Learning from comPA(I)Ring exemplars: Enhancing genre knowledge of argumentative texts

Tine Mombaers, Roos Van Gasse & Sven De Maeyer

University of Antwerp | Belgium

Abstract: Poor writing skills are problematic in today's society where writing expertise is essential in personal, academic and professional contexts. Students struggle most with argumentative writing. To write a good argumentative text, students need genre knowledge on this type of text. After all, genre knowledge has been proven to be related to writing quality. Considering its relevance, in this study we investigated whether learning from (comparing) text exemplars could be an effective method to enhance genre knowledge. This study aims to investigate whether learning from (comparing) text exemplars can enhance genre knowledge. A quasi-experimental study with 77 11th grade students was carried out to test the effects of four conditions on genre knowledge of argumentative texts. Findings show that genre knowledge increases through single and analogue text examples. In addition, learning from comparing text exemplars does not seem to increase genre knowledge more than learning from single, sequential exemplars.

Keywords: learning from (comparing) exemplars, genre knowledge, argumentative texts, comparative judgment

journal of WRITING RESEARCH

Mombaers, T., Van Gasse, R., & De Maeyer, S. (2024). Learning from comPA(I)Ring exemplars: Enhancing genre knowledge of argumentative texts. *Journal of Writing Research, volume*(issue), 163-193 DOI: https://doi.org/10.17239/jowr-2024.16.01.06

Contact: Tine Mombaers, University of Antwerp, Sint-Jacobstraat 2, 2000 Antwerp | Belgium tine.mombaers@uantwerpen.be. ORCID: 0000-0002-2545-555X

1. Introduction

In today's society, good argumentative writings skills are essential in academic, personal and professional contexts (Lee & Deakin, 2016; Pessoa et al., 2017). After all, writers' ability to analyze, compose, and judge an academically sound argument represents a key component of academic success (Muller Mirza & Perret-Clermont, 2009; Newell et al., 2011). Argumentative writing helps in developing thinking skills (Kuhn & Crowell, 2011), influencing others and it allows the debate of controversial issues next to promoting greater participation in social and democratic processes (Ferretti et al., 2009). In addition, students need good argumentative writings skills to be prepared for the modern workplace (Ferretti & De La Paz, 2011).

Though its importance and relevance has been proven, students have difficulties with this type of writing (Bacha, 2010). Students' argumentative writing proficiency appears to be poor in secondary school and in higher education (Graham & Perin, 2007; NCES, 2012; Ferretti & Lewis, 2013; Song and Ferretti, 2013).

Students struggle with recognizing and applying argumentative text structures (Chambliss & Murphy, 2002; Freedman & Pringle, 1984), generating evidence (Kuhn, 1991), offering relevant reasons (McCann, 1989) and producing counterarguments (e.g., Perkins et al., 1991; Stapleton, 2001; Nussbaum & Kardash, 2005). The lack of integrating counterarguments leads to poorly developed arguments that fail to consider alternative viewpoints. This tendency to consider only the side of the issue favored by the student was labeled as myside bias by Perkins et al. (1991).

But how can students know what the characteristics of a good argumentative text are and thus understand the importance of counterarguments or know what the correct text structure is for instance? This is where genre knowledge comes into play. Genre knowledge is the knowledge of the aim, content, structure (Martin, 2009), and the language style (Hyon 2001, 2002) of a specific text genre. Being able to understand, engage and shape genres is important in developing writing expertise (Bazerman, 1997; Miller, 1984). This theory was empirically proven by the study of Olinghouse et al. (2015) who found that genre knowledge was significantly correlated with writing quality and the use of genre elements in argumentative texts with 5th graders. These findings imply that genre knowledge can play an important role in writing qualitative argumentative texts.

Several empirical studies emphasize the importance of genre knowledge early on in elementary school and also in secondary school (e.g., Bigger, 2022; Olinghouse & Graham, 2009; Olinghouse et al., 2015) or intend to increase genre knowledge, within the elementary, university and doctoral context (e.g., De Smedt & Van Keer, 2018; Humphries Sandstrom, 2021; Yasuda, 2011);Humphries Sandstrom, 2021). These studies on increasing genre knowledge aim to enhance genre knowledge to improve writing quality. Hence, they measure writing quality, without including a pre- and post-test on genre knowledge. Their focus is on

enhancing genre knowledge as a means to improve writing quality, not on investigating genre knowledge separately. Investigating how genre knowledge in itself can be increased, may make an important contribution to the field.

In the field of cognitive psychology, learning from exemplars emerges as a promising way to generate conceptual knowledge change (Kruschke, 1992), which could be vital to increase people's writing skills. The use of exemplar comparisons in the classroom has been proven to be a strong predictor of general learning gains (Alfieri et al., 2013). However, learning from comparing designated text exemplars has hardly been investigated in educational writing research.

Generally, learning from exemplars can take two forms: (1) presenting people with single exemplars, and (2) presenting them exemplars in pairs (Gentner, 1983). These pairs can be analogue, 'near miss' or contrastive. Most research attention has been paid to comparisons of exemplars that are very similar, i.e. analogue comparisons (e.g., Gentner et al., 2009; Alfieri et al., 2013). However, there is evidence that comparison can not only aid in abstraction, but also in differentiation (Smith & Gentner, 2014). Therefore, comparing two near miss exemplars, which are identical except for a crucial structural difference and comparing contrastive exemplars, which have some overlap and many differences, are also effective to enhance learning (Hammer et al., 2008; Smith & Gentner, 2014).

While comparing exemplars is considered more effective in different research domains than presenting students with single exemplars (Alfieri et al, 2013), it is not clear if this also applies to writing. In addition, to the authors' knowledge, research investigating the effect of the three types of comparisons in one single study does not currently exist.

Since there are no studies available on how to best enhance students' genre knowledge of argumentative texts, this study aims to investigate to what extent learning from exemplars improves students' genre knowledge of argumentative texts and what types of comparisons are most effective.

2. Literature Review

2.1 Argumentative Writing

Argumentative writing involves identification of a claim, supportive evidence (empirical or experiential), and the assessment of warrants that connect the claim, evidence and situation constituting an argument (Newell et al., 2011). In order to be persuasive, an argumentative essay must contain a good surface structure by including alternative viewpoints and showing their weaknesses, but it must also support claims with excellent quality reasons that convince others (Kuhn, 1999, 2005; Stapleton & Wu, 2015).

Previous research shows that students find it hard to recognize and to apply argumentative text structures (Chambliss & Murphy, 2002), to generate evidence

(Kuhn, 1991), and to provide relevant reasons, counterarguments, and rebuttals (McCann, 1989). Students have the tendency to ignore arguments opposing those of the writer, which is termed "myside bias", first named by Perkins (1985), and this phenomenon has been widely discussed and described by many other researchers after him (e.g., Baron, 1995; Nussbaum & Kardash, 2005; Wolfe, 2012). Several studies report that students and adults struggle with behaving strategically in argumentative communication, as they are likely to use few counterarguments, and do not adapt what they say to respond to adversaries (Felton & Kuhn, 2001; Kuhn & Udell, 2007). Moreover, students tend to neglect information that does not reinforce their point of view (Perkins et al., 1991). This tendency to not include counterarguments is unfortunate, since two-sided messages (i.e., messages including counterarguments and rebuttals) are more persuasive than those that disregard counterarguments (O'Keefe, 1999).

Myside bias can be explained by several possible reasons, such as a high cognitive load (Coirier et al., 1999), the desire to sustain cognitive consistency (Simon & Holyoak, 2002) and the fact that students often do not know that considering and rebutting another viewpoint increases the persuasiveness of their argument (Nussbaum & Kardash, 2005).

Despite the difficulties that students endure, several studies proved that argumentative writing can be taught and improved (e.g., Granado-Peinado et al., 2019; Nussbaum & Schraw, 2007; Prata et al., 2019).

The structure of good argumentation can be found in Toulmin's work (1958, 2003), who has been frequently praised and widely cited for his framework that was able to isolate the elements of good argumentation: claim, data, warrant, backing, qualifier, and rebuttal. In Toulmin's work, *claim* is the initially stated conclusion; the *data* are the facts that underpin the claim; the *warrants* provide connections between the data and the claim; *backings* state the assumptions on which the warrants are based; *qualifiers* limit the strength of the claim and *rebuttals* are arguments that refute or form exceptions to the elements of the argument. This framework and its variations have been used in many studies and in teaching to determine the strength of students' ability to argue (Qin & Karabacak, 2010; Stapleton & Wu, 2015). Though Toulmin's framework has been crucial in emphasizing the need to consider alternative positions, it has received some criticism from other researchers in the field of argumentation. First, Sampson and Clark (2008) criticized the overemphasis on structural elements of argumentation, sacrificing the quality of logic and evidence. Second, several researchers have encountered difficulties applying the Toulmin model, because the arguments students write can often be allocated to more than one element (Sampson & Clark, 2008; Simon, 2008). Third, Nussbaum and Kardash (2005) questioned the complexity of the model, which is often redundant for analytic purposes. Because of these concerns about the complexity and usability of the model, a number of researchers

adapted or simplified the Toulmin framework to ensure reliable classification of argumentative elements (e.g., Nussbaum & Kardash, 2005; Qin & Karabacak, 2010; Stapleton, 2001; Stapleton & Wu, 2015).

