: "If we want our students to achieve learning outcome X, then we should involve them in strategy, activity, approach ... Y" (2019, p. 303). By basing instructional design principles for creative writing instruction on the findings of the meta-analysis, we ensure that they are likely to be effective.

In conclusion, the following questions were the main focus of this meta-analysis on both creative writing performance and processes:

1. What is the overall effectiveness of interventions designed to improve students' creative writing?
2. To what extent does the overall effectiveness depend on
 - a. study variables?
 - b. research methodology & assessment method?
 - c. the treatment (focus & content)?
 - d. the instructional approach?

More information on the variables mentioned in research question 2 can be found in section 2.4 Coding procedure.

2 METHOD

To explore the field, we first conducted a broad search for experimental and quasi-experimental research on creative writing covering a wide range of educational levels: from fourth grade up to university level. We did not limit the search to a specific time period or to narrative writing, but also included expressive writing, poetry writing and other forms of creative writing, in all languages. In this first phase, we searched for all publications, including those that were not

peer-reviewed. We then selected articles on secondary education that were peer-reviewed, as a mark of quality for the selected studies.

2.1 *Inclusion criteria*

Studies had to meet five inclusion criteria to be included in the review (see Table 5.1). For the first criterion, aim, we also included studies in which students did not write during the intervention, but did other activities such as drama, or creating a story in a game.

For the second criterion, research design, we included all kinds of designs: switching replications, interrupted time series, and so on. However, they had to include a control group, because that is necessary to determine whether the effect is due to the intervention or to other factors.

In terms of participants, criterion three, we deliberately chose to include students with special needs in this review, ranging from gifted students to struggling writers to autistic students. We assumed that the design principles that apply to such samples, might be generalizable to other samples.

For criterion four, we only included studies that were conducted in an educational context. Since our goal was to gain insight into effective design principles for teaching creative writing, we decided to exclude studies conducted in non-educational contexts. Therefore, we excluded studies in which creative writing was not the primary purpose of the study (e.g., studies in which creative writing was used to improve health rather than obtain a specific educational outcome).

Regarding the target language, criterion five, we decided to include studies that addressed both first and/or second language acquisition. Although second language acquisition often involves different learning goals, critical components of creative writing instruction in a second language may also apply to creative writing instruction in a first language.

Finally, we excluded studies which did not include writing product or process as outcome measures. Both analytic and holistic data were analyzed when it was possible to calculate a weighted effect size (Cohen's d). Writing processes were measured in a variety of ways, including text-based measures such as production rate via word count and registering processes via a keystroke logging program.

Table 5.1 Inclusion criteria

1. Aim	Studies designed to improve students' creative writing skills and/or creative writing process.
2. Design	Experimental and quasi-experimental studies: The minimum requirement was a design with two conditions: experimental and control. The control condition could consist of a no-treatment control condition or a comparison condition with an alternative approach or a second experimental condition.
3. Participants	Secondary education (ages 12-18 years), including all types of participants, ranging from special needs to regular students. If a study spanned multiple age categories, such as 10 to 12 or 18 to 21, we did not include these categories because they do not fall fully within secondary education.
4. Context	Studies conducted in an educational context, including both in-class and out-of-class studies, were included in the analysis.
5. Language	L1 and L2.
6. Dependent measures	Creative writing proficiency and/or writing processes.

2.2 *Search procedure*

The search terms and search strings are listed in Appendix Q. In November 2020 we searched five databases: PsycINFO, Education Resources Information Center (ERIC), Linguistic and Language Behavior Abstracts (LLBA), Web of Science and Scopus (October 2020). In doing so, we included one behavioral database, one educational database, one linguistic database, and two multidisciplinary databases. In this first phase, we searched for studies carried out in fourth grade to university and in all types of publications, including non-peer-reviewed studies. This resulted in 5.944 initial hits. We then exported the results to Zotero for de-duplication (Corporation for Digital Scholarship, n.d.), which left 4.131 hits.

2.3 *Screening in Rayyan*

Two researchers screened all articles in Rayyan (Ouzzani et al., 2016) to determine whether the article met the five inclusion criteria, based on its title and abstract. In case of disagreement the decision was made in consultation. In case of doubt, the article was included. The reasons for excluding articles in this round

