Synthesis writing

A comparison of instructional approaches and process measuring methods in English as a foreign language

Büyüktas Kara, M.

Publication date
2024

Document Version
Final published version

SYNTHESIS WRITING

A COMPARISON OF INSTRUCTIONAL APPROACHES AND PROCESS MEASURING METHODS IN ENGLISH AS A FOREIGN LANGUAGE

Müjgan Büyüktaş Kara

SYNTHESIS WRITING

A COMPARISON OF INSTRUCTIONAL APPROACHES AND PROCESS MEASURING METHODS IN ENGLISH AS A FOREIGN LANGUAGE

UNIVERSITY OF AMSTERDAM
Research Institute of Child Development
and Education

Synthesis writing.
A comparison of instructional approaches and process measuring methods in English as a foreign language
ISBN: 978-94-6473-445-4
Cover design: Mujgan Buyuktas Kara, Istanbul
Printing: Ipskamp Printing Enschede

SYNTHESIS WRITING

A Comparison of Instructional Approaches and Process Measuring Methods in English as a Foreign Language

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het college voor promoties ingestelde
commissie, in het openbaar te verdedigen
in de Agnietenkapel der Universiteit

op maandag 29 april, te 14.00 uur

door Müjgan Büyüktaş Kara
Born in Eskişehir

Promotor:	Prof. dr. G.C.W. Rijlaarsdam Universiteit van Amsterdam
Tweede promotor:	Prof. Dr. E. Van Steendam Katholieke Universiteit Leuven
Overige leden:	Prof. dr. D. Atay Bahçeşehir University
	Prof. dr. L. Sercu Katholieke Universiteit Leuven
	Prof. dr. F. Kuiken Universiteit van Amsterdam
	Prof. dr. S.J. Andringa Universiteit van Amsterdam
	Dr. D. van Weijen Universiteit van Amsterdam

Faculteit der Maatschappij- en Gedragswetenschappen

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The research studies in this dissertation were initiated by a needs analysis meeting organised at the English preparatory school of a private university, where I was working as an English Language instructor and a pre-faculty level test writer. The goal of the meeting was to align the writing curricula of the English preparatory program with the writing requirements of the departmental courses. The consensus was that there was a gap in the preparatory school's curriculum, specifically in the area of source-based writing, which was reported to be practically the only type of writing requirement at the departmental courses. I was informed about this mismatch through needs analysis studies that I reviewed as part of my Master's thesis, for designing a content-based writing course to enhance (English) literature students' proficiency development. The needs analyses, until then, had not been put into practice by those institutions. This time, probably because the university (hence the preparatory school) was in the early stages of establishment, theory met practice, and synthesis writing tasks began to be incorporated into the English preparatory program's writing curricula, and to the best of my knowledge, for the first time in Turkey.

In a previous year, taking an EU-funded course at the University of Amsterdam, I had the privilege of meeting Prof. Dr. Gert Rijlaarsdam and subsequently also the research team at the Graduate School of Child Development and Education. The research interests of the team including synthesis writing, strategy instruction and observational learning and the intervention studies that they designed focusing both on writing processes and products, was an inspiration for my own research endeavour. Consequently, under the supervision of Gert Rijlaarsdam and Elke van Steendam, I had the opportunity to conduct a series of studies aimed at enhancing synthesis writing instruction at my institution, while at the same time contributing to the existing body of research on writing instruction. Therefore, this dissertation embodies the first set of empirical studies aimed at improving the writing curricula of preparatory schools at Turkish universities, with a focus on synthesis writing in particular.

In this introduction, we offer an extensive explanation of the contextual factors that characterise the specific research setting. We do so with the understanding that the reader may not be well-acquainted with these factors that are probably specific to Turkey. Accordingly, we present an overview of the national context, L1 and L2 writing instruction at various levels of education, university education in Turkey, the English preparatory schools, the specific institutional context where the studies took place and the instructional components that we investigated for the purposes of the studies in this dissertation. Finally, we outline the overall structure of the dissertation and the individual studies that constitute it.

1 THE NATIONAL CONTEXT

As a developing nation, Turkey has been deeply affected by the global expansion of economies and cultures of the last decades. Fluency in English as a second language (L2) has become a necessity for applying for jobs in the private sector. Many multinational companies provide promotion opportunities only to those with an advanced level of L2 proficiency. However, the Turkish education system has not always been able to invest in the education sector to keep up with the requirements of the globalising world (Gedikoglu, 2005). The fact that public funding is based on taxes is a major problem in underdeveloped and developing countries and the Turkish government started seeing the capital stock and the experience of the private sector as a solution for funding neglected areas of public service (Ozyilmaz, 2008). Thus, for the last two decades, the education sector has been at the forefront of the privatisation policies of the Turkish Government.

In 2012 compulsory education was extended from 8 years of continuous education to a 12-years gradual three-step model consisting of 4 + 4 years of primary and 4 years of secondary education. With this policy change, between 2012 and 2019, the number of students studying at private institutions across all grades increased dramatically, that is, 53% for pre-schools; 40% for the first 4 years and 96% for the second 4 years of primary schools, and 305% for secondary schools[1]. Similarly, the number of private universities almost tripled between 2006 and 2019, growing from 25 to 73. By the year 2021, 38% of all the universities were private with 156 thousand students studying at private universities

[1] *https://sgb.meb.gov.tr/www/dokumanlar/icerik/30*

compared to 1 million 332 thousand students studying at state-run universities[1]. The private universities are clustered in 11 relatively more developed cities out of the 81 cities of the country, and 60% of these universities are situated in Istanbul. This leads to a fierce competition among the private education providers at all levels of education in Turkey.

1.1 *L1 Writing Instruction*

L1 instruction generally maintains a consistent approach across different private and state educational institutions at all levels. In secondary schools, L1 writing is incorporated into the curriculum as part of the Turkish Language and Literature course, with an average of five hours per week allocated to it. The primary goals are to enhance students' reading skills and grammar knowledge, using writing and oral communication as tools to reflect on the reading material. There are five learning units ranging from two to eight weeks in duration, spread over the 17-week-semester, and two semesters each year. In each unit there are several reading texts on a specific genre (i.e., poetry, drama, biography, letter, literary criticism, memoirs from the different periods of Ottoman/early Turkish Literature), followed by comprehension questions, vocabulary activities, literary analysis, and oral/written reflection in response to open-ended questions about the reading texts, or at the end of some selected units with writing tasks that are based on the topic of the unit.

1.2 *L2 Writing Instruction*

There is a substantial variation in L2 instruction across different private and state educational institutions at all levels. In private schools L2 instruction starts as early as pre-kindergarten (pre-K) with varying numbers of hours throughout school years. This variation is contingent on factors such as whether a private school is following a standard or internationally accredited program, and whether they prioritise English or another language as the primary foreign language. In state schools, L2 instruction commences in the 2nd grade at the primary level, with a weekly allocation of two hours. This increases to three hours in the 5th and 6th grades, and to four hours in the 7th and 8th grades.

The variation in approach to L2 instruction in the state schools is greater at the secondary level. Students are admitted to a secondary level education provider based on two criteria: (1) their performance on the high school entrance

[1] *www.yok.gov.tr*

exam and (2) their proximity to a high school in their residential area. The first criterion leads to the placement of 10% of all students in a top state school or in private high schools, according to their specific entry score requirements. The rest of the students are assigned to high schools based on the second criterion. Students are free to move between private and state schools at any level of their primary or secondary education.

Less than half of the students who are placed in state schools based on their high school entrance exam scores undergo a year of intensive English preparatory programme, consisting of 20 hours of English lessons each week out of a total of 28 hours, followed by six hours of weekly English lessons in subsequent school years. The majority of students placed in the nearest high school to their residential area, unless they intend to pursue English language-related programs at the university (e.g., English Language and Literature, English Translation Studies, or English Language Teaching) receive four hours of English instruction at all levels throughout high school.

As for the L2 curriculum of primary and secondary state schools, more emphasis is placed on listening and speaking skills to practice communication as in real-life. Through grades 1-6, writing is covered on word-level and with "limited" focus. Starting from 7th grade, as proficiency builds up, writing is at paragraph-level with speaking and listening still as the main foci. Although the English Teaching Program discourages emphasising language structures, the type of writing tasks that students are asked to complete are comparing two things, writing about a past event or writing about a future plan. These prompts often result in responses in the form of a list of grammatically correct sentences. There are also tasks requiring writing a thesis statement, reasons and topic sentences; however, to what extent this is properly covered by language teachers, is questionable. At 12th grade, students are asked to write an opinion essay, a descriptive text, an argumentative essay and a 'for and against essay'. These tasks need to be covered along with the other three language skills during the weekly four-hour L2 lessons[1]. This barely leaves time for collaborative or independent practice, feedback or multiple drafts, which have proved effective in teaching writing skills.

---

1. *https://mufredat.meb.gov.tr/Programlar.aspx*

Turkish students graduate from high schools usually with a low proficiency in English. For example, at Middle East Technical University, a state higher education provider which only accepts the top 1% of all the applicants taking the university exam, 74.5% of all male students and 68.4% of all female students were unable to pass the proficiency exam held before the start of the academic year, hence were subsequently placed in the English preparatory program of the university (Dayioglu & Asik, 2007). This study attested to the fact that academic success in high school is not predictive of L2 proficiency, which points to problems in L2 writing instruction in Turkey, in general. These problems typically arise due to various factors, including the quality of instructional materials and testing and assessment practices, the predominant conventional methods in L2 instruction that were prevalent in the field, the use of unsuitable language teaching techniques that align with specific student groups (Isik, 2008), an overemphasis on grammar and a heavy reliance on course books (Gomleksiz, 2011), and language teacher education programs that are theory-oriented (Ozturk & Aydin, 2019).

1.3 *University Education in Turkey*

The National Matriculation Exam for university placement is conducted every June in three separate sessions over two consecutive days for high school graduates. Turkish students take this exam to achieve the required scores set by universities and the Higher Education Council, which enable them to enroll in their chosen faculties and universities. International students seeking admission to universities in Turkey are typically required to submit their national high school exit exam or university entry exam scores to the admission committee of the respective universities they are applying to. Alternatively, international students can opt to take the foreign student matriculation exam (YÖS), individually prepared by each university, available in both English and Turkish. They can also submit scores from internationally recognised exams such as the Scholastic Assessment Test (SAT). The specific minimum score requirements vary according to the discretion of each university's admission committee (www.yok.gov.tr).

Private universities, except for the few established prior to the privatisation concerns of the last two decades, often admit students with lower grades compared to the admission requirements of state universities. However, they charge considerably higher annual enrolment fees. Consequently, with a few exceptions, the majority of private universities maintain similar admission grade requirements and student fees, leading to intense competition among them. To attract applicants, private universities strive to address the current demands of various

business sectors within their departments, employ top instructors and academicians in their field, facilitate international university exchange programs, and establish student internships through partnerships with multinational companies. In addition, private universities offer instruction in English as the medium of instruction in more departments than state universities, and they commit to enhancing the English proficiency of all students with varying levels of L2 proficiency before commencing their studies.

Students who do not meet the L2 language requirements of their chosen departments are enrolled in a year-long English Preparatory Program before beginning their studies in their respective departments. The goal of these preparatory programs is to ensure that students can effectively comprehend lectures, engage in discussions with peers and instructors, communicate their own ideas and those of others in both written and spoken language, and express themselves on campus-related social matters. To progress from the preparatory program to the faculty departments, students are generally required to demonstrate a minimum of B2 level English language proficiency according to the Common European Framework of Reference (CEFR). This corresponds to a score of 65 to 80 on the TOEFL IBT or 5.5 to 6.5 on the IELTS, with the specific score requirements varying depending on the university's educational standards. Alternatively, students can achieve an equivalent score on the English proficiency exams administered by the testing units of the English preparatory programs. The proficiency exam, which assesses reading, listening, writing, and speaking skills, typically accounts for 60% of the final passing grade, while other assessment components like quizzes, midterms, writing exams, process writing evaluations, and presentations throughout the year make up the remaining 40%. A total score of 65 to 70 out of 100 is typically considered a passing grade for transitioning from the preparatory program to the faculty departments. Students who provide the required scores from internationally accredited exams like TOEFL and IELTS at any point during the preparatory program may be exempt from other assessment components and the 60-70% attendance requirement of the preparatory schools.

1.4 *The Institutional Context*

During the academic year 2013-2014, the entry score requirements for Istanbul Sehir University, where two of the three studies in this dissertation were conducted, were notably higher than those of most private universities and even

exceeded the requirements of many state universities. The university also stood out for its scholarship availability. Among the 2,080 students enrolled at various faculty departments during that year (including 391 international students), nearly all received scholarships. Specifically, 25% of students received full scholarships, 15% received 75% scholarships, 35% received 50% scholarships, and 20% received 25% scholarships. Only 5% of students paid their university fees in full. Additionally, the university provided various other benefits, such as free laptops to all students, and access to dormitories, books, and food, depending on the degree of scholarship each student was entitled to. This scholarship availability promoted equality of opportunities among students and significantly enhanced the quality of education.

Furthermore, all departments at the university used English as the medium of instruction. Upon enrolment, students had to take a placement test comprising 80 multiple-choice questions covering reading, listening, vocabulary, and grammar. Based on their placement test scores, students scoring below 65 were assigned to different levels of the English preparatory program, ranging from elementary to upper-intermediate. Those who scored above 65 were eligible to take the proficiency test. Students with scores below 65 in the proficiency test were placed in upper-intermediate level classes within the preparatory program. Upon successful completion, they could advance to the pre-faculty course, an in-house level available only to upper-intermediate level achievers. For students who successfully completed the pre-faculty course, the minimum passing score requirement to enter the faculties was set at 5.5 from IELTS, with a minimum score of 5.0 for each of the four language skills. Prior to commencing the preparatory program, students who scored above 60-65 in the proficiency test (graded on a curve), or achieved a score of 6.0 on the IELTS with a minimum score of 5.5 for each language skill, or attained a score of 70 on the TOEFL iBT, were allowed to proceed directly to faculty courses.

The preparatory program encompassed five language levels, including the pre-faculty course, organised within a modular system consisting of five modules, each spanning seven weeks during one academic year. Students who commenced at the elementary level in Module 1 had the opportunity to complete the pre-faculty course by the conclusion of Module 5, provided they successfully progressed through the levels. However, there was an exception in the form of a combined module lasting 14 weeks, which was dedicated to the pre-faculty course. This deviation was necessitated by the fact that the pre-faculty level marked the endpoint of the preparatory program, and the conclusion of Modules 1 and 3 did not align with the start of either of the two semesters. Consequently, students who completed the upper-intermediate level in the 5th

Module of the previous year but did not pass the proficiency test were placed in the 14-week combined pre-faculty course in Module 1 of the following year. This arrangement ensured that these students did not lose seven weeks before commencing their faculty courses. See Table 1 for the distribution of the available pre-faculty courses throughout the academic year.

Table 1
The Available Pre-Faculty Courses in an Academic Year

Modules	1	2		3	4	5
14-week pre-faculty course	A		Semester Break	B		
7-week pre-faculty course		a			b	c

Students in the 7-week pre-faculty courses were generally better-performing students compared to the students in the 14-week combined programs, particularly in Module 1. The latter group, often referred to as *repeat students*, consisted of students who began the preparatory program at the elementary level but failed to progress successfully at least once during the previous academic year. Their challenges arose from either an inability to meet the in-module assessment criteria, which accounted for 50% of their passing grade, such as active class participation, completion of electronic assignments, presentation delivery, and involvement in discussions, or their inability to pass the end-of-module test (MET), which also constituted 50% of their passing grade. On the other hand, the student profile of the 7-week pre-faculty programs generally consisted of well-performing students who studied the pre-faculty program in Modules 2, 4, and 5 based on their entrance level. In the case of Module 4, some students began at an intermediate level and faced a setback at some point throughout the modules. The different student profiles of the 14-week combined pre-faculty course in Module 1 and the 7-week pre-faculty course in Module 4 were the participants in Study 1 and Study 2 of this dissertation, respectively.

The pre-faculty course comprised 20 English lessons per week, with four lessons lasting 50 minutes each scheduled daily from Monday to Friday. Instructors were required to dedicate a total of four lessons every week to teaching writing skills, using materials prepared by a selected panel of instructors prior to the start of the academic year. In the 14-week program, the focus of the writing lessons was structured based on the following objectives:

- Weeks 1-5 emphasised critical thinking and argumentation skills, library skills, paraphrasing, summarising, referencing, and citing sources in the APA style, as well as writing an argumentative essay.
- Weeks 6-7 were dedicated to teaching the skill of writing a synthesis text.
- Weeks 7-14 focused on guiding students through the process of writing an argumentative research paper with a process writing approach.

The 7-week program followed a condensed version of the same curriculum, with half the time allocated to each objective. This arrangement was based on the presumption that well-performing students would require less time for each task compared to the repeat students.

2 THE WRITING TASKS IN THE TOEFL IBT AND IELTS EXAMS

In both TOEFL IBT (IBT: Internet-based test) and IELTS, which are widely recognised internationally as English proficiency exams and are accepted by all state and private universities in Turkey, each of the four language skills (i.e., reading, listening, writing, speaking) are assessed in separate sections. The final score is derived from the average of the four individual sub-test scores, making each language skill contribute 25% to the overall exam grade. One common writing task in both exams is composing a well-organised argumentative essay[1], in IELTS a minimum of 250 words in 40 minutes; while in TOEFL, 300 words in 30 minutes. This task requires responding to a prompt by drawing on personal experiences and prior knowledge. Argumentative writing tasks are also frequently assigned in the proficiency exams of preparatory programs to evaluate students' writing skills.

The second writing task in IELTS and TOEFL introduces the use of secondary sources, such as graph writing in IELTS and integrated listening and reading for writing in TOEFL. In graph writing, candidates are tasked with describing or summarising information presented in a provided graph, chart, or diagram. This should be done with 150 to 225 words, typically within a recommended time frame of 20-minutes in pen and paper conditions. Candidates are expected to compare data, elucidate processes, or explain the workings of something using their own words. The score from this task contributes one-third to the overall writing score. In TOEFL, the integrated writing task[2] requires candidates to compose a 300-word response within an allocated 20 minutes on a computer. This

---

[1] *Should employees be encouraged to work from home? Discuss in a well-organised essay.*

[2] *Summarise the points made in the lecture, being sure to explain how they cast doubt on specific points made in the reading passage.*

task involves synthesising key points from two different sources. First, candidates read a text of approximately 300 words displayed on their screen for three minutes. This is followed by a two-minute audio lecture presenting a different perspective on the topic, also available on the screen throughout the writing task. The sub-skills required include synthesising by summarising and paraphrasing, which are considered to be the most complex writing skills in the academic domain, particularly in L2. The integrated writing task in TOEFL IBT contributes as much to the writing score as the argumentative writing task[1].

3 THE DISCREPANCY BETWEEN THE WRITING INSTRUCTION IN THE PREPARATORY PROGRAMS AND THE EXPECTATIONS IN THE DEPARTMENTAL COURSES

The fact that source-based writing tasks are included in both IELTS and TOEFL IBT writing sections, underscores the importance of synthesis writing tasks in academic discourse. Writing remains the primary medium for showcasing content knowledge at universities (Hinkel, 2015). Hybrid reading-and-writing tasks provide university lecturers with a practical and equitable way to assess a student's grasp of a particular subject. Similarly, the type of writing that students engage in at Turkish universities primarily involves demonstrating comprehension through exam responses or research projects (Sule, 1996). Numerous needs analyses highlight the acknowledgement of the paramount role of reading skills in academic success at Turkish universities, in both L1 and L2 contexts (Arik, 2002; Guler, 2004; Canbay, 2006). Generally reading tasks and writing based on reference and expository texts are the most common tasks at the university level. Consequently, the ability to understand the main idea or the author's perspective and deduce vocabulary from context are the most frequently required L2 skills for content comprehension, especially in departments where English is the medium of instruction (Eroglu, 2005; Sahbaz, 2005). Therefore, an important recommendation for preparatory programs is to enhance reading skills and integrate content into writing in English language instruction, both for students' linguistic development and success in their respective departments (Canbay, 2006). However, there exists a discrepancy between the type of writing tasks in preparatory programs and departmental studies. Source-based writing is not

[1] www.ielts.org; www.ets.org/toefl

integrated into preparatory program curricula. Some preparatory programs provide students with prompts (e.g., main/supporting ideas) during writing exams to ease the cognitive load associated with critical thinking in an exam setting.

There are several reasons for this discrepancy. The first one is the belief that once students are introduced to the academic writing discourse, they will transfer their knowledge to more demanding tasks that require source management skills later in their departments. However, transfer and maintenance of skills, and their generalisation to other content areas do not happen automatically; on the contrary it requires a lot of scaffolding and practice (Harris et al., 2008). Another reason for the discrepancy is the time-constraints of the English preparatory programs that take a toll on the writing instruction. The preparatory schools already have the responsibility of preparing students for their proficiency exams within a limited time of eight months. The preparatory schools contend the belief that their responsibility remains limited to preparing the students to faculty departments merely in terms of general English proficiency, hence they adopt a very systematic approach to teaching writing, in a way to align it with students' L2 development.

The above-mentioned reasons are understandable considering that training students for a reading-and-writing integrated task before they acquire a decent level of English language proficiency is an ambitious pursuit. However, there are several levels of complexity to synthesis tasks. For example, writing an integrated summary from a single/multiple sources is cognitively less demanding than synthesis writing that requires elaboration and integration of ideas, or source management skills (Mateos & Solé; 2009). Phasing synthesis writing instruction by introducing it with simpler tasks and gradually progressing to more complex tasks has been recommended for the effectiveness of instructional practices (Lee et al., 2018; Leijten et al., 2019). Of course, this requires diligent instructional planning, but it has also been shown to be possible even at the university writing courses, all the while maintaining the regular curriculum of the institution (Zhang, 2013).

For preparatory schools in Turkey, this could be a good starting point in developing curricula for teaching academic writing skills and adapting their practices to meet the actual needs of students and the requirements of the faculty departments. Istanbul Sehir University English Preparatory Program was unique in that it included reading-and-writing integrated tasks in its curriculum. Students started working with sources at the intermediate level, where they employed direct quotations as hooks in their introduction paragraphs or to support their ideas in process writing tasks. In the upper-intermediate level, students were required to write opinion essays based on two provided sources, in a

process writing approach. They did so by paraphrasing and summarising the relevant information and integrating it into their essays while adhering to basic APA citation rules. At the pre-faculty level, students further developed their synthesis writing skills. They learned to write an integrated summary based on three source extracts, which was tested in a subsequent timed-writing exam, and they also wrote an argumentative research paper by finding relevant sources to support their arguments using the electronic library catalogues.

3.1 *Modelling Mode of Instruction*

Social Cognitive Theory (Bandura, 1986) posits that observing modelled experiences can stimulate imitative functions by encouraging students to engage in self-monitoring, ultimately leading to self-regulation and metacognition, which fosters observational learning. Observing modelled experiences enhances students' self-efficacy, hence discovering their own capacity to perform similar tasks, especially when they can relate to the model, a concept known as model-observer similarity. This heightened self-efficacy, in turn, can improve their performance in the given task (Bandura, 1997).

Observing modelled experiences alleviates students from the dual task of learning to write and producing the text, allowing them to allocate their cognitive resources to focus on the former, bolstering their capacity as learners (Rijlaarsdam & Couzijn, 2000). Once declarative and procedural knowledge on instructional strategies are internalised, students can apply them in similar future situations and ideally generalise them to other contexts. This self-regulatory behaviour empowers students to take control of their own learning processes and may contribute to their development as academically autonomous learners (Schunk, 1991). Generalising skills to other content areas is particularly important in a preparatory program where the content material is used primarily as a means of enhancing students' L2 proficiency, while students are expected to internalise and apply synthesising strategies for use in their future academic pursuits.