An example of a study in which the authors simplified Toulmin's framework is the research of Stapleton and Wu (2015). They studied the perceived quality of reasoning in students' essays to explore the (lack of) connections between two aspects of argumentation, namely surface structure (e.g., claims, counterarguments, etc.) and substance (quality of reasoning). To determine how well argumentative structure was implemented in students' writing, the researchers developed a rubric that followed a modified Toulmin model based on Nussbaum and Kardash (2005), Nussbaum and Schraw (2007), and Qin and Karabacak (2010). Three main argumentation elements were included in this rubric: claim, counterclaim and rebuttal – each accompanied by their supporting reasons or associated data. The quality of the substance of the argumentative elements was judged by PhD students in Education (acting as proxies for teachers, contrasting the many studies in which researchers themselves assessed the students' texts). This study's findings show that in spite of a good surface structure, many claims and data written by the students were considered weak, which implies that good reasoning cannot always be associated with good surface structure. As a result, Stapleton and Wu (2015) created an Analytic Scoring Rubric for Argumentative Writing (ASRAW) that can be used to assess both argumentative structural elements and reasoning quality. On the one hand, the rubric is constructed with the surface structure based on Toulmin-like elements in the rows, such as claim, claim data, counterargument claim, counterargument claim data, rebuttal and rebuttal data. On the other hand, the quality of the supporting reasons is described in the columns (see Figure 1). Since its development, it has often been used by other researchers to assess or score argumentative texts in their studies (e.g., Abdollahzadeh et al., 2017; Mohsen & Qassem, 2020; Allagui, 2021).

This rubric holds a lot of specific elements of argumentative writing, so it will be a useful tool to use to look for genre elements that students incorporate in their argumentative writing.

Analytic Scoring Rubric for Argumentative Writing (ASRAW)

1.Claim(s) (5%)	Score: 5			Score: 0	
	States point(s) of view			Does not state point(s) of view	
2. Data (25%)	Score: 25	Score: 20	Score: 15	Score: 10	Score: 0
	a. Provides multiple reasons for the claim(s), and b. All reasons are sound/acceptable and free of irrelevancies	a. Provides multiple reasons for the claim(s), and b. Most reasons are sound/acceptable and free of irrelevancies, but one or two are weak	a. Provides one to two reasons for the claim(s), and b. Some reasons are sound/acceptable, but some are weak or irrelevant	a. Provides only one reason for the claim(s), or b. The reason provided is weak or irrelevant	a. No reasons are provided for the claim(s); or b. None of the reasons are relevant to/support the claim(s)
3. Counterargument Claim(s)/Alternative Point(s) of View(10%)	Score: 10			Score: 0	
	Provides counterargument claim(s)/alternative view(s)			Does not provide counterargument claim(s)/alternative view(s)	
4. Counterargument Data/Supporting Reasons for Alternative Point(s) of View(25%)	Score: 25	Score: 20	Score: 15	Score: 10	Score: 0
	a. Provides multiple reasons for the counterargument claim(s) /alternative view(s), and b. All counterarguments/reasons for the alternative view(s) are sound/acceptable and free of irrelevancies	a. Provides multiple reasons for the counterargument claim(s)/alternative view(s), and b. Most counterarguments/reasons for the alternative view(s) are sound/acceptable and free of irrelevancies, but one or two are weak	a. Provides one to two reasons for the counterargument claim(s) /alternative view(s), and b. Some counterarguments/reasons for the alternative view(s) are sound/acceptable, but some are weak or irrelevant	a. Provides only one reason for the counterargument claim(s)/alternative view(s), or b. The counterargument/reason for the alternative view is weak or irrelevant	a. No reasons are provided for the counterargument claim(s)/alternative view(s); or b. None of the reasons are relevant to/support the counterargument claim(s)/alternative view(s)
5. Rebuttal Claim(s) (10%)	Score: 10			Score: 0	
	Provides rebuttal claim(s)			Does not provide rebuttal claim(s)	
6. Rebuttal Data[a] (25%)	Score: 25	Score: 20	Score: 15	Score: 10	Score: 0
	a. Refutes/points out the weaknesses of all the counterarguments, and b. All rebuttals are sound/acceptable c. The reasoning quality of all the rebuttals are stronger than that of the counterarguments	a. Refutes/points out the weaknesses of all the counterarguments, and b. Most rebuttals are sound/acceptable, but one or two are weak c. The reasoning quality of most rebuttals are stronger than that of the counterarguments, while one or two are equal to that of the counterarguments	a. Refutes/points out the weaknesses of all the counterarguments, and b. Some rebuttals are sound/acceptable, but some are weak c. The reasoning quality of some rebuttals are stronger than that of the counterarguments, while some are weaker than that of the counterarguments	a. Refutes/points out the weaknesses of some counterarguments, or b. Few of the rebuttals are sound/acceptable; most of them are weak, or c. The reasoning quality of most rebuttals are weaker than that of the counterarguments	a. No rebuttals are provided; or b. None of the rebuttals can refute the counterarguments

Note. [a]An implicit requirement of rebuttal data is subsumed under the requirements of row 4 "Counterargument Data" , that is, each piece of rebuttal data should be aligned with each piece of counterargument data in terms of both quantity and logic.

Figure 1: Analytic Scoring Rubric for Argumentative Writing (ASRAW).

2.2 Genre knowledge

McCutchen (1986) defined genre knowledge as the characteristics of a specific genre. More specifically, genre knowledge is the knowledge of the aim, content and structure of a specific text genre (Martin, 2009). Next to these three elements in the definition of genre knowledge, Hyon (2001, 2002) also included language style (i.e., linguistic signals that shape a genre) to operationalize genre knowledge. Next to its core definition, the lens through which we will see genre knowledge in this study, genre knowledge has a relative importance towards writing contexts. Genre knowledge can be identified as one of the mental schema that writers invoke when they analyze new writing tasks in new contexts, referring to a conceptual frame that can bridge rhetorical and social knowledge (Beaufort, 2007; Tardy, 2009). Therefore, genre knowledge can aid students in recognizing and adapting more effectively and critically to new writing contexts (Miller, 2009).

Genre knowledge has a positive influence on writing quality. This assumption was first described in theoretic contemplations of several authors (e.g., Donovan & Smolkin, 2006; Gillespie et al., 2013; Olinghouse & Graham, 2009; Saddler & Graham, 2007). Students who have more knowledge of the basic genre elements or are more familiar with a genre can use this knowledge to define the writing assignment, guide and retrieve relevant information (such as ideas and vocabulary), and confirm the appropriateness of the retrieved ideas. This should lead to qualitatively better papers with appropriate genre elements (Olive et al., 2009; Donovan & Smolkin, 2006; Saddler & Graham, 2007).

These theoretic assumptions were, later on, confirmed by empirical research.

Olinghouse et al. (2015) found that knowledge of genre elements was statistically and positively related to writing outcomes in three writing genres, namely narrative, informative and persuasive writing. The authors showed that genre knowledge had a positive effect on students' (holistic) writing quality and also on the number of genre elements in their persuasive, story and informational texts. The positive influence of genre knowledge on argumentative writing was confirmed later on by Ferretti and Lewis (2019). Students who possessed more knowledge about persuasion and persuasive writing wrote higher quality persuasive essays than those with less knowledge. Furthermore, research has shown that teaching genre elements (within strategy instruction) has a positive effect on the quality of student writing (Graham et al., 2012).

Over the past few decades, numerous studies have underscored the significance of genre knowledge (e.g., De Smedt et al., 2018; Graham et al., 2012; Olinghouse et al. 2015). Consequently, it proves valuable to explore effective methods for its improvement.