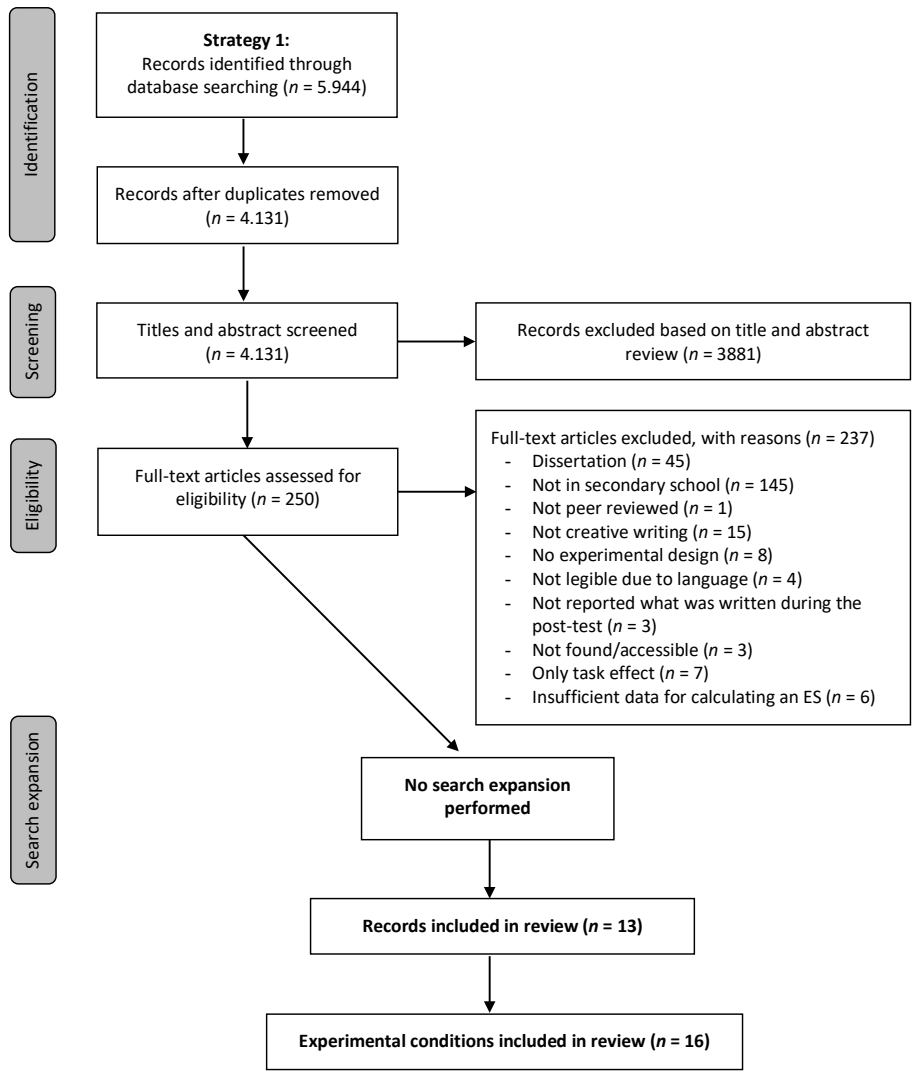
varied, but most often the studies were not (quasi)-experimental, were not situated in an educational context (criterion 4) or did not measure creative writing (criterion 6).

In the second screening round, the remaining 250 full-text articles were screened for eligibility. This eligibility screening was primarily performed by the first author, while unclear cases were jointly assessed by two authors to determine whether the articles could be included.

At this stage we considered two criteria. We selected peer-reviewed articles, which would offer the advantage of quality control. The disadvantage, however, is that articles with non-significant results are often not published, which may bias the results somewhat. In the end, we excluded 46 sources for this reason: 1 article and 45 dissertations. We also selected studies in secondary education. Because we were not sure whether there was enough peer-reviewed research on creative writing in secondary education, we initially took a broad view. When it became clear that there was a sufficient number of studies for a meta-analysis, we limited the analysis to our target group: secondary schools.

Of the 250 papers that were reviewed in full, thirteen met the inclusion criteria. Figure 1 provides a PRISMA flowchart (Moher et al., 2009) for the literature search and screening with a complete overview of the reasons why the remaining 237 articles were excluded during the eligibility screening.

Figure 5.1 Flow chart of the search procedure for suitable studies



2.4 *Coding procedure*

The coding scheme included seven components: 1. paper information, 2. study information, 3. treatment focus and content, 4. type of instruction, 5. assessment method, 6. research methodology, and 7. study quality indicators (see Appendix R for an overview of all variables). In what follows, we briefly explain the scheme.

2.4.1 Paper information

For the paper information we coded title, author, publication year, journal, and impact factor.

2.4.2 Study information

The study characteristics included total number of participants, type of student, grade, mean age, and continent. We also included the intensity of the intervention by coding the number of sessions (excluding the pre- and posttest), the duration of the intervention in minutes and the number of texts students wrote. We coded the number of teachers involved in the intervention (control and experimental classes), whether the intervention was aimed at first language (L1) or second or foreign language (L2), and the genre students wrote during the intervention and at the pre- and posttest.

2.4.3 Focus and content of the treatment

For the content of the intervention, we first coded the focus of the intervention: the creative product (the text) and/or the creative process. If the focus was on the product, we coded the text level: the text as a whole or a part of it, such as sentence or grammar. If the focus was on the process, we distinguished two foci in terms of content: a focus on creative thinking processes (e.g., activities that stimulate and support this process, such as brainstorming, associative thinking, and divergent thinking) and a focus on the writing process (e.g., activities such as revising, prewriting, writing, etc.), or both.