3.2 *Self-Regulated Strategy-Development (SRSD)*

Self-Regulated Strategy Development (SRSD) is a strategy instruction program that involves explicit teaching of strategies through activating and developing background knowledge, discussion, modelling, and memorisation of the strategies. It is practiced through collaborative and independent practice (Harris &

Graham, 1996). Typically, the entire program spans a few weeks and should be customized to meet the specific needs of individual student groups (Harris & Graham, 2009). SRSD is effective in enhancing the writing skills of both struggling and typically performing students at various levels in primary and secondary education, as demonstrated by numerous empirical studies (see Graham & Harris, 2003; Graham, 2006).

Integrating sources into writing augments the cognitive load of writing, which is already a complex and demanding task involving the orchestration of several cognitive operations simultaneously (De la Paz & Graham, 2002). In the context of this dissertation, this complexity is further heightened by the challenge of performing the task in L2 that the students are actively learning. Empirical research has indicated that SRSD is effective in teaching strategies for integrated tasks in L1, such as informative essays and summary writing, as well as synthesis tasks (Mason, 2013; Mason et al., 2012; Mason et al., 2006; Rogevich & Perin, 2008; Martínez et al., 2015). We propose that SRSD can be effectively used to teach L2 synthesising skills to Turkish students, as it has the potential to reduce the cognitive load in academically demanding tasks and strengthen self-regulatory functions through its core elements, presentational (direct) instruction and modelling.

Graham and Perin's meta-analyses (2007) demonstrate that explicit strategy instruction is more effective than non-explicit instruction in enhancing the planning, revising, and editing skills of struggling and typically achieving adolescent students in L1 writing. Similarly, modelling has proved effective in L1 (Dutch) for teaching argumentative texts (Braaksma et al, 2004; Couzijn, 1999) and synthesis texts (Raedts et al, 2007); and in L2 for teaching revision strategies (Van Steendam et al., 2010). Given the time constraints of the preparatory program, it is challenging to implement the entire SRSD package comprehensively. To the best of our knowledge, one empirical study seeked out to determine which mode of instruction – direct teaching or modelling – held greater potential in enhancing students' writing skills, which showed that modelling outperformed presentational instruction of explicit strategies (Fidalgo et al., 2015). Thus, the research conducted in this dissertation aims to ascertain the more promising component of SRSD within the outlined context.

4 THE STRUCTURE OF THIS DISSERTATION

There are five chapters in this dissertation including three empirical studies. Chapter 2, 3 and 4 represent Studies 1, 2 and 3. Chapter 5 is the general

discussion section. Chapter 2 has been published in an international journal, publications based on chapters 3 and 4 are in preparation.

In Chapter 2, we investigate the effects of explicit strategy instruction on synthesis writing instruction at the English preparatory program of a private university in Turkey. The participant profile is academically struggling students (i.e., repeat students). We compare explicit strategy instruction to the regular curriculum practices of the preparatory school where no strategies are incorporated into synthesis writing instruction. We also investigate the differential effects of two modes of explicit strategy instruction, a modelling mode and a presentational mode of instruction. We expected that explicit strategy instruction would be more effective than a no-strategy instruction condition on students' synthesis text quality and writing processes. Additionally, we anticipated that the modelling mode of instruction that is used to deliver strategies would have a distinctive impact beyond the presentational mode of instruction and the regular curriculum practice.

Chapter 3 is a replication and extension, on a larger scale, of Study 1 in Chapter 2 and is designed to further investigate the effects of explicit strategy instruction in synthesis writing, and the modelling and presentational modes of instruction that are used to deliver the writing strategies. The student profile is normally-achieving preparatory program students. We expected the results of this study to corroborate findings from the previous one, that is, modelling mode of instruction to outperform the two instructional mode conditions in effects on the writing products and the processes.

In Chapter 4, we conducted a methodological study, investigating the reactivity of Time-Sampled Self-Reporting, an online process measure that we used in Study 1 and Study 2 to register students' writing processes. The alternate form test-retest, employed in the posttest design of these studies, appeared to be responsive to the training module used to teach the TSSR-method. This might suggest a positive reactivity of TSSR on students' writing processes and products, as it seemed to be the case in Study 2. Such reactivity of the TSSR-method might also affect the online process data that is collected using the technique. We conducted this methodological study at a Belgian university by manipulating the reporting mode (TSSR vs. no-TSSR) in the posttest and instruction mode (modelling vs. presentational) used in the TSSR-training session, in a 2 by 2 between-subjects factorial design and investigated the component of TSSR that possibly caused the reactivity that we observed.

In Chapter 5, we synthesise the conclusions of the three empirical studies with a comprehensive discussion of the primary findings from the studies presented. Furthermore, we offer insights into the implications of our research for theory on writing research, educational practices, and methodological issues. We conclude the chapter by proposing directions and recommendations for future studies.

Chapters 2, 3 and 4 each represent an empirical study that has been prepared for submission to international journals for publication. While this approach allows for individual chapters to be read independently, it has also led to some noticeable overlap in the introduction and method sections across the various chapters.

CHAPTER 2

THE EFFECT OF TWO MODES OF INSTRUCTION: MODELLING VS. PRESENTATIONAL[1]

Abstract
We implemented an intervention of four lessons and tested the effects of two instructional modes as compared to the regular curriculum practices for completing a synthesis task at the preparatory program of a Turkish university. Participants were 48 upper-intermediate English as a Foreign Language learners (mean age = 18) assigned to three conditions. The presentational condition received direct strategy instruction supported by mnemonics; the modelling condition observed a video of a peer doing the task using the same strategies mnemonic. In the control condition, there was no explicit reference to strategies; rather, students inferred the necessary information about writing an effective synthesis text from the instruction and the lesson materials. We hypothesized that both of the experimental conditions would have a positive effect on students' synthesis text quality and writing processes and that modelling of explicit strategy use would have an effect over and above the other conditions. Results showed that students in the modelling condition improved their source use skills significantly more than students in the presentational condition, which was maintained in the delayed posttest four weeks later. No statistically significant condition effect was observed for content and authenticity of students' texts. The modelling condition also showed and reported a more process-oriented approach to writing.

Keywords: L2 writing, higher education, strategy instruction, modelling

---

1. *This chapter was published as Buyuktas Kara, M., Van Steendam, E., Rijlaarsdam, G., & Kuru, H. (2018). The effect of two modes of strategy instruction: Modelling vs. Presentational. International Online Journal of Education and Teaching, 5(2), 460-495. We have converted the US spelling to British spelling for consistency throughout the dissertation.*

1 INTRODUCTION

Integrating content in teaching foreign/second language (L2) writing skills is rare (Hinkel, 2015) and is definitely a challenging experience for instructors and curriculum developers in time-constrained language-teaching programs. English preparatory schools of Turkish universities are no exception. Students are almost always asked to write persuasive or argumentative essays using prior knowledge and personal experience rather than synthesis texts. However, the primary means of receiving information in an academic context is reading (Grabe & Stoller, 2001). Students often write to demonstrate competence across the curriculum in response to content material (Hinkel, 2015). Hence, segregated writing tasks fail to represent academic expectations and are not as effective in improving linguistic skills or contributing to students' intellectual growth as much as integrated writing tasks (Leki & Carson, 1997).

Content-integrated writing should ideally promote the processes described in the "knowledge-transforming" model of writing. In this model, writers actively generate and evaluate content and organise information in a more sophisticated manner than in the "knowledge-telling" model of writing (Bereiter & Scardamalia, 1987). On the other hand, integrating content into writing definitely adds up to the already overwhelming cognitive load that is commonly associated with the writing activity (Mateos & Solé, 2009). Writing a synthesis text entails critical evaluation of information in accordance with genre-specific features and the audience. This multifaceted structure of synthesis tasks necessitates decomposing the several different aspects of the complex reading-and-writing task and addressing each one separately and explicitly to foster performance.

Flower and Hayes (1981) described writing as "a goal directed thinking process" (p. 366) and the role of the writer as that of a problem solver. Solving problems entails employing heuristic strategies (i.e., optional techniques to approach the task at hand). Having conscious access to a repertoire of heuristics (i.e., procedures for writing) can make the writing process considerably easier for the writer; and, fortunately, these heuristic strategies can be translated into teachable techniques (Flower & Hayes, 1977). This is especially important in an EFL (English as a Foreign Language) context, in which the majority of learning takes place in instructional settings.

1.1 *Current Instructional Strategies in EFL Instruction in Turkey*

In 1997 and in 2008, the Turkish Ministry of National Education adopted several policy changes in an effort to reform Teaching English as a Foreign Language (TEFL) practices in Turkey. The primary objective for teaching English at

secondary schools was defined as development of learners' communicative capacity (Kirkgoz, 2007). In this vein, the "communicative" approach to TEFL was introduced at a national level for the first time (Kirkgoz, 2005). Curricular issues, such as selection of teaching materials, curriculum design, and the role of the teacher in the classroom were also defined in line with the main goals of Communicative Language Teaching (CLT) (Ozsevik, 2010).

CLT can be defined broadly as an approach to TEFL that emphasises the participation of the learner in meaningful L2 interaction with a focus on functional and communicative aspects of the language. It has been associated with *implicit learning*, since it provides a naturalistic view of learning the language, that is "acquiring skills and knowledge without conscious awareness", similar to learning the mother tongue. Explicit learning, on the other hand, refers to "learner's conscious and deliberate attempt to master some material or solve a problem" (Dörnyei, 2009, p.3), which is associated with more traditional approaches that dominated the TEFL domain until the beginning of the 1970s (Dörnyei, 2009). Thus, with the introduction of CLT approach, there has been a shift towards implicit instruction of all four skills of the language (i.e., reading, listening, speaking and writing) in TEFL settings in Turkey. However, this created a discrepancy between TEFL practices in Turkey and evidence-based writing practices, particularly in L1.

1.2 *Current Instructional Strategies in L1 Writing Instruction*

Empirical research studies both with learning disabled students and normally performing adolescent students in L1 proved that interventions with explicit teaching of strategies for planning, revising and/or editing text are much more effective in improving writing skills than non-explicit instruction conditions (i.e., teaching text structures, the process writing approach, traditional instruction, practise writing and literature study) (Graham & Perin, 2007). Among strategy instruction regimens, Self-Regulated Strategy Development (SRSD) designed by Harris and Graham (1996) is distinct with its explicit teaching of writing strategies. It is an all-encompassing training program tailored to the needs of the students in terms of its recursive nature and the time allocation of different stages during which strategies are presented, discussed, modelled, memorised, scaffolded and practiced collaboratively and individually. SRSD has proved effective in a myriad of research studies (see Brunstein & Glaser, 2011; De La Paz, 2005), including two meta-analyses (Graham & Harris, 2003; Graham, 2006), as well as in teaching reading-and-writing hybrid tasks (Mason, Hickey Snyder, Sukhram, & Kedem,

2006; Martínez, Mateos, Martin & Rijlaarsdam, 2015). SRSD is especially potent because three very effective components in writing instruction are intertwined in its instruction: promoting self-regulation, direct instruction and modelling of the strategies.

1.2.1 Self-regulation

Deliberately employing strategies involves self-regulation functions, which include self-monitoring, self-instruction, goal-setting, and self-reinforcement (Sawyer, Graham, & Harris, 1992). Once learnt, self-regulation skills can be internalised and maintained for use in similar future circumstances (Schunk & Zimmerman, 1997). This autonomy helps learners control their learning processes (1998). Self-regulation is not necessarily a stand-alone component, since it can also be triggered through other components of strategy-focused instruction.

1.2.2 Direct instruction

Direct writing instruction is a deductive approach to learning. It includes the explanation of rules followed by controlled practice and delivery of explicit feedback (Manchón, 2009). Instruction is teacher-led and conveyed in a presentational mode, especially in the initial stages of the training program. It also draws upon the use of strategies through mnemonics to help create a representational system (Reber, 1976). This representational system enables the individual use of the strategies by gradually releasing the control from the instructor to the student, also known as scaffolding. Memorisation helps fasten this process and is enabled primarily with the use of mnemonics, but also through graphic organisers, think-sheets and/or prompt cards (Baker, Gersten, & Graham, 2003). Mnemonics reduce the task requirements to a single chunk in the required order (Worthen, & Hunt, 2011) and facilitate retrieval by locating information in memory with associations (Malhotra, 1991). Hence, they may alleviate the cognitive load of learning to write complex writing tasks such as synthesis tasks. Mnemonics might be especially helpful if students are "cramming" for an exam (McPherson, 2000), which is a typical circumstance in time constrained EFL programs, as is also the case in this study.

1.2.3. Modelling

A key component of SRSD programs that has been exclusively studied is *modelling*, or its mirrored learning activity, that is, *observational learning* (hereby used interchangeably to refer to the same concept). Several factors have

contributed to the growing body of research in this domain. First, observing modelled experiences revives imitative functions; and, arguably, triggers self-monitoring, self-judgment, and self-reaction, which are the three pillars of self-regulation (Bandura, 1986). Vicarious experiences improve learners' self-efficacy beliefs (Bandura, 1986), which promotes positive learning behaviour (Schunk, 1996; Zimmerman, 1995). When learners observe peers completing tasks successfully, they may form outcome expectations, which in turn motivate behaviour towards achieving the desirable outcome (Zimmerman, 1977).

The simultaneous orchestration of several cognitive strategies is especially difficult for novice writers, as they have not yet acquired skills needed to manage the process of writing that strong writers have (Breetvelt, Van den Bergh & Rijlaarsdam, 1994; De la Paz & Graham, 2002). Observing modelled experiences helps students direct their limited cognitive resources to learning-to-write instead of producing a text and, thus, counteracts the challenge of the "dual agenda" of having to zero in on both (Rijlaarsdam & Couzijn, 2000). Students who observe their peers on task also adopt a more recursive approach to writing and delay the executive writing activities to the later stages of the writing process and engage in more metacognitive activities, such as goal-orientation, and analysing in the initial stages of writing. Students in a non-observation condition, on the other hand, adopt a more linear approach to writing with transcribing during the initial stages and with formulating spread throughout the whole writing process. Thus, observation encourages a purposeful temporal organisation of cognitive activities and this has a positive effect on the quality of final written products (see Braaksma, Rijlaarsdam, Van den Bergh, & Van Hout-Wolters, 2004; Groenendijk, Janssen, Rijlaarsdam, & Van den Bergh, 2008).

We have reason to assume that peer modelling of explicit strategy use may be beneficial in teaching L2 synthesis tasks (the type of task in the present study) to Turkish university students. Writing is an academic requirement that both low- and high-achieving Turkish students find difficult and view as something to "persevere through in order to pass certain exams" (Yavuz & Genç, 1998, as cited in Erkan & Saban, 2011). This notion stems from negative student attitudes and writing apprehension, as well as low self-efficacy in writing (Erkan & Saban, 2011). As observation improves self-efficacy, it can have an activating power towards positive learning behaviour and result in writing success.

Synthesis writing is cognitively more demanding than most other writing tasks (see Mateos & Solé, 2009), as it brings together task requirements such as organising, selecting and connecting (Spivey, 1997). This complexity becomes daunting when the task is in L2, as in this case language proficiency level also

comes into play (Plakans, 2009). Arguably, as observation may activate the "learner" capacity in a learning-to-write activity, students can use more of their cognitive resources for metacognitive and procedural knowledge instead of focusing predominantly on the production process, (Rijlaarsdam, Braaksma, Couzijn, Janssen, Kieft, & Broekkamp, 2005). This may be an effective strategy for them to complete synthesis tasks, where the "procedures" (i.e., organising, selecting and connecting) are more burdensome than in the other writing tasks.

In Turkey students are discouraged to critically question text information (Turkkollu, 1994; Clachar, 2000). However, writing a synthesis requires engagement with the text at a critical level (Mateos & Solé, 2009). For learning to write a synthesis text, students need to familiarize themselves with this new discourse, as well as to produce a new text. In such a complex writing task as synthesising, students may benefit from observation as it has the potential to alleviate the "dual-agenda" (see Rijlaarsdam & Couzijn, 2000) of learning-to-write and producing a text at the same time. In this way, it may also stimulate Turkish students' engagement with the task at a critical level.

Although the three components of strategy-focused instruction (i.e., self-regulation, direct instruction and modelling) have been tested separately and in combination with each other in several studies, only one study has compared the effects of direct instruction and modelling (see Fidalgo, Torrance, Rijlaarsdam, Van den Bergh, & Álvarez, 2015). However, the differences in the participant profile, the tasks and especially the design issue certain caveats in the comparability of the two studies (see Conclusion section).

Another issue to consider is that although we expect effects of peer modelling on synthesis writing, individual differences between learners also affect the results of observation. In Zimmerman and Kitsantas's study (2002), college students benefited more from observation of a coping model than a mastery model. Braaksma, Rijlaarsdam and Van den Bergh (2002), in their study with students in secondary education, showed that when the task is novel, struggling writers benefit more from observing struggling models; and stronger writers from observing a stronger model. This shows that in studies looking into the effects of observation, individual differences should also be taken into consideration, which will be controlled for in this study.

2 THE PRESENT STUDY

We set out to improve the synthesis writing performance in the EFL program of a private Turkish university. Therefore, we tested the effects of a strategy-focused instructional design based on the principles of observational learning by

comparing three treatment groups. In a modelling condition, students observed their peers modelling the use of strategies for completing a synthesis task. In a presentational condition, students received direct strategy instruction without modelling. Thus, the distinguishing feature of the two strategy instruction conditions is the mode in which the strategy instruction was conveyed (i.e., through modelling mode in the modelling condition compared to a direct, presentational format in the presentational condition). In the control condition, instruction was not strategy-focused, so there was no explicit presentation or modelling of a strategy. Teaching in the control condition took place more on the implicit rather than explicit end of the instructional scale, in line with the CLT approach to TEFL. In the control condition, students had to work out the task requirements (i.e., their own heuristic strategies) from the given materials in accordance with the guidance provided during the training session.
The hypotheses of the study are:

1. Hypothesis 1: writing performance. Modelling of strategies results in qualitatively better synthesis texts compared to presentation of strategies (1A), while presentation of strategies results in qualitatively better synthesis texts compared to a control condition (1B).
2. Hypothesis 2: writing processes. Modelling of strategies leads to improved synthesis writing processes compared to presentation of strategies (i.e., more meta-cognitive activities in the initial stages and more executional activities in the later stages of writing) (2A) and presentation of strategies leads to improved synthesis writing processes compared to a control condition (2B).

We also investigate whether the students' motivational orientation is a confounding variable in the analysis of the results and, for generalisation purposes, whether our hypotheses apply to students with different learner characteristics in terms of initial levels of motivation and writing performance.

3 METHOD

3.1 *Participants*

Participants were 48 (54% male; mean age: 18) pre-faculty course students in Module 1 of the 14-week combined program at a private Turkish university. They were a homogeneous group of students in their reading, listening, writing and speaking skills, measured at the end of module tests, which were prepared in

line with IELTS international exam specifications by the test office of the institution. Students' L1 was mainly Turkish except for six international participants (evenly distributed over the three conditions) admitted studying their entire degree at the university (two Syrian, one Afghan, one Iraqi, one Moldavian and one Macedonian). Students did not differ in their motivational orientation prior to the study (see Table 8) but did differ at pretest for text quality (see Table 9), which was corrected for using the pretest scores as covariate in the final analysis. The training and the tests were part of the curriculum apart from a summary task assigned as one of the pretests and a synthesis task assigned as one of the posttests (i.e., the delayed posttest). We informed the participants about the study before the delayed posttest, which would replace their previous grade should they get a higher score in the delayed posttest. Participants could ask for the removal of their data after that time until the end of the Module. All the participants agreed to take part in the study.

Newly enrolled students were not eligible for studying in the pre-faculty course, so all of the students in this combined 14-week program were so-called "repeat students" who had failed at least once at any level in the previous academic year. The reasons for failure were mostly failing to meet academic standards, such as completing assignments, following ethical principles in writing and research and/or attending lessons regularly. Because of the distinct student profile, in Module 1, the management chose instructors with experience in the pre-faculty course with the particular (14-week) group of students.

3.2 *Design*

The number of lessons at the pre-faculty course is 20 lessons per week. There are four lessons of 50 minutes every day. Every week instructors have to allocate a total of four lessons to teaching writing skills with the materials prepared by a selected panel of instructors prior to the start of the academic year. Through weeks 1 – 5, the focus of the writing lessons is on critical thinking and argumentation skills, library skills, paraphrasing, summarising, referencing and citing sources in APA style and writing an argumentative essay. Through weeks 6 – 7, the focus is on writing a synthesis text; and through weeks 7 – 14, students write an argumentative research paper through a process writing approach.

The study was conducted in nine sessions of 50 minutes each. Four sessions were reserved for training in Week 6 (Sessions 1-4) and five sessions for pre- and posttest administration, distributed over weeks 4 and 13: One session to collect baseline data about motivational orientation and text quality in the form of summary writing of a single source (Pretest-session), three additional sessions for

posttest 1 (Session 5), writing log training (Session 6) and delayed posttest administered with the process registration measure (Session 7). Motivation questionnaire and learner report data were collected in a post-session. See Table 1 for the distribution of the sessions across weeks and pre- and posttests of the study.

Table 1
Distribution of the Sessions Across Weeks and Pre and Posttests of the Study

	Session	Week	Activities	Text Genre
Pretests	Test session	4	MSLQ	ARG & SSS
	1-4	6	Training	
	5	7		SYN
Posttests	6- 7	11	Log Training Writing Logs	SYN
	Post-session	13	MSLQ and Learner Reports	

ARG: Argumentative essay; SSS: Summary of a single source; SYN: Synthesis task

We compared the effects of modelling and direct instruction of the use of strategies supported by mnemonics as opposed to a control condition on participants' synthesis writing performance and writing processes in an experimental pretest-posttest design. Participants were randomly assigned to three classes. These classes were assigned to three conditions, with the researcher's class appointed to the modelling condition because the majority of instruction in this condition is conveyed through the peer videos, so her instructional contact with the students would be minimal. The two other classes were randomly assigned to either the presentational or control conditions. All three instructors had around six years of experience in teaching EFL. Two of the instructors were female native speakers of the Turkish language, with English language and literature degrees and both PhD candidates. The other instructor was male, native speaker of English of Canadian origin, with an MA in ELT.

3.3 *Materials*

The focus of this study is the effect of different approaches to teaching a strategy. Two different strategy instruction conditions are compared to a control group. In both experimental conditions a multicomponent strategy to write a

synthesis text was taught, but via a different instructional format. The materials were the same except for the ones immediately related to each instructional format, that is, peer videos, and the two materials for the instruction of the strategies: a slide on the introductory PowerPoint presentation (PPT) with the strategies mnemonic (i.e., TRAMPOLINE) and the accompanying handout.

3.3.1 TRAMPOLINE strategies

We adapted TRAP IDEAS reading-and-writing strategies for summarising (Mason, Reid & Hagaman, 2012) into TRAMPOLINE strategies to write a synthesis text (see Table 2 for the TRAP IDEAS and Table 3 for the TRAMPOLINE strategies). The researcher and a colleague piloted the TRAP IDEAS strategies simultaneously in two pre-faculty level classes in the last module of the academic year preceding the actual experiment. The two trainers then liaised for the adaptations necessitated by the differences in the tasks (summary vs. synthesis), the contexts (L1 vs. L2) and the academic writing conventions for synthesis writing.