2.3 Learning from exemplars / comparisons

In cognitive psychology, a lot of effort has been made to investigate how people develop schemata or conceptual knowledge. A promising way to do so is learning from exemplars (Alfieri et al., 2013; Gentner, 1983). Exemplars are typical examples or appropriate models for learning. The idea is that learners are presented with exemplars of different quality. Subsequently, aspects of the exemplars are assessed by learners in terms of quality. Convinced of its opportunities in education, many researchers have started to examine the potential benefits of explicit case comparisons for academic learning across a variety of contexts (Alfieri et al., 2013). These studies include research in science, such as problem solving within Physics (Gadgil & Nokes, 2009), recognizing the process of heat transfer (Kurtz et al., 2001) and assessment in student learning (Nagarajan & Hmelo-Silver, 2006). Moreover, comparisons are also investigated in the field of Math (comparing solution procedures by Rittle-Johnson & Star, 2007) and business contract negotiation between disputing parties (Gentner et al., 2004). These studies prove that learning from comparisons can be an adequate manner to acquire more complex skills, implying that genre knowledge could also be enhanced by this type of learning.

Learning from exemplars can take two forms: presenting single exemplars (whether or not sequentially) and presenting exemplars in comparisons (Alfieri et al., 2013; Gentner, 1983; Rittle-Johnson & Star, 2009). Different knowledge domains (e.g., problem solving or the negotiation of contracts) found that people learn more from comparing exemplars than from processing single exemplars (see meta-analysis of Alfieri et al., 2013). The underpinning of these findings can be found in Gentner's (1983) structure-mapping theory. This theory states that the salience of common aspects is increased by comparing exemplars resulting in abstraction, the

learner is invited to infer from one exemplar to the other and the alignable differences are highlighted.

Building upon these theories, the first hypothesis in this study is: *Learning from (comparing) exemplars enhances genre knowledge (H1)*.

Structure-mapping theory has led to research in comparing exemplars focusing on two forms: learning from comparing analogue exemplars (Alfieri et al., 2013) (cases that belong to the same category) and more recently learning from comparing contrasting exemplars (cases that do not belong to the same category) (Rittle-Johnson & Star, 2009; Smith & Gentner, 2014). Category abstractions are inferred/built through aligning common features and recognizing interrelations (Gentner & Namy, 1999; Namy et al., 2007). These category abstractions are called schemas and these schemas are cognitive representations of the structures of relational systems shared by cases/exemplars that have been highlighted through structural alignments (Gentner, 1983, 2010: Gick & Holyoak, 1983; Markman & Gentner, 1993; Namy & Gentner, 2002). For example, Namy et al. (2007) showed that 4-year-olds learned better when provided with two exemplars (two hats) for a given label than with one. In the post-test, the children in the two-exemplar condition could extend the label to a sombrero (a type of hat) than to an igloo (similar in shape and color, but not function). Thus, what students learn from comparisons goes beyond a surface (perceptual/object) representation and leads to a deeper relational understanding of the category (conceptual) (Alfieri et al., 2013).

Cognitive psychology research on learning from comparing exemplars has mainly focused on learning from analogue exemplars (Smith & Gentner, 2014). In their meta-analysis, Alfieri et al. (2013) demonstrated that learning from analogue comparisons clearly outperforms learning from other learning situations (such as single cases, traditional instruction, sequential study task, etc.) (medium effect size) and that this effect depends on the use of supportive prompts that were given to the learner (see further). Based on this empirical evidence, the following hypothesis is put forward: *Learning from analogue exemplars improves genre knowledge (H2)*. Hammer et al. (2008) theoretically underpinned that learning from contrasting exemplars induces different cognitive processes than learning from same-class or analogue comparisons. Moreover, learning from contrasting exemplars also depends on the degree of difference between the exemplars: it is most effective when the exemplars are rather equivalent, except for a key aspect that is necessary to discriminate between both categories (Hammer et al., 2008), also called 'near miss' exemplars (Smith & Gentner, 2014). When comparing 'near miss' exemplars, the comparison becomes more self-aligning: crucial aspects become salient to the learning itself when aligning the similarities and differences between exemplars (Smith & Gentner, 2014). We can thus hypothesize that *learning from near miss exemplars improves genre knowledge (H3)*.

Nevertheless, as Smith and Gentner (2014) argued, much category learning involves learning from pairs that show moderate similarity, with some overlap and many differences (further named as 'contrastive' exemplars). Consequently, not all important distinctions can be illustrated with 'near miss' pairs. Therefore, the next hypothesis can be formulated: *Learning from contrastive exemplars improves genre knowledge (H4)*.

The different types of comparisons described above benefit from divergent prompts that are most effective for students when comparing. For analogue comparisons, according to Alfieri et al. (2013) letting learners focus on similarities is the most effective, as it helps students focus on finding the critical features of the cases. Identifying both similarities and differences is still effective, but to a lesser extent. Finally, only identifying differences is found to be contra-productive. Focusing on differences only could underline the superficial similarities that are irrelevant for finding the target features. Moreover, focusing on superficial differences may also lead to a higher cognitive load compared to focusing on only similarities, which could potentially make it more difficult for students to encode the common features of the target content. Gentner and Gunn (2001) found that people were more likely to observe differences (typically, alignable differences, therefore applicable to near miss comparisons) between a pair of concepts if they had previously listed commonalities for that pair. This is also the case for contrastive exemplars since Smith and Gentner (2014) stated that more critical differences were discovered after focusing on commonalities.

Comparison is found to be critical for relational abstraction (Christie & Gentner, 2010), but it is also possible to learn from single exemplars that are presented to students sequentially. Though literature on learning from single exemplars is not very extensive, we know that both adults and children can align sequentially presented examples (e.g., Childers, 2008; Reed; 1987; Ross & Kennedy, 1990). Such learning depends on the learner's ability to align the sequential representations (Christie & Gentner, 2010).

Previous studies have used single exemplars that were presented sequentially instead of using a control group (e.g., Christie & Gentner, 2010; Gentner & Gunn, 2001; Rittle-Johnson & Star, 2009). According to Rittle-Johnson and Star (2009) sequential conditions are favorable to control groups because they differ from the comparison conditions only by being studied in succession. In these sequential conditions, students are not asked to compare both texts. Though we suspect that students can also learn by studying single text exemplars sequentially, we expect that students will learn more from comparing exemplars than studying single exemplars. Therefore, the following hypothesis is put forward: *Learning from comparing exemplars enhances genre knowledge more than learning from single, sequential exemplars (H5)*.

3. This study

In the current study, we examine the effect of looking at or comparing text exemplars on genre knowledge of argumentative texts. To the authors' knowledge, this is the first study to investigate the effects of learning from (comparing) text exemplars within writing education.

Unlike previous studies, we will not contrast one type of comparison with looking at single, sequential exemplars. This will be the first study in the field of learning from (comparing) exemplars that will compare all three types of comparisons (analogue, near miss and contrastive) and sequentially studied exemplars in one study, based on students' outcomes. In this case, students' outcomes represent their level of genre knowledge of argumentative texts. ASRAW text elements, as described in the literature review, will play an essential role in the scoring of genre knowledge, as these elements are found to be critical in defining an argumentative text according to literature on the subject.

4. Methodology

An intervention study with four conditions (single, analogue, near miss and contrastive) in an authentic classroom setting was set up to test the hypotheses above. Data collection took place in October and November 2021 during school hours, in the students' classroom.

4.1 Participants and setting

Through power analysis the desirable number of participants was calculated through G*Power (Faul et al., 2007). With an effect size of 0.20, four groups and two measurements, the sample size should consist of 76 participants to obtain a power of 0.80.

The attainment goals imposed by the Flemish Ministry of Education on argumentative writing are situated within the curriculum of 11th and 12th grade students in general and technical education in Belgium. Therefore, 11th grade students in these education levels were selected to participate in this study. Students from two catholic schools in Flanders, the Dutch-speaking region of Belgium, participated in the study. School A was a school with technical study profiles, whereas school B provided more general education (see Table 1). Selecting schools with different study profiles ensured heterogeneity in the sample. We ran a pilot to verify appropriate timing and to ensure cognitive validity of the materials used. The pilot was conducted in one compiled class group in school A of 16 students studying Construction Techniques, Agriculture, and Horticulture.

Due to the COVID19 pandemic, reaching the required number of participants for the main study was challenging. By selecting an extra class group to fill in for students who were willing to participate, but who were not at school at the day of

data collection (due to illness or quarantine related to COVID19), the required number of participants was reached.

As a result, 77 eleventh grade students from different fields of study participated in the main study, as shown in table 1. 43,59 % of them were girls and 56,41 % of them were boys. Their age ranged from 15 to 17 years old.

Students were randomly assigned to the four conditions within each class group: single exemplars (N= 20), analogue (N= 21), near miss (N= 20) and contrastive comparisons (N= 17). The booklets for students were randomly distributed to the students' desks before they entered the classroom. When the students came in, they could choose where to sit. Within each class group, we strived to equally divide the students into the four conditions.