Secondly, we coded the *specific learning content* of the intervention (e.g., creative thinking, narrative writing techniques, grammar, storytelling, fiction-based, motivation, etc.), where combinations were also possible.

2.4.4 Instruction

Instruction consisted of three major instructional activities: instruction, feedback, and revision. Table 5.2 summarizes the coding identifiers for instruction type and the accompanying explanations.

Table 5.2 Coding identifiers for type of instruction and descriptions

	Coding identifier	Description
Instruction	Comprehensiveness	How extensive was the instruction during the intervention? Options: no instruction, minimal, extensive instruction or unknown.
	Focus	Was the focus of the instruction on the process or the product or both?
	Implicit or explicit	Was the instruction implicit (the learner reaches understanding and rules through watching, analyzing, comparing et cetera) or explicit?
	Approach	Through which instructional approach did students need to acquire the content (e.g., problem-centered, activation, task representation, demonstration, application, peer assistance or integration)?
	Practice	How extensive was the practice during the intervention?
	Scaffolding	Was scaffolding present?
Feedback	Feedback Provision	Was feedback provided? Who provided the feedback: teacher(s), researcher(s), peers, peer & teacher, self-feedback?
	Focus	Global or focused on a specific component such as plot or style?
	Tools/criteria	Were instruments or criteria used to provide feedback?
Revision	Revision	Did students revise?
	Feedback based?	Did students revise based on feedback?

2.4.5 Assessment

We coded whether the assessment of the text was analytical or holistic, what assessment tools were used, the number of assessors and who the assessor was.

2.4.6 Research methodology

We coded the type of experiment – experimental or quasi-experimental – and the number of experimental groups. We also coded whether learner variables were included in the analysis and whether they were related to writing and/or creative thinking. The dependent variables were coded as text length, text quality, writing process or originality/creativity of the text.

2.4.7 Study quality indicators

For study quality indicators, we used Graham & Perin's (2007) measures and added researcher allegiance as an additional indicator. Table 5.3 provides an overview of the study quality indicators.

Table 5.3 Coding identifiers for study quality indicators

Coding identifier	Description
Assignment to conditions	The strategy to assign participants to conditions.
Mortality equivalence	Criterion was met if 90% or more of the students completed the study.
Ceiling/floor effects	Criterion was met if the mean of the posttest was less than one standard deviation from the lowest and highest scores.
Pretest equivalence	Criterion was met if there was no difference in text quality and/or writing processes between conditions before the start of the intervention.
Instructor training	Criterion was met if the study described how teachers were trained for the intervention.
Type of control condition	Distinguished were (1) an alternative treatment, (2) an unspecified treatment or (3) a no-treatment condition (business as usual).
Hawthorne effect	Criterion was met if condition effects on attention and time were checked.
Treatment fidelity	Criterion was met if at least 80% of the experimental condition was performed as intended.
Teacher effects	Criterion was met if teachers taught both conditions or were randomly assigned to conditions.
Researcher allegiance	Criterion was met if the researcher was the developer of the intervention.

The development of the coding scheme took place in stages. Based on the content of the articles, the categorization in the resulting coding scheme was further refined by the first, second and third authors during several meetings. Two authors independently fully coded all thirteen studies, after which coding differences were jointly resolved.

2.5 *Calculation of effect sizes and statistical analysis*

We calculated effect sizes for writing quality for all thirteen studies and for process measures for four of them. When available, we used the reported holistic score for writing quality. Some studies reported various aspects of writing quality, such as creativity, novelty, and imagination (see Appendix R, Table R2). In such cases, we calculated the effect size per aspect and then averaged them.

We calculated the standardized mean difference effect size (ES; Hedges' g), using the ES tool on the Psychometrica website (Lenhard & Lenhard, 2016). Input for the calculations were either the pre- and posttest means, standard deviations, and sample sizes (Morris, 2008) or other reported values such as F or t values. If a study only reported posttest means and standard deviations, we used these to calculate the effect size. Six studies did not report means and standard deviations. Therefore, we had to use reported inferential statistics to calculate the effect size.

The studies represent various study designs. Eight studies implemented a pre-posttest control group design. One study implemented a time series study and four studies consisted of a post-test-only design. Three studies in the sample had multiple treatment groups (see Appendix R, Table R2). For the Groenendijk study (2013) and the Dolmaz & Kaya study (2018), we calculated effect sizes for each of these groups, as the treatments in each group were sufficiently different. For the study by Teo & Tan (2003) however, we averaged the effect sizes of each of the groups into one effect size because the treatment was identical for the three groups that the authors distinguished in the report. The only difference between the groups was the participants' prior knowledge of a style figure (e.g., Biyu).