3.3.2 Videos

We shot three separate videos using the Camtasia Screencast Program (Techsmith, 2016) for the three sub-strategies: TRAM, P (Paraphrasing) and OLINE. The duration of the videos was 9, 17 and 12 minutes, respectively. All of the videos featured the same model: a female freshman student who had studied the preparatory program in the previous year. The videos were controlled think-aloud protocols written and performed in the English language. We prepared the framework of the script based on a more or less ideal student performance, that is, using the TRAMPOLINE strategies during task execution; but with occasional instances of the most common student mistakes in writing a synthesis text based on an error analysis we did with the instructors teaching in the pre-faculty course. The model refrained from adopting a prescriptive tone. She was asked to mimic an actual account of completing the writing task with the help of the strategies. To ensure authenticity, the model did not follow the script very strictly.

Table 2
TRAP IDEAS Reading-and-writing Strategies for Summarising

Think before reading.
Read the paragraph.
Ask: "What is the paragraph mostly about?"
Paraphrase the important information.
Identify important details to support the main idea.
Delete trivial details.
Eliminate redundant details.
Add a term for a list of words or concepts.
Summarise.

Table 3
TRAMPOLINE Strategies for Writing a Synthesis Text

	Think – Three steps
a.	Before reading the first extract: Think about the purpose why you are given different extracts on the same topic?
	After reading the second and the third extracts: What is the relationship of this extract to the previous one?
b.	What do you expect to learn from the extract?
c.	What do you already know about the general topic/the focus of the extracts?
	Read the extracts
	Ask - What is the main idea?
	Mark the important details
	Paraphrase the main idea and the important details
	(Repeat TRAMP for each extract and OLINE for the whole summary)
	Organise the paraphrased ideas BY
	Linking the ideas with appropriate linkers
	Including APA
	Nesting reporting verbs
	Edit your summary

3.4 *Conditions*

Both conditions are based on the stages of the SRSD program, with memorisation (through the use of mnemonics) present in both of the experimental conditions. The distinctive feature in the two strategy instruction conditions is the mode in which the instruction is conveyed: modelling versus a verbal presentational format manipulated in the experimental conditions. All other features appearing in the stages of SRSD are applied to the lesson plans at a micro level, rather than spread over a long period of time, as is the case in SRSD programs. Thus, the focus of the study is not SRSD as the time allocated for teaching synthesising was predetermined by the administration to be four lessons, which is rather short for an SRSD program to be implemented. Table 4 provides an overview of the distinctive characteristics of the instructional conditions and Table 5 of the training session for the experimental conditions.

Table 4
Characteristics of Instructional Conditions

Component	Modelling	Presentational	Control
Presentational Mode	-	+	-
Peer Modelling	+	-	-
Strategies (mnemonic) for synthesising	+	+	-
Direct Instruction	+	+	-
Collaborative practice	+	+	+
Individual practice	+	+	+

3.4.1 Experimental condition 1: Modelling condition

In Session 1, students discussed a controversial topic aimed at creating a meaningful context for introducing the task and the strategies. Using a Power Point Presentation (PPT.) and a complementary worksheet, the instructor showed participants two extracts from different articles about the topic and then a weak and a strong sample of student syntheses of these two extracts. The comparison of the two syntheses enabled generating common knowledge for completing the task, which to some extent corresponded with some of the strategies in the TRAMPOLINE mnemonic. Next, the instructor introduced the TRAMPOLINE strategies mnemonic and gave each student a TRAMPOLINE handout detailing the steps in the strategy for self-reference. The instructor showed examples of each strategy step by referring to the sample summaries and checked

comprehension through a one-item exercise for each strategy. Finally, the instructor checked memorisation of the mnemonic with a whole class drill. In session 2, students received a synthesis task with extracts from three different articles, observed the model in a video completing the TRAM strategies for extract 1 while thinking-aloud. Subsequently, participants emulated the strategies individually (finding the main idea and the important details) for extracts 2 and 3. Then, the participants watched the model paraphrasing (P of TRAMPOLINE) the main idea and the important details in extract 1. In this video, the sub-skills of effective paraphrasing were shown. In session 3, in groups of four, participants paraphrased the main ideas and the important details of extracts 2 and 3 (subsequent collaborative emulation). The instructor supervised the activity, provided scaffolding and showed possible responses on the board. Finally, students observed the video model showing the OLINE strategies for organising the ideas by using linkers, including APA, nesting reporting verbs and editing text. In Session 4, participants completed a new synthesis task individually with minimal support for practicing purposes. The videos were not available to the students after the screening, but they were free to refer to the TRAMPOLINE mnemonic handout and could ask for minor assistance from the trainer.

3.4.2 Experimental condition 2: Presentational condition

The main difference between the two experimental conditions is that the observation tasks in the modelling condition are replaced with the teacher presentation of the TRAMPOLINE strategies. The instructor taught the strategies in, what can be defined as, presentational mode (Hillocks, 1984), with occasional teacher–led whole class question-and-answer episodes. The content of the first and the last sessions and the sequence of learning content in the 2nd and the 3rd sessions, was the same in both conditions: TRAM and Paraphrasing strategies in the 2nd session and Paraphrasing practice and the OLINE strategies in the 3rd session. In Session 2, the instructor gave the synthesising task to participants, and instead of showing the video she presented the TRAM strategies through extract 1 via a teacher-led question-and-answer session and subsequent individual practice of the strategies by the participants on extracts 2 and 3. This was followed by presentation of P by the instructor. In Session 3, the participants practiced the strategies collaboratively and the instructor supervised, provided scaffolding and showed sample paraphrased sentences to the students. Finally, the instructor presented the OLINE strategies. Session 4 was the same as in the modelling condition.

3.4.3 Control condition

This is the regular curriculum practice of the institution. In this condition we adapted and used a lesson plan previously prepared by an instructor and observed and approved by the administration as part of yearly course and instructor evaluations. This lesson plan also set the premise for the lesson plans used in the experimental conditions, by making concise adaptations; we differentiated the training sessions for the three conditions. The main difference between the experimental conditions and the control condition was that there was no explicit and systematic strategy instruction in the latter. The learning content was the same as in the experimental conditions except for the materials related to the explicit presentation of strategies (i.e., the TRAMPOLINE strategy practice slides on the introductory PPT, the TRAMPOLINE handout and the peer videos. The content of the first session was the same as in the experimental conditions except for the brief introduction to TRAMPOLINE strategies. Giving participants more time for self-discovery of the task requirements filled this absence. The only explicit task requirements on the PPT were: "underline the key points" and "paraphrase," mentioned prescriptively, as well as some sentence-level paraphrasing practice. As participants in all conditions had already studied APA in-text referencing, reporting verbs and linking words in the previous weeks, there was a brief reference to that on the PPT, but no explicit instruction was provided with alternative structures as in the experimental conditions. In sessions 2 and 3, students worked on the same synthesis task as the students in the other conditions. Taking the weak and strong synthesis samples as reference tasks, the instructor asked guided questions to elicit task requirements that are similar to the strategies in the other conditions, that is, finding the main idea, supporting ideas and the formalities of effective paraphrasing in the second session and more surface-level concerns such as reporting verbs, linkers, APA conventions in the third session. Some guided questions were: "What would you include in your synthesis?" "Why did you choose that sentence?" and so on. Although the instructor followed a plan for the overall session, the questions needed to be partially improvised according to the answer of the previous question. Each cluster of task requirements was followed by individual practice, whole class-check and collaborative practice. Students read the extracts and underlined the key points individually. After checking the answers, the participants paraphrased the underlined points in groups of four (collaborative practice). The instructor supervised the activity and showed sample paraphrased sentences, which were reported using various reporting verbs, APA conventions, combined with linkers and required some editing, which students were expected to discover and mention. Session 4

was the same as in the experimental conditions (see Table 5 for the training session for two experimental conditions).

Table 5
Training Session for Two Experimental Conditions. (MC refers to Modelling Condition, PC to Presentational Condition)

	Aim(s)	Contents	Instructor / Student Activities	Teaching Techniques	Materials
			SESSION 1		
MC = PC	Creating a meaningful context for introducing the task and the strategies Developing background knowledge about the task & task requirements	Discussing a controversial topic to create context. Reading the extracts. Comparing weak and strong sample student synthesis texts. Finding task requirements in students' sample synthesis texts. Doing exercises about the task requirements	Instructor facilitates and moderates discussion to set the context and introduces the task with a PPT. Instructor and students discuss the strong and weak points of student samples. Instructor elicits the strategies and gives the TRAMPOLINE handout to the students. Student complete exercises for each strategy. Instructor checks for memorisation of the strategies.	Direct Instruction Brainstorming Guided questions (to generate common knowledge and retrieve joint experiences) Elicitation Awareness Raising Joint reflection Whole class drill	PPT and complementary worksheet with the two extracts, strong and weak student sample syntheses. TRAMPOLINE strategies on PPT TRAMPOLINE handout.

SESSION 2

	Aim(s)	Contents	Instructor / Student Activities	Teaching Techniques	Materials
MC	Modelling of the use of TRAM strategies Scaffolding the use of strategies	TRAM Video and subsequent individual practice Paraphrasing (P) Video	Students read the extracts of a synthesis task and look up the vocabulary using online and paperback dictionaries. Students watch video 1 for TRAM strategies. Individual student emulation. Instructor projects the answer key on the board; gives whole class feedback. Students watch video 2 for Paraphrasing strategies	Peer modelling via video	A synthesis task: *Reasons for the increase in divorce rate* TRAMPOLINE strategies handout TRAM and P videos
PC	Presenting TRAM strategies Scaffolding the use of strategies	TRAM Strategies and subsequent individual practice Paraphrasing (P) Strategies	Students read the extracts of a synthesis task and look up the vocabulary. Instructor explains how to use the TRAM strategies. Individual student emulation. Instructor projects the answer key on the board and gives whole-class feedback. Instructor explains how to use Paraphrasing strategies.	Teacher presentation	A synthesis task: *Reasons for the increase in divorce rate-* TRAMPOLINE strategies handout

SESSION 3

	Aim(s)	Contents	Instructor / Student Activities	Teaching Techniques	Materials
MC	Practicing paraphrasing strategies	Paraphrasing	In groups of four, students paraphrase the main ideas and the details in all the extracts that they worked on previously. Instructor supervises the activity, provides scaffolding and shows sample answers. Students watch video 3 for OLINE strategies.	Collaborative practice Monitoring Scaffolding Feedback	A synthesis task: *Reasons for the increase in divorce rate* -TRAMPOLINE strategies handout - OLINE video
PC			In groups of four, students paraphrase the main ideas and the details in all the extracts that they worked on previously. Instructor supervises the activity, provides scaffolding and shows sample answers. Instructor presents the OLINE strategies.		A synthesis task: *Reasons for the increase in divorce rate* -TRAMPOLINE strategies handout
			SESSION 4		
MC = PC	Enabling independent practice	Completing a synthesis task	Instructor gives the practice worksheet and walks around the students, monitors each student and provides individual help when students ask for it.	Individual Practice Monitoring Scaffolding Feedback	A synthesis task: *Factors that play a role in academic success*

3.5 *Training Delivery and Intervention Fidelity*

We trained the instructors of presentational and control conditions in a one-hour training session; and provided detailed lesson plans and the materials organised in a folder in the order to be followed. As an implementation check, instructors rated their integrity in implementing each of the critical steps of the intervention on a 100-point scale, and we observed each of the two instructors in one session of the experiment. Both the treatment fidelity scores (mean score: 90) and our observations yielded satisfactory results.

3.6 *Measures*

Table 1 shows the list of pre- and posttest measures of the study and their distribution across weeks. We measured text quality with students' exam papers, writing processes with the use of writing logs and motivational orientation with the use of a questionnaire adapted from Motivated Strategies for Learning Questionnaire (MSLQ).

3.6.1 Product measures: Motivation questionnaire data and writing performance

We used an adapted version of the MSLQ questionnaire developed by Pintrich, Smith, Garcia, and McKeachie (1991) by selecting relevant items from the instrument in its original language, that is, English. We selected a total of 17 questions from three scales of the instrument (i.e., six items for task value, eight items for self-efficacy and three items for intrinsic goal-orientation) measured on a 7-point Likert scale, pre- and posttest. The reliability of the whole scale was .93 for the pretest and .98 for the posttest.

We evaluated text quality through three different genres and four different tasks: an argumentative essay and a summary as pretests and two synthesis tasks as posttests. Participants took the tests simultaneously in pen and paper written exam conditions on the dates preset by the directorate. Students wrote on A4 papers with the exam prompts written at the top. Except for the summary, all the tasks were compulsory assessment components of the course counting towards students' General Point Average.

In the argumentative task, students were asked to write an essay of about 350 words in response to a 50-word prompt prepared by the testing unit (See Appendix A for the list of writing prompts). The other sections of the exam (i.e.,

listening, reading and writing) were clustered around one theme selected from previously covered topics. Students received the writing exam paper with the writing prompt in the last 60 minutes of the exam after all other test materials were taken from them. The task was to write a well-organised argumentative essay for or against the given prompt.

In the summary task, students were asked to summarise a textbook article in 150 words. In both synthesis tasks, students were asked to write a synthesis of 150-200 words integrating extracts from three different articles (each one paragraph), hereby referred to as sources, in response to a writing prompt which was around 25 words including the instruction. The tasks were identical to the tasks used for teaching and practicing in all three conditions of the intervention, but the students saw the content of the exam materials including the sources and the prompt for the first time in the exams. To prevent a possible distracting effect of completing the writing logs in the delayed posttest (see Table 5) from putting students at a disadvantage, we extended the duration of the delayed posttest an extra 10 minutes (i.e., 60 minutes as opposed to 50) and gave students easier extracts to synthesise (i.e., 9.8 on Flesch Kincaid readability tests in the delayed posttest, as opposed to 12.2 in posttest 1).

Rating Procedure. For rating purposes, the handwritten student papers were typed (on word documents) to eliminate any possible negative effect of student handwriting on the raters (Klein & Taub, 2005). We trained an outside panel of seven raters for rating the papers. The papers were divided over the seven raters, that is, the raters rated the texts in panels of two or three. All raters were second- or third-year bachelor students of English Language and Culture at the University of Groningen, the Netherlands; 20 to 23 years of age, with Dutch as their mother tongue.

We assessed the argumentative texts on four traits: (1) structural organisation, (2) strength of the argumentation, (3) lexical richness, and (4) range and accuracy of grammatical structures Therefore, we used benchmark essays, since earlier research has demonstrated the positive effects of this rating procedure on rater reliability (see Schoonen, 2005; Tillema, Van den Bergh, Rijlaarsdam, & Sanders, 2012). For each trait, we used as benchmark essay a text from the 48 student texts written for pretest 1 that was of average quality with respect to that feature. The benchmark texts got an arbitrary score of 100. A total of three raters scored the other argumentative essays in comparison with the benchmark texts. If a text was considered twice as good as the benchmark text, it was scored as 200, if it was half as good, it was scored as 50, and so on. Each benchmark text was enriched with a list of the weak and strong points of the text with regard to the

feature that had to be assessed with it, in order to help the raters focus on the right aspects when scoring a text on a particular trait.

Unlike the argumentative texts where students used personal experience and knowledge, in summaries and synthesis texts students worked with sources. For rating of text comprehensibility of the summary and the synthesis tasks, we did not inform the raters about the nature of the task to ensure that they were able to evaluate it for readers not acquainted with the sources. Additionally, the summary and synthesis texts were examined with regard to the incorporation of main ideas, supporting ideas and examples of the sources to summarise/synthesise. Two raters received lists with the main ideas, supporting ideas and examples of the source(s), and had to determine independently of one another the percentage of (1) the total number of main ideas, (2) supporting ideas and (3) examples in each student texts for both the summary and the synthesis tasks.

The synthesis texts were also analysed on authenticity, source comprehension, and source use. For the last two analyses, the same kind of holistic scoring procedure was used as for the grading of the argumentative texts. Authenticity was identified with the function 'Compare and Merge documents' in Microsoft Word. After comparing and merging the original texts with the student texts, the parts of the student texts that overlapped with the source(s) were highlighted. Two raters subsequently calculated how many words of each student text were highlighted on a scale of 0 to 100%. The overlap percentage was subtracted from 100 and the corresponding value constituted the authenticity score of the student. If the overlap was 60%, the student received a score of 40. A higher score from authenticity meant less plagiarism.

Before the raters individually rated the argumentative texts, summaries and syntheses on the different dimensions, they had practiced the rating method together in a short training session, during which they received the benchmark texts, lists with the main ideas, supporting ideas and examples of the original texts, and/or the rating scale for the authenticity assessment. They read them carefully and used them to individually score six argumentative texts, summaries, or syntheses on a particular trait. When the raters differed in their scoring, they discussed possible reasons and solutions for their disagreement.

Table 6 shows the reliabilities of the rating of the texts written for pretest and posttests with benchmark essays. Consequently, we opted to use (the same method and) the same benchmark essays to score the texts of the delayed posttest as the ones in posttest 1. Tillema (2012) and Bouwer, Béguin, Sanders, & Van den Berg (2015) suggest that different tasks in the same genre can reliably be assessed with the same benchmark essays. Therefore, we hypothesized that the

quality difference of the synthesis texts written for posttest 1 and the delayed posttest could reliably be determined with rating scales that were developed for the scoring of posttest 1. The same raters assessed posttest 1 and the delayed posttest. Their scoring of the texts of the delayed posttest appeared to be reliable (see Table 6).

Finally, we calculated the mean of the scores the raters had given to the student texts on the particular traits. We determined the effect of our intervention with these mean scores.

Table 6
Reliability in Cronbach's Alpha of the Text Scoring on the Different Traits (2 to 3 raters per text and 2 to 3 items for Cronbach's alpha)

	Pretests		Posttests	
	Argumentative essay	Summary	Synthesis text 1	Synthesis text 2
Structural organisation	.88			
Argumentation strength	.83			
Lexical richness	.84			
Grammar & punctuation	.90			
Main ideas		.84	.70	.77
Supporting ideas		.82	.86	.67
Examples		.77	.87	.72
Comprehensibility		.94	.87	.82
Correct paraphrases			.79	.77
Source presentation			.93	.91
Authenticity			.97	.96

3.6.2 Process measures: Writing logs and learner reports

We tested the effects of the training on the writing processes of the students through the time sampled self-report method (i.e., writing logs) that Fidalgo, Torrance and Garcia (2008) implemented in their studies with six graders. In this method, students hear a bleep sound at regular intervals of 1-2 minutes during the writing and they are supposed to tick a box on the writing logs indicating the activity they are engaged in at that moment. We administered the writing log measure where participants responded to 45 bleep sounds concurrent with the delayed posttest. We showed the exact numbers of the bleeps projected on

the board in case they lost track of the order of the bleep sounds. The eight activity categories in the writing logs were adapted from the original (Fidalgo et al., 2008) taking into consideration the possible activities that students would do when completing a synthesis task. We also included a simple graphic representation next to each activity category to help students locate the activity on paper easily. The activities in writing logs were categorised and defined as follows:

1. I am reading the sources: I am trying to understand the sources.
2. I am paraphrasing: I am writing the sentences in my own words.
3. I am working on the sources: I am trying to find the main idea important details, writer, year of publication, etc.
4. I am editing: I am making changes to the writing: correcting spelling mistakes, changing/-adding words.
5. I am writing my text: I am writing my synthesis text.
6. I am reading my text: I am reading through part or all of my text.
7. Other: I am doing something unrelated: looking for a pen, looking out of the window.
8. Finished writing.

We adopted Torrance, Fidalgo and Garcia's (2007) strategy to train the students on how to complete a writing log prior to the actual practice (50 minutes in total). First, the participants went through the activity names, their explanations and the graphic representations. Then, they watched a video of a student model doing the synthesis task interrupted with occasional bleeps at different moments. Students were asked on a demo version of the writing log to tick the box, which showed the writing activity category that the model was engaged in at the moment they heard the bleep sound. We made sure that the students were able to distinguish between the different writing categories. We checked and discussed the answers with the students. Subsequently, we simulated the exam conditions and gave the participants a synthesis task similar to the one they would do in the exam and the writing logs. Finally, in the delayed posttest, after a quick reminder of the different categories of writing activities on the writing logs, students wrote their synthesis texts and filled out the writing logs simultaneously.

To increase validity of the results, we used multiple process measuring methods (Schellings & van Hout-Wolters, 2011), so combined online self-reported data of the writing logs with the offline learner report method (De Groot, 1980). The latter was used to provide insight into learners' experiences and identify the extent of conscious knowledge that students were able to retain after the training prompted by open-ended questions (See Appendix B for the questions). After an instruction and a standardization session, two coders (i.e., the researcher

and another instructor) independently coded the responses. The two coders found and underlined the predetermined key words in student reports (i.e., main idea, details, paraphrasing, summarising, organisation, linkers, APA, think, edit, steps/stages, time-management). As a second step, they clustered the key words and relevant semantic units under four categories: main ideas and details under *content management*; paraphrasing and summarising under *synthesising skills*; organisation, linkers, APA under *source use skills*; think, edit, steps/stages and time management under *process knowledge*. The inter-rater reliability was .70 based on a sample of 10 cases.

4 ANALYSES

To test Hypothesis 1, that is, whether peer modelling results in qualitatively better synthesis texts compared to presentation of strategies (1A) and presentation results in qualitatively better synthesis texts compared to a control condition (1B), we used students' summaries and argumentative writings as two pretests and two synthesis texts as posttests. There was a positive correlation between the four subscores of pretest 1 Summary of a single source, that is, main ideas, supporting ideas, examples and text comprehensibility, ranging from (r = .436, p = .003) between main ideas and text comprehensibility to (r = .550, p < .001) for examples and text comprehensibility, and Cronbach's alpha reliability coefficient yielding highly satisfactory results (.77). Thus, we merged these four subscores and created a composite score, hereby referred to as *pretest score 1 for Summary of a single source.* In the same way, a statistically significant positive correlation was observed between the four subscores of pretest 2 Argumentative writing, that is, organisation, strength of argumentation, lexis and grammar ranging from (r = .387, p = .007) between lexis and organisation to (r = .641, p <.001) for strength of argumentation and organisation. Reliability analysis for the different measures of pretest 2 yielded a highly satisfactorily alpha coefficient (α = .823). In subsequent analyses we used a composite score for pretest 2 Argumentative Writing, hereby referred to as *pretest score 2*. These two composite pretest scores provide us with a strong and valid (i.e., as more generalisable) measure of students' initial level of writing skill in two different genres. Controlling for students' initial writing skills across genres by including the two pretest measures as covariates in the analyses will increase the validity of our results (i.e., effect after having balanced out pretest differences in writing skill).

In the two (synthesis) posttests, that is, posttest 1 and the delayed posttest, we had seven subscores, that is, *main ideas*, *supporting ideas*, *examples, text comprehensibility*, *source comprehension, source use and plagiarism.* There was

a strong correlation between the subscores of *main ideas*, *supporting ideas*, *text comprehensibility* and *source comprehension*, with Cronbach's alpha reliability coefficient .88 for posttest 1 and .75 for the delayed posttest. Hence, we merged these subscores and calculated a composite score, subsequently referred to as *content*. Although there was a low correlation between supporting ideas and text comprehensibility in the delayed posttest (r = .264, p = . 070), there was high correlation between the rest of the values ranging between .408 and .680, which indicated a good internal consistency for merging of the sub-scores, so we followed the same procedure for standardization purposes. The examples subscale was expected to belong to the content composite score, but there was no correlation with any of the subscores for posttest 1. For the delayed posttest, there was low correlation with supporting ideas (r = .322, p < .05), but not with other subscores, so we eliminated the score from the final analysis. Therefore, in the final analysis we had three aspects for the quality of the two posttests (i.e., synthesis texts), that is, *content*, *authenticity* and *source use* (see Rating Procedure).