Table 1. Overview of study participants and their field of study

Field of study	n students / condition				total n students	school
	single	analogue	near miss	contrastive		
Carpentry	3	3	4	4	14	A
Electrical Techniques	2	2	2	1	7	A
Biotecnical Sciences	2	2	1	2	7	A
Industrial Sciences	3	3	2	1	9	A
Modern Languages & Science	0	4	3	1	8	B
Humanities	5	1	3	4	13	B
Sciences & Mathematics	5	6	5	4	19	B
TOTAL					77	

Note. Carpentry and Electrical Techniques formed one class group. The same applies for Humanities and Sciences & Mathematics.

4.2 Procedure

Preparing text material

For this study, we used a sample of 165 texts from another argumentative writing study in Flanders, conducted by researchers at Ghent University (Landrieu et al., 2022).

The 22 texts that were used to create exemplars for the intervention were purposefully selected (based on their analytical score and holistic ranking) and meticulously manipulated. Certain texts were manipulated to try to avoid students being distracted by, for example, spelling, style, structure, word choice, etc. Texts

of varying quality were used since we also needed texts that did not include all genre elements to be compared by students.
For the single exemplars condition, texts with an analytic score of > 65% (mean score = 73,33%) and holistic rankings between 9 and 95 (out of 165 texts) were selected. Texts for the analogue selection had a total analytical score of 70% or 75% (mean score = 74.17%) and scores on the different ASRAW aspects were very similar. Holistic rankings for the texts in the analogue condition ranged from 1 to 85. Selection for the near miss condition contained texts with holistic ranking varying from 15 to 130 and analytical scores from 45% until 75% (mean score = 64.17%), with a clear difference in scoring of one or two ASRAW components (claim data, rebuttal claim and data and counterargument claim and data). Texts with very divergent scores on ASRAW aspects were selected for the contrastive condition as we wanted these texts to strongly contrast. Their total analytic scores ranged from 30% till 85% (mean score = 54.17%), with holistic rankings from 18 to 158. The mean analytic scores of the weaker texts in the near miss and contrastive condition were respectively 55% and 33.3%. The better texts in these conditions had a mean analytic score of 73.3% (near miss) and 75% (contrastive).

Appendix A provides a more detailed description of the selection and manipulation of the text exemplars.

Instrument to test genre knowledge

An instrument that consisted of a pretest and a posttest for genre knowledge was developed, since genre knowledge is central to this study as a dependent variable. To measure genre knowledge, students were asked in a 10-minute pre-test to give advice to a friend to write a good argumentative text. Giving advice to a friend who is a novice in a certain writing genre, has been proven to be an effective method to measure students' writing knowledge (Schoonen & de Glopper, 1996) and has been used in several studies (e.g., van Drie et al., 2021; Bouwer & Koster, 2016).

The following instruction was given: *"Suppose your best friend has to write an argumentative text for school, but he/she has never done this. You give him/her advice about what genre elements must be present in his/her text, so he/she is able to write a good argumentative text.*

What genre elements should a good argumentative text contain in your opinion? Write down as many genre elements (of an argumentative text) as possible so you can deliver this list to your friend." Genre elements were explained to the students: "Genre elements are characteristics of a certain type of text (=genre) that are very typical/characterizing for that type of text (in this case an argumentative text)." Students did not have to know the correct words for these genre elements; a good description of a counterargument, for example, was also scored as an ASRAW element. An example of the scoring of one of the students' pretest can be found in Appendix C.

Consecutively, students were asked to write an argumentative text about keeping animals in a zoo. They received the following instruction: "*a) Write an argumentative text about (not) keeping animals in zoos. Take a position (for or against) and argue this position. You can use the draft paper to write a draft version. Your final text is minimum one page in length. b) You get two source texts. You can select possible relevant information for your own text from these source texts. You are also allowed to use your own ideas and visions in your text.*" Students had 45 minutes to complete the text.

In this study, the argumentative texts that students wrote will be merely used as a reflection tool for students when they reflect upon genre elements that they have used themselves (see further below).

After the intervention in which students received text exemplars to study, students received a posttest about genre knowledge. First, students were asked a very general question: *"What have you learned from looking at/reading these six texts?"* in the single condition and *"What have you learned from making these comparisons?"* in the comparison conditions.

Second, students were asked whether, after comparing the texts, they would change the advice they gave to their friend. Students got their pretests back to answer this question. If their answer was positive, they were then asked which genre elements they would add to their list, and/or which genre elements they would adapt or delete. Third, students were asked which genre elements that they gave as advice, could be found in their own argumentative text. Finally, students were asked what changes they would make to their own argumentative text to improve it. These last two questions were added to collect implicit genre knowledge. Students might not have thought of particular elements as 'genre knowledge' and thus would not write down these genre elements in their adjusted advice. But when going through this data, there were barely any students whose answers to these two questions contained more information than in the adjusted advice. So, the answers to those questions, were not included in the analyses. Time allocation for reflection and posttest was 15 minutes.

The bundles that were used for each condition are freely accessible on OSF in Dutch (https://osf.io/pjn9k/) and English (https://osf.io/nd5s9/).

Intervention

As mentioned earlier, students were randomly assigned to four conditions: single exemplars, analogue comparisons, near miss comparisons and contrastive comparisons. As the researcher conducted the intervention herself, the intervention in the different class groups took place on different days and different school hours. Duration of the intervention was 30 minutes.

Before students came into the classroom, the bundles that they had to fill in were randomly distributed throughout the desks, with each condition in a different

color. Students could choose where to sit. Before having a look at and reading the six texts, students were instructed to particularly pay attention to genre elements and not to give heed to text length, spelling, sentence structure and paragraph structure. These instructions were repeated with each text (pair). That way, we hoped that they would focus more on genre elements while looking at or comparing the texts. The questions that students received for different types of text comparisons, were based on previous literature on learning from exemplars (Alfieri et al., 2013; Gentner & Gunn, 2001; Smith & Gentner, 2014).

In the first condition, single exemplars, students were asked to read each text. After reading each text, students were asked the following two questions: *"What is good about this text?"* and *"What is not (so) good about this text?".*

Students in the second, analogue condition were asked after each comparison of a text pair: *"What similarities do you see in both texts?",* because previous research shows that letting students focus on similarities in analogue exemplars is most effective (Alfieri et al., 2013).

In the third and fourth condition, near miss and contrastive comparisons, students received two questions after comparing each pair: "*What similarities do you see in both texts*?" and "*What differences do you see between both texts*?" These questions have proven to be most effective for these types of contrasting comparisons (Gentner & Gunn, 2001; Smith & Gentner, 2014).

The first author conducted all data collections. Participants had 100 minutes to complete all the assignments (pretest, intervention and posttest).

All essential treatment fidelity aspects described by Sanetti et al. (2021) in order to conduct a good quality intervention, were complied with in this research. These essential aspects include adherence, consistency, dosage and exposure. Regarding adherence, the intervention steps were implemented as planned. The first author designed the intervention and also followed the intervention protocol rigorously while implementing the intervention, resulting in a very consistent intervention. All requirements with regard to dosage (i.e., the frequency with and duration of the intervention is delivered) were met since all class groups exactly received 100 minutes for the assignments and all class groups participated in a one-time intervention only. As a result, recipients in all conditions received the one-time intervention for the same duration, which met the exposure requirements. During the intervention, participants were able to ask the first author questions if they did not understand something that was in the assignments, so she was sensitive to students' questions and remarks. Moreover, the first author paid a lot of attention to class management to make sure that every participant could fully focus on the task at hand. Hence, we can say that the intervention steps were well implemented, resulting in an intervention of good quality (Sanetti et al., 2021).

A second researcher and the students' Dutch teacher were present for logistical and practical assistance only.

4.3 Instruments

Pre and posttest of genre knowledge were scored, considering different categories of genre knowledge (see Table 2 and for a more extensive overview with examples see Appendix B, C and D). These categories and their scoring weights were developed in close collaboration with an expert in argumentative writing.

ASRAW elements were scored the highest because these elements are critical in argumentative texts. Claim and claim data received a score of 3 points, whereas counterargument claim & data and rebuttal claim & data were given a 4 because these aspects were expected to be more difficult to detect by the pupils. Apart from the ASRAW elements, other genre elements were also incorporated (including reinforcing argumentation, general text structure, IME structure (Introduction-Middle or Body-Ending), language use, text goal).

The first author scored all students' pre and post-tests on genre knowledge. The scores of each student for each category of genre knowledge was summed up to get a single score of genre knowledge for each student.

In the posttest, students could add genre elements, make changes to the elements that they had already listed and/or delete genre elements. Adding more correct elements meant an increase in their posttest score compared to their pretest. Students' changes to genre elements could result in an increase of decrease of their score compared to their pretest. Deleting genre elements that were not correct in their pretest did not affect their score, since they did not receive any points for these elements in the first place. But deleting correct genre elements, resulted in a decrease of their score. An example of the scoring of a posttest compared to the pretest can be found in Appendix D.

