



## UvA-DARE (Digital Academic Repository)

### Synthesis writing

*Teaching high school students how to read, plan, draft, and revise*

van Ockenburg, L.

### Publication date

2022

[Link to publication](#)

### Citation for published version (APA):

van Ockenburg, L. (2022). *Synthesis writing: Teaching high school students how to read, plan, draft, and revise*. [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, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

## CHAPTER 3

### LEARNING HOW TO SYNTHESIZE

#### Design and Evaluation of a Reading-Writing Learning Unit for High-School Students\*

*Abstract.* This paper reports on the systematic design and evaluation process of a learning unit for ninth-grade students, aimed at learning to write synthesis texts. The unit was based on design principles derived from a review of effective synthesis writing interventions and general principles of effective learning. To evaluate the quality of this unit, we considered three aspects: the validity, feasibility, and effectiveness of its design. The design as construct was based on state-of-the-art knowledge which ensured construct validity. Furthermore, user data indicated that the construct's operationalization was valid: the content and structure of the unit reflected the construct. In addition, teachers were generally positive about the unit's overall feasibility and teacher logs indicated that the lessons were mostly taught as intended. Furthermore, student data indicated that the unit was feasible for students, as well. Finally, the effectiveness of the unit was confirmed by the outcomes of an intervention in five classes, using a switching replication design. Some options for further improvement of the design are also proposed.

*Keywords:* writing education, synthesis texts, instructional design, design study, design principles

#### 1 INTRODUCTION

Nowadays there is a continuous explosive increase in the availability and accessibility of information, which increasingly influences all aspects of our lives (Sociaal en Cultureel Planbureau [The Netherlands Institute for Social Research], 2004). Education must teach students how to deal with this wealth of information. As the demands on multiple text reading comprehension and writing skills become more severe, the current state of affairs seems troublesome. Most OECD countries saw a decline in students' reading comprehension (OECD, 2019), which focuses on building knowledge, critical thinking, and informed judgment. At the same time, there are concerns about students' overall writing skills (Bonset, 2010) and the writing skills taught in secondary education do not match the

---

\* This chapter is a somewhat adapted version of Van Ockenburg, L., Van Weijen, D., & Rijlaarsdam, G. (2021). *Learning how to synthesize: The design and evaluation of a reading-writing learning unit for high-school students*. L1-Educational Studies in Language and Literature, Volume 21, 1-33. <https://doi.org/10.17239/L1ESLL-2021.21.01.06>

required skills in higher education, such as writing source-based research papers (Meestringa, 2011; Meestringa & Ravesloot, 2013).

Therefore, it is necessary to pay specific attention to the development of information processing skills. This can be done, among other things, by introducing hybrid tasks in secondary education, such as information synthesizing tasks, which require the integration of information from different sources. Producing written syntheses requires continuously alternating between reading and writing: exploratory reading and understanding the sources, selecting relevant information, arranging the selected information in a logical way and integrating it into a new text (Klein & Boscolo, 2016). Synthesis writing induces complex processes that call for recursivity and mediation (Vandermeulen et al., 2020b).

In this paper we report on the design and evaluation process of a learning unit for ninth-grade students (14-15-years old) who were taught a new skill that integrates reading and writing: to produce a written synthesis. This unit is intended as a first introduction to hybrid reading-and-writing tasks for students, who usually do not receive instruction about such tasks in Dutch language classes before ninth grade. Therefore, we designed a learning unit to provide students with relatively simple synthesis tasks. Our position, from a curricular point of view, is that they should first be allowed to fully focus on the three basic processes of synthesizing, i.e., selecting, organizing and connecting source information (Spivey & King, 1989), before trying to master additional skills such as independently searching for sources, while simultaneously assessing their reliability, and weighing conflicting information.

### 1.1 *Validating designs of learning units*

With the present design study, we join the growing group of L1-researchers (Elving-Heida, 2019; Koster & Bouwer, 2018; Schrijvers et al., 2019; see also Luger, 2020) who responded to a call from Rijlaarsdam et al. (2017), proposing the use of design principles as the basis for the structure of learning units. Design principles are heuristic statements that describe the conditional relationship between a desired outcome (Y) and an instructional unit (X). For example:

"If we involve students in activity/approach/strategy X, then they will achieve learning outcome Y."

or

"If you want to achieve learning outcome Y, students must be involved in learning activity X, that can be stimulated by instruction I."

A design principle thus represents a theoretical insight on a relation between a learning activity, instruction, and the intended outcome. We expect that a such a principle has content validity, as it represents state-of-the art knowledge. A set of design principles form the construct of a learning unit, similar to a set of sub-skills that define the construct writing skill. Design principles substantiate the choices for components of a unit, i.e., which elements from the current knowledge base should be included in the design. Construct validity is achieved when these components are consistently linked to ensure that the unit forms a coherent ensemble of functionally related elements as opposed to a collection of individual elements (Nieveen, 1999).

When discussing the quality of a learning unit, we discuss and compare the construct-as-intended, and the construct-as-implemented. However, there might be a large discrepancy between the blueprint of the unit and its implementation, for all kinds of reasons such as the operationalization in materials or the way the unit was implemented in the classroom. For a thorough evaluation, a unit's feasibility and effectiveness must also be tested (Van den Akker, 1999). Van den Akker uses the term practicality when referring to its feasibility, which refers to the extent that both teachers and students consider the unit appealing and usable in 'normal' conditions. A unit is considered effective when it results in the outcomes it aimed to achieve.

## 1.2 *Research questions*

We aim to systematically evaluate a specific learning unit for the acquisition of L1 synthesis writing skills. The outcomes may lead to a redesign of the unit, on several levels: the construct-as-intended, the construct-as-implemented, content validity and feasibility. All outcomes of a design study can lead to revisions, which are necessary before the design can be tested on a larger scale in authentic classroom settings. We investigated three aspects: (a) validity, (b) feasibility, and (c) effectiveness of the unit to answer the main question:

Which elements of the unit should be modified in a redesign to further improve its' quality in three aspects: validity, feasibility, and effect?

To this end we will deal with three sub questions:

1. To what extent does the unit for L1 synthesis writing meet the requirements regarding both content and construct validity?

We expected that we could achieve a high degree of content validity by drawing up design principles based on an analysis of proven effective synthesis writing interventions we found through a literature review. We expected that a high degree of construct validity could be achieved by adhering to general principles for effective learning (e.g., Merrill, 2002) and by regularly consulting an expert in language and literature teaching during the design process.

2. To what extent does the unit meet the requirements regarding feasibility?

To answer this question, we field tested the unit. We expected that in a feasible learning unit, (1) teachers would be able to carry out at least 90% of the learning activities in their classroom practice, (2) that most students would be able to participate in the lessons and complete the tasks within the allocated time, (3) that teachers would find the unit attractive to work with, and (4) that students would report feeling that their participation in the lesson unit was worthwhile.

3. To what extent does the unit meet the requirements regarding effectiveness?

We expected that participating in the unit would improve students' synthesis text writing sustainably compared to students who did not participate.