The studies in our sample varied both in terms of content and approach. Therefore, we used a random effects model for the analyses. Such a model does not estimate a single true effect size, but rather the mean of a distribution of effects. To test for homogeneity, we used a random effects model to test for heterogeneity due to the large variability in our sample. In the final step of the analysis, a mixed effects moderator analysis determined whether and to what

extent variability could be explained by moderators such as treatment content, instruction, or continent.

2.6 *Description of studies included in the analysis*

Appendix R provides descriptive information for each study. The thirteen included studies represent 1,042 participants. We calculated sixteen effect sizes for overall writing quality and eight effect sizes for processes: twenty-four effect sizes across thirteen studies. All studies were published in peer-reviewed journals. Ten studies involved regular students and eight were carried out in L1 classes. However, the descriptive information was far from complete: many elements of our coding scheme were not reported in the articles, an issue which we will return to in the discussion.

In the end, we realized that we coded more variables than we ultimately analyzed (see Appendix R for an overview). We had to exclude variables for which too many studies did not provide information, for instance about the study variables (see Appendix R, Table R1). Appendix R shows that many studies did not provide specifications for many variables (indicated by an asterisk in the column headings).

3 RESULTS

In this section, we present the results in two main sections: first the effects on text quality, then on processes.

3.1 *Text quality*

We computed effect sizes for text quality for all thirteen studies. We included effect sizes from Groenendijk (2013), for two conditions, and three effect sizes from the three different experimental groups in the Dolmaz & Kaya (2018) study. As a result, the n for sample size in this section corresponds to the number of effect sizes rather than the number of studies. The total number of effect sizes in the analyses is sixteen.

The overall effect size across all studies is large ($g = 0.88$, $se = 0.21$, $p < .001$, $CI = 0.44 - 1.32$). The confidence interval indicates considerable variation. The heterogeneity test confirmed that a common effect size for this sample is not a reliable index ($Q = 262.89$, $df = 15$, $p < .001$). Given the observed variance in effect sizes, and the variation in content and instructional approaches, we will examine whether moderators can explain the variance in effect sizes. Such analyses may reveal relevant variations in content and instructional approaches.

These moderators could contribute to the definition of interventions in future research.

We report the results in the following order: study information, research methodology & assessment method, followed by the content of the intervention: treatment and instruction. For all these variables, an intercept model was compared with a model that included the intercept and a specific moderator. As all articles were peer-reviewed, there was no need to test for possible publication bias. If fewer than two conditions had different scores, we omitted the moderator analysis.

3.1.1 Study information

We tested six out of nine study information variables (see Appendix R, Table R1), none of which explained the variance in effect sizes. There were no significant effects of student type ($\Delta\chi^2(1) = 0.41, p = .520$), continent ($\Delta\chi^2(4) = 7.15, p = .128$), target language (L1/L2) ($\Delta\chi^2(1) = 0.34, p = .558$), mean age ($\Delta\chi^2(1) = 1.40, p = .238$), number of sessions during the intervention ($\Delta\chi^2(1) = 0.00, p = .985$), and number of minutes during the intervention ($\Delta\chi^2(1) = 0.3, p = .864$).

3.1.2 Research methodology & assessment method

Two research designs were represented: experimental ($n = 7$) and quasi-experimental ($n = 9$). No significant effect of research design on effect size was observed ($\Delta\chi^2(1) = 2.63, p = .105$), nor an effect of study design: time series ($n = 1$), pre- and posttest with control group ($n = 11$), and posttest with control group ($n = 4$) ($\Delta\chi^2(2) = 1.66, p = .437$).

The studies differed in how writing quality was assessed. We analyzed whether there was a difference between analytical ($n = 12$) versus holistic ($n = 3$)/unspecified ($n = 1$) assessment. Assessment method did not significantly affect effect size ($\Delta\chi^2(1) = 3.74, p = .053$).

3.1.3 Treatment

Treatment focus

Interventions could focus on the process (e.g., writing process or creative process) ($n = 3$), the product (e.g., poems or stories) ($n = 3$), both ($n = 6$), or unspecified ($n = 4$). The presence of a process focus does not make a significant difference ($\Delta\chi^2(1) = .392, p = .53$), nor does the presence of a product focus ($\Delta\chi^2$

(1) = 1.694, $p = .193$) or the focus on both the product and process ($\Delta\chi^2(1) = 3.694$, $p = .075$).

When we zoom in on the type of process focus, we distinguish focus on the writing process ($n = 4$), a combined focus on the creative thinking and writing processes ($n = 5$), or an unspecified focus ($n = 7$). When the process focus was specified, the model fit improved ($\Delta\chi^2(1) = 12.884$, $p < .001$), with a significantly larger effect of focus on the writing process ($\Delta g = 1.13$, $se = 0.31$, $p = .009$), showing a large positive effect on writing quality ($g = 1.48$, $se = 0.25$, $p < .001$).