Table 7
A Cross Match of the TRAMPOLINE Strategies, Text Quality Aspects and the (Composite) Scores

Strategies	Text Quality Aspects	(Composite) Scores
Think Read Ask – What is the main idea? Mark the important details	Main ideas Supporting ideas Text comprehensibility Source comprehension	Content
Paraphrase the main idea and the details	Plagiarism	Plagiarism
Organise the ideas BY Linking the sentences Including APA Nesting reporting verbs	Source use skills	Source use kills
Edit your text	(Not rated in text quality Measured with writing logs)	

The triadic subset of TRAMPOLINE strategies corresponds with the three aspects for rating the quality of student texts and the composite scores (See Table 7).

Thus, we expected to observe mastery in the related set of strategies in the corresponding quality of paper, and hence, the corresponding (composite) scores.

To explore whether one of the learning conditions resulted in better scores for a particular group of participants than another condition, we analyzed interactions between the three conditions and two learner variables: motivation and writing skills, both based on pretest scores, on three aspects of posttest text quality: quality of content, source use and authenticity. We applied Hayes moderator regression analyses (Hayes, 2013), as add-in in SPSS, which allowed us to estimate the regions within the moderator variable in which differences between the learning condition were statistically significant, using the Johnson-Neyman procedure. We present the explorations per posttest variable.

To test Hypothesis 2, that is, whether peer modelling leads to improved synthesis writing processes compared to presentation of the strategies (i.e., more meta-cognitive activities in the initial stages and more executional activities in the later stages of writing) (2A) and presentation of the strategies leads to improved synthesis writing processes compared to a control condition (2B), we used writing logs and students' learner reports. To analyse the writing logs, we divided the process time of each individual student (calculated by the total number of responses in the log) into three phases (i.e., Phase 1, Phase 2 and Phase 3) of equal length, based on the number of responses to the bleeps (max. 45, each bleep occurring on average every 90 seconds, at random intervals of between 60 and 120, a practice in line with previous studies for generalisability purposes). It is customary to divide the writing process into stages to interpret the temporal distribution of the cognitive effort over the writing process (Rijlaarsdam, Van den Bergh & Van Steendam, 2015). When the result of the division of the number of responses into three phases was not a whole number, we transferred the surplus value in the decimals to Phase 3 (e.g., for 13 responses for student x, four were allocated to phase 1, four to phase 2, and then five to phase 3). We calculated the frequency of writing activities for each Phase calculated by the number of activity items reported in the logs.

To analyse the learner reports, we counted the number of categorical statements of all conditions and calculated proportions for each condition (by dividing the total number of statements by the number of students in that condition). We applied analysis of variance to analyse the learner reports with the writing pretest as a covariate.

As an implementation check, we tested the improvements in students' motivational orientation over time through MSLQ, by applying analysis of variance, with the measurements as within factor and conditions as between factors.

5 RESULTS

We hypothesized that strategy instruction would be more effective in improving students' text quality and writing processes compared to regular curriculum instruction (i.e., the Control Condition) and that modelling would have an effect over and above the two conditions.

5.1 *Preliminary Analyses*

We checked students' motivational orientation as an implementation check through MSLQ and did not find initial differences between conditions ($F(2,42) = 1.596$, $p = .214$, $\eta^2 = .066$). There was an effect of time ($F(1,42) = 11.724$, $p = .001$, $\eta^2 = .207$) indicating a progress over time, but no interaction between time and condition ($F(2,42) = 0.697$, $p = .503$, $\eta^2 = .030$). See Table 8 for the mean pre- and posttest scores for motivational orientation.

Table 8
Motivational Orientation in Pre- and Posttest (Means and Standard Deviations)

	Pretest		Posttest	
Condition	*M*	*SD*	*M*	*SD*
Modelling	4.80	0.76	5.40	0.92
Presentational	5.09	0.69	5.35	0.74
Control Condition	4.51	1.08	5.21	1.11

(7-Point scale 1: Strongly Disagree, 7: Strongly Agree)

5.2 *Text Quality*

Table 9 shows mean text quality scores for the pretests for the three conditions. For pretest score 2, no statistically significant differences were observed between the three conditions ($F(2,44) = 0.460$, $p = .634$, $\eta^2 = .020$). For pretest score 1, however, an analysis of variance (ANOVA) showed a significant difference between the conditions ($F(2,44) = 3.964$, $p = .027$, $\eta^2 = .162$), so we used Tukey post hoc tests to identify sample means that are significantly different from each other. This analysis showed that there were initial differences between control and the presentational conditions for pretest 1 (MD -21.14, $p = .047$) with the control condition scoring lowest. No differences were found between the presentational and modelling conditions (MD = 3,028, p = 1.00*)* or between

modelling and the control conditions (*MD* = 18,109, p = .062). Consequently, we will include both pretest scores as covariates to adjust for initial level of writing skill.

Table 9
Text Quality Scores at Pretests for Three Conditions (Mean, Standard Deviaton)

	Modelling			Presentational			Control		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
SSS	19	52.01	20.59	12	55.04	22.14	13	33.90	20.27
ARG	18	81.37	15.84	13	86.20	17.75	16	81.20	13.72

ARG: Argumentative essay, SSS: Summary of a single source

Table 10 A shows text quality scores for writing performance in posttest 1 for all three conditions. A multivariate analysis of covariance including the two pretests as covariates with the three text quality measures for posttest 1 (content, authenticity, and source use) showed a statistically significant condition effect: Λ = 0.682, F (6, 74) = 2.566, p = .026, η^2 = .176). Separate follow-up ANOVAs on the outcome variables showed no effect of condition for content and plagiarism was observed (F(2,38) = 2.301, p = .114, η^2 = .108, and (F(2,38) = 0.537, p = .589, η^2 = .027 respectively). For source use, a condition effect was observed in favour of Modelling (F(2,38) = 3.905, p = .029, η^2 = .170).

Table 10 B shows text quality scores for writing performance in the delayed posttest for all three conditions. A multivariate analysis including the two pretest composite scores and also partialling out the effect of the content, authenticity and source use scores of posttest 1 (including 5 covariates), using Wilks's Lambda showed a statistically significant main effect for condition Λ = 0.612, F (6, 66) = 3.057, p = .011, η^2 = .217. For content and plagiarism, no effect of condition was observed (F(2,35) = 0.175, p = .840, η^2 = .010, and F(2,35) = 0.844, p = .439, η^2 = .046 respectively). For source use, a condition effect was observed in favour of Modelling (F(2,35) = 9.426, p = .001, η^2 = .350).

Table 10
Mean Text Quality Scores for Synthesis Texts for Three Conditions (Posttest & Delayed Posttest)

A. Posttest 1

	Modelling			Presentational			Control		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Content	18	67.33	27.59	12	57.06	20.56	13	67.42	21.01
Authenticity	18	85.11	22.58	12	91.25	10.48	13	91.61	12.39
Source Use	18	123.44	51.46	12	76.25	32.20	13	93.77	54.70

B. The Delayed Posttest

	Modelling			Presentational			Control		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Content	18	84.21	18.28	12	87.02	23.34	13	76.44	18.95
Authenticity	18	88.83	16.20	12	87.50	17.52	13	97.00	6.22
Source Use	18	155.17	30.91	12	73.50	41.48	13	120.15	40.71

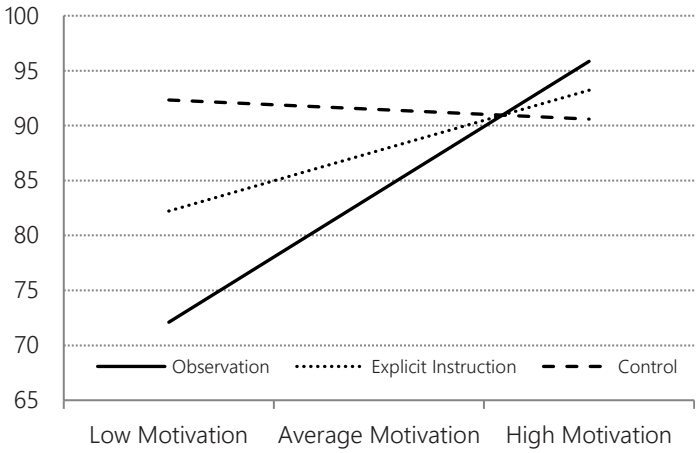
5.3 *Explorations: Interactions Between Learner Characteristics and Learning Condition*

5.3.1 Effect of initial levels of motivation

When testing for moderator effects for students' initial level of motivation (pretest based) and with Hayes moderator regression analyses (Hayes, 2013), a statistically significant interaction effect was shown on source use and authenticity of posttest 1. For the content aspect of the text quality of posttest 1 no statistically significant interaction effect was observed. For source use, the regression between pretest motivation scores and the modelling condition was strongest. Next to main effects of pretest motivation scores (t = 2.36, p = .02) an interaction between motivation and learning condition was found (t = -2.18, p = .03). The

effect was statistically significant for the 35% most motivated participants. In the modelling condition the most highly motivated students produced the best texts whereas the most motivated students in the control condition wrote the poorest texts. The condition did not affect the source use scores of the other 65% of the participants.

Figure 1
Interaction Between Learners' Initial Levels of Motivation and Learning Condition on Authenticity of Synthesis Texts (Y-axis)



For authenticity, we observed a main effect of pretest motivation (t = 3.09, p = .001), and condition (t = 3.06, p = .001) as well as an interaction between motivation and authenticity scores (t = -2.75, p = .01) (See Figure 1). Here the interaction holds for the lower scoring group (37%) on pretest motivation: participants in this group scored significantly lower on authenticity in the modelling condition than in the control condition. The general pattern observed is that the modelling condition is more sensitive to motivation than the control condition (See Figure 1).

5.3.2 Effect of initial levels of writing performance

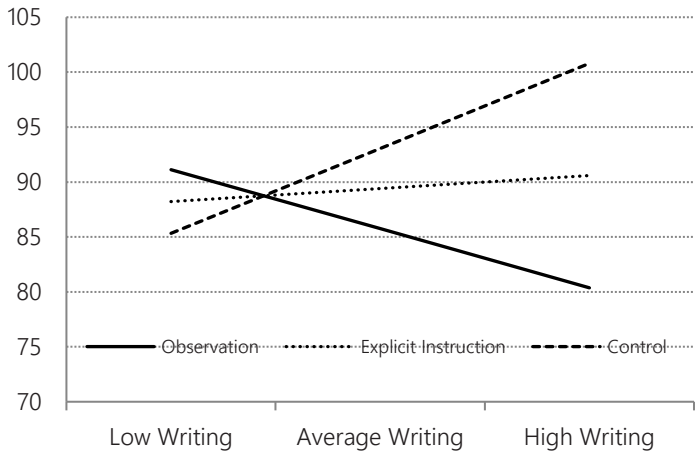
For the pretest Writing Performance, we had two pretest scores that did not correlate to a statistically significant degree (r = .26, p = .086): pretest score 1 and pretest score 2.

For *source use* no interaction effects were observed. For c*ontent* we found an interaction between pretest score 1 and learning condition (t = -3.30, p = < .01) next to a main effect of the pretest scores (t = 2.44, p = .02). The interaction

effect is significant for the 25% lowest scores on the pretest, as well as for the 25% highest scoring group of participants. The effect is strongest in the modelling condition, and non-significant in the control condition. Participants who have a relatively high score in Pretest score 1, score best in the modelling condition; while the participants scoring lowest in the pretest score 1, score lowest in this condition.

Figure 2

Interaction Between Learners' Initial Levels of Writing Performance and Learning Condition on Authenticity of Synthesis Texts (Y-axis)



For the scores on *authenticity* at the posttest, we found an interaction between the pretest score 2 and learning condition ($t = 2.29$, $p = .03$). The effect is significant for the 26% highest scoring group of participants in the pretest. Again, the strongest effect of pretest scores is in the modelling condition: participants with relatively high pretest scores scored significantly lower on posttest *authenticity* in this condition than the participants that started with lower pretest scores. This effect is not observed in the control condition (See Figure 2).

5.4 *Writing Processes*

Table 11 shows the process data for the three conditions. The numbers in the table stand for the average number of the reported activities, that is, the frequency of each activity per phase for each condition. To have an overall picture

of the most frequently reported activity in each condition, we also calculated percentages for each phase by adding up the averages for each one of the eight activity items and dividing it by the average for the relevant activity item.

Table 11
Process Results: Frequency of Activities per Condition and Phase

	Modelling		Presentational		Control	
	M	*S.D.*	*M*	*S.D.*	*M*	*S.D.*
Phase 1						
Reading the sources	2.68	1.77	2.00	1.15	3.81	1.68
Paraphrasing	2.63	2.39	2.23	1.88	3.12	2.03
Working on the sources	3.05	2.20	1.85	1.52	1.87	1.41
Editing	.32	.58	.38	.51	.19	.54
Writing my text	3.16	2.59	4.15	3.18	3.00	2.83
Reading my text	.32	.67	.46	.88	.13	.50
Other	.11	.32	.08	.28	.56	1.26
Phase 2						
Reading the sources	1.68	1.42	1.85	1.91	1.00	1.41
Paraphrasing	2.95	2.70	2.08	2.25	2.19	2.17
Working on the sources	2.89	2.83	.85	1.07	2.19	2.32
Editing	.42	.61	.92	.86	1.31	2.75
Writing my text	3.58	2.69	4.38	2.63	4.75	2.35
Reading my text	.21	.42	.77	1.36	.81	2.17
Other	.53	1.02	.31	.63	.44	.63
Phase 3						
Reading the sources	.32	.75	.38	.77	.94	1.88
Paraphrasing	1.11	1.63	1.46	1.98	1.00	1.21
Working on the sources	1.47	1.81	.77	.93	.56	1.15
Editing	2.11	1.59	1.92	2.43	1.13	2.13
Writing my text	3.68	2.08	3.15	1.86	6.50	3.16
Reading my text	2.47	1.65	2.46	2.11	1.87	1.86
Other	.63	.90	.69	.95	.88	1.31

Results showed that students reported spending most of their exam time, that is, 45 bleeps/60 minutes, *writing* their summaries (i.e., for 31% of the time).
In Phase 1 of the writing process, the modelling and the presentational conditions wrote their syntheses 26% and 38% of the time, respectively, whereas the control condition reported reading the sources 32% of the time. There is a significant difference between conditions in the activity *reading the sources*: the control condition reported spending more time on the activity than the other

two conditions (F(2,42) 4.743, p = .014, η^2 = .184). In Phase 2, the modelling condition and the presentational condition reported writing their syntheses for 31% and 41% of the time respectively, whereas the control condition reported reading their own synthesis texts 35% of the time. A condition effect was observed for *working on the sources* with the modelling condition spending more time on this activity than the presentational condition (F(2,42) 2.358, p = .048, η^2 = .101). In Phase 3, the proportion of total writing time spent in the activity *writing my synthesis* was 33%, 32% and 55%, respectively. The control condition reported spending more time on this activity in this phase than the experimental conditions (F(2,42) 6.678, p = .003, η^2 = .241).

5.5 *Learning Experiences*

Table 12 shows the results from the learner reports for the three conditions (see Appendix B for a sample learner report). The numbers in the table show the percentage of the average number of categorical terms mentioned in each condition.

Table 12
Learner Reports: Mean Percentage and Standard Deviation of Statements per Condition and Category

	Modelling				Presentational				Control			
	Weak		Strong		Weak		Strong		Weak		Strong	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*D*
Content management	3.4	.036	1.2	.026	1.9	.022	1.7	.020	1.7	.020	3.1	.032
Synthesising skills	1.7	.015	2.4	.015	3.4	.010	2.0	.017	1.6	.017	2.4	.016
Source use skills	2.7	.029	3.1	.025	0.0	.000	1.9	.025	1.6	.024	2.1	.019
Process knowledge	3.6	.018	3.2	.039	2.7	.034	.08	.016	1.0	.017	1.4	.020

The majority of the participants in the modelling condition reported learning experiences mostly about process knowledge, in the presentational condition about synthesising skills and in the control condition about content management. No effect of condition was observed for content management (F(2,42) = .171, p = .843, η^2 = .008), synthesising (F(2,42) = .757, p = .476, η^2 = .035) and source use skills (F(2,42) = 2.348, p = .108, η^2 = .101). The only category that

seemed to be sensitive to conditions was process knowledge with an effect of ($F(2,42) = 3.418$, $p = .042$, $\eta^2 = .140$), in favour of modelling condition.

Mean percentages are calculated according to the occurrence of key phrases per student by multiplying the total number of responses by the number of participants in each condition divided by total number of participants in the study. Students in the experimental conditions expressed improved process knowledge, especially about managing the stages and/or steps and using their time more efficiently in the timed-writing synthesising tasks of the Posttests with sentences such as:

> 'I learnt if I want to synthesise, I should paraphrase first.' Student no 8
>
> 'Synthesising is not as hard as I thought. I noticed the ways to write a synthesis in an orderly way.' Student no 10
>
> 'When I looked at the exam paper (before the training) everything interfered. But now I know what I am doing. There are some processes for writing something.' Student no 12
>
> 'Before the training I read all the extracts at once, now I learnt that I have to work on one extract at a time.' Student no 18
>
> 'It works like processing (any) work. You can do everything step by step.' Student no 32

Hence, the learner reports pointed to positive learning experiences and outcomes for students in PO.

6 DISCUSSION

We examined the effects of modelling and presentational modes for strategy instruction as compared to a control condition, which was the regular curriculum instruction of the institution on students' synthesis text quality, synthesis writing processes and learning experiences. This study is unique in that it tests the effectiveness of the separate components in strategy-focused instruction and in an L2 context. We partially confirmed hypothesis (1A). Modelling condition resulted in better synthesis texts in the posttest and the delayed posttest than the presentational condition for *source use* (i.e., knowledge of the available sources, using a variety of citation techniques, correct use of APA conventions, including a variety of reporting verbs and linkers), one of the three aspects of text quality, but not for *content* and *plagiarism*. However, we did not confirm hypothesis (1B). Presentational condition did not differ in effects from the control condition for synthesis text quality.

The success of the modelling condition in two consecutive posttests in the source use aspect of text quality and not in content and authenticity was an intriguing finding. It has been shown in previous studies (Braaksma et. al, 2002), that observation works best for novel tasks. Unlike discursive essays with similar organisational and thematic requirements that the students are mostly familiarized with throughout their preparatory year, writing from sources is a novel task both within the L1 and L2 writing curricula. Source use skills are predominantly related with the specifics of this novel task rather than the common skills across different writing tasks such as paragraphing, argumentation, etc., so this unfamiliarity with the novel features of the task might have contributed to students' success, all of which have been rated under the source use skills aspect of text quality. Another reason for the relatively lower gains of the modelling condition in content and authenticity might be the underlying linguistic skills needed to improve these aspects of text quality, which tap into students' reading comprehension, content selection, paraphrasing, elaboration, and writing formulation skills. These skills are cognitively demanding "higher order" skills (Mateos & Solé, 2009), and they are, to a great extent, dependent on L2 proficiency. Thus, a strategy-focused instruction session conveyed through modelling or presentation may not have sufficed to improve formulation skills in L2 that are required to independently apply the strategies for improving content and authenticity. Improvement in these skills may be possible in long-term strategy training programs, or in an L1 context where it has been shown that speed of access to linguistic features, (i.e., fluency) poses less of a hindrance to writing performance than is the case in an L2 context (Schoonen, et al., 2003).

Preliminary analyses on the motivation pretest showed (see Results) no differential effects of condition on motivation and contrary to our expectations, we did not find an effect of modelling on students' motivational orientation, including their self-efficacy beliefs (Bandura, 1986). Thus, the higher performance of the modelling condition in the source use aspect of text quality cannot be attributed to differences in motivation. However, it was interesting to observe in the analyses for initially more and less motivated students (with initial level of motivation as a moderator variable) that 35% of the most motivated students fared best in the modelling condition for the source use aspect of their text quality; whereas, the most motivated students in the control condition wrote the poorest texts for the same aspect. Hence, for success in mastering the novel aspects of the task (source use in this task), it seems that a higher level of initial motivation combined with modelling yields better results than the regular curriculum instruction in this institution. Another interesting finding was that 37%

of the least motivated participants scored significantly lower in the modelling condition in the authenticity aspect of their text quality. This provides further evidence that modelling is the most sensitive condition to motivation and that in future studies assigning students on the basis of motivation to an instructional condition can be an interesting exploration.

For generalisation purposes, we also looked into the effects of conditions on students with differing levels of initial writing performance (with initial level of writing performance as a moderator variable). Analyses showed that 25% of the highest scoring participants in pretest score 1 (summary of a single source) scored highest in posttest *content* in the modelling condition; and accordingly, 25% of the lowest scoring participants wrote the poorest texts in this condition. However, a reverse effect is observed for the 26% of the highest scoring participants in pretest score 2 (argumentative essay), who scored significantly lower on posttest *authenticity* in the modelling condition than the participants that started with lower pretest scores. The differing condition effects on the success of the students in the posttests could be the result of the different nature of the tasks, that is, summary of a single source in the former and argumentative essay in the latter. As being able to summarise a single source is a prerequisite in a synthesis task, it is presented as one of the strategy steps in the teaching of the synthesis task in the presentational and the modelling conditions. It may explain students' higher scores on the content aspect of the posttests in the modelling condition, which was not observed for the students who scored highest on the pretest score 2. The fact that the results were only observed in the modelling condition and not in the presentational condition is further evidence to the added benefit of observation in teaching of the strategies. These results are in line with the results of previous studies, which have demonstrated different condition effects on learners with different aptitude levels in composing written texts (Braaksma et al., 2002) and on learners with different creativity levels in creating design products (Groenendijk, Janssen, Rijlaarsdam & Van den Bergh, 2013).

We used a multiple methods approach (i.e., writing logs and learner reports) in measuring the writing process to answer the question whether strategy-focused conditions lead to more effective temporal organisation of cognitive activities (Hypotheses 2A and 2B) than the regular curriculum instruction. We evaluated the writing log data according to the parameters we set on the basis of previous studies. Accordingly, instead of executive writing activities spread over the whole writing process in a linear manner, we expected more metacognitive, or planning activities (i.e., reading the sources, working on the sources) in the initial stages and more executive activities (paraphrasing, reading the synthesis text, writing the synthesis text) in the later stages of the writing task. We also

expected two different ways of dealing with the sources: reading to understand the sources and strategic reading of the sources (i.e., finding, underlining and/or highlighting the main ideas, supporting ideas, etc.), as is labeled in the writing logs as *working on the sources*. Thus, we expected writers with effective temporal organisation of cognitive activities to spare a minimum amount of time on *reading the sources* and more time on *working on the sources* especially in Phases 1 and 2 of the writing process. In Phase 3, more executive activities should ideally follow.

The analysis of writing log data could only partially be explained with the aforementioned parameters. In Phase 1, the control condition reported being engaged in the activity *reading the sources* (trying to understand the sources). Their reading time may have been extended due to a failure to strategically work on the sources. However, modelling and presentational conditions did not prove to have better strategies as they were predominantly engaged in writing their texts, an activity ideally expected as the dominant activity in Phase 3. In Phase 2, the modelling condition reported being engaged in working on the sources, a very important indication of effective temporal organisation of cognitive activities, in a way following the TRAMPOLINE strategy steps that were shown to them in the peer videos, that is, planning before writing. In Phase 3, all three conditions reported writing their texts, as would ideally be expected.