A report on a design study, like this one, falls into two research categories. It is both a theoretical, argumentative account of the content and construct validity, as well as an empirical study on feasibility and effect. Therefore, in this paper we present two sets of data. In Part 1: Theoretical Background: Validity we present the data for the discussion on validity issues, the background of the design principles and the operationalization of those principles in an instructional design. In Part 2: Field Trial, we report on feasibility and effect data from a quasi-experimental intervention study.

## 2 PART 1. THEORETICAL BACKGROUND: VALIDITY

In this section, we describe the design principles on which we based the unit. Following Koster and Bouwer (2018), we classified the design principles into two categories: (a) principles for the learning content, i.e., what is being taught or learned (2.1), and (b) principles for the mode of instruction, i.e., how the content is being taught or learned (2.2). Finally, in section 2.3 we describe how we operationalized the design principles in a unit.

## 2.1 *Design principles for learning content*

The design is based on an analysis of the six most effective interventions for synthesis writing ( $ES > .80$ ) found after a systematic review of the literature (see Chapter 2). This resulted in three design principles.

### 2.1.1 Design principle 1: Focus on Cognitive Strategies

All the studies on synthesis writing we analysed, focused on teaching all three basic processes: selecting information from sources, organizing, and connecting that information (Spivey & King, 1989), which together form a global strategy for task decomposition. Such strategies help prevent cognitive overload, an inherent characteristic of the complex nature of the writing process (Kellogg, 1996). Each of the three basic processes relies on specific cognitive strategies. From the studies we analysed, we selected the strategies that were used to teach selecting, organizing, and connecting source information. This selection contained the potential learning content for each task component, and encompassed strategies such as comparing-and-contrasting sample synthesis texts (e.g., Kirkpatrick & Klein, 2009) or source content (e.g., Martinez et al., 2015), schematizing source content (e.g., Barzilai & Ka'adan, 2017) and performing linking exercises (Zhang, 2013). Therefore, as our first design principle we chose to focus on cognitive strategies:

If students learn to apply strategies for selecting, organizing, and connecting, then they will write better synthesis texts.

### 2.1.2 Design principle 2: Focus on Writing Routines

Students develop personal routines to prevent cognitive overload inherent to complex tasks such as writing. They decompose the writing process into "steps" that they perform in a certain order (Kieft & Rijlaarsdam, 2005). In models of the writing process such routines are called 'task schemas' (Hayes et al., 1987), which are stored in the long-term memory. During writing, writers monitor the execution of a schema (Hayes, 2012), especially in cases when the execution is so complicated that the default task schema is insufficient, for example when tasks are difficult or new. Most adolescents do not write texts based on sources: at school they usually 'just write', mainly using a knowledge telling strategy (Scardamalia & Bereiter, 1987). If sources are available, they may pick a few elements or citations that fit in their text-written-so-far. That strategy becomes insufficient when we confront them with a new genre such as a synthesis text. Then students must

learn to create a text that respectfully represents the source materials, shifting from a knowledge-telling to a knowledge-transforming strategy.

Earlier research has shown that writing routines can vary greatly in our target group, ninth-grade students. They generally prefer a fast-writing process without much post-writing revision (Van Ockenburg et al., 2018). They tend to have low Preplanning tendencies and even lower Drafting tendencies. These writing routines can play a role when implementing writing instruction. In studies with 10th-grade students, instruction of two different strategies—Preplanning by outlining or planning by Drafting—resulted in different learning outcomes, depending on students' initial writing routine (Kieft et al., 2007; Kieft & Rijlaarsdam, 2008). These outcomes indicated that writing routines moderate the effectiveness of that instruction. Thus, our second design principle is:

If students are offered strategies for writing synthesis texts which match their personal writing routine, they will write better synthesis texts.

### 2.1.3 Design principle 3: Activate Metacognition by Choice

The third design principle originated from a study by Robledo-Ramon (2016), which aimed to develop personal writing strategies, with a focus on the development of metacognition. She based her intervention on the Strategic Learning Content (SCL) model (Butler, 1998), which encourages students to generate personalized strategies instead of teaching them specific, predefined strategies. The SCL-model aims for students to acquire productive metacognitive knowledge about tasks, strategies, and themselves as learners and about how these three factors interact to affect the course and outcome of cognitive activities (Flavell, 1979). Similar to Escorcia et al. (2017), we define metacognitive knowledge in writing as personal knowledge about one's own cognitive functioning in writing, involving the use of specific strategies to regulate writing.

To construct metacognitive knowledge of task strategies, we opted to implement choices from strategies on how to proceed. This required a flexible learning unit, so that students can try out different strategies. When students must choose, their metacognition is activated, which creates opportunities to monitor and evaluate the suitability of the choices made. To ensure students' choices are well informed, the unit should pay explicit attention to the writing process and various writing routines, while working towards the development of a successful, personal writing strategy. Goal setting is a crucial element in this process (Schunk, 2003); without setting goals when a choice is offered, monitoring and evaluation of progress and success are impossible, and no learning will occur. To conclude, our third design principle is:

If students are aware of the different strategies that are offered, so that they can decide which one might suit them best given their personal writing routine, they will write better synthesis texts.

## 2.2 *Design principles for mode of learning*

In this section, we present the design principles that guided our choices for the mode of learning. Observing, practicing, and evaluating are the three key modes of learning in the unit. These principles are supplemented by Merrill's (2002) general principles for effective instruction related to: (1) real-world problem solving, (2) activating prior knowledge, (3) task demonstration, (4) applying new knowledge, and (5) internalizing new knowledge and applying it in new ways (Merrill, 2002). We chose to adhere to these general principles, to ensure that the lesson unit did not consist of a collection of individual learning activities, but rather that all learning activities were consistently linked in means-end relations.

### 2.2.1 Design principle A: Observational learning

Observational learning was a key learning activity in several synthesis writing units (Kirkpatrick & Klein, 2009; Martinez et al., 2015) and in most effective instruction units focused on reading and writing strategies (Harris & Graham, 2017). Observational learning occurs when learners learn by watching and evaluating others performing tasks they too must carry out (Bandura, 1986). For instance, models (teachers or fellow students) perform a task (live or on video) while thinking out loud to induce certain behaviour in an observer (a learner, a student). Learning by observing offers the possibility to catch sight of the task processes (Zimmerman & Risemberg, 1997). This is in line with Merrill's third principle: "Learning is promoted when new knowledge is demonstrated to the learner" (Merrill, 2002, p. 45).

In addition, observing someone who performs a new, unknown task is cognitively less demanding than having to perform this new task yourself. As a result, one's working memory is less burdened, leaving more cognitive room for learning. Therefore, observational learning can be effective when learning a new and cognitively complex task, such as writing (Braaksma et al., 2002). Consequently, our first design principle for the modes through which we offer the learning content in the unit is:

If students acquire learning content as described in the three focus principles through observational learning, this will contribute to the effectiveness of the unit.

### 2.2.2 Design principle B: Evaluating

Most effective synthesis writing units included evaluation activities, often in a combination of small group discussions (two to five students) followed by whole class discussion (Kirkpatrick & Klein, 2009; Robledo-Ramón, 2016; Zhang, 2013). In our unit, we intended to distinguish three types of learning activities to stimulate students to evaluate their learning: reflection, learner dialogues, and peer feedback.