When the treatment was focused on the product, it could focus on either the text level ($n = 4$), the sentence or a lower level ($n = 3$), or both ($n = 5$). However, this moderator did not improve the model fit ($\Delta\chi^2(2) = .51$, $p = .775$).

Treatment content

We distinguished eleven content variables. We could not include three of these (*unspecified*, *non-creative writing*, and *other*) in the analyses because we did not have enough data to reliably calculate the effect. The estimated effect sizes, standard errors, significance values, and confidence intervals are shown in Table 5.4.

Table 5.4 Treatment content. Estimated effect sizes (g), standard error (se), significance values (p), and 95% confidence intervals (ci)

Content of the treatment	n	g	se	p	Confidence intervals	
					ci. lb	ci. ub
Creative thinking	6	0.27	0.27	.334	-.31	0.86
Fiction-based	4	1.98	0.30	< .001	1.34	2.61
Writing strategy	3	-0.03	0.42	.948	-0.93	0.88
Motivation	2	-0.15	0.50	.769	-1.23	0.93

Four content categories are not significantly associated with differences in effect size: a focus on creative writing ($n = 9$) ($\Delta\chi^2(1) = 1.76$, $p = .185$), storytelling ($n = 6$) ($\Delta\chi^2(1) = 0.48$, $p = .487$), narrative writing techniques ($n = 3$) ($\Delta\chi^2(1) = .06$, $p = .801$), and grammar for storytelling ($n = 2$) ($\Delta\chi^2(1) = .168$, $p = .682$). However, the model fit improved for three moderators: creative thinking ($n = 6$) ($\Delta\chi^2(1) = 7.62$, $p = .006$), writing strategy ($n = 3$) ($\Delta\chi^2(1) = 5.46$, $p = .020$), and motivation ($n = 2$) ($\Delta\chi^2(1) = 4.721$, $p = .030$). Within these moderators, we observed significantly smaller effect sizes for creative thinking ($\Delta g = 0.96$, $se = 0.35$, $p = .015$),

writing strategy ($\Delta g = 1.09$, $se = 0.46$, $p = .035$), and motivation ($\Delta g = 1.17$, $se = 0.54$, $p = .047$). However, the effect sizes were not significantly different from zero (see Table 5.4). A fourth moderator, fiction-based, also improved the fit of the model ($n = 4$) ($\Delta\chi^2(1) = 18.79$, $p < .001$), indicating a larger effect size when the treatment was fiction-based ($\Delta g = -1.46$, $se = 0.34$, $p < .001$), which was a large effect ($g = 1.98$).

3.1.4 Instruction

Of the six instructional variables we coded, we could not analyze the effect of instructional focus and the effect of the number of texts written during the intervention because we had too little data to calculate a reliable effect. Three of the instructional variables were not significantly associated with differences in effect size: whether instruction was provided implicitly or explicitly ($\Delta\chi^2(2) = 1.63$, $p = .443$), the presence of scaffolding ($\Delta\chi^2(1) = 2.63$, $p = .105$), and the presence of feedback ($\Delta\chi^2(1) = 3.49$, $p = .062$). Only extensive instruction ($n = 8$) improved the model fit: ($\Delta\chi^2(1) = 4.92$, $p = .027$), with a significantly larger effect of extensive instruction ($\Delta g = 1.037$, $se = 0.47$, $p = .027$) compared to no instruction ($n = 2$), showing a large positive effect of extensive instruction ($g = 0.85$, $se = 0.20$, $p = .003$).

3.2 Processes

Four studies measured processes, using different measures. Groenendijk et al. (2013) reported two process measures – process time and revision ratio – collected in two measurement sessions (two different poems), in two different experimental conditions, resulting in eight effect sizes. We aggregated the effect sizes across the two tasks per measurement occasion, yielding a total of four effect sizes. Felix & Lawson (1994) and Christensen (2005) measured productivity (number of words). Felix & Lawson (1994) also reported fluency rate (number of words per time unit). Alias et al. (2013) reported problem solving scores. In total, eight effect sizes for process measures were included in the analyses.

The overall effect size for writing processes was significant according to a random effects model ($g = 0.68$, $se = 0.17$, $p = .005$, $CI = 0.28 - 1.07$). The homogeneity test indicated a significant variability in the effect sizes of the included studies ($Q = 18.31$, $df = 7$, $p < .01$). Moderator analyses may show a reduction in this variability. Due to the small number of effect sizes we could examine only a few moderators, due to lack of variability. If fewer than two conditions differed, we refrained from the moderator analysis.

3.2.1 Treatment

Focus

Whether conditions focused on process ($n = 5$) or on process and product ($n = 3$), did not explain variance in effect size ($\Delta\chi^2(1) = .01, p = .927$), nor did the moderator for process content (writing process only ($n = 2$), creative process only (no study), or focus on both ($n = 4$)) ($\Delta\chi^2(1) = 2.22, p = .134$). However, the model fit improved in favor of the focus on writing only ($\Delta\chi^2(1) = 4.40, p = .04$), with significant effect sizes for writing only ($g = 1.02, se = 0.22, p = .004, n = 4$) and the combination of writing and creative processes ($g = .45, se = 0.16, p = .029, n = 4$).