In none of the conditions did students report editing their texts. One reason might be that editing was only minimally modelled in the videos with mostly sentence level corrections; thus, it may not have taken up writing process time at a statistically significant level. Previous studies also showed that even in higher education only a minority of students improved revision skills with instruction in L1 (Torrance, Thomas & Robinson, 1999); and in L2, less experienced writers only detected surface level corrections instead of making global revisions (see Van Steendam, Rijlaarsdam, Sercu, & Van den Bergh, 2010). It is also possible that the effects of the training did not emerge within the time span of the training program, because editing emerges later than planning in developing writers in L1 (Berninger & Swanson, 1994), which may also be the case in L2 learners. Additionally, an improved planning phase may have cut down on the time devoted to editing, so there might have been a trade-off in favour of planning (Torrance et al., 2007; Martínez, et al., 2015).

The results of the learner reports showed that the modelling condition reported to have more process knowledge required for writing a synthesis text. They reported learning how to deal with the steps of writing a synthesis text and/or, attributed their success to this process knowledge, a finding in line with

previous research (see Groenendijk, Janssen, Rijlaarsdam & Van den Bergh, 2013). Why this is not entirely reflected in the process registration technique (i.e., writing logs) we used is an issue to consider. The discrepancy might have been a result of low concurrent validity of online versus offline process registration techniques (see Veenman, 2011). As we distributed the learner reports two weeks after the actual writing task, the writing process might have been reconstructed in the writer's memory with possible memory failure and distortions (Schellings & van Hout-Wolters, 2011). Also, at preparatory schools in Turkey, it is customary to write in pen and paper conditions in exams. Thus, for ecological validity reasons, we were not able to combine self-reported data with online process registration techniques such as keystroke logging to record the writing activity in real time and provide data from different perspectives, which give more leeway for interpretation. In future studies, it might be worthwhile to use multiple online registration techniques to get more insight into the writing processes of students.

In evaluating the results, some issues of validity should be discussed. Several studies showed that within a certain participant profile, initially stronger writers (in terms of aptitude) benefit more from observation than initially weaker writers (Groenendijk et al., 2013). Also, strong writers benefit more from observing mastery models and weak writers from coping models, also known as model-observer similarity (Braaksma et al., 2002). Although we looked into the effects of the training on initially weaker and stronger writers within this specific group, and reported the results, we may expect the effects to be different on a group of academically high achieving students. The model in this study was designed to demonstrate a standard student performance; however, considering the low-achieving student profile, we may not have been able to create the best environment for model-observer similarity. In a future study, special attention should be paid to student profile and heterogeneity of the participants in terms of competence and motivation that could affect the results of the study.

Previous research (Braaksma, Van den Bergh, Rijlaarsdam & Couzijn, 2001; Sonnenschein & Whitehurst, 1984) has shown that observation results in larger learning gains when it is set with evaluation and elaboration tasks. In our study, such a task was absent; instead, we opted for presenting the target behaviour in digestible chunks of knowledge (Bandura, 1986), as in the presentation of the three subsets of strategies (i.e., TRAM, P, OLINE) which is, arguably, another way of channeling learner's attention to the modelled activity. Nevertheless, in future studies, it is worthwhile to include evaluation and elaboration tasks in the observation activity to ensure students' utmost engagement with the modelled behaviour.

Another issue to consider is the assignment of the classes to conditions. Participants were randomly assigned to three classes by the administration prior to the training, but the assignment of intact classes to conditions had to be arranged in line with institutional priorities. The main instructor of each class had to give the training in their own classes because having a totally different instructor teach in each class would mean interfering with the natural course of the module, which would not be favoured by the students or supported by the administration. We did not have the researcher give the training in each class because she was the main instructor in one of the classes that was assigned to a condition and in that case her class would have had the privilege to be instructed by their own instructor and the others would not. To control for a possible trainer effect on the results, the researcher's class was appointed to the modelling condition where the interaction between the instructor and the students was minimal compared to the other conditions, because the key element of the intervention, that is, the strategy instruction, was administered through peer videos. Thus, the assignment of the modelling condition was not random, but the presentational and the control conditions were randomly assigned to intact classes. We believe that in future studies, the same instructor should train the students in all conditions or a design with random assignment of students to conditions should be implemented.

Another limitation with regard to the trainers might be their different profiles. Although the trainers in the experimental conditions held a similar profile in terms of nationality, educational background and gender (i.e., Turkish, with an MA in literature, PhD candidates and female), the trainer in the control condition was Canadian, with an MA in ELT and male. This difference may seem to be in favour of the control condition since having a native speaker as the instructor might have improved students' motivation and possibly overall language proficiency. However, it is a strict requirement at Turkish preparatory schools that the lessons are conducted in English (British Council, 2015, p. 93), especially at this university for standardization purposes since half of the instructors were native speakers of English at the time. Therefore, any possible effect of nationality might have been counteracted by the rules of the institution.

We should also point out the fact that students are placed at pre-faculty level classes according to certain criteria. In the previous academic year, they either failed the pre-faculty level exit exam; or passed the upper-intermediate level exit exam, but failed another level previously. Thus, the participants are a homogenous group of learners with an Upper-intermediate initial level of language proficiency, tested through a four-hour exam prepared by the testing specialists of

the university in four skills (i.e., reading, listening, speaking and writing) according to carefully designed testing specifications of the level. Students' initial motivation and level of writing skill in L2 was tested as well, to control for pretest differences. Initial differences in writing were taken into account by running analyses of covariance.

It should also be mentioned that, although we staged our sessions similar to an SRSD program especially focusing on two of its main stages, that is, *modelling* and *memorisation* (through mnemonics), this intervention is not a strategy training program, but an attempt to improve the regular curriculum instruction of a specific task through strategy-focused instruction. Although full strategy training programs are likely to yield better results, we found it worthwhile to decompose the training program and address each component separately to see the main acting agents in the training. This study is also an example of how training programs can be tailored to fit in a realistic time slot allocated for the actual teaching of a specific task. Although this is one reason why the results are not generalisable to studies that operationalise a form of SRSD, it is also a strong aspect of the study towards ecological validity as trainings in hectic EFL programs need to be carefully fit in the time slot allocated for specific tasks for reasons of practicality.

It has been suggested that to be able to reach generalisable conclusions in L2 research with secondary level students, there should be a total of 3 to 4 assignments per student each rated by two raters (Schoonen, 2005). In this study, we measured students' text quality in three different genres (i.e., summarising, synthesising and argumentative writing) with four different tasks, two pretests and two posttests. Thus, the results of text quality are generalisable within the strategy-focused writing intervention programs and provide grounds for future strategy-focused interventions in L2 writing.

7 CONCLUSION

Our results are complementary to previous studies, which proved the effectiveness of the modelling component of strategy-focused instruction on certain aspects of writing performance. There are two unique aspects of this study that, to the best of our knowledge, are not present in other studies. First, it separately investigates the effectiveness of two distinct instructional modes both of which are typically present in strategy-focused instruction studies, that is, *modelling* and *direct instruction.* The results of one study that singles out the separate components of a strategy-focused programme by Fidalgo et al., (2015), differs from our study in that in the latter (1) participants are 6 grade Spanish students instead of L2 students in higher education and (2) L1 tasks are in a different genre

(i.e., compare-contrast and opinion essays instead of synthesis writing). Additionally, (3) the differences in the designs of the two study seriously affect the comparability of the two studies. In the study by Fidalgo et al. (2015), direct instruction follows modelling, and the type of modelling is implicit; whereas, in our study direct instruction precedes modelling, and the type of modelling is explicitly supported by mnemonics.

Secondly, we conducted this study in ecologically valid circumstances and the effectiveness of the intervention was tested compared to regular curriculum practice, which tends to be standard across universities in Turkey. Thus, it has the potential to be insightful for similar teaching environments to revisit their practices and encourage further research. Although this specific university is one of the few institutions, which includes synthesis writing in their preparatory school curriculum, in the departments of other universities where the medium of instruction is English, synthesis writing is a common task in compulsory academic writing courses starting in the freshman year. Synthesis writing is also a task that is commonly encountered in faculty departments, both in L1 and in L2 contexts. This expands the scope of the results of this study to educational practices in other universities in Turkey and other comparable institutions around the world.

CHAPTER 3

THE EFFECT OF STRATEGY INSTRUCTION AND MODE OF INSTRUCTION ON BOTH WRITING PROCESS AND PRODUCT IN A FOREIGN LANGUAGE: MODELLING VS. PRESENTATIONAL

The majority of communication in a literate society happens through reading and writing (Molfese & Breznitz, 2012), and unlike the theoretical conception that the two are separate skills, in practice they are closely intertwined and mutually reinforcing (Graham & Hebert, 2010). In higher education, reading is the primary means of receiving information (Grabe & Stoller, 2001), and writing of demonstrating competence in response to content material (Hinkel, 2015). However, tasks that require the integration of both, that is, synthesis writing, are also the cognitively most demanding ones for the students to master (Spivey, 1997), and for the teachers to teach (Mateos & Solé, 2009). The complexity of these tasks also increases "the higher one goes up the educational ladder" (Mateos & Solé, 2009, p. 436). Unfortunately, the same cannot be argued for the mastery level of writers: novice writers do not necessarily evolve into expert writers as they mature; nor does the nature of their writing transform from the linear "knowledge-telling" model into the more sophisticated "knowledge-transforming" (Bereiter & Scardamalia, 1987).

The challenge is further increased when the reading-and-writing hybrid task is in L2, and more so, when students are developing learners in both English as a foreign language (EFL) and the academic writing discourse (Gao et al., 2021; Lee et al., 2018; Leijten et al., 2019; Plakans, 2009b). The additional linguistic requirements posed by the non-native language (Gao et al., 2021; Lee et al., 2018; Plakans, 2009b) and the relatively slow speed of access to these features (Schoonen et al., 2003) make writing in L2 difficult in its own right. Also, synthesis writing both in L1 (Spivey, 1997; Spivey & King, 1989) and in L2 (Plakans, 2009a), is largely dependent on students' level of reading comprehension, thereby harbouring an interplay of a reading problem and a language problem (Alderson, 1985).

These factors augment the complexities of learning-to-write; thereby posing a hindrance to the improvement of writing performance. Overcoming obstacles in L2 requires a strong purposeful effort because the implicit learning possibilities in L1 are not present in L2 (Dörnyei, 2009), which is typically learnt in instructional settings. Hence, effective instructional practices are crucial for teaching L2 synthesis writing. A consensus in the literature is the need to phase instruction of synthesis tasks and scaffolding each competence, both for effectiveness of instructional practice (Lee et al., 2018; Leijten et al., 2019), and for ensuring the validity and reliability of their assessment (Bennett et al., 2016).

An *integrated summary of multiple sources* is one such task that is used for phasing synthesis writing instruction in EFL contexts (see Bennett et al., 2016). Students read the provided source extracts to select relevant ideas on a given topic and subsequently organise them in a linear organisational format using proper citation in form and function (see Buyuktas Kara et al., 2018, Chapter 2). Charles (1996), in a corpus study for phraseological patterning in citations in the graduate theses of L1 English writers, documented the key pattern that occurred, that is, an integral citation using a reporting -that clause and the present tense reporting verb -*argue* (i.e., "Author/date argues that ...."). The reporting verb determines the rhetorical function of the subsequent sentence (e.g., agree, disagree, suggest, add, etc.), thus consolidating both the form and the function of the citation within a single phrase. A linking device can also be integrated in the formulation for establishing coherence in the text: "Author/date argues that another reason for x is y", or "Another reason for x, as argued by Author/date, is y". The follow-on sentence is termed after the linguistic skill category that it entails, that is, a paraphrase or summary of the original idea drawn from the source text (Gao et al., 2021). Using phraseology is an example to strategy instruction which enables scaffolding of citation use by way of a manageable step that ideally progresses into a more expansive repertoire of reporting forms and understanding of their various rhetorical functions (Mansourizadeh & Ahmad, 2011; Petrić, 2007; Thompson & Tribble, 2001). Hence, what is regarded as a lower-level aspect of text quality (see Zhang et al., 2020), can evolve into the high-level dimension that it duly deserves in discourse synthesis, in a way to represent the development of the writer from a knowledge-telling to a knowledge-transforming one (Lee et al., 2018). It also helps the instructors to identify the root cause for (often unintentional) plagiarism, that is, whether it results from inaccurate citations or inadequate paraphrasing and summarising, which are usually placed in the same category (Pecorari et al., 2014). Phasing synthesis writing instruction by incorporating explicit strategies into the L2 writing courses at university, without disruption to the regular curriculum of the institution, is possible and

effective, even in relatively short courses (see Buyuktas Kara et al., 2018; Zhang, 2013). Within the L2 literature, the positive effects of strategy instruction have been documented in its different dimensions such as undergraduate students with different proficiency levels (De Silva & Graham; 2015), and metacognitive self-regulation in improving undergraduate writing performance (Nguyen & Gu, 2013). A substantial body of research has also demonstrated the benefits of strategy instruction in L1, with learning-disabled students and normally performing adolescent students for planning, revising and/or editing in improving students' writing skills (see Graham & Perin, 2007, for a more extensive review of literature). Strategy instruction has shown significant effectiveness in numerous L1 studies, enhancing the quality of synthesis texts among secondary school students in the form of compare-contrast essays (Hammann & Stevens, 2003; Kirkpatrick & Klein, 2009), undergraduate students' argumentative synthesis texts (Boscolo et al., 2007), as well as in-service teachers' literature reviews and their writing processes (Segev-Miller, 2004).

Self-Regulated Strategy Development (SRSD) by Harris and Graham (1996) is a strategy instruction regimen, and is also the instructional model of the present study. SRSD aims at enabling independent performance of the strategies and self-regulatory functions involved in the deliberate employment of strategies (i.e., self-monitoring, self-instruction, goal-setting, and self-reinforcement) by gradually releasing the control from the instructor to the learner, also known as scaffolding, a critical element in strategy instruction and the implementation of SRSD (Harris & Graham, 1996). Scaffolding is operationalised in six stages, that is, instruction of strategies via (1) activating and developing background knowledge, (2) discussing, (3) modelling, (4) memorising; and practice of strategies via (5) collaborative practice and (6) independent performance. The memorisation stage marks an important threshold in scaffolding, signalling that learners are ready to proceed to the practicing stage. It is enabled primarily with the use of mnemonics, which reduce the strategy steps to a single chunk in the required order (Worthen, & Hunt, 2011) and facilitate retrieval by locating information in memory with associations (Malhotra, 1991). The stages of SRSD are recursive; any stage can be revisited anytime during instruction and the time allocation to particular stages can be tailored to the needs of different student populations. SRSD has proved effective in L1 writing, in a myriad of research studies (see De La Paz, 2005; Brunstein & Glaser, 2011; Teng, 2019), as also emerges from meta-analyses (Graham, 2006; Graham & Harris, 2003); also in teaching reading-and-writing hybrid tasks, in improving written products (Mason et al., 2006), as well as writing processes (Martínez et al., 2015). It also proved effective when adapted to an L2 context and used with undergraduate students

and proved effective in improving students' writing performance, self-regulation, and self-efficacy (Teng, 2022; Teng & Zhang, 2020).

In the instruction phase of SRSD (i.e., stages 1-4), explicit strategies for specific writing tasks are directly presented and overtly memorised. Instruction is mainly teacher-led and the mode of instruction can be argued to be presentational in nature, except for the modelling stage (i.e., stage 3) that is aimed at triggering observational learning. Observation lies at the heart of Bandura's Social Learning Theory (1977) and suggests that children learn by observing others and the outcomes of their behaviours, which later serves as a "guide for action" (p. 22). Observational learning depends on the occurrence of four interrelated processes: attention, retention, reproduction and motivation. The attention component includes several techniques such as the use of evaluative activities, and/or elimination of distractive factors to involve students in the modelled behaviour. Retention refers to storing of the observed behaviour in memory through retrievable codes, so that it is available for use in similar future circumstances. Reproduction refers to practicing the behaviour, ideally several times, to help consolidate the retrieval of the encoded behaviour. Finally, motivation denotes the incentive to do the behaviour; which can be reinforced with rewards or discouraged with punishments (Bandura, 1986; see also Raedts et al., 2017). Modelling, as an instructional mode, or its learner complement, observational learning, has proved effective in teaching cognitively challenging writing tasks. It was successfully implemented as compared to more traditional practicing conditions (i.e., learning by doing) in L1 (Dutch) for teaching strategies to write argumentative texts (Braaksma et al., 2004; Couzijn, 1999) and synthesis texts (Raedts et al., 2007); and in L2 to teach revision strategies (Van Steendam et al., 2010, 2014).

In teaching such a complex task as synthesising in L2, as is the case in the context of this study, alleviating students' cognitive load in the teaching environment may yield positive results. A typical writing lesson calls for two interdependent competencies, that is, writing and learning to write. Hence, a good instructional design should allow the students to attend to the writing task with not only their writer, but also learner capacities; and arguably, this is enabled by observation (Rijlaarsdam et al., 2008). In observation studies in writing (see Rijlaarsdam et al., 2008), observational learning is prompted by cognitive modelling, in which models vocalize their mental task processes of the "multiple examples of problem-solving actions" (Zimmerman, 2013, p.136). Solving problems entails goal-directed (i.e., strategic) thinking, and in the same vein, having access to a repertoire of strategies can make the writing process considerably easier for the writer. However, what may be intuitive for one writer may not be immediately accessible to another (Flower & Hayes, 1977) since novice writers have not yet

acquired skills needed to manage the process of writing that strong writers have (Breetvelt et al., 1994). Thus, in observational learning, the "think-aloud" of the model consists of their strategies for solving the "writing problem", which enables the student-writers to use more of their cognitive resources for metacognitive and procedural knowledge instead of focusing predominantly on the production process (Rijlaarsdam et al., 2005). This shift in allocation of cognitive resources has repercussions for both the writing products and the writing processes. In a study with 14-year-old students, Braaksma et al. (2004) for example showed that students who were strategy-trained via modelling for content development in argumentative writing, delayed the executive writing activities to the later stages of the writing process and engaged in more metacognitive activities such as goal-orientation, and analysing in the initial stages of writing, compared to students in the practice only condition, who adopted a more "monotonous" approach with executive activities such as transcribing during the initial stages and formulating spread throughout the whole writing process. Students' improved writing processes were also positively related to the quality of certain aspects of their written products. This kind of temporal organisation of cognitive activities during a writing task has the potential to be effective in completing synthesis tasks, where the "procedures" are more burdensome than in other writing tasks (Buyuktas Kara et al., 2018, Chapter 2).

Observation can also enhance self-efficacy (Bandura, 1997). Two sources that the students derive self-efficacy beliefs from are first-hand (i.e., enactive) mastery experiences and observed (i.e., vicarious) experiences, that is, when learners complete a task successfully and get rewarded, or observe peers do so, they form positive notions about their own potential to deal with future occurrences of a similar nature. In vicarious learning experiences, the perceived similarity of the model positively affects self-efficacy of the observer, thereby improving gains from observation, a condition also known as model-observer similarity (Bandura, 1997). Therefore, observation with a peer as the model (i.e., peer observation) may be more beneficial than teacher modelling, especially "when children hold self-doubts about their learning or performance capabilities" (Schunk, 1987, p.166). For student populations with low self-efficacy beliefs as a major hindrance to their learning, peer observation may be especially effective. For Turkish EFL students this might be the case. Writing in English is an academic requirement that both low- and high-achieving Turkish students find difficult and view as something to "persevere through in order to pass certain exams" (Yavuz & Genç, 1998, as cited in Erkan & Saban, 2011, p.166). This perception stems from negative student attitudes and writing apprehension, as well as low self-efficacy in writing (Erkan & Saban, 2011). As observation improves self-

efficacy, and peer observation more so, it can help Turkish students form positive learning behaviours and result in writing success.

The two modes of instruction that can be defined in instructing strategies in SRSD are *presentational* and *modelling*. For research purposes, it is tempting to isolate the instructional components to investigate the mode of instruction (i.e., presentational vs. modelling) with a greater impact in the teaching of writing strategies, which, to the best of our knowledge, has not yet been fully investigated. Fidalgo et al. (2015), with a similar research objective, investigated the separate effects of modelling and shared reflection, direct instruction, peer feedback and solo practice in strategy-focused instruction to teach *compare and contrast*, and *opinion* essays to Spanish 6th-graders in L1. Results showed that modelling combined with shared reflection is effective as a stand-alone component compared to the other instructional components. However, in the study by Fidalgo et al. (2015), modelling precedes direct-instruction and a reversion of the sequence of these two components could also point to similar benefits in students' writing performance, hence the mode of instruction with the greater impact remained unclear and pointed to the need for a direct comparison of the two (Fidalgo et al., 2015), which we undertake in this study. It should also be pointed out that in the study by Fidalgo et al. (2015), observation was prompted by *teacher modelling* of strategies, whereas in the present study we use *peer modelling* of strategies supported by mnemonics.

Based on the findings of previous studies, we expect the modelling mode to be effective for all types of students, especially for students with an initially lower writing proficiency level. Van Steendam et al. (2014) investigated the effects of observation of collaborative writing on revision as compared to a more traditional practicing condition and showed that observation is effective for all types of ability pairs (e.g., homogeneous vs. heterogeneous). Similarly, in a study by Braaksma et al. (2018), although complex patterns of interactions between learning conditions and pretest variables (i.e., content knowledge, writing proficiency, self-efficacy) emerged, almost no differences in effects could be observed for the observational learning mode of writing and initial levels of writing performance. In another study Braaksma et al. (2002) investigated the interaction between instruction (i.e., observation vs. direct writing) and aptitude levels (initially lower vs. higher aptitude), with a further manipulation in the observational learning condition in regard to the type of model (i.e., weak vs. strong). Results showed that initially low-aptitude students benefited more from observation when they focused on weak-models (rather than a strong model), than by performing writing tasks. Hence, in the present study we checked whether the

condition effects on synthesis writing quality are generalisable for both levels (i.e., lower vs. higher) of initial writing proficiency.

1 THE PRESENT STUDY

In English preparatory programs in Turkey, reading is rarely integrated in the writing curricula due to its complexity, and the time constraints of the programs, which is a challenge both for the instructors and students. In this study, we implemented an intervention that is aimed at improving the regular curriculum instruction for a synthesis task, in one of the rare preparatory programs that incorporated synthesis writing in its writing curricula. Our main research objective was to determine the impact of explicit strategy instruction in teaching synthesis writing, and the effect of two modes of instruction used in explicit strategy instruction, that is, presentational and modelling modes of instruction, on students' writing products and processes. Therefore, we investigated the following research questions.

1.1 *Research Questions*

RQ1. Effect on synthesis text quality:

RQ1A. Does explicit instruction of strategies (i.e., the strategy conditions) result in qualitatively better synthesis texts as compared to a regular course (i.e., no- strategy)?

RQ1B. Does explicit instruction of strategies via modelling mode result in qualitatively better synthesis texts as compared to presentational mode?

A previous study by Couzijn and Rijlaarsdam (1996) has shown that observation has a stronger transfer effect to another mode of communication (i.e., writing to reading, reading to writing) than regular courses in which students apply theory in exercises. Similarly, a transfer of writing strategies has been demonstrated in the study by Fidalgo et al. (2015), in which explicit instruction of strategies via modelling for compare-contrast texts is transferred to argumentative writing. Therefore, we added a second research question.

RQ2. Effect on transfer to argumentative texts:

RQ2A. Does explicit instruction of strategies for teaching a synthesis task result in qualitatively better transfer to argumentative writing as compared to a regular course?