*Reflection.* Research by Braaksma et al. (2001) showed that carrying out reflective activities to compare and contrast models while observing contributed positively to learning effects. Observational learning should therefore include a reflective component (Rijlaarsdam et al., 2008). Furthermore, students develop metacognitive skills by elaborating on their observations. Research in young children shows that observing models can produce greater learning effects than practicing that same task. In addition, the learning effect was much greater when the children had to evaluate the models after observing them (Sonnenschein & Whitehurst, 1984). In our intervention we intend to include short writing assignments, which stimulate students to make their thoughts explicit and then enable them to compare them with others in an external dialogue.

*External dialogues.* Toorenaar and Rijlaarsdam (2011) showed that students' learning can be encouraged by sharing and comparing their thoughts and experiences with the thoughts and experiences of others in a learner dialogue: a verbal and cognitive collaboration between students. These dialogues can take place in pairs, small groups or in a teacher-led whole class discussion. Such a plenary discussion phase might enrich students' thinking, because even if the collaboration within small groups was not optimal, students still have access to what other groups have discussed. For teachers, concluding their lessons by means of a brief inventory of the group discussions provides them with the opportunity to gain insight in students' learning processes before moving on to the next learning activity.

*Peer feedback.* Meta-analyses (Graham & Perin, 2007; Koster et al., 2015) have shown that peer feedback can produce large effect sizes related to students' writing performance, respectively 0.75 and 0.88. However, the number of studies on feedback is relatively small, so the effect sizes are imprecise. In addition, feedback can take many forms (e.g., peer feedback vs. teacher feedback) and can be applied in various ways (e.g., product-focused vs. process-focused, and criteria-driven vs. spontaneous). Nevertheless, as all the available studies found positive effects of peer feedback, giving and receiving peer feedback has become an

indispensable part of learning to write and was included in several of the effective interventions we analysed (e.g., Zhang, 2013).

Finally, the importance of evaluating is also emphasized by Merrill's principle related to internalizing new knowledge, whereby learners evaluate what they have learnt and search for ways to apply that knowledge in future tasks (Merrill, 2002). Thus, our second design principle for mode of learning is:

If students are encouraged to evaluate through various types of learning activities, with the aim of developing task knowledge, this will contribute to the effectiveness of the unit.

### 2.2.3 Design principle C: Practicing

Many effective synthesis writing interventions provide repetitive practice to allow students to automatize the strategies under study (Britt & Aglinskas, 2002; Zhang, 2013). These repetitions are often implemented through scaffolded, collaborative practice followed by individual practice (Barzilai & Ka'adan, 2017; Kirkpatrick & Klein, 2009; Martinez et al., 2015; Robledo-Ramón, 2016). Not surprisingly, developing (synthesis) writing skills appears to require a fair amount of practice. Merrill also emphasizes providing sufficient opportunity for students to apply their new knowledge or skills. Furthermore, the importance of practicing is also linked, albeit somewhat more indirectly, to Merrill's principles regarding problem-centred learning, activation of prior knowledge and application. Merrill suggests that showing students the real-world problem or task they will be able to solve after completing the learning unit and allowing them to experience this task as a whole prior to instruction, can help make instruction more effective. Furthermore, experiencing such a task can also help activate students' relevant prior knowledge (Merrill, 2002). Therefore, our third principle of mode is:

If students have enough opportunity to apply what they have learned, first through scaffolded practice, and then gradually working towards independent, self-regulated task execution, this will contribute to the effectiveness of the unit.

### 2.3 *Operationalization of the design principles*

The design phase resulted in a unit that included three main components: a pre-flection, an instruction, and an evaluation phase. In the first phase, the emphasis was on 'pre-flection': previewing and experiencing the whole task that would be the object of the upcoming learning unit, moving in Lesson 1 to elaboration of

the task representation. The second phase, Lessons 2 through 4, was an instruction phase, in which a new strategy was added to the same task during each lesson. In Lesson 5, students had to apply the strategies instructed so far in a new task. The third phase, Lesson 6, emphasized 'reflection' via the evaluation of the texts written in Lesson 5; students looked back on what they had learned during this unit.

Table 3.1 shows the overarching structure of the unit. Each row describes the unit number, phase, mode, focus, specification of the key learning activity, and whether the learning activity was performed individually or together with others.

*Table 3.1 General lesson plan of the unit*

Phase	Mode	DP <sup>a</sup>	Specification	Level <sup>b</sup>
Pretest (Whole task experience)				
i	Observing	3	Observing a 3-minute animated video providing information about integrating source information in synthesis texts.	I
ii	Practicing	3	Performing a synthesis pre-task to experience the synthesizing process as a whole.	I
Lessons				
1 (pre-reflection)				
i	Evaluating	3	Ranking four model texts (on the same topic as the pre-task) according to quality, to build task representation for synthesis texts, by noticing differences in text quality.	I
ii	Evaluating	3	Discussing the quality of these four model texts to refine and elaborate task representation for text quality.	G
2-4 (similarly structured)				
i	Observing	1	Comparing-and-contrasting two strong peer models who perform a synthesis subtask while employing different strategies to retrieve prior knowledge about task representation, to discover that different strategies can work, while noticing similarities and differences.	I

Phase	Mode	DP <sup>a</sup>	Specification	Level <sup>b</sup>
ii	Choosing	2	Discussing the modelled strategies to determine which strategy one prefers and why.	G
iii	Practicing	1	Applying chosen strategy to a subtask to experience one's preferred strategy.	I
iv	Evaluating	3	Discussing written product(s) to evaluate the result of applying the chosen strategy and estimate whether the result was satisfactory.	G
<hr/>				
5				
i	Choosing	2	Considering whether to keep or change the strategies chosen in earlier lessons and why and determine which topic to choose and why.	I
ii	Practicing	1	Performing a new synthesis task to experience the use of the chosen strategies in a whole-task assignment.	I
<hr/>				
6				
i	Evaluating	3	Considering the quality of each student's own text, using a text scale to determine to what extent the chosen strategies helped to meet the quality criteria agreed on	I
ii	Evaluating	3	Giving peers feedback, based on previously discussed text quality criteria (in Lesson 1) to experience the effect of one's text on a reader and determine to what extent the chosen strategy was adequate.	I, G

a: DP = Design Principle: 1 = Cognitive Strategies; 2 = Writing Routine; 3 = Metacognition: Choice; b: I = individually; G = first in groups of approx. four students, followed by a short (teacher-led) whole class discussion

### 2.3.1 The design-as-constructed

*Whole Task Experience.* Because students had little prior knowledge about synthesis tasks, we showed them an instructional video before they started the whole-task assignment that preceded the unit. Figure 3.1 shows a still of this 3-minute animated video that provided students with information about integrating source information in synthesis texts and showed what they had to do

(product task representation), but not yet how to do it (strategies). For more information about the content and effects of the video, see Chapter 4. The way students deal with a new task is often determined by their own interpretation of the writing task. By providing them with both visual and verbal information, we tried to ensure that all students started the whole-task assignment with similar prior task knowledge and a valid task representation, in line with Merrill's principles for engaging with real-world problems and activating prior knowledge (Merrill, 2002).