The expert who collaborated in setting up this scoring instrument coded and scored the pre and posttest of 10 students. Interrater reliability was calculated through Cohen's kappa, which was 92%, indicating a high level of agreement (Cohen, 1960).

Table 2. Categories of genre knowledge and their scores

Category	Score in points
ASRAW elements	3 or 4
reinforcement of argumentation	2
text goal	1
IME structure	2
general text structure	2
language use	1

4.4 Analyses

The data were statistically analyzed through a two-way repeated measures ANOVA, including post-hoc tests, using RStudio (R Core Team, 2021). We used the packages dplyr, lsr, ggplot2, emmeans, ggpubr, reshape, rstatix, tidyverse (Wickham et al., 2022; Navarro, 2015; Wickham, 2016; Lenth, 2022; Kassambara 2020; Wickham, 2007; Kassambara, 2021; Wickham et al., 2019). The data files and R code used to perform the analyses in this study, can be found on OSF (https://osf.io/d73z6/ for the data files and https://osf.io/86ucb/ for the R code).

In a set of preliminary analyses, we tested out if our data met the required assumptions to carry out regular repeated measures ANOVA. First, we checked for extreme outliers by using the Shapiro test and boxplot methods. No data was deleted since there were no extreme outliers present. Second, we looked at the distribution of the data. Visual inspection of the QQ-plots indicated that the data did not deviate from a normal distribution. As there were no extreme outliers and the data was normally distributed, regular repeated measures analyses could be performed on the data.

The repeated measures ANOVA was conducted to identify possible differences between conditions regarding genre knowledge with the factor time (pre versus posttest) as the within subjects variable and condition (single, analogue, near miss and contrastive) as the between subjects factor. Descriptive statistics of the variables can be found in table 3. Figure 2 shows boxplots for conditions across measurement times.

5. Results

Table 4 presents the mean scores of genre knowledge at pre and posttest and their standard deviations for each condition, as well as the outcomes of the repeated measures analyses.

The repeated measures analyses revealed that the measurement occasion effect was significant, F (1, 148)=20.35; p=.000, with a small effect (partial 2 = 0.12). This means that across conditions, students' posttest improved compared to their pretest. There was no significant condition effect found for the four conditions.

Table 3. Descriptive statistics

variable	mean	sd	min.	max.
genre knowledge pre	8.74	4.12	0	18
genre knowledge post	12.01	4.85	3	24

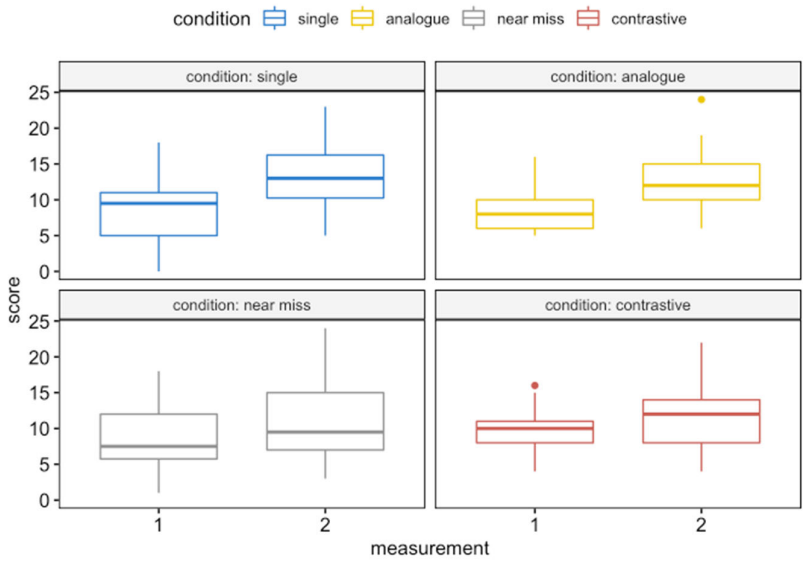

Figure 2: Boxplots per condition across measurement times.

The interaction effect between measurement occasion and condition was small (partial $^2 = 0.2$), but not significant.

Next, the procedure for pairwise comparisons for a non-significant two-way interaction was followed. These pairwise comparisons showed significant differences in measurement occasions, but insignificant differences in mean scores between the four conditions.

Table 4. Overview of outcomes for repeated measures analyses for genre knowledge

Variable	Measure-ment occasion	Single condition	Analogue condition	Near miss condition	Contrastive condition	Measure-ment occasion effect	Condition effect	Interaction effect
		(N = 20)	(N = 21)	(N = 20)	(N = 17)			
		mean (sd)	mean (sd)	mean (sd)	mean (sd)			
genre knowledge	pre-test	8.10 (4.97)	8.57 (3.04)	8.75 (4.81)	9.71 (3.42)	$F = 20.35$**	$F = 0.32$	$F = 1.05$
	post-test	12.95 (4.99)	12.67 (4.36)	10.80 (5.25)	11.53 (4.82)	$^2 = 0.12$	$^2 = 0.01$	$^2 = 0.2$

* = p < .05; ** = p < .001
Interpretation of eta squared (2) effect sizes: 0.2 = small, 0.5 = medium, 0.8 = large.

Post-hoc analyses showed that there were no significant differences between the conditions prior to the intervention. Nor were any significant differences noticeable between conditions in the posttest (see table 5 and 6).

Table 5. Overview of outcomes for post-hoc analyses on the differences between conditions for measurement occasion 1

Condition	Difference	SE	p-value
single - analogue	-0.47	1.41	.99
single - near miss	-0.65	1.43	.97
single - contrastive	-1.61	1.49	.70
analogue - near miss	-0.18	1.41	1.00
analogue - contrastive	-1.13	1.48	.87
near miss - contrastive	-0.96	1.49	.92

Table 6. Overview of outcomes for post-hoc analyses on the differences between conditions for measurement occasion 2

Condition	Difference	SE	p-value
single - analogue	0.28	1.41	1.00
single - near miss	2.15	1.43	.44
single - contrastive	1.42	1.49	.78
analogue - near miss	1.87	1.41	.56
analogue - contrastive	1.14	1.48	.87
near miss - contrastive	-0.73	1.49	.97

Also, through post-hoc analyses we examined the differences between measurement occasions within conditions, which provided us insight in the learning gains within each condition (see table 7). Results show significant effects for the single condition (p=0.001) and for the analogue condition (p=.004). Effects for the near miss and the contrastive condition were both non-significant (p= .154 and p=.242 respectively)

Table 7. Overview of outcomes for post-hoc analyses on the difference between pre and posttest per condition

Condition	Difference	SE	p-value
single	4.85	1.43	<.001
analogue	4.10	1.40	.004
near miss	2.05	1.43	.154
contrastive	1.82	1.55	.242

6. Conclusion & discussion

Given the importance of genre knowledge in writing quality it is essential to look for effective ways that improve genre knowledge. That is exactly what we did in this study.

We investigated the effects of learning from (comparing) text exemplars on genre knowledge of argumentative texts with 11th grade students.

With respect to the first hypothesis, 'Learning from (comparing) exemplars enhances genre knowledge', we found significant differences in the main effect between the pre and posttest scores of students across the four conditions. In contrast, post-hoc tests only showed significant differences for pre and posttest within the single and analogue condition.

Therefore, we can conclude that students learn from looking at exemplars when they receive them sequentially without making comparisons, as well as when they make analogue comparisons. We can thus confirm this hypothesis, but only for single exemplars and analogue comparisons. Regarding the other two conditions, we have no sufficient evidence that there might be an increase in genre knowledge.

These findings correspond with Christie and Gentner (2010) who stated that learners are able to align sequential representations. Childers (2008), Reed (1987) and Ross and Kennedy (1990) already stated that adults and children can align sequentially presented examples. This study confirms that also adolescents can learn from single, sequentially presented texts.

Moreover, this study offers support to the statement that learning from comparing analogue exemplars enhances knowledge (Alfieri et al., 2013; Gentner, 1983). All these studies on learning from (comparing) exemplars mentioned above were conducted in other domains such as vocabulary learning, formative assessment and Mathematics. Based on the results of this study, we can move further and add another domain in which learning from single exemplars and making analogue comparisons is effective, namely learning genre knowledge of argumentative texts.