RQ2B. Does explicit instruction of strategies via modelling mode result in qualitatively better transfer to argumentative writing as compared to presentational mode?

In the study conducted by Braaksma et al. (2004), it was demonstrated that students in the observational learning conditions adopted a more purposeful temporal management of the writing process. In the context of L1 argumentative writing, it entailed an increased focus on metacognitive and planning activities during the initial stages and a greater emphasis on executive activities in the later stages of the writing process. In writing an integrated summary from multiple source texts, an effective writing process was defined as *a recursive use of reading* for producing the target text, which was defined from the writing processes of students who wrote the better texts in the dual task with think-aloud protocols. We wanted to test if these results hold true for L2 synthesis writing as well, that is, a recursive use of reading for producing the target text, mediated within the procedural dynamics of this task, that is, reading and working on the sources dominating the initial stages; and paraphrasing, writing and reading the target text in the later stages of the writing process. Therefore, we added a third research question.

RQ3. Effect on writing processes (3A/B) and learning experiences (3C)

RQ3A. Does explicit instruction of strategies via presentational mode result in a different temporal distribution of writing activities as compared to a regular course?

RQ3B. Does explicit instruction of strategies via modelling mode result in a different temporal distribution of writing activities as compared to presentational mode?

In addition, we would like to know students' reported learning experiences which could explain effects on writing process behaviours:

RQ3C. Were students' learning experiences affected by the instructional conditions?

2 METHOD

2.1 *Participants*

Participants were 155 advanced level EFL learners (53.5% male, mean age: 19.9) at an eight-week pre-faculty course in the preparatory program at an English-

medium private university in Turkey. Their L1 was mainly Turkish except for 17 international students (six Syrian, three German, two Serbian, one Afghan, one Azerbaijani, one Chinese, one Jordanian, one Moroccan and one Saudi Arabian) enrolled at the university for their entire degree, and evenly and randomly distributed over the three conditions. Participants were prospective students of various departments. They were required to show English proficiency in all four skills of the language (i.e., reading, listening, writing, speaking) to commence their departmental studies. They were a relatively homogenous group of learners with regard to language proficiency who had recently passed the upper-intermediate level test prepared by the test office of the institution. Participants signed a consent form and could ask for the removal of their data until the end of the pre-faculty course. All the students agreed to take part in the study in return for an opportunity to sit a second timed-writing exam (also the second posttest of the study), which would replace their grades in the previous exam if they received a higher grade, not in the context of the study, but only counting towards their Grade Point Averages (GPAs).

2.2 *Research Design*

We tested the effects of two instructional modes, presentational and modelling, on participants' text quality and writing processes, as compared to a control condition, which was the regular instructional practice of the institution, in an experimental pretest-posttest design, including a transfer test.

Students were randomly assigned to 11 classes. These classes were randomly assigned to three conditions – Eight classes divided by two and placed in one of the two experimental conditions (i.e., four intact classes in each) and three classes were placed in the control condition. The distribution of gender was not significantly different across the conditions ($X^2(3, N = 155) = 0.318$, $p = .853$), with the percentage of female students varying from 44 to 49%. We also checked for possible confounding effects of differential effects across conditions for motivation. Conditions did not differ in students' motivational orientation, or in text quality as measured by pretests prior to the study (see preliminary analyses on pretest differences).

Each class had a main instructor, who served as the trainer in the study (subsequently called instructor). Via a randomized block design, the 11 main instructors were assigned to conditions prior to the start of the module, according to the instructors' gender and L1 (Turkish vs. all other). See Table 1 for the distribution of instructors across the three conditions.

Table 1
Distribution of Instructors Across Condition

Modelling	Presentational	Control
1 Turkish female 1 Turkish male 1 Foreign female (Slovak) 1 Foreign male (American)	2 Turkish females 2 Foreign males (American and Syrian)	2 Turkish females 1 Foreign male (Canadian)

2.3 *Procedure*

As part of their writing course, students receive four lessons of synthesis writing instruction and are expected to complete a follow-up timed-writing exam. We implemented the intervention to improve this instruction, which was, together with the pre- and posttests, already a part of the pre-faculty course curriculum. Table 2 presents the distribution of the sessions and administration of pre- and posttests of the study.

Table 2
Distribution of the Sessions Across Weeks and Pre- and Posttests of the Study

Week	Session	Activity	Variable	Writing Task*
1	Test	Pretests	Motivation	ARG_1**
			Text quality	SUM
2	1-4	Instruction		SYN
3	5	Posttest 1	Text quality	SYN_1
5	6	Writing Log Training		SYN
5	7	Posttest 2	Writing process	SYN_2
6	8		Motivation & Learning experiences	
7				
8	9	Transfer test	Text quality	ARG_2

*ARG: Argumentative essay; SUM: Summary of a single source; SYN: Synthesis task.
** Data taken from the preceding course

In total, the study was conducted in a total of 10 sessions (including the Test-session) each ranging between 50 - 60 minutes, spread out over the eight-week pre-faculty course: four sessions were reserved for training and six sessions for

pre- and posttest administration, incorporated into the regular course practices. In the test-session, we collected baseline data for motivational orientation and text quality in the form of a summary of a single source (hereby, SUM). We implemented the intervention in four sessions in the 2nd week on two consecutive days simultaneously in all classes (Sessions 1-4). Five sessions were reserved for the following: the synthesis timed-writing task as Posttest 1 (hereby, SYN_1) at Session 5 to measure Text Quality, writing log training at Session 6, synthesis timed-writing task as Posttest 2 (hereby, SYN_2) at Session 7 to measure the writing process, and collecting motivation questionnaires and learner reports at Session 8. At Session 9, we administered an argumentative writing task (ARG_2) to measure text quality transfer to another genre.

2.4 *Independent Variable: Courses*

2.4.1 Strategies

TRAMPOLINE Mnemonic. In the experimental conditions (i.e., modelling & presentational), we used the TRAMPOLINE strategies mnemonic for teaching synthesis writing. It was used on the introductory PowerPoint presentation and the accompanying handout, as well as in the peer videos. We adapted TRAMPOLINE synthesis writing strategies from the TRAP IDEAS reading-and-writing strategies for summarising (Mason et al., 2012). The researcher and a colleague piloted the TRAP IDEAS strategies simultaneously in two pre-faculty level classes in the last module of the academic year preceding the actual experiment, with none of the currently participating students involved. The two instructors then liaised for the adaptations necessitated by the differences between the study by Mason et al. (2012) and the present study, such as the contexts, L1 vs. L2 and the type of writing tasks, that is summary of a single source vs. multiple sources respectively. Also, citation use and the ensuing task requirements were also included in TRAMPOLINE such as using reporting verbs, linkers, in-text citations and referencing as per APA conventions, which were not present in the TRAP IDEAS strategies. See Appendix A for the content of TRAP IDEAS and TRAMPOLINE strategies.

Videos. For modelling the use of TRAMPOLINE strategies, we shot three separate videos for TRAM, P and OLINE strategies of 9, 17 and 12 minutes respectively, using the Camtasia Screencast Program (Techsmith, 2016). This program allows users to create educational videos by simultaneously capturing their computer screens and recording their face and/or voice enabling narration during demonstration. With this software, we shot controlled think-aloud protocols of a model

acting out an actual account of a student completing the synthesis task with the help of the strategies, which was scripted by the researcher. The model was a female freshman who had studied the preparatory program in the previous year. We prepared the framework of the script based on a good student performance. Hence, the model used the strategies during task execution; but with occasional instances of the most common student mistakes in writing a synthesis text based on an error analysis conducted with the pre-faculty course instructors. To ensure authenticity, the model did not follow the script very strictly nor adopt a prescriptive tone.

2.4.2 Course Design

In the two experimental conditions of this study, the lesson plans were adapted from the regular course practice of the institution for teaching an introductory level synthesis task that entails summarising from multiple sources, in a way to include explicit strategies. We devised the TRAMPOLINE mnemonic for presenting the explicit strategies for the instruction of the task and used it in the experimental conditions, or the *strategy conditions*, subsequently used interchangeably to refer to the same learner group. The regular course of the institution with no teaching of explicit strategies (i.e., no-strategy condition) is also controlled for the experimental purposes of this study. Hence, the terms *regular course* and *no-strategy condition* refer to the control condition and will be subsequently used interchangeably to refer to the same learner group.

The difference between the two experimental conditions lies in the instructional mode, that is, modelling vs. presentational, that is used in the instruction of the TRAMPOLINE strategies for synthesis writing,

Modelling. The modelling condition involved a brief introduction to the TRAMPOLINE mnemonic by the instructor followed by students observing a peer model explicitly using the TRAMPOLINE strategies. In Session 1, students were familiarized with the theme, the task and the TRAMPOLINE strategies mnemonic for synthesising. In Sessions 2 and 3, students observed videos of a peer modelling the use of strategies (in-class projection), and emulated the strategies through collaborative practice. Session 4 consisted of independent practice with the assistance of the instructor on students' demand.

Presentational. In the presentational condition, explicit strategies for synthesis writing were delivered via the TRAMPOLINE mnemonic. The mode of instruction can be defined as presentational, with occasional teacher–led whole class

question-and-answer episodes. In Session 1, the instructor introduced the theme, the task and presented the TRAMPOLINE strategies mnemonic. In Sessions 2 and 3, the instructor elaborated on each strategy step of the TRAMPOLINE strategies and instructed students on how to apply them. Students then practiced the strategies through collaborative practice. Session 4 consisted of independent practice with the assistance of the instructor on students' demand.

Thus, in the two strategy conditions, Session 1 and Session 4 were the same. However, in Sessions 2 and 3, the sequence of learning content was the same, that is, instruction and practice of TRAM, P and OLINE strategies; yet the mode of delivery was different in each condition, that is, via presentational and modelling mode.

Control Condition (no-Strategy). The regular instructional practice for teaching the synthesis writing task (and other writing tasks to that purpose) is that the instructors follow the lesson plans that are prepared by the institution as part of their annual course observations and instructor evaluations. In line with this practise, the directorate had observed and approved the instruction on this task in the previous year, as an appropriate way to teach the synthesis task. In the control condition, no changes were made to this regular course and is followed verbatim; however, the lesson plans followed in the experimental conditions were adapted from this baseline lesson plan, in a way to reflect the instructional practices of the modelling and presentational modes of instruction in their respective conditions.

In the control condition, the explicit, strategy-focused practices of the experimental conditions were absent (hence, the no-strategy condition). Instruction was mainly operationalised by the guided questions of the instructor enabling student self-discovery. Students received the same amount of collaborative and individual writing practice, both of which enables scaffolding without teacher interference. The materials and their content were the same, except for those used in the experimental conditions that are immediately related to the instructional mode (see Materials section, supra). See Appendix B for a detailed description of the training procedure across the three conditions.

Students in all conditions needed to sit a timed-writing exam in the week following the training session, which counted towards their GPAs, so instructors in all three conditions had to make sure that all task and learning objectives were met, including task knowledge and process mastery. Thus, the lesson objectives were the same in all the three conditions; only the way in which the instruction was

delivered differed. In the presentational condition, students received explicit strategy instruction supported by mnemonics. In the modelling condition, students watched videos of a peer using the mnemonics and completing the task. In the control condition, students followed the regular instructional practices of the institution of inferring the task requirements from the learning materials through the guiding questions of the instructor. In all conditions the practicing stages were present. Students practiced collaboratively and individually to enable independent performance. Table 3 provides an overview of the distinctive features of the three instructional conditions.

Table 3
Distinctive Features of the Instructional Conditions

Component	Modelling	Presentational	Control
Strategies (and mnemonics) for synthesising	+	+	-
Collaborative practise	+	+	+
Individual practise	+	+	+

2.5 *Instrumentation*

2.5.1 Motivational Orientation

We created a questionnaire by selecting a total of 17 questions from three scales of the Motivated Strategies for Learning Questionnaire (MSLQ) developed by Pintrich et al. (1991) in the English language. The adapted version of the questionnaire included six items for task value, eight items for self-efficacy and three items for intrinsic goal orientation, measured on a 7-point Likert scale. The reliability of the whole scale was .93 (pretest) and .98 (posttest). See Appendix C for the adapted version of MSLQ that was used in this study.

2.5.2 Writing Process Measures

Multiple process measuring methods increase the validity of the results (Schellings & van Hout-Wolters, 2011). In this study, we combined an online Time-Sampled Self-Report (TSSR) method developed by Torrance et al. (2007) with an offline Learner Report method developed by De Groot (1980). TSSR delivers data that is reported by the writer concurrent to the writing task, Learner Report

indicates the learning experiences of the participants after the writing task has been completed.

Time-Sampled Self-Report. This is an online process registration method providing insight into students' cognitive activities that they are engaged in throughout the writing process. During a writing task, students hear a bleep sound at regular intervals, and they tick a box on the given writing logs indicating the activity they are engaged in at that moment. A total of 45 bleeps were played during the execution of the second posttest (i.e., SYN_2), at a variable interval schedule of 60 and 120 seconds (on average every 90 seconds), a practice in line with previous studies. The number sequence of the bleeps was also projected on the board for students to keep track of. The eight activity categories in the writing logs were adapted from the original (Torrance et al., 2007) to include possible activities in a synthesis task. Simple graphic representations for each activity category were used to help students locate the activity on paper easily. The activities on the TSSR-writing logs that we used in the present study were categorised and defined as follows:

1. I am reading the sources: I am trying to understand the sources.
2. I am paraphrasing: I am writing an idea from the text in my own words.
3. I am working on the sources: I am trying to find the main idea, important details, the author, year of publication, etc.
4. I am editing: I am making changes to my text: correcting spelling mistakes, changing/adding words.
5. I am writing my text: I am writing my synthesis text.
6. I am reading my text: I am reading through parts or all of my text.
7. Other: I am doing something unrelated such as looking for a pen, looking out the window, etc.
8. Finished.

We trained the students in all three conditions on how to complete a writing log before the actual practice (see Torrance et al., 2007). We first introduced the activity names, their explanations and graphic representations so that the students can easily locate the activity categories on the writing logs during the dual task of writing and self-reporting. Then we showed students a short training video, that is, a compilation of nine fragmentary footages of a peer model thinking aloud while completing a synthesis task, with each fragment representing an activity category as they appeared on the writing logs. Video fragments were paused with a bleep sound during which the students had to tick a box indicating the model's writing activity on a demo version of the writing log. The instructors checked and discussed the answers with the students. Subsequently, we

simulated the timed-writing test conditions of SYN_2, that is, completing the writing task concurrent to TSSR-implementation. On the actual exam day, we briefed the students one more time about the activity categories on the writing logs of TSSR. See Appendix D for the writing log of TSSR used in this study.

Learner Report. This method is implemented after the intervention has been completed (*i.e.,* Session 8) by prompting the students via open-ended questions to report their learning experiences about the training. The data obtained helps the researcher gain further insight into the writing process. See Appendix E for the Learner Report used in this study.

2.5.3 Text Quality Measures

We implemented three types of writing tasks, all of which were prepared by the testing unit of the institution. Tasks were, a summary of a single source (SUM), a synthesis task (SYN) and an argumentative essay (ARG). (1) SUM was implemented at pretest only for text quality, (2) SYN_1 and SYN_2 were both implemented at posttest; SYN_1 for text quality under regular writing conditions and SYN_2 under concurrent self-reporting writing process conditions (TSSR), and (3) ARG_1 and ARG_2 were implemented and used as pretest and transfer test for text quality, respectively.

SUM, ARG and SYN writing tasks were similar to those used for teaching and practicing in all three conditions of the intervention, but the students saw the content of the exam materials including the prompts in all writing tasks and the sources in synthesis tasks for the first time in the exam.

Summary Task. Since the synthesis task of this study is summarising from multiple sources, we purported that summarising from a single source would yield valid information that is generalisable across the two tasks. Thus, as a pretest, students wrote a summary of a single source (SUM) in a pretest-session. We provided the students with a reading text taken from a course book and asked them to summarise it in one paragraph using no more than 150 words. See Appendix F for the list of writing prompts used in this study.

Argumentative Task. In both argumentative writing tasks (ARG), students were asked to write a well-organised essay of about 300 words in response to a 50-word prompt. The prompt in ARG_1 asked students to discuss their opinions about a given statement in the form of a cause-and-effect essay, and in ARG_2, a for-or-against essay, by including at least one counterargument and refutation.

Synthesis Task. In both synthesis tasks (SYN_1 & SYN_2), students were asked to write a 150–200-word summary of one-paragraph-extracts from three different sources, in response to a writing prompt which was around 25 words including the instruction. Students were discouraged from integrating their own ideas in their texts but were expected to use APA in-text citation conventions. This introductory level task was mainly aimed at introducing students to proper citation use in discourse synthesis (see Appendix F).

2.6 *Test Procedures*

Except for SUM, all tasks were compulsory assessment components of the institution counting towards students' GPAs. The writing tasks of SYN_1, and ARG_2, were compulsory assessment components that counted towards students' passing grades in the pre-faculty course, and ARG_2, in the preceding upper intermediate course. Tests were executed simultaneously in all classes, in paper and pen written exam conditions on dates pre-set by the directorate. Students wrote on A4 papers with the exam prompts written at the top. All the writing tasks were implemented under timed-writing exam conditions.

ARG_1 and ARG_2 were part of an adjunct module-end test (MET), for the upper-intermediate and pre-faculty level courses, respectively. They were prerequisite assessment components, among others, for commencing to a higher English language proficiency level, that are designed within the curriculum as eight-week modules. They were implemented one week before the start of the following module, hence are nine weeks apart. In both ARG_1 and ARG_2, students received the writing exam paper with the writing prompt in the last 60 minutes of MET, after all other test materials belonging to other language skills (i.e., listening and reading) were collected by the examiner.

SYN_1 and SYN_2 were both implemented at posttest, but for different purposes, hence under different test conditions. SYN_1 was implemented as a measure for text quality, whereas SYN_2 was implemented for collecting writing process data concurrently with TSSR. We allocated 50 minutes to SYN_1 and 60 minutes to SYN_2, extending its duration a total of 10 minutes, to compensate for a possible latency effect that can be observed in concurrent self-reported techniques, without decrements in text quality (Janssen et al., 1996). We also gave students the opportunity to replace their score in SYN_1 and SYN_2 in the case that they increased their scores in between the two measurement occasions, not in the context of this study, but only counting towards their Grade Point Averages (GPAs). The two tests were implemented two weeks apart.

2.7 *Rating Procedures*

2.7.1 Text Quality

The handwritten student papers were typed (on word documents) to eliminate any possible negative effect of student handwriting on the raters (Klein & Taub, 2005). Raters were second- or third-year bachelor students of English Language and Culture at the University of Groningen, the Netherlands; 20 to 23 years of age, with Dutch as their mother tongue. A total of 11 raters were involved, six raters had experience as they also rated in our first study; five raters were new to the project. For each text quality aspect there was at least one rater that had performed the assessment in a previous study. The available raters were distributed using a design of overlapping rater teams (Van den Bergh & Eiting, 1989), which means that the texts were randomly distributed among three raters, and each text quality aspect was rated separately (e.g., comprehensibility, source use, content, etc.), by two raters. That is, every aspect in a student's text received a score from two independent raters and the mean of the scores given by different raters on a particular aspect determined the final scores for that aspect of text quality. Each rater rated about 28 to 42 texts per task depending on the size of the panel of raters.

In this study, we adopted benchmarking, a holistic scoring technique that entails comparing each quality aspect of a task to the mean of a predefined interval scale. Earlier research on writing has demonstrated that benchmarking is a reliable and valid method for rating text quality (see Bouwer et al., 2023; Schoonen, 2005; Tillema et al., 2012; Vandermeulen et al., 2020). For each aspect, we selected as benchmark a text from the available student texts. Raters scored certain aspects of the texts in comparison with the benchmark texts, each of which had a standard score of 100. For example, if a text was considered twice as good in quality for that aspect, it was scored 200; if it was half as good, it was scored 50, and so on. We also added annotations for the benchmarks including weak and strong points that justify the accompanying score, to help the raters focus on these traits when scoring a text on a particular quality aspect.

Previous research has also shown the reliability of using the same benchmark texts for assessing different tasks in the same genre (Bouwer et al., 2023; Tillema, 2012; Vandermeulen et al., 2020). In another study (see Buyuktas Kara et al., 2018, Chapter 2), we successfully applied the same benchmarks for two tasks of the same genre, but with different writing prompts. Therefore, in this study, we used the benchmark texts from the previous study for all aspects of text quality that were scored via benchmarking (see infra for details, see also Appendix G for the rating scales we used in this study).

Table 4
Rating Reliability in Cronbach's Alpha

	Pretests		Posttests		
	SUM	ARG_1	SYN_1	SYN_2	ARG_2 (Transfer)
Examples	.86				
Comprehensibility	.96				
Main ideas	.81		.69	.70	
Supporting ideas	.82		.81	.87	
Source Comprehension			.74	.77	
Text Comprehensibility			.77	.72	
Source Use			.91	.81	
Authenticity			.99	.96	
Organisation		.60			.63
Content		.66			.58
Grammar		.51			.78
Lexis		.49			.74

Note. ARG: Argumentative essay; SUM: Summary of a single source; SYN: Synthesis task.

We held a standardization session prior to the rating session, to ensure that raters focused on the relevant traits when scoring a particular aspect of text quality. In this session, raters received the benchmark texts, the lists with the idea units drawn from the source texts (i.e., main ideas, supporting ideas and examples) and the rating scale for the assessment of authenticity. Raters individually practiced rating each aspect of text quality from a total of six student texts from each one of SUM, ARG_1, ARG_2, SYN_1 and SYN_2. When the raters differed in their scoring, they discussed possible reasons and solutions for their disagreement. The raters' scoring of the texts for the pre- and posttests appeared to be reliable. See Table 4 for the reliabilities between the raters in scoring the text quality aspects of the study.

SUM. The prompt for SUM was the same as in the previous study (Buyuktas Kara et al., 2008*, Chapter 2*), so we used the same benchmark text and the same list of idea units that was prepared a priori via source text analysis. We rated SUM on two aspects: (1) comprehensibility, with holistic scoring via benchmarking, and (2) content, which was operationalised by the correct selection of main ideas, supporting ideas, and examples. Comprehensibility was rated without access to the source text, which means it was assessed from the viewpoint of a reader who

is not familiar with the original source. Content was rated by counting the number of correct idea units and calculating the percentage of their occurrence in student texts.

ARG. We scored ARG on four quality aspects: (1) organisation, (2) content, (3) grammar, and (4) lexis. The organisation was related to the overall structure of the essay, whereas content was operationalised via strength of argumentation. Range and accuracy of level appropriate grammatical structures were rated under grammar and lexical richness was rated under the lexis aspect.

SYN. Synthesis texts were scored in six aspects, (1) main ideas, (2) supporting ideas, (3) source comprehension, (4) text comprehensibility, (5) source use, and (6) authenticity. Main ideas and supporting ideas aspects were scored similar to SUM, that is, based on the lists we prepared a priori via source text analysis. The raters scored each aspect at a time and independently of one another, by counting the number of correct idea units, and the subsequent student scores were determined by the percentage of their occurrence in student texts.

The quality aspects of text comprehensibility, source comprehension and source use were scored using the holistic scoring procedure with the benchmarking technique.

In text comprehensibility, we were interested in how students themselves represented the source texts in their texts, so we did not analyse the ideas drawn from the text in relation to the source text. Raters scored the extent to which a reader without access to source texts understands the scope of the topic under discussion. Hence, they scored the text comprehensibility aspect first, to ensure that they could assess it for readers not acquainted with the sources.