Content

Four distinctions in learning content were not included in the analyses, as less than two conditions represented such a moderator: focus on communicative writing, storytelling, narrative techniques, and fiction-based. Three moderators were linearly interdependent: when the condition focused on creative thinking, the condition also focused on writing strategies and self-motivation. We considered revising these moderators into one compound moderator, creative thinking and writing instruction ($n = 4$). However, this moderator would completely overlap with the moderator 'providing instruction' (see below).

The only remaining moderator for content was whether participants actually wrote creative texts ($n = 6$), which had no effect ($\Delta\chi^2(1) = 2.47, p = .12$). There were also interventions in which students did not write creative texts because, for example, they observed a model writing a creative text (Groenendijk et al., 2013).

3.2.2 Instruction

Providing instruction was a significant moderator ($\Delta\chi^2(1) = 15.00, p < .001$), with instruction ($n = 3$) ($\Delta g = 1.11, se = 0.28, p < .012$) being more effective for students' writing processes than no instruction, e.g., observation only ($n = 4$). Other distinctions for instruction such as providing scaffolding and feedback completely overlapped with the compound variable we presented before the distinction, and thus were not significant moderators.

4 DISCUSSION

The aim of this meta-analysis was to gain insight into effective interventions for improving students' creative writing in secondary education, so that we could

define instructional design principles for instruction in creative writing. In this study, we included effects on writing performance and writing process. Next, we will discuss our findings and their implications for creative writing research and educational practice.

The overall effect sizes we reported of $.88$ for writing quality and $.68$ for writing process suggest that instruction helped to improve both creative writing performance and process. These findings are consistent with previous research on creative pedagogies (Cremin & Chappell, 2021; Karwowski et al., 2022; Liu & Chang, 2017), writing instruction (Graham & Perin, 2007; Graham & Sandmel, 2011; Graham et al., 2023) and creative idea generation in adolescence (Kleibeuker et al. 2016; Stevenson et al., 2014). It should be noted, however, that this result based on overall effect size should be treated with caution, as journals tend to publish studies with significant results more often than those with no effect. Nevertheless, these results show that instruction has a large effect on creative writing quality and a moderate effect on the creative writing process. It is therefore a promising indication that we can teach students to write better creative texts with well-designed instruction. Thus, creative writing, like creative thinking in general, is not just an innate talent, but a skill that can be improved.

Our second question was the extent to which overall effectiveness depended on study variables, research methodology & assessment method, the treatment (focus & content), and the instructional approach. We discuss all four questions for both writing quality and writing processes.

4.1 *Study variables*

We found that none of the study variables contributed to the effect size of the interventions, either for writing quality or for writing processes. However, there is a difference between these findings because we examined six variables for writing quality (e.g., student type, continent, target language, mean age, number of sessions during the intervention, and number of minutes during the intervention), but there were not enough data for writing processes to calculate an effect. The fact that we found no difference between L1 and L2 is consistent with previous research showing that creative writing in L2 classes is effective not only for students' language skills but also for their creative thinking skills (e.g., Dai, 2010; Liao et al., 2018; Wang, 2019; Yeh, 2017).

4.2 *Research methodology & assessment method*

As with the study variables, we omitted moderator analysis for Research Methodology & Assessment Method for writing processes due to the small number of available effect sizes. For writing quality, we found no difference in the effectiveness of research methodology (e.g., research design and study design). Nor did it matter whether the evaluation was holistic or analytical.

4.3 *Treatment*

4.3.1 Treatment focus

For writing quality and writing processes, there was no significant difference between an intervention that focused on process, product, or both. However, when the intervention focused on process, writing quality was more positively affected when instruction focused on the writing process than when instruction did not. Writing processes were more positively affected when instruction focused on the writing process or a combination of the writing and creative processes.

For both the creative process and the writing process, our findings are consistent with previous research. For the creative process, previous research has shown that explicit instruction to build knowledge of creative strategies has a positive effect on students' creative processes (Van de Kamp et al., 2015, 2016). In addition, a recent meta-analysis by Graham et al. (2023) showed that teaching critical/creative thinking skills improved the quality of students' writing.

For the writing process, Graham & Perin (2007) reported an overall effect size of 0.32 for a process writing approach that was significant compared to no process approach. In a recent meta-analysis Graham et al. found an effect size of 0.47 for comprehensive writing programs, which included the process approach (2023). Similarly, Graham & Sandmel (2011) found that for general education students, process-focused writing instruction resulted in statistically significant but relatively modest improvements in overall writing quality, with an overall effect size nearly identical to Graham & Perin's analysis (ES = 0.34). Thus, both analyses showed that writing process instruction positively affected writing quality; our meta-analysis indicates that this also explicitly applies to creative writing quality. When the treatment was product-focused, we found no difference in writing quality for the different text levels (e.g., text level or sentence / lower level), and we were unable to analyze product focus for the writing process due to lack of data.