Regarding the second hypothesis, 'Learning from analogue exemplars improves genre knowledge', post-hoc analyses to investigate the differences between measurement occasions within conditions showed that there was a significant effect in the analogue condition. Thus, based on the post-hoc test, we can confirm this second hypothesis. Learning from analogue text exemplars improves students' genre knowledge. This suggests that also when comparing texts, abstractions are built through aligning common features and recognizing interrelations (Gentner & Namy, 1999; Namy et al., 2007). In this study, students were able to align aspects of genre knowledge that were similar in both texts. Hence, the fact that learning from analogue comparisons increases knowledge (Alfieri et al., 2013), can also be applied to learning from analogue texts.

As for the third and fourth hypothesis, 'Learning from near miss and contrastive exemplars improves genre knowledge', post-hoc analyses for differences between measurement occasions within conditions did not reveal any significant differences between the pre and the posttest scores on genre knowledge. So, in accordance with the post-hoc test, the third hypothesis cannot be confirmed unambiguously. Based on our findings we cannot confirm that students learn from near miss and contrastive comparisons. Therefore, we cannot apply Smith and Gentner's (2014) statement that learning from rather equivalent exemplars, but different in one key aspect, enhances knowledge, to learning from comparing near miss exemplar texts. Nor can we confirm that learning from text exemplars with many differences (Smith & Gentner, 2014) enhances genre knowledge. So, comparing text exemplars that are different, does not seem to improve students' genre knowledge, unlike what previous research has said. There are several possible explanations why these results deviate from earlier studies. Asking for differences may be the least effective method to make comparisons. It might be easier for students to learn when they (only) focus on similarities.

Moreover, the text quality of the text exemplars might have had an effect on the comparisons students made. In the near miss and contrastive condition, one text of high quality had to be compared to a text of lower quality, because texts had to differ in certain aspects of genre knowledge being present or not. That means that in these conditions, and especially in the contrastive condition, students received one text with many genre elements that they had to compare to a text with lesser genre elements and thus of lower text quality. Presenting students with two texts, one of them being of lower quality, might have influenced their learning process. Nonetheless, learning from poorly written text exemplars can enhance students' writing knowledge according to Holliway and McCutchen (2004), but maybe this does not apply when comparing a poorly written text to a text of higher quality. Maybe if students were explicitly told to compare a good and a weak paper, it might have improved their understanding of genre elements.

Concerning the fifth and last hypothesis, 'Learning from comparing exemplars enhances genre knowledge more than learning from single, sequential exemplars', analyses outcomes did not indicate a significant interaction effect between conditions and pre and posttest. This points to the fact that being in a certain condition, does not enhance students' genre knowledge in a different way. In other words, we have no evidence to suggest variations in effectiveness between the four conditions. As a result, based on the findings of this study, we cannot confirm that students learn more from comparisons than from looking at single text exemplars sequentially. This is an unexpected result, since all relevant literature on learning from comparisons states that learning from comparisons is a more effective method to enhance cognitive knowledge than learning from single exemplars (Alfieri et al., 2013). Though being divergent from literature on the subject, there are several

plausible explanations to be found. First, the cognitive load of comparing texts might be too high for students. Comparing texts will probably ask more from students' working memory than comparing word pairs, as is the case in many studies on structural alignment (e.g., Christie & Gentner, 2010; Gentner & Gunn, 2001). The working memory might struggle with these complicated tasks. Second, the level of abstraction might be too high for students. Comparing texts on genre knowledge requires abstract thinking, which can prove to be strenuous for students. This may lead to students comparing more 'concrete' aspects of texts such as content, word choice, etc., although they were explicitly asked not to. Third, the instructions that students received in the single and comparison conditions were different. In the comparison conditions students were asked about similarities only, or about similarities and differences. In contrast, in the single condition, students were asked what they found good and not so good about the text, making them form some kind of judgment about the text, something they were not asked to do in the comparison conditions. The findings might have been different when students in the comparison conditions were also asked to form some kind of judgment about the text. Nevertheless, giving students in the single condition the same instructions as in the comparison conditions was not an option, because they were not allowed to make comparisons. In the future, however, the instruction could be more 'general' in a single condition, for example: "Which genre elements are present in this text?". Using this instruction, students do not have to form a judgment about the text. In a follow-up study, the effects of different types of instructions can be tested. Last, students in the single condition may have consciously or unconsciously compared the texts that they looked at, despite receiving the explicit instruction not to compare the texts. In their minds, they could have compared the next text with the previous one for example. This might also be an explanation why students in the comparison groups did not learn more from the text exemplars opposed to students in the single condition.

6.1 Limitations and further research

Despite its relevance to the field of argumentative writing and learning from exemplars, this study also contains some limitations. The first limitation is that students did not receive a delayed post-test. For some students who need more time to process when learning, outcomes might have been different. Moreover, students were not asked to write an argumentative text after the posttest. By letting them write an argumentative text after the posttest, implicit knowledge that students may have learned through the comparison process might have been present in these texts. In the posttest, only explicit knowledge emerged. The third limitation entails that we do not have insight in why certain students did not adjust their advice to a friend. Some students did not add genre elements to the posttest compared to the pretest. These students could possibly be labeled as careless

responders who did not fully engage in the posttest or maybe they were not motivated to do better. The absence of a control condition is a fourth limitation to the study. As researchers, we did not opt for a control condition, since sequential conditions are favorable to control groups because they differ from the comparison condition only in that cases are studied in succession (Rittle-Johnson & Star, 2007). Nevertheless, the lack of a control condition should be considered when interpreting the results of this study.

A final limitation is that this study, like all studies on learning from (comparing) exemplars, focused on student outcomes and thus the effects of the intervention. Focusing on the learning process during the intervention and in the different conditions, could provide explanations why the findings of this study are divergent from literature on learning from comparisons. Therefore, future studies should also include a focus on what goes on during students' learning.

Though this study holds several limitations, it undeniably makes a substantial contribution to the realms of learning from (comparing) exemplars and argumentative writing research. Within the sphere of learning from (comparing) exemplars, it stands out as the first study to include several types of comparisons. Furthermore, to the author's knowledge, it pioneers the application of learning from (comparing) exemplars within the domain of writing education. Within the context of argumentative writing, this study contributes to finding effective means to increase students' comprehension of argumentative text genres. This study proves that students can improve their genre knowledge of argumentative writing from learning text exemplars. Possibly, depending on the type of learning, different types of text exemplars might be appropriate.

6.2 Implications for practice

The findings of this study hold several practical implications for education. First, when teachers want to enhance students' genre knowledge of argumentative texts, single text exemplars and analogue text comparisons are the best types of exemplars they can use. Second, students do not need a lot of instruction to improve their genre knowledge. The instructions that were given in the intervention were very limited, but even then we saw an improvement in genre knowledge for the single and analogue condition.

Author note

Acknowledgements

We would like to acknowledge the FWO (Fonds voor Wetenschappelijk Onderzoek Vlaanderen – Flemish Research Fund) for their financial support to this project (1SE4823N).

Also, we would like to thank our colleagues at the University of Gent (Yana Landrieu, Fien De Smedt, Hilde Van Keer and Bram De Wever) for sharing their (raw) data, including argumentative texts.

OSF

All necessary documents and data to replicate this study are available on OSF: https://osf.io/hynur/.

References

Abdollahzadeh, E., Farsani, M. A., & Beikmohammadi, M. (2017). Argumentative Writing Behavior of Graduate EFL Learners, *Argumentation, 31*(4), 641-661. https://doi.org/10.1007/s10503-016-9415-5

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning Through Case Comparisons: A Meta-Analytic Review. *Educational Psychologist*, *48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

Allagui, B. (2021). Writing a descriptive paragraph using an augmented reality application: an evaluation of students' performance and attitudes. *Tech Know Learn, 26*, 687-710. https://doi.org/10.1007/s10758-019-09429-2

Bazermann, C. (1997). The life of genre in the classroom, the life in the classrom. *Genre and writing: Issues, arguments, alternatives*, 19-26.

Bigger, A. T. (2022). *Secondary Students' Proficiency with and Knowledge of Process and Genre Approaches to Writing.* (Doctoral dissertation, University of Pittsburgh).