For source comprehension, the raters scored the student texts considering the extent to which a reader is correctly informed about the important content in the source texts, taking the sources as point of reference for comprehension. The aspect of source use was operationalised using a variety of correct reporting verbs and linkers, and also correct in-text citation techniques as per the APA principles.

For scoring authenticity, we operated the function 'Compare and Merge documents' in Microsoft Word, which automatically highlights the parts of the student texts that overlap with the uploaded source(s). Two raters subsequently calculated how many words of each student text were highlighted on a scale of 0 to 100%. The overlap percentage was subtracted from 100, and the corresponding value constituted the authenticity score of the student. For example, in the case of a calculated overlap of 60%, the text received a score of 40 on authenticity.

Table 5
A Cross-Match of the TRAMPOLINE Strategies, Text Quality Aspects and the (Composite) Scores

Strategies	Text Quality Aspects	(Composite) Scores
Think Read Ask – What is the main idea Mark the important details	(Correct selection of) Main & Supporting ideas Text comprehensibility Source comprehension	Content
Paraphrase the main idea and the details	Authenticity	Authenticity
Organise the ideas (by) Linking the sentences Including APA Nesting reporting verbs	Source Use	Source Use
Edit your text	(Not rated in text quality/ Measured in writing logs)	

Each strategy in the TRAMPOLINE mnemonic was aimed at improving a certain text quality aspect and rated accordingly. Table 5 shows the cross-match of TRAMPOLINE strategies with the quality aspects of the synthesis texts in clusters, as well as the aspects that were merged under the content composite score. See also the preliminary analysis on reliabilities for creating composite scores for each text quality measure (see infra).

2.7.2 Learner Reports

Following a standardization session, the researcher and another instructor independently coded the responses by identifying the predetermined key words and relevant semantic units in student reports, clustering them under four categories: main ideas and details under *content management*, paraphrasing and summarising under *synthesising skills*, organisation, linkers, APA under *source use*, think, edit, steps/stages and time-management under *process knowledge*. Interrater reliability was .70 based on a sample of 34 units in 10 student cases.

2.8 *Training Delivery and Intervention Fidelity*

We provided the instructors of all three conditions with the materials and lesson plans (organised in a folder) in the order to be followed, which is the usual

procedure for following the course content and level objectives of the different modules at the institution. The researcher trained the instructors of the experimental conditions (i.e., modelling, and presentational) in a two-hour session prior to the start of the module. To avoid contamination of conditions, the instructors of the control condition, reflecting the regular course practice, did not participate in the training. As an implementation check, instructors rated their fidelity in implementing each of the critical steps of the intervention out of a hundred. In addition, we observed one randomly chosen instructor from each one of the experimental conditions in one session of the intervention, and scored their fidelity based on our observation using the same rating scale. Both the treatment fidelity scores (M = 85, SD = 11.7) and our observation scores (M = 92.5, SD = 5.7) of the instructors in the experimental conditions yielded satisfactory results.

2.9 *Analyses*

2.9.1 Preliminary Analyses

Reliabilities. For SUM, the four scores formed one factor (principal axis factoring, direct oblimin, 50% variance explained) and were therefore combined in one composite total score. Cronbach alpha was .76, with item rest correlations varying from .33 for Main Ideas) to .76 for Comprehensibility.

For text quality of SYN, we aimed to mirror the components for the content composite score of SUM, that is, main ideas, supporting ideas, and comprehensibility (that includes the quality traits for both source comprehension and text comprehensibility). The content composite score proved to be homogeneous (Cronbach alpha .71 pretest, .78 posttest, with item-rest correlations varying from .40 to .70 (SYN_1) and from .44 to .69 (SYN_2). The correlations between three quality indicators for SYN, that is, content, authenticity and source use were absent or reached a maximum of .225 (source use and authenticity in SYN_1.)

For ARG_1 and ARG_2, we first ran factor analyses (principal axis factoring, direct oblimin rotation), resulting in two factors as expected, explaining 72% and 61% of the variance respectively, corresponding to a content & organisation score (hereby, content) and a grammar & lexis score (hereby, language use). Cronbach alpha for the composite scores were satisfactory (content .80 and .79 for ARG_1 and ARG_2 respectively, and language use .80 and .62).

Pretest Differences. We checked for initial differences between conditions on pretest measures: motivational orientation and text quality (SUM & ARG_1). We applied a p value of >.20 to avoid a Type II error, a false negative (Table 6).

Table 6
Pretest and Posttest Means and Standard Deviations for Motivational Orientation

		Pretest		*Posttest*	
Condition	*N*	*M*	*SD*	*M*	*SD*
Modelling	57	4,99	0,97	5,26	0,87
Presentational	57	4,79	0,85	4,94	0,8
Control	39	4,72	0,96	4,96	0,93

Note. On a 7 Point scale: 1: Strongly Disagree; 7: Strongly Agree

No initial differences were observed for motivational orientation ($F(2,152) = 1.543$, $p = .217$, $\eta^2 = .011$) or text quality (Table 7; $\Lambda = 0.46$, $F(6,298) = 1,116$, $p = .323$, $\eta^2 = .023$).

Table 7
Means and Standard Deviations for the Pretest Text Quality Scores (SUM and ARG_1)

	Modelling n = 57		Presentational n = 57		Control n = 41	
Test	*M*	*SD*	*M*	*SD*	*M*	*SD*
SUM	52,44	19,20	56,62	22,55	56,09	19,85
ARG_{1_CO}	102,08	14,54	102,53	15,58	103,52	12,80
ARG_{1_LU}	100,13	15,71	102,08	16,09	108,35	17,23

Note. SUM: Summary of a single source; ARG_1: Argumentative task at pretest: CO: Content, LU: Language Use (Grammar and Lexis.

Effects on Motivational Orientation. To check for a differential effect of the conditions on motivation, which could be a confounding variable, we ran an analysis of variance, with time (two measurement occasions: pretest and posttest as within variable and conditions as a between-variable, after having selected outliers (one case for each measurement occasion). No confounding effect of motivation was observed. A repeated measures ANOVA showed an effect of time ($F(1,152) = 6.890$, $p = .01$, $\eta^2 = .044$), indicating a progress over time in motivation, but no interaction between time and condition ($F(2,152) = 0.161$, $p = .851$, $\eta^2 = .002$). We may conclude that motivation increased during the intervention, but

conditions did not contribute to the increase in motivation for writing differently. The average score in all conditions is well above 4, the midpoint of the scale. See Table 6 for the means and standard deviations for pretest and posttest scores for motivational orientation.

Effect of Test Condition on Text Quality. We administered two posttests for synthesis writing: SYN_1 and SYN_2 (see Table 2). SYN_1 was a regular writing test, under normal conditions, aimed at collecting text quality data. SYN_2 was implemented to collect writing process data via TSSR, a methodology that was used in various studies (Fidalgo et al., 2008; Fidalgo et al., 2015; Torrance et al., 2007). It includes a training in reporting categories followed by the test during which the writing process is interrupted by auditory signals prompting the participants to report the writing activity they are engaged in at that moment. Although in the current study participants in all three conditions participated in the same test conditions, one may expect that the effect of this test condition may affect the instructional conditions (i.e., modelling, presentational and regular course) differently. When the text quality scores of the synthesis text are differently affected by the writing process test condition, we should be cautious in interpreting the writing process data that we collected. Hence, we inquire into the extent to which writing under TSSR-condition (i.e., test condition) affected text quality.

Therefore, we ran a multivariate analysis with the three indicators of text quality, learning condition as between-variable, and test condition as within-subject variable. The multivariate (three dependent variables for SYN text quality, that is, content, source use and authenticity, subsequently called aspects of text quality) repeated measures (SYN_1 and SYN_2) showed a main effect of test condition (Λ = 0.67, F(1, 152) = 74, 25, $p < .001$, $\eta^2 = .33$). This in itself does not indicate an effect of test condition – SYN_1 vs. SYN_2 – while the two writing prompts were similar, but the topics were different, moreover, there was a time lapse between the two tasks. The learning condition effect (F(1, 152) = 8,27, $p < .001$, $\eta^2 = .098$) is not affected by the task (i.e., no statistically significant interaction effect (F(2, 152) = 2,73, $p = .068$, $\eta^2 = .035$). The effect does not vary for the three text quality indicators (F(4, 304) = 1,88, $p = .114$, $\eta^2 = .035$). Note that the p value reported is smaller than .20, which is usually set as border for no effect hypotheses to avoid a Type II error, a false negative. Strictly applied, we must decide that we cannot prove that the scores of the two test situations generated do not differ.

From Table 8 we may infer that overall, students performed quite well or even better on SYN_2, but not for all variables. Whether the differences between SYN_1 and SYN_2 are due to the test condition, the topic, and/or the time lapse cannot be decided. We will use the SYN_2 task to interpret the effects of learning

conditions on writing process behaviours as planned but will refrain from using the text quality scores as indicator for delayed effects. When we want to test the effects of test conditions on text quality, another research design is to be preferred.

The subsequent univariate analyses confirmed this finding: no interactions between test condition and instructional conditions were significant at $p < .05$ (SYN content $p = .80$, $\eta^2 = .003$, source use $p = .06$, $\eta^2 = 04$, authenticity $p = .09$, $\eta^2 = 03$). We can conclude that the data collected on writing processes at SYN_2 can be validly used as process data for synthesis writing.

2.9.2 Analytic Strategy

Text Quality: Condition Effects (RQ1-2). For the effect on the quality of SYN_1 (i.e., RQ1) we ran multivariate analyses of variance for three dependent variables (i.e., content, source use, and authenticity) with the (pretest) scores for SUM, $ARG_{1\text{-}CO}$ and $ARG_{1\text{-}LU}$ (correlations are small but significant, and vary from = .16, p = .048 and .279, p < . 001) as covariates.

Generalisability. The findings of the study by Braaksma et al. (2002) showed an interaction between the instruction conditions and the initial levels of aptitude. Students with initially lower aptitude levels performed better in observational learning conditions (focusing on weak models) than by performing writing tasks. Hence, to explore for RQ1, whether one of the instructional conditions resulted in better synthesis text quality scores for a particular group of participants than another condition, we analysed interactions between the three conditions and two learner variables: motivation and writing skills, both based on pretest scores, on three aspects of posttest text quality: content, source use and authenticity. We applied Hayes moderator regression analyses (Hayes, 2013, version v3.3), as add-in in SPSS.

For the effect on ARG_2, that is, transfer to argumentative writing (RQ2), we ran multivariate analyses of variance with scores on content and language use as dependent variables, and covariates, after selecting outliers (two and four cases in respectively pre- and posttest scores).

Writing Process: Condition Effects (RQ3A/B). The effect of conditions on writing process activities was measured online by Time-Sampled Self-Reporting (TSSR). We analysed the TSSR data first by comparing the mean duration of the process and the proportions of the activities. Second, to study the effect on the distribution of process activities across the writing process, we divided the writing

process into three intervals (Rijlaarsdam et al., 2015). Hence, we divided the process time of each individual student (calculated by the total number of responses in the log) into three phases (i.e., phase 1, phase 2 and phase 3) of equal length, based on the number of responses to the bleeps. We calculated the frequency of writing activities for each phase by counting the number of activity items reported in the writing logs and calculated proportions. We analysed the data for each of the six process activities with repeated measures, with the three intervals as within variables and condition as between-variables.

Writing Process: Condition Effects on Learning Experiences (RQ3C). The effect of conditions on students' reported learning experiences was measured retrospectively via Learner Reports. To analyse the Learner Report data, we counted the number of categorical statements (i.e., per category) of all conditions and calculated proportions for each condition by dividing the total number of statements by the number of students in that condition. We applied analysis of variance to analyse the coded data from the Learner Report with the writing pretest as a covariate.

3 RESULTS

We hypothesized that in teaching strategies for writing, instruction would be more effective if it is conducted via modelling rather than a presentational mode and that both instructional modes (i.e., strategy conditions) would prove more effective than the regular curriculum instruction (i.e., no-strategy condition).

3.1 *Text Quality*

3.1.1 Condition Effects (RQ1-2)

A multivariate analysis of covariance including the two pretests (i.e., SUM and ARG_1) as covariates with the three text quality measures (content, authenticity, and source use) showed statistically significant condition effects for SYN_1 (Λ = 0.800, (F(6, 294) = 5.65, p < .001, η^2 = .102). Follow-up univariate analyses showed no effect of condition for content (F(2,153) = 0.066, p = .936, η^2 = .001). An effect of condition was observed for authenticity (F(2,153) = 6.695, p = .002, η^2 = .083) and source use (F(2,153) = 8.878, p < .001, η^2 = .108). Pairwise comparisons (Bonferroni) showed that for source use, the modelling condition outperformed the presentational condition (p = .003) and the control condition (p < .001). For authenticity the modelling condition performed better than the presentational

condition (p = .003), while the control condition performed better than the presentational condition (p = .02). Table 8 presents the descriptive data.

For ARG_2 (i.e., transfer test) scores the analysis showed a statistically significant condition effect: Λ = 0.929, (F(4,292) = 2.712, p = .029, η^2 = .031. Separate follow-up univariate analyses showed no effect of condition for language use (F(1,152) = 0.219, p = .748, η^2 = .004). For content, a condition effect was observed (F(1,148) = 3.904, p = .022, η^2 = .050): the strategy condition Presentational outperformed the no-strategy condition (p = .02), the effect of the Modelling condition did not outperform the control condition (p = .099) No difference was observed between the two experimental conditions (p = 1.000).

Table 8
Text Qualities for the Posttests: Synthesis (SYN_t) and Transfer task (ARG_2). Means and Standard Deviations

		Modelling n = 57		Presentational n = 57		no-Strategy n = 41	
		M	*SD*	*M*	*SD*	*M*	*SD*
SYN_1							
	Content	60,48	18,32	59,72	17,31	62,26	16,19
	Source Use	129,68	40,07	109,11	48,37	93,38	29,66
	Authenticity	98,19	6,84	89,64	20,77	97,34	14,45
SYN_2							
	Content	84,90	15,17	83,11	19,22	87,43	14,50
	Source Use	132,37	28,94	130,61	39,59	113,75	25,84
	Authenticity	97,12	6,90	94,61	10,67	95,90	9,23
ARG_2							
	Content	93,93	12,07	95,42	14,50	89,21	12,34
	Language Use	95,73	12,65	96,50	12,09	100,61	22,19

To summarise, the effects of strategy training was not independent of the mode of instruction in SYN_1. What is demonstrated is an interaction between strategy training and instructional mode: In SYN_1, for source use and authenticity, the modelling condition was more effective than presentational condition. For authenticity, the control condition performed better than the presentational condition. However, regarding the transfer effect in ARG_2, the hypothesis that the strategy conditions would outperform the no-strategy condition is confirmed for only one of the two text quality scores, for only one of the instruction modes. That is, content scores are highest in the presentational strategy condition.

Generalisability. Subsequently, we successively analysed whether the three pretest variables (SUM, ARG_1/content, and ARG_1/language use; subsequently called the pretest score) interacted with condition and the three dependent variables of SYN_1. Only one out of the nine analyses showed an interaction effect: the pretest score indicating the quality aspect of language use in ARG_1 interacted with the conditions ($F(5,146) = 4,58$), $p < .001$, $R^2 = .14$). In the lower regions of the pretest scores ($B_{\text{instruction-mode}} = 34,70$ ($t(146) = 3.26$, $p < .001$), indicating that the modelling condition outperformed the other conditions. The effect of the strategy factor was not significant ($p = .42$). At the midpoint of the pretest scores, instruction mode was still significant ($B_{\textit{instruction mode}} = 17,05$ ($t(146) = 2,22$, $p = .03$), while the strategy factor gained weight in the prediction, although significant at the boundary ($B_{\text{strategy}} = 17,45$ ($t(146) = 1,95$, $p = .05$). Then, in the higher regions of the pretest scores, the strategy factor was significant ($B_{\text{strategy}} = 23,57$ ($t(146) = 2.09$, $p = .03$), while the instruction mode factor did not contribute to the prediction ($p = .99$). Figure 1 shows the patterns.

Figure 1
Interaction Effect of Pretest Language Use and Conditions on Synthesis Text Quality

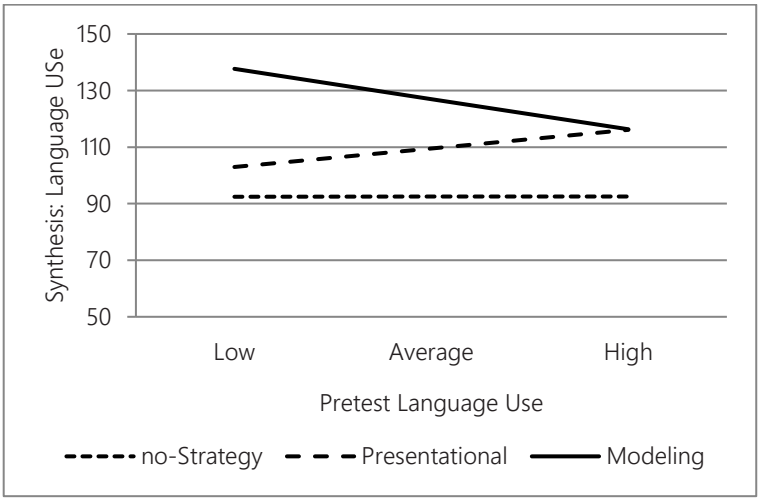


Figure 1 shows that within the no-strategy condition, the effect of the instructional condition does not depend on the pretest scores, while in the strategy conditions such an effect exists. For the weaker writers, the modelling condition seems to be most beneficial, while for the stronger writers both strategy conditions would be a good choice. The difference between the modelling and the

presentational condition for the weaker writers is large (ES > .80). The effect between the strategy conditions and the no-strategy condition for the stronger writers is .60.

3.2 *Writing Process*

3.2.1 Condition Effects on Process Activities (RQ3)

We collected data on online process activities via TSSR. The mean number of bleep responses was 31 (out of a set max of 45) (*SD* = 7). There were no differences between the three conditions in the duration of the process ($F(2, 144) = .84$, $p = .44$). The conditions did not differ in the number of the activity category 'Other' ($F(2,154) = 2,594$, $p = .078$), therefore here we focus on a total of six activities, as listed in Table 9.

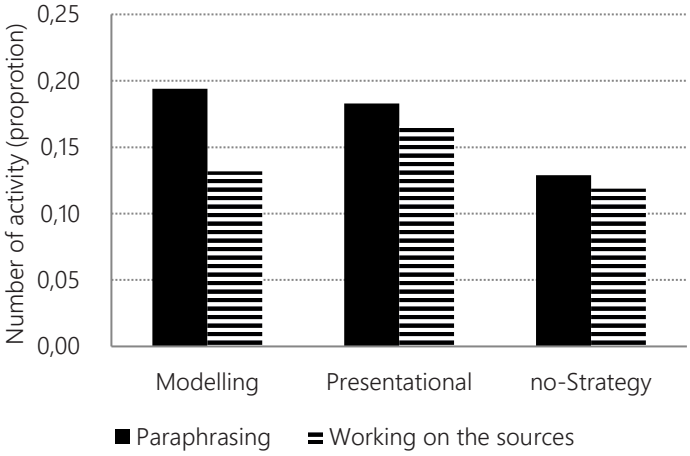
Table 9
Distribution of Process Activities Across the Three Conditions

	Modelling		Presentational		no-Strategy		
	M	*SD*	*M*	*SD*	*M*	*SD*	*Effect*
Reading the sources	.18	.11	.14	.08	.14	.11	
Paraphrasing	.19	.12	.18	.11	.13	.11	M/P>NoS
Working on the sources	.13	.09	.17	.08	.12	.09	P>M/NoS
Editing the target text	.09	.07	.10	.08	.12	.10	
Writing the target text	.32	.18	.33	.17	.38	.16	
Reading the target text	.09	.07	.08	.06	.11	.07	
Total source-related	.31		.31		.26		
Total target-text-related	.69		.69		.74		

We grouped each of the six activity categories under two broad activity clusters: *source-related* and *target-text-related* activities. The source-related activities included reading the sources and working on the sources, and the target-text-related activities included editing the target text, reading the target text, writing the target text and paraphrasing (source content). Paraphrasing is an activity between source and text production, between mentally and actually reformulating the ideas that were drawn from the sources. According to our definition in the TSSR-writing logs distributed to the participants, and the operationalisation

of the strategy in this study, the focus is on the actual formulation of the written text. Therefore, we included paraphrasing in the target-text-related activities.
Subsequent univariate analyses reveal that these effects are established in two activities: Paraphrasing (F(2,154) = 4,28, p = .02, η^2 = 0,053) and Working on the sources (F(2,154) = 4,03, p = .02, η^2 = 0,05). Pairwise comparisons showed that for Paraphrasing, the strategy conditions were proportionally more active (Modelling>Control: p = .01; Presentational>Control: p = .02). For Working on the sources, the presentational condition showed more actions (Presentational>Modelling: p = .04; Presentational>Control: p = .01). See Table 9 for the proportions per activity and condition, and Figure 2 for interaction effects for Paraphrasing and Working on the sources.

Figure 2
Interaction Effects for Paraphrasing and Working on the Sources (from Table 9)



The distribution of engagement in activities varied across conditions; Λ = 0.89, (F(10, 296) = 2.35, p < .01, η^2 = 0.07). Subsequent univariate analyses reveal that these effects are established in two activities: paraphrasing (F(2,154) = 4,28, p = .02, η^2 = 0,053) and working on the sources (F(2,154) = 4,03, p = .02, η^2 = 0,05). Pairwise comparisons showed that for paraphrasing, the strategy conditions were proportionally more active (Modelling>Control: p = .01; Presentational>Control: p = .02). For working on the sources, the presentational condition showed more actions (Presentational>Modelling: p = .04; Presentational>Control: p = .01). See Table 9 for the proportions per activity and

condition and Figure 2 for interaction effects for paraphrasing and working on the sources.

The difference lies in the extent to which participants divided their time in activities in source-related and target-text-related activities. In the strategy conditions, 45% of the time is spent on source-related activities; whereas in the no-strategy condition, it is 35% as is shown in Table 9. Another somewhat striking difference between the strategy conditions and the no-strategy condition is the variation in engagement in the process activities expressed in a standard deviation for the five activities, which varies from .05 (the modelling condition) to .01 (no-strategy condition), except for one activity (i.e., writing the target text). The variation between the activities in both strategy conditions is larger, which may indicate a more dynamic process (Braaksma et al., 2004).

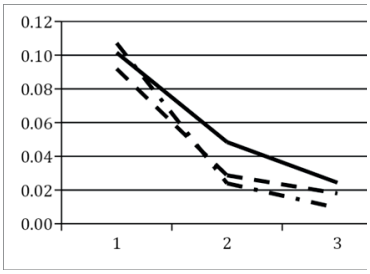
The next perspective to compare the processes is the distribution of the activities over time, exploring a purposeful temporal organisation of writing activities (Braaksma et al., 2004). See Table 10.