*Preflection.* The purpose of lesson 1 was to create a definition of a well-written synthesis text. In phase (i) of lesson 1 (see Table 3.1), students compared four texts written by anonymous peers in response to the same assignment as the one the students had received for the whole-task assignment. In doing so, they focused on one out of four quality aspects: (a) selection of relevant source information, (b) integration, (c) structure, or (d) style, and subsequently ranked the texts according to their quality. In phase (ii), the students refined and elaborated their task representation through a group discussion in groups of four. Each group member assessed the texts on another aspect and together the group determined a joint, final ranking. We opted for this method to ensure that students would discuss the various quality aspects in relation to each other as well as each aspect's contribution to text quality in general. Finally, in a teacher-led discussion, the class collectively determined what the characteristics of a good synthesis text are. The teacher put these characteristics on the whiteboard and the students copied them in their workbooks.

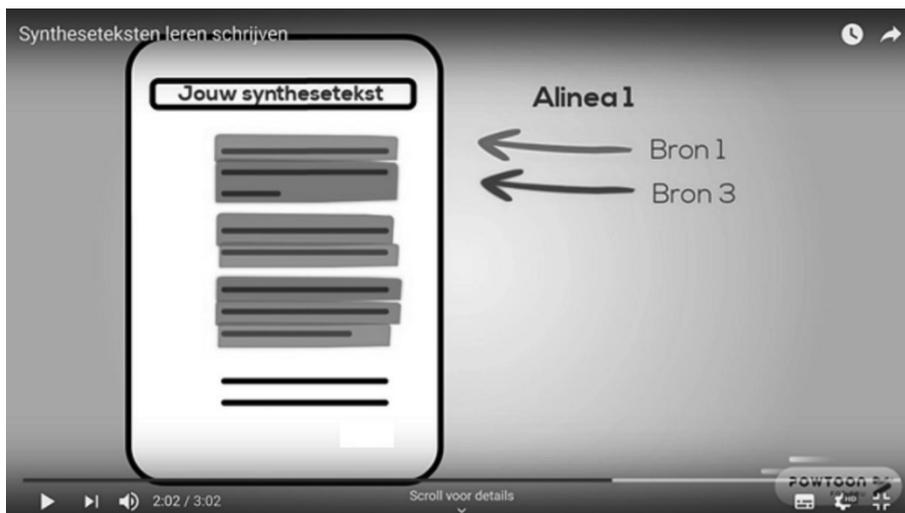
*Strategy-instruction.* Lessons 2 through 4 all followed a similar pattern (see Table 3.1). Here students were presented with strategies to successfully complete the synthesizing sub-processes by observing models in film clips. Subsequently they discussed and then practiced these strategies, adding a strategy in each subsequent lesson to what they had learned in previous lessons.

Each strategy was instructed through four phases: (i) observational learning by comparing and contrasting different strategies deployed by two peer models, (ii) discussing what had been observed, first in small groups and then in class discussion, and then choosing which strategy to apply, (iii) applying the chosen strategy individually, and (iv) follow-up discussions in groups about the results of their lesson assignments, first in small groups, followed by a final plenary classroom discussion.

Different strategies and choices were offered implicitly through modeling in each lesson, but in Lessons 3 and 5, we invited students to choose explicitly by writing down and explaining their choice in their workbooks, to stimulate their

writing process awareness (Principle 2). We designed two different planning strategies: Preplanning by outlining or planning by Drafting, based on research by Kieft et al. (2007), and Kieft and Rijlaarsdam (2008). The first strategy involves writing down the most important information from every source text, in key words, on Post-it notes (different colours for the different source texts), and then organizing the notes (with information on the same subtopics, possibly from different sources), before writing a first draft. This strategy is based on the Color-Coding Method (Darowski et al., 2016; Lundstrom et al., 2015) and is most likely to meet the needs of students with a preference for Preplanning. The second strategy involves immediately writing a rough draft, while simultaneously clustering and organizing information from different sources. This strategy is likely to meet the needs of students with a preference for Drafting. In Lessons 3 and 5 students could choose between these two strategies. A detailed description of all strategies the models show in the film clips and the contents of Lessons 2 through 4 can be found in Appendix B.

Figure 3.1 Still from the 3-minute animated video referring to integration



*Note.* Translations for the Dutch words: alinea = paragraph; bron = source; jouw synthesetekst = your synthesis text

*Task integration.* Lesson 5 provided students with the opportunity to practice the recently observed strategies once more in a new, whole-task synthesis assignment, in line with Merrill's integration principle (Merrill, 2002). At the start of this lesson, students again decided which of the two planning strategies they

wanted to apply in the new task. Then they selected a topic to write about: regional dialects or animal language.

*Evaluating.* Lesson 6 was entirely devoted to evaluating the text written during Lesson 5. In the first phase of the lesson, students compared the text they wrote in Lesson 5 to model texts, using a text scale, which consisted of three anchor texts that represented examples of synthesis texts of lower, average, and higher quality. Each anchor text was accompanied by an elaborate explanation for each of the four quality aspects identified in Lesson 1: information, integration, structure, and style. Students determined the quality of their text by positioning it on the scale. In the second phase of the lesson, students provided peers with independent feedback in three consecutive rounds using a single point rubric. This feedback from a reader's perspective helped to confirm or nuance the quality as determined by the student himself by using the text scale.

### 3 PART 2. FIELD TRIAL: FEASIBILITY AND EFFECTIVENESS

In this section we present the data from the quasi-experiment we carried out to evaluate the feasibility and effectiveness of the unit described in Part 1 above. Following this part, we provide an evaluation of the design's validity, feasibility, and effectiveness in an overarching Discussion section.

#### 3.1 *Research design*

To empirically evaluate the operationalization of the design, we conducted a quasi-experiment with switching panels (Shadish et al., 2002) and three measurement occasions (M1, M2, M3), as shown in Table 3.2. We assigned five intact ninth-grade classes non-randomly to two panels: the teacher-researcher (two classes) taught Panel 1 (EC), the other teachers Panel 2 (CE). Tests were administered simultaneously in all classes.

*Table 3.2 Quasi-experimental design with switching panels*

	M <sub>1</sub>	E	C	M <sub>2</sub>	E	C	M <sub>3</sub>
Panel 1	o	x		o		x	o
Panel 2	o		x	o	x		o

*Note.* *M* = Measurement occasion (*o* = present),  
*E* = Experimental condition, *C* = Control condition (*x* = present).

During the first iteration (M1 to M2), the unit was implemented in Panel 1, in two classes as part of the regular curriculum for two consecutive weeks, in three 50-minute lessons per week. The remaining three classes formed Panel 2 and served as a control group during this period. They continued with their regular curriculum but did not receive any writing instruction. After the M2 test, the conditions were alternated, and the unit was implemented in Panel 2. The whole intervention took five weeks to complete.