Chambliss, M. J., & Murphy, P.K. (2010). Fourth and Fifth Graders Representing the Argument Structure in Written Texts, D*iscourse Processes, 34*(1), 91-115. https://doi.org/10.1207/S15326950DP3401_4

Childers, J. B. (2008). The Structural Alignment and Comparison of Events in Verb Acquisition *Proceedings of the Annual Meeting of the Cognitive Science Society*, 30. Retrieved from https://escholarship.org/uc/item/2jz052xr

Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, *11*(3), 356–373. https://doi.org/10.1080/15248371003700015

Coirier, P., Andriessen, J., & Chanquoy, L. (1999). From planning to translating: The specificity of argumentative writing. in Coirier & Andriessen (Eds.), *Foundations of argumentative text processing* (pp. 1-28). Amsterdam, The Netherlands: University Press. https://doi.org/10.5117/9789053563403

De Smedt, F., & Van Keer, H. (2018). Fostering writing in upper primary grades: a study into the distinct and combined impact of explicit instruction and peer assistance. *Reading and Writing*, *31*(2), 325–354. https://doi.org/10.1007/s11145-017-9787-4

Felton, M., & Kuhn, D. (2001). The development of argumentative discourse skill. *Discourse Processes, 32*(2-3), 135-153. http://dx.doi.org/10.1207/S15326950DP3202&3_03

Ferretti, R. P., & De La Paz, S. (2011). On the comprehension and production of written texts: Instruc- tional activities that support content-area literacy. In R. O'Connor & P. Vadasy (Eds.), *Handbook of reading interventions* (pp. 326–355). New York, NY: Guilford.

Ferretti, R. P., and Lewis, W. E. (2013). "Best practices in teaching argumentative writing," in *Best practices in writing instruction* 2nd ed, eds S. Graham, C. A. MacArthur, and J. Fitzgerald (New York, NY: Guilford Press), 113–140.

Freedman, A., & Pringle, I. (1984). Why Students Can't Write Arguments, English in Education, 18(2), 73-84. https://doi.org/10.1111/j.1754-8845.1984.tb00668.x

Gadgil, S., & Nokes, T. (2009). Analogical Scaffolding in Collaborative Learning. Proceedings of the Annual Meeting of the Cognitive Science Society, 31. Retrieved from https://escholarship.org/uc/item/5b74x3jv

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science, 7*(2), 155-170. https://doi.org/10.1207/s15516709cog0702_3

Gentner, D., & Gunn, V. (2001). Structural alignment facilitates the noticing of differences. *Memory & Cognition, 29*(4), 565-577. https://doi.org/10.3758/BF03200458

Gentner, D., & Namy L. L. (1999). Comparison in the development of categories. *Cognitive Development, 14*(4), 487-513). https://doi.org/10.1016/S0885-2014(99)00016-7

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. (2012). A Meta-Analysis of Writing Instruction for Students in the Elementary Grades. *Journal of Educational Psychology, 104*(4), 879-896. https://psycnet.apa.org/doi/10.1037/a0029185

Graham, S., & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools.* New York: Carnegie Corporation.

Granado-Peinado, M., Mateos, M., Martín, E., & Cuevas, I. (2019). Teaching to write collaborative argumentative syntheses in higher education. *Reading and Writing*, *32*(8), 2037–2058. https://doi.org/10.1007/s11145-019-09939-6

Hammer, R., Bar-Hillel, A., Hertz, T., Weinshall, D., & Hochstein, S. (2008). Comparison processes in category learning: From theory to behavior. *Brain Research*, *1225*, 102–118. https://doi.org/10.1016/j.brainres.2008.04.079

Holliway, D, & McCutchen, D. (2004). Audience perspective in young writers' composing and revising. In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes* (pp. 87-101). Boston: Kluwer. https://doi.org/10.1007/978-94-007-1048-1_6

Humphries Sandstrom, K. (2021). Building genre-knowledge through peer review: L2 doctoral students' provision of feedback in the natural sciences, Journal of Writing Research, 12(2), 257-283. https://doi.org/10.17239/jowr-2021.13.02.03

Hyon, S. (2001). Long-term effects of genre-based instruction: a follow-up study of an EAP reading course, *English for Specific Purposes, 20*(1), 417-438. https://doi.org/10.1016/S0889-4906(01)00019-9

Hyon, S. (2002). Genre and ESL reading: A classroom study. In John, A.M. editor, Genre in the classroom. Mahwah: Lawrence Earlbaum Associates Publishers.

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

Landrieu, Y., Smedt, F. De, Keer, H. Van, & Wever, B. De. (2022). *Assessing the Quality of Argumentative Texts : Examining the General Agreement Between Different Rating Procedures and Exploring Inferences of (Dis) agreement Cases Rating Procedures to Assess Text Quality. 7*(May). https://doi.org/10.3389/feduc.2022.784261

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*(1), 22-44. https://psycnet.apa.org/doi/10.1037/0033-295X.99.1.22

Kuhn, D. (1991). *The skills of argument.* Cambridge University Press. https://doi.org/10.1017/CBO9780511571350

Kuhn, D., & Udell, W. (2007). Coordinating own and other perspectives in argument. *Thinking & Reasoning, 13*(2), 90-104. https://doi.org/10.1080/13546780600625447

Kurtz, K. J., Miao, C., & Gentner, D. (2001-). Learning by Analogical Bootstrapping. The *Journal of the Learning Sciences, 10*(4), 417-446. https://doi.org/10.1207/S15327809JLS1004new_2

Martin, J. R. (2009). Genre and language learning: A social semiotic perspective, *Linguistics and Education, 20*(1), 10-21. https://doi.org/10.1016/j.linged.2009.01.003

McCann, T. M. (1989). Student Argumentative Writing Knowledge and Ability at the Three Grade Levels, *Research in the Teaching of English, 23*(1), 62-76.

McCutchen, D. (1986). Domain knowledge and linguistic knowledge in the development of writing ability. Journal of Memory and Language, 25(4). 431-444. https://doi.org/10.1016/0749-596X(86)90036-7 Miller, J. G. (1984). Culture and the

development of everyday social explanation. *Journal of Personality and Social Psychology, 46*(5), 961-978. https://psycnet.apa.org/doi/10.1037/0022-3514.46.5.961

Mohsen, M. A., & Qassem, M. (2020). Analyses of L2 Learners' Text Writing Strategy: Process-Oriented Perspective, Journal of Psycholinguistic Research, 49(1), 435-451. https://doi.org/10.1007/s10936-020-09693-9

Nagarajan, A., & Hmelo-Silver, C. (2006). Scaffolding Learning from Contrasting Video Cases. InBarab, S. A., Hay, K. E., & Hickey, D. T. (Eds.), *The International Conference of the Learning Sciences*: Indiana University 2006. Proceedings of ICLS 2006, Volume 1 (pp. 495-501). Bloomington, Indiana, USA: International Society of the Learning Sciences.

Namy, L. L., Gentner, D., & Clepper, L. E. (2007). How close is too close? Alignment and perceptual similarity in children's categorization. *Cognition, Brain, Behavior, 11*, 647–659.

NCES (2012). *The Nation's Report Card: Writing 2011.* Washington, DC: NCES.

Newell, G. E., Beach, R., Smith, J., & Vanderheide, J. (2011). Review of research: Teaching and learning argumentative reading and writing: A review of research. *Reading Research Quarterly*, *46*(3), 273–304. https://doi.org/10.1598/RRQ.46.3.4

Nussbaum, E. M., & Kardash, C. M. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology*, *97*(2), 157–169. https://doi.org/10.1037/0022-0663.97.2.157

Nussbaum, E. M., & Schraw, G. (2007). Promoting argument-counterargument integration in students' writing. *Journal of Experimental Education*, *76*(1), 59–92. https://doi.org/10.3200/JEXE.76.1.59-92

Olinghouse, N. G., & Graham, S. (2009). The relationship between the discourse knowledge and the writing performance of elementary-grade students. *Journal of Educational Psychology, 101*(1), 37-50. https://doi.org/10.1037/a0013462

Olinghouse, N. G., Graham, S., & Gillespie, A. (2015). The relationship of discourse and topic knowledge to fifth graders' writing performance. *Journal of Educational Psychology, 107*(2), 391-406. https://psycnet.apa.org/doi/10.1037/a0037549

O'Keefe, D. J. (2016). How to Handle Opposing Arguments in Persuasive Messages: A Meta-Analytic Review of the Effects of One-Sided and Two-Sided Messages, *Annals of the International Communication Association, 22*(1), 209-249. https://doi.org/10.1080/23808985.1999.11678963

Perkins, D. N., Farady, M., & Bushey, B. (1991). Everyday reasoning and the roots of intelligence. In J. Voss, D. Perkins, & J. Segal (Eds.), *Informal reasoning and education* (pp. 83–105). Hillsdale, NJ: Erlbaum.

Prata, M. J., de Sousa, B., Festas, I., & Oliveira, A. L. (2019). Cooperative methods and self-regulated strategies development for argumentative writing. *Journal of Educational Research*, *112*(1), 12–27. https://doi.org/10.1080/00220671.2018.1427037

Qin, J., & Karabacak, E. (2010). The analysis of Toulmin elements in Chinese EFL university argumentative writing. *System*, *38*(3), 444–456. https://doi.org/10.1016/j.system.2010.06.012

Reed, S.K. (1987). A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 124-139.