4.3.2 Treatment content

For writing processes, we could hardly measure the effect of content moderators, either because there was too little data or because the variables were linearly dependent on each other. The only remaining content moderator was whether participants wrote creative texts during the intervention, as opposed to an intervention in which no creative texts were written but, for example, only a model was observed (Groenendijk et al., 2013). However, this did not moderate the effect of instruction on the process. Thus, we must conclude that only a treatment based on the use of fictional texts (e.g., stories or novels) was effective in improving the quality of students' creative writing. However, these findings must be interpreted with caution because they are based on only two studies (e.g., Bartan 2017; Dolmaz & Kaya, 2018).

4.4 *Instruction*

For instruction, we found a significantly larger effect of extensive instruction compared to no instruction on writing quality. For writing process, we found that providing instruction was more effective than no instruction. Although these results are limited, they show that instruction improves both students' creative writing quality and writing processes. For the writing process, however, these results should be interpreted with caution. Although there were four no-instruction effect sizes, they were all from the same study: Groenendijk (2013). In this study, students observed a model and then wrote a short poem. However, rather than having students observe a model without further instruction on the writing process, it seems preferable to provide students with instruction on the writing process if we want to influence this process. Other studies have found observational learning to be effective for argumentative/expository writing (Braaksma et al., 2002; Van Ockenburg et al., 2021), although Groenendijk et al.'s (2013) study did not find observational learning to be effective for creative writing.

The results suggest that providing instruction helps to improve both creative writing performance and process, with text quality improving when that instruction is extensive. What that instruction should consist of was not fully elucidated by this meta-analysis, so the appropriate content still remains somewhat of a black box. However, we did find that writing quality improves when instruction is based on fictional texts and when attention is paid to the writing process. For the writing process, instruction is most effective when it focuses on the writing process or a combination of writing and creative processes. Thus, we can formulate two design principles for improving creative writing quality and one principle if our goal is to affect the writing process:

- 1 If we want our students to improve their creative writing, then ...
 - ... instruction should focus on the writing process.
 - ... we should use fictional texts (excerpts) as stimuli for learning tasks.
- 2 If we want to influence our students' writing process, then....
 - ... instruction should focus on the writing process or a combination of the writing and creative processes.

4.5 *Limitations*

The main purpose of this analysis was to formulate key components of effective instruction to improve the quality of students' creative writing. First, we limited our research to secondary schools. This was our target group because there are hardly any teaching materials available for secondary schools in the Netherlands, yet there is an increasing demand for them. In addition, we limited our analysis to peer-reviewed results, so we may have missed some interesting studies, but an important advantage is that the peer-review process ensures that the studies are of sufficient quality.

Second, it is possible that the small number of studies in our sample (13 studies) contributed to a greater degree of heterogeneity between studies. In addition, our studies varied widely in terms of treatment content and instructional approach. Many studies lacked relevant information about this content and instruction, which is also a possible explanation for the high degree of heterogeneity among these studies.

Third, some instructional procedures and treatment content have been studied more than others. For example, a focus on creative thinking or a focus on the writing process may be an effective way to improve students' creative writing process, but there is currently not enough information in our analysis to draw even preliminary conclusions about their effects. For several treatments, we relied on a very small number of studies, so we have less confidence in the reliability of these average effect sizes. For example, the findings for the fiction-based interventions rely on only two studies (Bartan 2017; Dolmaz & Kaya, 2018) and should therefore be treated with caution.

Fourth, not only this analysis but also the studies we included have limitations. This applies of course to all research, but it was particularly striking in this meta-analysis. Even though we restricted our analysis to peer-reviewed studies, many of the limitations we encountered were due to the studies' mediocre reporting. First, we had to exclude up to six articles that met all our inclusion criteria because we could not calculate an effect size using the data reported in the paper. Second, the content of the intervention was not clearly reported in many of the

studies. For example, of the study information variables, the mean age was not known for seven of the thirteen studies, and we did not know how many teachers participated in the intervention for nine studies. Even more problematic, we barely knew the content of the treatment from two studies, and from these studies we could not code any of the eight instructional variables because they were simply not reported.

The underreporting is most pronounced for treatment content, but also for instructional modality, the very variables that are most important in formulating key components of effective instruction. As many as 15 percent of the content variables were unspecified, and as many as 51 percent of the instructional variables were unspecified. For example, in many studies, we do not know whether students had practice revising their writing and whether scaffolding was used, both of which we know can improve student writing. Similarly, as many as 42% of the study quality indicators were unspecified or noncompliant. Thus, we can conclude that underreporting occurred in the majority of the studies. We return to this in the section with recommendations for future research.