To keep the students equally motivated for all measurement occasions, they could choose which text, written during one of the three measurement occasions, was submitted for marking by their teacher.

### 3.2 *Participants*

Three teachers, one of whom was the teacher-researcher, participated in the study. Two were female, and one was male. They were all certified, varying in years of experience (2-23 years). They taught the unit to 152 ninth-grade pre-university students from five classes at the same school. Class sizes varied from 29 to 32 students ( $M = 30.4$ ,  $SD = 1.02$ ).

During the first iteration of the intervention, the teacher-researcher, who designed the lessons and materials, taught her regular classes (Panel 1,  $n = 60$ ; age  $m = 14.00$ ,  $sd = 0.48$ ). She used her experience in this first iteration to maximize the lesson unit's feasibility for both students and teachers in the second iteration.

Before the start of the second iteration, the teacher-researcher familiarized the two other teachers with the theoretical-empirical background of the intervention, the design of the unit and the use of student and teacher materials. Prior to this training session, the two teachers had studied the materials so that they could ask clarification questions.

During the second iteration the teacher-researcher's colleagues taught the unit in three classes (Panel 2,  $N = 92$ ; age  $M = 14.01$ ,  $SD = 0.52$ ). Students in both panels did not differ in terms of gender (63.3% female in Panel 1, 64.1% in Panel 2;  $\chi^2(2) = 0.10$ ,  $p = .953$ ) or age ( $M = 14.00$  years,  $SD = 0.49$  in Panel 1;  $M = 14.01$  years,  $SD = 0.52$  in Panel 2;  $t(150) = -0.13$ ,  $p = .988$ ). The students' parents received an informed consent e-mail and could object to their child's participation in the study, which one parent did.

### 3.3 Measures

#### 3.3.1 Feasibility

We collected data on student feasibility through analysis of their workbooks and through an evaluation questionnaire in both panels. In addition, we designed an online teacher log to keep track of the implementation of the unit for use during the second iteration and to identify problems, based on the teacher-researcher's experience in the first panel.

*Teacher Logs (second iteration).* The teachers completed a log after each lesson, indicating the extent to which they had completed key lesson activities (fully, partially, or not). The response rate was high (98%). We collected ratings for 50 key lesson activities. If the teachers indicated they "fully" or "partially" completed an activity, they also evaluated the activity's feasibility on five items: (a) how much order and discipline they experienced in the classroom during that activity, (b) how clear it was what was expected of the teacher, (c) to what extent the students understood what they had to do, (d) how interested/engaged the students were, and (e) how feasible this phase was for the teacher to carry out. The feasibility items formed a reliable scale ( $\alpha = .89$ ). If the teachers did not do or complete an activity, they could indicate the reason (e.g., 'not enough time', 'I deliberately left it out') and provide an explanation. We analysed how many key lesson activities were "fully", "partially", or "not" completed and how their feasibility was rated by the teachers.

*Student Workbooks.* We analysed workbooks (Lesson 3, 5) to gain insight in the feasibility of the choices between strategies. Available data were the strategy students' chose and the reasons for their choice. We also analysed the reasons they gave for substantiating their choice; based on a selection of one third of all responses, from one class in the first and another from the second panel ( $n = 58$ ; with six participants generating missing data). Via bottom-up coding we distinguished 17 categories, four of which were clearly indicative of awareness of one's own writing process:

1. evaluating one's own writing process,
2. referring to one's own writing process,
3. referring to self-efficacy, and
4. referring to a transformative writing belief.

A fifth category, in which students stated a learning-oriented goal, could indicate writing process awareness, but not in all cases. We considered comments in the other categories as indicating that students were not (sufficiently) aware of the influence of their choices on their writing process, because their responses did

not clearly demonstrate metacognitive insights (see Appendix C for all categories with examples of the students' responses).

*Student Questionnaire.* Students filled in an online questionnaire to provide feedback on the unit. The first set of questions concerned the unit as a whole and students' sense of safety during the learner dialogues. Students scored evaluative words that followed the sentence "I found the unit ...", for example "useful", "dull", "clear", etc. (10 items). They scored all items on a 5-point scale with 1 stating complete disagreement and 5 complete agreement. We recoded negatively formulated items so that they matched the positively formulated ones before further analysis. A subsequent principal component analysis with Direct Oblimin rotation (SME = .78; Bartlett's test  $p < .001$ ) revealed that three components explained 64% of the total variance: engagement (37.8% variance), including words such as "nice", "interesting" and "dull", overall clarity and comprehensibility (15.4% variance), including "clear", "difficult" and "confusing", and overall relevance (10.7% variance), including "meaningful" and "innovative". Therefore, we will report findings related to these three main categories in the results section.

Additionally, the students evaluated the synthesis tasks which they could choose between in Lesson 5, so we could determine their feasibility: did students experience them as equivalent in terms of their difficulty and did they find their topics appealing? Finally, students evaluated four (recurring) key learning activities, indicated how much they thought their synthesis writing skills had improved, and provided us with what they experienced as the most striking strength and the most urgent suggestion for improvement of the unit. They reported on strengths and possible improvements in response to an open question, so their answers were categorized for further analysis.

### 3.3.2 Effectiveness

*Measurement Tasks.* Students wrote an informative synthesis text of approximately 200 words, with three different tasks, each nested in a specific measurement occasion. To ensure the validity of the tasks, we used tasks that were designed and tested as part of a national assessment project on synthesis writing (Vandermeulen et al., 2020a). The topics were related to current affairs (i.e., the human-wildlife conflict in Africa, self-driving cars, and food additives) and included three short, informative source texts (roughly 190 words per text) per task, that partly complemented each other content-wise and partly overlapped.

*Text Quality.* We instructed raters to provide a holistic score, based on four foci: (a) representation of source content, (b) integration of source information,

(c) structure, and (d) style/language. These four aspects together reflect what is generally regarded as the basis for determining the quality of synthesis texts (e.g., Boscolo et al., 2007; Martínez et al., 2015; Segev-Miller, 2004; Spivey & King, 1989; Vandermeulen et al., 2020a).

To allow the quality of all 432 texts to be assessed by three independent raters, we designed a procedure with overlapping rater teams. For each measurement occasion, we created portions of about fifteen anonymized texts, anonymized for order and condition. Each rater received three different portions and each portion was rated by three independent raters. The jury consisted of 21 raters in total (teachers, preservice teacher students, and former teachers). Interrater reliability (correlation coefficient following Van den Bergh and Eiting, 1989) was moderate ( $\rho = .71$ ).

Each rater received written instructions before the actual assessment task: (1) students' assignments and source texts, and (2) a benchmark scale, that showed the possible variation in quality by offering a range of sample texts, including explanations of their characteristics. The use of benchmark scales is known to increase the validity of holistic judgement (Pollitt, 2012) and ensures raters will be less likely to adapt their judgements as the benchmarks serve as fixed reference points (Bouwer et al., 2016). Raters scored the four quality aspects on a scale from 1 to 5, and then judged them holistically on a scale from 1 to 100.