Rittle-Johnson, B., & Star, J. R. (2007). Does Comparing Solution Methods Facilitate Conceptual and Procedural Knowledge? An Experimental Study on Learning to Solve Equations. *Journal of Educational Psychology*, *99*(3), 561–574. https://doi.org/10.1037/0022-0663.99.3.561

Rittle-Johnson, B., & Star, J. R. (2009). Compared With What? The Effects of Different Comparisons on Conceptual Knowledge and Procedural Flexibility for Equation Solving. *Journal of Educational Psychology*, *101*(3), 529–544. https://doi.org/10.1037/a0014224

Ross, B. H., & Kennedy, P.T. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 42-55.

Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions, *Science Education, 93*(3), 447-472. https://doi.org/10.1002/sce.20276

Simon, D., Holyoak, K. J. (2002). Structural dynamics of cognition: From consistency theories to constraint satisfaction. *Personality and Social Psychology Review, 6*(4), 283-294. https://doi.org/10.1207/S15327957PSPR0604_03

Smith, L. A., & Gentner, D. (2014). *UC Merced Proceedings of the Annual Meeting of the Cognitive Science The Role of Difference - Detection in Learning Contrastive Categories*.

Stapleton, P. (2001). Assessing critical thinking in the writing of Japanese university students: Insights about assumptions and content familiarity. *Written Communication, 18*(4), 506–548. https://doi.org/10.1177/0741088301018004004

Stapleton, P., & Wu, Y. (Amy). (2015). Assessing the quality of arguments in students' persuasive writing: A case study analyzing the relationship between surface structure and substance. *Journal of English for Academic Purposes*, *17*, 12–23. https://doi.org/10.1016/j.jeap.2014.11.006

Song, Y., and Ferretti, R. P. (2013). Teaching critical questions about argumentation through the revising process: effects of strategy instruction on college students' argumentative essays. *Reading and Writing, 26*, 67–90. doi: 10.1007/ s11145-012-9381-8

Toulmin, S. (1958). *The uses of argument.* Cambridge: Cambridge University Press.

Toulmin, S. (2003). *The uses of argument, 2nd edution,* Cambridge University Press. https://doi.org/10.1017/CBO9780511840005

Yasuda, S. (2011). Genre-based tasks in foreign language writing: Developing writers' genre awareness, linguistic knowledge, and writing competence. *Journal of Second Language Writing*, *20*(2), 111–133. https://doi.org/10.1016/j.jslw.2011.03.001

Wickham, H. (2007) Reshaping data with the reshape package. *Journal of Statistical Software, 21*(12). https://doi.org/10.18637/jss.v021.i12

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. https://doi.org/10.1007/978-3-319-24277-4

Wickham, H. et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., & Müller, K (2022). dplyr: A Grammar of Data Manipulation. R package version 1.0.9. https://CRAN.R-project.org/package=dplyr

Wolfe, C. R. (2012). Individual differences in the "myside bias" in reasoning and written argumentation. *Written Communication, 29*(4), 477-501. https://doi.org/10.1177/0741088312457909

Appendix A: Description of the selection and manipulation of the text material

First, the main researcher selected texts to use in the different conditions for comparison. These texts, written by 11th graders, were selected from a more extensive sample of 165 texts from another argumentative writing study in Flanders, conducted by researchers at Ghent University. The subject of the texts was keeping animals in a zoo. Texts were selected to create pairs for the four conditions, based on the analytical scoring by means of the ASRAW scoring rubric (Stapleton & Wu, 2015) and holistic ranking (calculated via the tool for comparative judgment Comproved, www.comproved.com).

In a first manipulation round the selected texts (N=22) were manipulated (i.e., improved and/or corrected) in terms of style, spelling, punctuation, grammar, word use and sentence & paragraph structure. In some texts, a title, introduction or conclusion was added to improve similarity across the texts. These manipulations were made to prevent students from getting distracted while comparing texts, by for example misspelled words or an inconsistent style. Next, eight experts in argumentative writing were asked to code the selected texts. These experts had ample experience as a Dutch teacher and teaching experience in grades 11 and 12 and thus experience in argumentative writing instruction. Most of them were (ex-)teachers (N=7), whereas one participant was a PhD student in Rhetorics and Persuasion and was also responsible for academic writing courses. This PhD student was later on replaced by a retired Dutch teacher. All ex-teachers still had a job related to teaching Dutch or were retired.

These experts used the ASRAW scoring rubric (Stapleton & Wu, 2015) to code the texts according to the genre elements in this rubric. They were not asked to assess or score the texts. Introduction and conclusion were added as a possible code, as these are also typical elements of an argumentative text. Experts could also indicate 'miscellaneous' if they felt that a text fragment did not belong to the 8 possible codes. Each text was coded by four experts. To verify interrater reliability, Fleiss' kappa was calculated for each text. Mean of the 22 kappas was 0.54, with 8 texts with a kappa between 0.61 and 0.80 (substantial agreement), 11 texts with a kappa between 0.41 and 0.60 (moderate agreement), 3 texts with a kappa between 0.21 and 0.40 (fair agreement) (Landis & Koch, 1977).

Member-checking was done with five of the eight experts to clarify why they made certain coding choices. Experts whose coding were deviant or non-deviant in terms of coding similarities were chosen. Several text fragments with a low interrater reliability were selected to discuss. Not all text fragments could be part of the selection, because of the set time limit for the interview (around 30 minutes). The participating experts were asked why they assigned a particular code and why, whether they would change it and why or whether they would double code the text fragment. Based on the information in these member-checking interviews, additional manipulations to the texts were made to enhance the clarity in codes.

Subsequently, the definitive coding of the experts was compared to the original coding of the (UGent) researchers. When certain codes given by researchers and experts did not match, the first author decided upon the definitive coding for that text fragment, based on the theoretical framework of Stapleton & Wu (2015) and/or the explanations of experts.

In a next phase, the ASRAW scores for the 22 texts were recalculated, because several codes had been changed in comparison with the original scoring. Due to the not too rigorous character of the ASRAW scoring rubric, most texts received the same ASRAW score, despite modest scoring differences. A third manipulation round was set up to alter some adjusted scores back to the original scores. That way, the selected text pairs that were made for the four conditions could be preserved. Manipulations in this round were minor; changes were made in only two texts.

Appendix B: Categories of genre knowledge and their score

1) ASRAW — 3 points
   - position (for or against) (claim)
   - argumentation (claim data)
     - clear arguments
     - different arguments
     - arguments to support position
   - counterargument (counterargument claim) — 4 points
   - argumenting counterargument (counterargument data)
   - refuting counterargument (rebuttal claim)
   - arguments for rebuttal (rebuttal data)

2) reinforcement of argumentation — 2 points
   - fact/opinion
   - describing pros and cons
   - giving examples
   - mentioning sources
   - providing proof
   - critical
   - objective / subjective

3) text goal — 1 point
   - persuading

4) IME structure — 2 points
   - introduction
   - middle (body of the text)
   - ending / conclusion

5) general text structure — 2 points
   - paragraphs
   - use of signal words
   - clearly structured
   - (good, attractive) title
   - blank lines between paragraphs

6) language use — 1 point
   - formal language
   - proper Dutch (no dialect)
   - written in first person (I)
   - being aware of the audience that you are writing to
   - full sentences
   - clear & short sentences

Appendix C: Example of pretest scoring of student *to2610*

What student wrote	coded as	score
start with objective information on the topic	/	0
Give your point of view	claim	3
argument this point of view as much as possible	claim data	3
possibly with facts and be as concrete as possible	reinforcement of argumentation	2
add other viewpoints, of others	counterargument claim	4
compare to yours and the facts	rebuttal data	4
TOTAL SCORE		18

Appendix D: Example of posttest scoring of student *ti040*

Pre-test

What student wrote	Coded as	Score
writing in 'I perspective' a lot	Language use	1
Clear and short sentences	Language use	1
Good explanation why you choose that argument; need for different arguments	Claim data	3
Providing a solution to the problem	/	0
TOTAL SCORE		5

Post-test

- Would you change your advice? Student's answer: yes
- What would you add?

What student wrote	Coded as	Score
Describing advantages and disadvantages	Reinforcement of argumentation	2
In the end a clear conclusion	IME structure	2
TOTAL SCORE to add to pretest score		+ 4

- What would you change?

What student wrote	Coded as	Score
Writing in 'I perspective' a lot	/ This was not coded as it was not clear what the change was in relation to the same thing this student wrote down in the pretest.	/

- What would you delete?

What student wrote	Coded as	Score
Clear and short sentences	Language use	- 1
	* Student deleted this, however this was a correct genre element. Therefore, the score decreased by 1.	- 1

Scoring post-test: 5 + 4 – 1 = 8