Fifth, we had to make many decisions during the coding process about what exactly the treatment content and mode of instruction consisted of and how to code them, and as Graham and Perin (2007) also emphasized, we too have no doubt that other researchers might question the decisions we ultimately made or would have chosen to code differently. However, like Graham and Perin, we have tried to be as transparent as possible about our process and decisions.

Finally, we also had the problem of unequal control comparisons. We attired to determine the effect of a particular treatment, but the students in the control conditions in our studies were treated differently. In four of the thirteen studies, students received an alternative treatment, while in the remaining nine studies they followed the regular program. This makes it difficult to interpret the results because the studies were set up differently. The same goes for the effect of study design. If only post-test scores are available, the effect size may be larger than if pre-test scores can also be included in the analyses.

4.6 *Recommendations for further research on creative writing in education*

Based on the results presented in this paper, we would like to make several recommendations for further research on creative writing in education. We reiterate that more research is required, as evidenced by the small number of studies-included in our meta-analysis. Although several studies on creative writing in secondary education have recently been added (e.g., Banegas & Lowe, 2021; Ten Peze et al., 2021), the total number of studies on creative writing in secondary

schools remains very limited, especially when compared to the extensive research on improving the quality of student writing in secondary schools. The relative neglect of creativity in language education curricula (Wyse & Ferrari, 2015) does not seem an unlikely explanation.

In addition, it is important that the research that has been done is properly reported. In the present study, we found large heterogeneity between studies, which may be due to differences in the implementation of the interventions. However, because many details of the interventions were not reported, we could not include these moderators as explanatory variables in the analysis. To this end, recommendations for reporting writing interventions in schools have been made by several writing researchers (Graham & Harris, 2014; Koster & Bouwer, 2018; Rijlaarsdam et al., 2017). Koster & Bouwer (2018) emphasize that there are two main reasons to describe writing interventions in detail. First, such a description is essential for scientific purposes such as reliability and theory building. Rijlaarsdam and colleagues also noted that the lack of a complete and accurate description of the intervention, with all teaching and learning activities, "...hampers replication and concurrent studies, theory building, and communication about effective writing instruction" (2017, p. 280). Second, Koster & Bouwer (2018) noted that clear reporting is necessary for dissemination and effective implementation in educational practice. Likewise, Rijlaarsdam et al. (2017) emphasize that underreporting hinders the main goal of educational intervention research: translating effective interventions into classroom practice. After all, if we do not know the principles on which the instruction is based, we cannot replicate it, which means we are looking at a black box, so to speak, and we are thus in the dark about which components work and which do not.

Rijlaarsdam et al. (2017) proposed a scheme for reporting the content and structure of interventions in writing research, which is also ideally suited to future creative writing research. At a minimum, the reporting of such research should consist of a list of learning activities and learning content, with clear reporting of the phases of the instructional sequence in which they occur, and the instructional stimulus associated with them. Bouwer & De Smedt (2018) adapted this scheme into a general framework for describing writing intervention research and made additional recommendations for reporting. We summarize these elements very briefly. At a minimum, the description of an intervention should include: 1. design principles for the instructional focus, 2. design principles for the instructional mode, 3. how the focus and mode are integrated and operationalized into learning activities, 4. the teaching activities that support these learning activities, and 5. the training and guidance that teachers need. We recommend that future creative writing researchers use these findings and adapt their reports

accordingly so that both scholarship and educational practice will benefit from new knowledge. In short, in addition to further research on creative writing in education, we recommend reporting at least those elements that provide good insight into the content of a writing interview.

4.7 *Recommendations for educational practice*

Based on our findings, we conclude that instruction contributes to improving the quality of students' writing and the writing process in secondary schools. This is promising for improving a school environment that pays attention to types of texts beyond the standard communicative text. When attention is paid to creative writing, students actually become better at it. This not only improves their writing quality, but also positively affects their writing process.

Our ultimate goal for the present study was to formulate design principles for future interventions based on the content of effective interventions. Based on our analysis, and returning to our if-then statements (Reigeluth, 1999; Van den Akker, 1999), we formulated two design principles for improving creative writing quality and one principle if the goal is to affect the writing process:

- 1 If we want our students to improve their creative writing, then ...
 - ... instruction should focus on the writing process.
 - ... we should use fictional texts (excerpts) as stimuli for learning tasks.
1. If we want to influence our students' writing process, then ...
 - ... instruction should focus on the writing process or a combination of the writing and creative processes.

Because our meta-analysis allowed us to articulate only two key components of effective creative writing instruction, further research on this topic is obviously still needed and most welcome. We emphasize that not only the amount of research is currently limited, but the quality of reporting often leaves much to be desired. However, this reporting is crucial for determining exactly how to help students develop their creative writing skills. We believe that knowing more about the effective teaching of creative writing and about how to support the creative writing process, will help teachers to improve the teaching of writing in their schools and to give creative writing the attention it deserves.