*Analyses.* To analyse the data, we applied mixed model analyses with Student as a random factor to account for the dependencies of scores due to intact classrooms and repeated measurements. We started our analysis with a model in which only the mean holistic text quality and two variance components (within students and between students) were estimated (Model 0) after which we systematically added parameters to this model (Model 1: Time measurement occasion; Model 2: Condition; Model 3: Interaction between time and condition). The fit of these models was compared using a likelihood ratio test.

## 4 RESULTS

In this section we present the outcomes of the empirical part of the research in terms of feasibility (4.1) and effectiveness (4.2) of the intervention.

### 4.1 Feasibility

#### 4.1.1 Teacher logs

The two teachers that taught Panel 2 reported that they did not skip key lesson activities during the unit's execution. In general, the lessons were taught as

intended: 92.2% of the key lesson activities were fully completed and 7.8% partially, which were all part of the last lesson. In that lesson, students were meant to give each other independent feedback on their texts in three consecutive feedback rounds, but the teachers indicated that this learning activity was hard to organize and that students lacked time to complete all feedback rounds.

Only 2 out of 12 different key learning activities were found to be somewhat less feasible. Teachers reported that the students did not fully understand how to use the single point rubric we provided ( $M = 3.0$ ,  $SD = 0.66$ ) in Lessons 4 and 6 to give feedback and did not understand how to apply the text scale in Lesson 6 ( $M = 2.8$ ,  $SD = 0.34$ ). Finally, both teachers made it clear that students' conversations during the group discussions tended to drift off topic after a few minutes. Nevertheless, the teachers were generally positive about the overall feasibility of the lesson activities ( $M = 4.01$ ,  $SD = 0.92$ ).

#### 4.1.2 Student workbooks

To gain insight in how students dealt with the strategic choices we gave them in the unit, we analysed their workbooks. In both Lessons 3 and 5, they noted in their workbooks which information organization strategy they chose: (a) Preplanning (Post-it strategy) or (b) Drafting (direct drafting strategy). Table 3.3 shows how their choices were distributed.

*Table 3.3 Percentages of students' choices for information organization strategies (N=152)*

	Planning by		Unclear/absent
	Preplanning	Drafting	
Lesson 3	44.7	50.7	4.6
Lesson 5	43.4	43.3	13.2

Additionally, to gain insight in their metacognitive considerations, we analysed one third of all students' explanations of their choices. Almost all students who chose the Preplanning strategy (96%) provided a positive explanation, mostly indicating that they expected it would help facilitate the text production phase. Those who chose the Drafting strategy, mainly gave negative comments regarding the Preplanning strategy (63%), considering it to be too time consuming, while the other 37% chose the Drafting strategy based on a positive orientation

towards it. Overall, only 15.5% of the responses showed (some) metacognitive awareness of one's own writing process.

#### 4.1.3 Student questionnaire

After completing the intervention, students evaluated four aspects of the unit (see Table 3.4). Although the intervention did not score high on overall appreciation, it did score above average on the other three components. Based on ANOVA's we found no significant differences in appreciation between the two panels (see Appendix D).

*Table 3.4 Students' appreciation of the unit on a 5-point Likert scale (N=118)*

Component	Cronbach's alpha	no of items	Mean	SD
Overall appreciation	.78	4	2.6	0.7
Clarity and comprehensibility	.78	3	3.5	0.8
Overall relevance	.68	3	3.6	0.7
Sense of safety	.77	3	4.0	0.8

Additionally, students rated choosing between the topics dialect and animal language in the fifth lesson, as "meaningful" ( $M = 3.7$ ;  $SD = 0.9$ ), and "fun" ( $M = 3.7$ ;  $SD = 1.0$ ). They indicated that they did not find it hard to choose a topic ( $M = 2.1$ ;  $SD = 1.0$ ). The choice distribution and appreciation for both topics were reasonably balanced: 42% chose "animal language" and 53.8% chose "regional dialect". They rated the topics on a scale from 1 (not interesting at all) to 10 (very interesting): "animal language" ( $M = 6.7$ ;  $SD = 1.7$ ) and "regional dialect" ( $M = 6.1$ ;  $SD = 2.0$ ). An independent t-test indicated that the difference between these ratings was not statistically significant:  $t(112) = 1.67$ ,  $p = .09$ . Finally, almost three quarters of the students (74%) reported that they felt they had improved their synthesizing skills after the intervention (score > 3).

In total, students listed 117 strengths of the unit. The largest category of strengths (25.6%) pertained to the learning outcomes of the unit. Students indicated that they had learned something new, that they had learned a lot and had learned it thoroughly. 14.5% of the strengths pertained to the clarity and coherence of the unit. Students indicated that the structure, goals, and content of the lessons were clear. 10.2% of the strengths pertained to being given enough opportunity to practice. Other strong points students mentioned were related to the film clips (7.6%), group discussions (5.9%), giving and receiving feedback

(5.9%), the strategies offered (5.1%), usefulness (5.1%), diversion (5.1%), and 'other' (14.5%).

Students listed 110 suggestions for improvement, 29% of which pertained to the unit being found dull. In addition, 20% of the suggestions concerned the unit's relatively large number of lessons, while 15.4% of the suggestions pertained to lack of time to complete assignments, 13.6% to the film clips, and 8.1% to the learning materials that were still under development during the implementation in Panel 1. The "other" category (13.6%) contained answers that were mainly related to students' personal experiences such as "participating better in class", and "being allowed to make groups yourself".

#### 4.2 Effectiveness

Text quality scores were used as an indicator for the unit's effectiveness. We expected that after participating in the unit, the students would write texts of a significantly better overall quality than students in the control group. Table 3.5 presents model comparisons for the dependent variable holistic text quality, with parameter estimates for model fit.

*Table 3.5. Effect of condition and time on holistic text quality*

Model	Models	$\chi^2$	Npar	Mod-els	Comparison		
					$\chi^2$	df	$p$
0	Intercept plus random factor (ID)	8866.123	3				
1	Plus: Time	8769.850	5	0 vs 1	96.273	2	<.001
2	Plus: Condition	8765.974	6	1 vs 2	3.875	1	.049
3	Plus: Interaction Time*Condition	8648.362	8	2 vs 3	117.611	2	<.001

Model 3 fitted the data best, which indicates that the effect of time was dependent on condition. Note that tasks in all three measurement occasions were similar, but not equal, which implies that score differences between measurement occasions are difficult to interpret as improvements, unless the change in scores clearly differs between conditions.

Table 3.6 (also Figure 3.2) shows that text quality appears to improve to a great extent from M1 to M2 in the experimental condition 1,  $t(55) = -6.12$ ,  $p < .001$ , as compared to the control condition,  $t(83) = -1.20$ ,  $p = .23$ . Between M2

and M3 text quality in Panel 2, now the experimental condition, does not change significantly,  $t(86) = -1.51, p = .13$ , while in Panel 1, at this time the control condition, a significant decrease is observed,  $t(54) = 4.66, p < .001$ . However, paired samples t-tests indicated a significant increase in text quality from M1 to M3 in Panel 1,  $t(56) = -2.48, p = .016$ , as well as in Panel 2  $t(83) = -2.34, p = .022$ .

*Table 3.6 Means (and standard errors), estimated under model 3, for three measurement occasions (M1-M3) per panel*

C*	N	M1	M2	M3
Panel 1 E-C	60	57.87 (1.32)	68.88 (1.34)	62.49 (1.33)
Panel 2 C-E	92	58.78 (1.08)	60.22 (1.07)	63.51 (1.07)

*Note:* \*C = Condition;

E-C = during first iteration experimental group, during second iteration control group;

C-E = during first iteration control group, during second iteration experimental group

*Figure 3.2 Average holistic TQ-score at M1, M2 and M3 for both panels.*

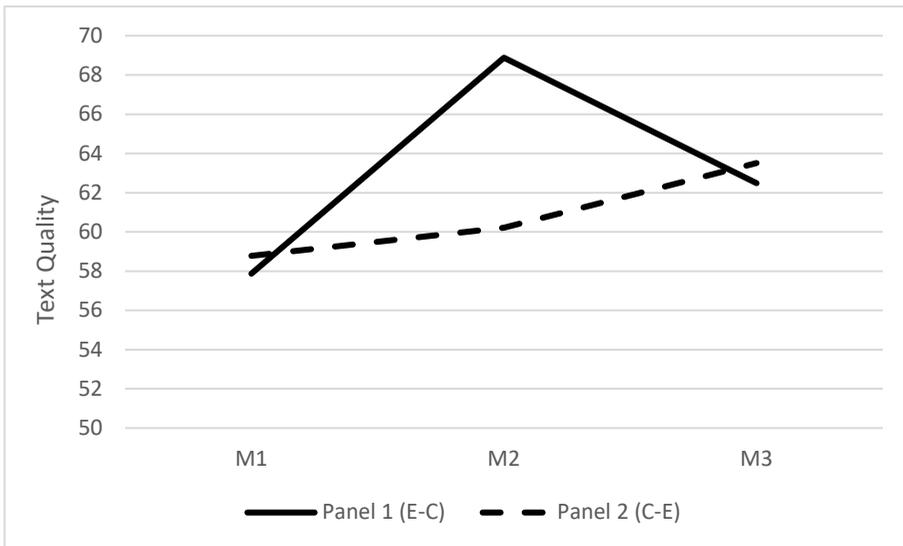


Figure 3.2 shows that at M1 there is no difference between the two groups ( $t(150) = 0.5, p = .59$ ) while at M2 the difference in text quality between Panels 1 and 2 is significant ( $t(150) = 5.0, p < .001$ ), but at M3 it is no longer significant ( $t(150) = 0.5, p = .55$ ). This indicates that the intervention ultimately produced the same

effect for both experimental groups. Additional analyses did not reveal any evidence of a teacher effect. We found no significant interaction between group and time within both panels (interaction in Panel 1 (EC) ( $F(2, 456,656) = 0,60, p = .552$ ); in Panel 2 (CE) ( $F(4, 701,774) = 0.55, p = .697$ )).

## 5 DISCUSSION

This study was set up to gain insight into the quality of the design of a unit which focused on how to write synthesis texts. In this section we reflect on three quality aspects as proposed by Van den Akker (1999), i.e., the validity, feasibility, and effectiveness of the unit, and discuss which adjustments to a potential redesign we suggest based on this reflection. Finally, we consider if the outcomes of this study justify the implementation of the unit in an effect study on a larger scale.

### 5.1 *Validity*

#### 5.1.1 Content validity

*Design Principle 1: Focus on Cognitive Strategies.* An analysis of the six most effective interventions found after a systematic review of the literature (see Chapter 2), indicated that effective synthesis writing interventions focus on cognitive strategies that target all three synthesizing sub-processes selecting, integrating, and structuring. The successful operationalization of this design principle was confirmed by the fact that in the evaluation questionnaire 74% of the students indicated that they felt that their synthesis writing skills had improved. This was further supported by the strengths of the unit they reported, the largest category of which pertained to positive learning outcomes. In a potential redesign, we would therefore choose to leave the operationalization of this design principle unchanged.

*Design Principle 2: Focus on Writing Routines.* Inspired by studies by Kieft et al. (2007), and Kieft and Rijlaarsdam (2008) with 10th-grade students in which instruction in different strategies resulted in different learning outcomes depending on students' initial writing routines, we designed two strategies: (a) Pre-planning (Post-it strategy) or (b) Drafting (direct drafting strategy). An analysis of students' choices showed a balanced distribution between the two options they were given. Both strategies were feasible choices for the students, which indicates that this principle would also remain unchanged in a potential redesign.

*Design Principle 3: Activate Metacognition by Choice.* Based on the study of Robledo-Ramon (2016), we anticipated that the effectiveness of an intervention can be increased if students are taught to make informed strategic choices based

on metacognitive knowledge of their own writing process. Students' responses in our study, however, showed little sign of insight in their own writing process: only 15.5%. We had expected a larger effect of this design principle, but perhaps students lacked a frame of reference and vocabulary to refer to their own writing processes.

For a redesign of the intervention, we therefore suggest creating opportunities to raise students' awareness of their writing processes, perhaps by adding an extra lesson after the first lesson, to provide students with general information about writing processes. Recent research by Van Steendam et al. (2020) explored if effective writing configurations could be distinguished for synthesis tasks. Four different writing configurations were found, of which a production-oriented profile proved to be the most effective, resulting in qualitatively higher texts than the others. Consequently, by comparing their own writing process to other students' writing processes, they might gain a clearer picture of how their processes relate to others', which in turn might give them more insight into the choices they can make during the strategy and practice lessons. Even though such an addition might result in a longer unit, we expect that students will find the unit more varied as a result. This is also likely to increase the unit's feasibility, as the extra lesson is not aimed at practicing strategies step-by-step, but rather at thinking about one's personal writing approach.

*Modes of learning.* The learning content of the unit was mainly offered by means of observing, practicing, and evaluating, in line with Merrill's (2002) general principles for effective instruction. The learning unit was designed to enable students to apply this content independently.

*Design Principle A: Observing.* In the questionnaire, students mentioned observational learning via film clips both as strengths of the unit as well as related to possible improvements. Some comments expressed the fact that students were not used to observing film clips with peer models, for example: "There's no need for film clips: I would rather have just discussed them (strategies) / worked independently on a task". Only a few comments focused on the strategies shown in those clips ("I thought it was good that you got to see those videos and saw that there are different ways of writing synthesis texts"). Most comments just expressed a clear but unfounded opinion regarding this learning activity: they either liked it ("Watching those film clips was fun") or did not ("Watching those film clips was boring") but did not explain why. These types of responses are in line with previous analyses of mixed learner judgments about observational learning through film clips (e.g., Elving-Heida, 2019, p. 102; Van der Loo et al., 2019).

Because observational learning is an important design principle, we would keep the film clips in a redesign, despite students' mixed reactions. However, we would supplement them by providing a clearer context for students and explaining how and why film clips can help them to improve their own writing.

*Design Principle B: Evaluating.* Students reported a relatively high sense of safety ( $M = 4.0$ ,  $SD = 0.8$ ), which is a prerequisite for successful evaluative learning activities. They indicated that they felt sufficiently at ease during class discussions as well as during small group discussion with peers to participate in the conversation and/or to share their work. In addition, two strong points mentioned relatively frequently in the student questionnaire, were the learner dialogues, and giving and receiving feedback. Overall, the outcomes appear positive enough to maintain this principle and its operationalization in a potential redesign.

*Design Principle C: Practicing.* Since applying knowledge and skills in new contexts or tasks is considered a necessary condition for effective learning (Merrill, 2002), which was also reflected in all the effective interventions we analysed (see Chapter 2), we provided the students with plenty of opportunities to practice the strategies they had observed and discussed. Students appeared to appreciate this, as relatively many of the strengths reported (10.2%) related to receiving sufficient opportunity to practice. However, some suggestions for improvement pertained to lack of time for completing the assignments. In a redesign, we intend to resolve this lack-of-time issue, by shortening the source texts for the assignments in Lessons 2 to 5 slightly. By doing so, we would maintain sufficient practice opportunities, while at the same time making completing the assignments more manageable for students.

### 5.1.2 Construct validity

In the student evaluation, the unit scored above average on clarity and comprehensibility. In addition, the largest category of strengths listed by the students (14.5%), pertained to the clarity and coherence of the unit. Responses such as: "Clear lessons and at the end time to look back on what you have learned" and "The sequence of the lessons and their goals were very clear", indicated that the structure, goals, and content of the unit were sufficient. Therefore, we see no reason to change the operationalization of this construct in a redesign.

## 5.2 Feasibility

We had several expectations regarding the unit's feasibility. First, we expected that a feasible learning unit would enable teachers to carry out at least 90% of the learning activities in their classroom practice. Teachers' logs indicated that the lessons were mostly taught as intended. Nearly all key learning activities (92.2%) were fully completed, except for one particular activity in the last lesson when students had to exchange synthesis texts three times with their peers to provide independent feedback. In the redesign we intend to form dyads for providing feedback, which should give students enough time to receive and evaluate the feedback.

Second, we expected that students would be able to understand the instructions and complete the tasks within the allocated time. However, both teachers in Panel 2 indicated that students did not fully understand how to use the single point rubric we provided for giving useful feedback. In a redesign, adding an example teachers can show the students before carrying out this task will hopefully clarify how to use a single point rubric for providing feedback. Furthermore, teachers noticed that students' conversations during the group discussions tended to drift off topic after a few minutes. To maintain this key-learning activity in the design, we will reduce the allotted discussion time from 5 to 3 minutes.

Third, we expected that teachers would find a feasible unit attractive to work with. Our results indicated that they were generally positive about the overall feasibility of the unit ( $M = 4.01$ ). Finally, we expected that the unit would be feasible if students reported feeling that their participation in the lesson unit was worthwhile. Unfortunately, the unit did not score high on overall appreciation in students' evaluations. This is not surprising, given the fact that other studies (e.g., Elving-Heida, 2019, p. 105) have shown that students generally do not tend to perceive intervention lessons as more or less attractive than regular lessons. However, teachers indicated that students generally participated actively during the lessons and were able to complete most of the tasks within the allotted time. Furthermore, students rated the overall relevance of the unit and their engagement with an above-average score. Moreover, they reported choosing between two whole task synthesis assignments in the fifth lesson as being reasonably meaningful and fun. Therefore, we concluded that the intervention is still feasible in terms of its attractiveness for students.

### 5.3 *Effectiveness*

#### 5.3.1 Research design

We had several reasons for choosing a quasi-experimental design with switching replications. By doing so, the teacher-researcher tested the unit herself in the first iteration and adjusted any unforeseen ambiguities or problems before the lessons were taught by the teachers in Panel 2. Second, the design was ethically justified, because it enabled all participants in the study to experience the unit and thus benefit equally from the unit. Third, this design offered the opportunity to replicate the study in Panel 2, and to administer a delayed posttest to Panel 1 at M3 to measure their level of retention (Shadish et al., 2002). Of course, the measurement of retention remains quite limited because the second iteration immediately followed the first in the present study. Panel 1 therefore did the delayed posttest just under three weeks after the completion of the intervention and M2. Therefore, we did not measure a real mid- or long-term effect. In addition, we cannot compare the results of M2 and M3, because the students performed different tasks each time and a difference in, for example, the degree of difficulty between the tasks cannot be excluded. We will discuss this issue further in the section below.

#### 5.3.2 Sequence of measurement tasks

The experimental condition resulted in higher scores on the posttest (M2) in the first iteration than the control condition, while these differences between conditions disappeared at the second posttest (M3), after panels had switched conditions. However, the drop in text quality between M2 and M3 is striking and could be due to a difference in difficulty between the different synthesis tasks. This assumption was supported by several assessors who indicated that they had noticed that the sources for the task at M3 were more difficult to integrate than those in the tasks completed at M1 and M2. In a redesign, we would therefore replace the task used for M3 with another synthesis task.

### 5.4 *Limitations and conclusion*

With this study we aimed to provide insight into the choices we made as educational designers. The analysis of various data sources, including teacher logs and student workbooks and evaluations, provided us with a wide range of both teachers' and students' experiences, from which we derived valuable information for further improving the unit. However, this study was carried out within a

certain context that is not representative of all educational contexts. The design was implemented at a single school with a fairly homogeneous group of students, i.e., all pre-university students, who personally knew the teacher-researcher as a teacher at their school. The question is, of course, whether the (re)design would be equally feasible and effective in schools with a more heterogeneous groups of students who do not know the teacher-researcher personally. In addition, it is possible that students' answers in the evaluation were influenced, positively or negatively, by their teacher-student relationships with the teacher-researcher. Their answers therefore probably reflected not only the feasibility of the design, but the context in which this unit took place, as well.

Furthermore, the teachers who provided the lessons in Panel 2 were the teacher-researcher's direct colleagues. They were personally and extensively instructed in a 1-hour training session before the lessons in Panel 2, but they could also ask questions before or after each lesson if they encountered problems. This would be more difficult if the unit is implemented at other schools and can have consequences for the effectiveness of the design if, for example, certain components are implemented differently than intended. It can also be difficult for teachers who do not have a direct collegial relationship with the teacher-researcher to carry out all lessons exactly as intended, as the lessons are quite intensive both in terms of preparation and execution. On the other hand, the three teachers who participated in this intervention, were the designer of the unit, an experienced and an inexperienced teacher. The fact that no teacher effect was found, despite the differences between them, suggests that the unit functions well regardless of teachers' level of experience. This is a promising outcome in view of its future dissemination to other teachers in other schools.

In conclusion, the outcomes of this study appear to justify the implementation of the unit in an effect study on a larger scale. The analysis of different empirical data sources provided valuable information for a potential redesign by showing which adjustments might further improve the unit. Therefore, our next step will be to implement the redesigned unit at various other schools and then compare the effects and evaluations of both versions, to shed further light on their quality and to contribute to the ongoing design research on source-based writing.