Table 10
Distribution of the Six Writing Activities over the Process

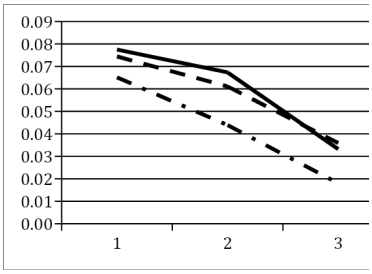
		Phase			Effects		
	Condition	1	2	3	Condition	Time	Interaction
Reading	Modelling	.10	.05	.02			
the source texts	Presentational	.09	.03	.02		1>2>3	No
	no-Strategy	.11	.02	.01			
Paraphrasing	Modelling	.08	.07	.03			
	Presentational	.07	.06	.04	M>C	1>2>3	No
	no-Strategy	.07	.04	.02			
Working	Modelling	.05	.05	.03			
on the sources	Presentational	.07	.06	.03	P>C	1>2>3	No
	no-Strategy	.06	.04	.02			
Editing	Modelling	.01	.02	.03			
the target text	Presentational	.01	.02	.04		1<2<3	No
	no-Strategy	.02	.03	.04			
Writing	Modelling	.07	.11	.10			
the target text	Presentational	.09	.13	.08		1<2>3	Yes
	no-Strategy	.07	.15	.12			
Reading	Modelling	.01	.02	.03			
the target text	Presentational	.00	.02	.04		1<2<3	No
	no-Strategy	.00	.02	.04			

Figure 3. A-F
Distribution of the Six Writing Activities
over Three Phases in Three Conditions

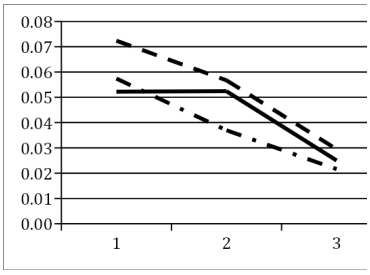
A. Reading the source texts



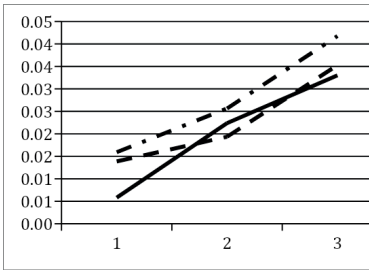
B. Paraphrasing



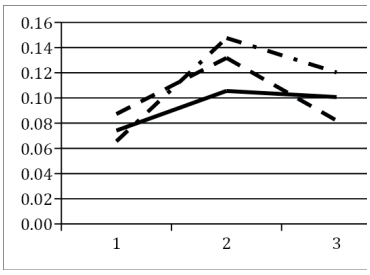
C. Working on the sources



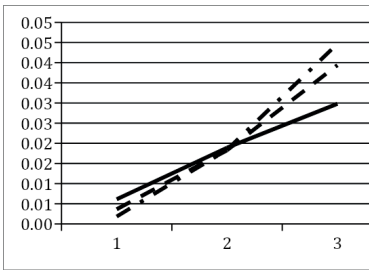
D. Editing the target text



E. Writing the target text



F. Reading the target text



Note. Presentational (line), Modelling (regular interrupted line), no-Strategy (irregular interrupted line). X-axis: three phases (1-2-3). Y-axis: Percentage of Activity.

Figure 3 illustrates that differences between the conditions are observed for paraphrasing and working on the sources. For both activities, one of the strategy conditions performed more of them. As for the effect of time, the two source-related activities show a linear decrease for time spent on these activities during the process, while the target-text-related activities show a linear increase, except for the activity 'writing the target text'. It seems that this development describes all but one pattern: for 'Writing the target text' an interaction between time and condition was observed, which indicates that the linear time-development pattern varied according to conditions (see Figure 3, E).

To sum up the results on the writing processes collected via TSSR, the first effect is related to the distribution of the reported activities. That is, the strategy conditions were reportedly engaged in the paraphrasing activity more than the no-strategy condition, whereas for working on the sources, the presentational condition reported more activities than the other conditions. Secondly, in the strategy conditions, process time that was allocated to source-related activities was more than in the no-strategy condition (i.e., %45 vs. %35 percent, respectively), which is an indication of recursivity and also their distribution of the process time across the different writing activities showed more variety than that of the no-strategy condition, which is an indication of a more dynamic writing process. Adding the time to the analysis as a within-subject variable helped us see the predominant activity that was reported in the different conditions over the writing process. Accordingly, the presentational condition was predominantly engaged in the activity working on the sources, significantly more than the no-strategy condition, whereas the modelling condition is predominantly engaged in paraphrasing the source content, significantly more than the no-strategy condition. Another significant effect was that we mostly observe a linear time development pattern, that is a linear decrease in students' engagement with the source-related activities, and a linear increase in their engagement with target-text-related activities, except for the activity writing the target text that was reported by the no-strategy condition in Phase 2, which means that the linear time development model varied according to conditions. For the activity of paraphrasing, the time development is a linear decrease, that is, in the reverse pattern than the other target-text-related activities, which was expected given that we considered the activity paraphrasing source content to be a transition between source-related and target-text-related activities, and in the final analysis it was categorised under the latter. Hence, this pattern was not an unexpected outcome.

3.2.2 Condition Effects of Learning Experiences (RQ3)

Table 11 shows the results of the Learner Reports for the three conditions. Participants generated 0 to 10 learning experiences, with a mean of 2,2 (*SD* 2.04). The number varied per condition ($F(2,152) = 3.306$, $p = .04$), with more experiences in the modelling condition than in the control condition ($p = .01$) and almost more than in the presentational condition ($p = .09$).

Table 11
Learner Reports: Mean Proportion of Responses (Standard Deviation) per Condition and Category

	Modelling		Presentational		no-Strategy	
(Main) Categories	*M*	*SD*	*M*	*SD*	*M*	*SD*
Content Management						
Main Ideas	0,77	0,43	0,71	0,46	0,71	0,51
Details	0,71	0,46	0,60	0,51	0,60	0,51
Synthesising Skills						
Paraphrasing	0,78	0,42	0,79	0,42	0,79	0,51
Summarising	0,85	0,36	0,85	0,36	0,85	0,47
Source Use						
Organisation	0,57	0,51	0,40	0,52	0,40	0,49
Linkers	0,57	0,51	0,40	0,52	0,40	0,33
APA Style	0,71	0,46	0,65	0,49	0,65	0,47
Reporting Verb	0,56	0,53	0,43	0,53	0,43	0,00
Process Knowledge						
Steps/Stages	0,63	0,50	0,57	0,51	0,57	0,47
Time- management	0,71	0,46	0,45	0,52	0,45	0,47
Think	0,45	0,52	0,25	0,46	0,25	0,47
Edit	0,14	0,38	0,25	0,46	0,25	0,00

The numbers in Table 11 show the proportion of the responses in a certain category. For two categories significant condition effects were observed. For paraphrasing ($F(2,68) = 3,59$, $p = .03$, $\eta^2 = .10$) and time- management ($F(2, 42) = 3,27$, $p = .048$, $\eta^2 = .14$), the modelling condition resulted in more responses than the no-strategy condition (for both comparisons $p = .02$), while for paraphrasing also the presentational condition resulted in more responses than the no-strategy condition ($p = .02$).

4 DISCUSSION

This study can be seen as a replication, on a larger scale, of a previously published study on the differential effects of a modelling mode and presentational mode in explicit strategy instruction that is typically present in strategy instruction programs, particularly in Self-Regulated Strategy Development (SRSD), which is also the instructional model of this study. In Chapter 2's study we demonstrated positive effects of modelling strategies on students' text quality, which were maintained in a second posttest four weeks later, also, on students' writing processes. To get a more stable understanding of the previously observed effects, we pursued a similar research endeavour with a larger sample size in this study (i.e., 155 participants vs. 48 participants in the previous study). The writing task was an integrated summary of three source extracts, which is an introductory level synthesis writing task used in instruction for developing learners in English as a second language (L2) and academic writing. The independent variables were the instructional conditions for teaching the synthesis task, that is, modelling mode, presentational mode, and the regular course. Dependent variables were text quality for synthesis texts, transfer (of text quality) to another genre, and writing processes.

4.1 *Text Quality: Condition Effects on Synthesis Writing (RQ1)*

Our first research question (RQ1A) investigated whether the effects on text quality were larger when the synthesis task was instructed via explicit strategy instruction (i.e., the strategy conditions or the experimental conditions, subsequently used interchangeably to refer to the same concept) as compared to a regular course with no-strategy instruction. Also, within the strategy conditions, we investigated (RQ1B) whether it is the modelling mode or the presentational mode of instruction that resulted in qualitatively better synthesis texts. The answer to RQ1A was negative. There is no overall statistically significant effect between the strategy conditions and the no-strategy condition. What we found was an effect of instructional mode, with the modelling mode of instruction resulting in qualitatively better texts than the presentational condition in two out of the three aspects that we used to rate the synthesis texts, that is, in source use and authenticity aspects of text quality. Hence, the answer to (RQ1B) was positive and corroborated the findings of the previous study (Buyuktas Kara et al., 2018, Chapter 2). The results also showed that the no-strategy condition (i.e., the regular course) wrote more authentic texts than the presentational condition which was a new finding when compared to the results of the previous study where no

significant differences between the presentational and the control conditions were found (see Buyuktas Kara et al., 2018, Chapter 2).

The finding that the modelling condition resulted in significant success compared to other conditions was expected based on the findings of the previous study (Buyuktas Kara et al., 2018, Chapter 2), and on the theoretical and empirical studies on observational learning (see literature review). We explain the success of the modelling mode in the source use aspect of text quality, by the novelty feature of the synthesis task. In Turkey, in L1 and L2 writing curricula at secondary level, students are mostly assigned argumentative tasks based on personal experience and knowledge. The technicalities of a synthesis task that we rated in the source use aspect of text quality in this study (i.e., using correct APA in-text citations conventions, a variety of reporting verbs, and a variety of linkers for coherence) are new to students. Previous studies have shown that observation works best for novel tasks (Braaksma et al., 2004), or for the novel features of a particular writing task (Buyuktas Kara et al., 2018, Chapter 2). Our findings seem to corroborate this for synthesis writing in L2. The explicit strategy we used for instruction in citation use also partly explains the success of the modelling mode in the authenticity aspect of text quality. The strategy for citation use was an elaborated version of a phraseological pattern from the corpus of graduate theses of L1 writers (i.e., "Author/date argues that …"; see Charles, 1996). It conveniently packs the function and the form of the citation, the linking devices, APA in-text citation conventions, but also the category of the language skill (i.e., summary/paraphrase) required for incorporating an idea from the source texts in the follow-on sentence (Gao et al., 2021). Paraphrasing avoids plagiarism, the absence of which is scored in the present study as authenticity. We handled paraphrasing as a distinct strategy step during the training as it is known to tax the linguistic skills of developing writers in discourse synthesis in an L2 context (Leki & Carson, 1997). The citation strategy, when modelled, seems to have helped students not only in the unfamiliar concept of citation use, both in terms of form and function, but also in the paraphrasing requirement of the follow-on sentence. We argue that this simple strategy of partitioning the citation formulation in smaller manageable steps, in line with the instructional recommendations for citation use and synthesis writing instruction, (see Literature Review), when modelled by a peer, helped with the time-management in the subsequent timed-writing task and also reduced the burden on linguistic proficiency. These two factors are interrelated, especially for the participant profile in our study who are still learning an L2 and are new to the academic writing discourse, which will be further explained in the discussion on the writing process, as data from the Learner Reports revealed complementary findings.

4.2 *Generalisability*

Regarding the effects on text quality that we demonstrated above, we were also interested in their generalisability across all levels of proficiency (RQ1C). Previous studies showed that within a specific participant profile initial differences (i.e., learner variables) based on a specific criterion for classification, may affect the results of observation. In the study by Braaksma et al. (2002), the learners were classified in two groups based on aptitude levels before the training (i.e., initial levels of aptitude), and also the model used in the observation (i.e., weak vs. strong models). The results of the study showed that initially low-aptitude students benefited more from observation (focusing on weak models) than initially high-aptitude students. In a later study (Braaksma et al., 2018), students were classified based on initial levels of writing performance, without the manipulation of the observational learning conditions as modelled by weak vs. strong models. This time, almost no differences in effects on text quality were observed. There is also research (Van Steendam et al., 2014) showing that observation is more effective than no observation, e.g., more traditional practicing conditions, for all types of writers (i.e., of all ability levels). Hence, we were unable to build a solid hypothesis on this issue based on the literature and included it as an exploratory question into the generalisability of the effects to two levels of initial writing proficiency. Our findings attested to the findings of the study by Braaksma et al. (2002). The modelling condition seems to be the most beneficial for initially lower-performing students, whereas for the initially higher-performing students, both strategy conditions would be a good choice.

4.3 *Text Quality: Condition Effects on Transfer to Argumentative Writing (RQ2)*

To investigate the second research question, we checked whether students' text quality in the synthesis posttest (SYN_1) transferred to an argumentative writing task (ARG_2) which was conducted five weeks later (RQ2). For exploring the transfer effects, we compared the strategy conditions to the no-strategy condition (RQ2A), and the modelling condition to the presentational condition (RQ2B). Argumentative texts were rated in two aspects pertinent to argumentative writing: a content composite score which was operationalised by the overall organisational structure and the strength of argumentation in the texts; and a language use composite score which was operationalised by the variety and the accuracy of the lexical and the grammatical structures used in the texts. The answer to (RQ2A) was partly positive. The transfer effect to the argumentative task for the content aspect of text quality was larger in the presentational strategy condition compared to the no-strategy condition; whereas the answer to (RQ2B) was

negative, there were no differential effects between the modelling condition and the presentational condition in effects on transfer.

The discussion of this finding is interesting when taken from a (cognitive) developmental perspective of writing skills (see Guo et al., 2020). Bennett et al. (2016) categorised summarising from multiple sources, which is the type of synthesis task examined in this study, as the initial step toward more cognitively complex synthesis tasks, such as argumentative writing based on sources. This classification aligns with an expert's problem-solving behaviour for argumentation when dealing with multifaceted tasks (Bennett et al., 2016). Therefore, the observed transfer of learning from synthesis writing to argumentative writing in the strategy conditions is well-founded. The aspects of quality transfer (i.e., content & organisation) are also in line with the instructional goals that are represented in synthesis writing. That is, creating a super-proposition regarding the content (Segev-Miller, 2004), a macrostructure for the organisation of ideas (Van Ockenburg et al., 2019), and an intertext that connects all sources (Boscolo et al., 2007). Of course, considering that the modelling condition outperformed the presentational condition in synthesis writing at posttest (i.e., SYN_1), we would expect transfer skills to represent the same trend. However, since this task is not a source-based writing task, the two learning gains of the modelling condition from the synthesis tasks, that is, authenticity (as the absence of plagiarism) and source use (proper citation use) were not operationalised in the argumentative writing task. We expect that if argumentation skills were tested in a source-based manner, as in an argumentative research paper for example, the modelling condition would outperform the presentational condition in the transfer test as well, considering that representing the processes of an expert's problem- solving skills would be best achieved via peer-modelling of strategy use.

4.4 *Writing Process: Condition Effects on Process Activities (RQ3A/B) and Learning Experiences (RQ3C)*

Our third research question inquired into which instructional conditions resulted in better writing processes (RQ3), by comparing the strategy conditions and the no-strategy condition (RQ3A), and the modelling and the presentational modes of instruction used in explicit strategy instruction (RQ3B). To address research questions RQ3A/B, we evaluated the TSSR-outcomes while considering the unique demands of the synthesis task in our current study (and the TRAMPOLINE strategies taught as part of the intervention). Additionally, we applied the quality criteria for effective temporal organisation of cognitive activities as previously defined in previous studies, for testing the impact of observational learning

(Braaksma et al., 2004) and for identifying the processes involved in writing a synthesis text (Solé et al., 2013). In the study by Braaksma et al. (2004), observational learning resulted in the distribution of metacognitive activities such as goal-orientation, analysing and planning in the initial phases of the writing process, whereas executive activities were delayed to the later stages of the process. In this study, the metacognitive activities pertain to (reading and working on) the given hierarchical argumentative idea units and the executive activities pertain to the (formulating, writing, editing) for writing an L1 argumentative task. Furthermore, a variation in engagement was noted across the various writing activities within the observational learning conditions. This variation suggested a more dynamic approach to writing when compared to the relatively consistent and less varied activities observed in the writing-only conditions, which indicated a more "monotonous" writing process. We expected to see similar results in the writing processes of the modelling condition. To identify the expected changes in processes, we also referred to the study by Solé et al. (2013) who used a similar type of synthesis task was used as in the present study, that is, an integrated summary based on multiple source texts. They demonstrated that the qualitatively better texts were the result of writing processes that incorporated a relatively more *recursive use of reading* for producing the target text. That is, using sources as reference throughout the writing process to mediate the several stages in writing, such as planning, writing and revising. Based on the parameters we devised from the two studies, we proved the recursivity hypothesis in favour of the strategy conditions over the no-strategy condition (RQ3A); and in favour of the modelling condition over the presentational condition (RQ3B), which corroborated findings of the previous study (Buyuktas Kara et al., 2018, Chapter 2). There were differences with regard to the particular process activities, which did not hamper the recursivity hypothesis (that we will report in the following paragraphs).

We also used the offline Learner Report method to complement the results of the online TSSR-method, which showed students' own perceptions about their learning (i.e., learning experiences). Hence, before proceeding to discussing the online writing behaviours of the students, we will report the results of the learner reports. The answer to RQ3C, which inquired whether the instructional conditions had distinct effects on students' learning experiences, was positive. The strategy conditions reported learning experiences that were coded under the activity paraphrasing significantly more than the no-strategy condition. Significant effects were also observed for the modelling condition, with students reporting more learning experiences about the various task requirements of the synthesis task as a result of the instructional intervention than other conditions,

and also significantly more learning experiences in the category time-management than the other conditions. This suggests that they adopted a more time-efficient approach to the synthesis writing task, which led to changes in their perceptions about their ability to meet the various task requirements within the time allocated for the timed-writing exam.

The majority of the effects on students' online writing processes (as measured by TSSR) were common for both modes of instruction in the strategy conditions. The strategy conditions were engaged in the activity paraphrasing proportionally more than the no-strategy condition. Their process activities also showed more variation throughout the writing process, and they allocated their process time to source-related activities more than the no-strategy condition (i.e., %45 vs. %35). The differences between the modelling and the presentational modes of instruction is that, the presentational condition reported being engaged in the activity working on the sources more than the other conditions. In fact, when the time variable was added to the analyses, the effect of being engaged in paraphrasing for the strategy conditions translated into significantly more paraphrasing in the modelling condition compared to the no-strategy condition, and significantly more working on the sources in the presentational condition compared to the no-strategy condition. A finding that was new to this study when compared to the previous study (Buyuktas Kara et al., 2018, Chapter 2) was that the time-development models showed a linear pattern over the course of writing. This means that the different conditions did not differ in their engagement with the particular writing activities in the separate phases of the writing process, with the exception of an interaction effect for the no-strategy condition and the activity writing the target text in Phase 2.

The type of synthesis task we used in this study is aimed at introducing students to discourse synthesis focusing on its novel features, specifically citation use. The instructional component that we used for instruction on citation use for source attribution was the explicit strategy (i.e., OLIN of TRAMPOLINE) that included the phraseology instruction, and corresponded to the source use aspect of quality. Paraphrasing source content is a distinct component in the TRAMPOLINE strategies (P of TRAMPOLINE, see Table 5), hence it was instructed in a separate session (see Table 3), and was scored under the authenticity aspect of text quality. However, citation use is composed of two distinct requirements, that is, source attribution and paraphrasing (the selected ideas taken from the source texts). Therefore, the phraseology strategy we used for the instruction of citation use also tapped into paraphrasing skills at a declarative level, that is, the necessity to employ diverse formulations for paraphrasing ideas taken from sources in the follow-on sentence (i.e., noun phrase, full sentence, complex sentence,

etc.), which is controlled by the phraseology of the citation that precedes it. See below:

Author (date) argues that another reason for x is y (follow-on paraphrased idea in the form of noun phrase)

Author (date) disagrees with Author (date) on the causes of x stating that although… (follow-on paraphrased idea in the form of a complex sentence using a subordinating conjunction)

This explicit strategy instruction on citation use resulted in changes in the process activities of the presentational condition. We can track these changes in their process behaviours, that is, working on the sources significantly more than the other conditions, and a higher probability of occurrence of different writing activities, as well as a more recursive use of reading (i.e., source-related activities) for producing their texts (i.e., target-text-related activities), as shown by the process time allocated to these activities in the strategy conditions. When we check the text quality scores, the process activities of the presentational condition translated into marginal success over the no-strategy condition in source use aspect of text quality, and the modelling condition outperformed both conditions in source use. Regarding authenticity, which refers to the absence of verbatim source use, both the modelling condition and the no-strategy condition performed better than the presentational condition. This is noteworthy because, in learner reports, students in the presentational condition indicated having more learning experiences related to paraphrasing compared to the no-strategy condition. These results show that the presentational condition increased their declarative knowledge on citation use which comprises source attribution and the need to paraphrase the original ideas, resulting in changes in their writing processes; however, these changes did not translate into higher scores in the corresponding aspects of text quality (except for a marginal success in source use compared to the no-strategy condition).

Based on these results, we argue that for the presentational condition an overall task representation was formed at a declarative level. However, because the students in the present study are developing learners in L2 and relatively naive to discourse synthesis, the recently learnt task requirements may have taxed their linguistic skills and overall writing competencies, and their time on task, especially considering the timed-writing conditions of the posttest. In the face of several task requirements, the presentational condition may have opted for formulation of source attribution, as it is more about the technicalities of the task, and less dependent on L2 proficiency, therefore, a possible trade-off may have occurred between the citation formulation operations and the

paraphrasing operations in the writing process. This could also elucidate why students in the no-strategy condition achieved better results in terms of authenticity than those receiving presentational instruction. It is possible that the coping strategy employed by the presentational condition for tackling a recently acquired skill, might have interfered with their performance during the writing task. Conversely, students in the no-strategy condition, likely due to the absence of an explicit strategy for source use, may have concentrated on a single overarching strategy applicable to the entire task, namely, the imperative need to construct sentences using their own words.

In an intervention it can be expected that the experimental conditions show a more time-dependent distribution indicating that the activities are adapted to the progressing stages of the task completion (Van den Bergh & Rijlaarsdam, 1996; Rijlaarsdam & Van den Bergh, 1996). However, this does not automatically lead to higher scores in every instructional condition, and not for every aspect of text quality. In fact, instruction can sometimes improve certain aspects of text quality but raise other issues in other aspects that are connected to it, which can indicate a possible trade-off between quality aspects. This was shown in an action research (Wette, 2010) on source use and citation. That is, although both declarative and procedural knowledge increased as a result of the intervention, other source-use problems emerged in the posttest, such as partial and/or inaccurate paraphrase and inappropriate source use. Hence, when the instruction on a particular writing task is phased via simpler tasks, learning may extend over the course of the intervention meaning that the corresponding quality aspects in students' texts may not (yet) be visible in the subsequent posttest. A timed-writing condition is also a contributing factor particularly on task-specific aspects that are attended to in the later stages of a writing process which may be cut short because of the time constraints (see Barkaoui, 2016), and may explain the lower text quality scores.

In the strategy conditions, as per the quality criteria we established for effective synthesis writing processes (as discussed in Solé et al., 2013), students were instructed to read all three sources before proceeding to paraphrase the selected ideas (as outlined in the TRAMPOLINE strategies in Appendix A). However, in the absence of procedural knowledge regarding how to efficiently orchestrate these operations, students may have delayed the paraphrasing activity to the point where they were unable to complete the synthesis writing process within the time constraints of a timed-writing environment. On the other hand, the modelling condition proved to be effective in helping students manage their time more efficiently, enabling them to allocate it to the task requirements that are linguistically more demanding, such as the need for paraphrasing in citation

use. The data from offline learner reports support our assumptions, as the results indicate that while both conditions reported learning experiences related to paraphrasing, the modelling condition also reported significantly more learning experiences in time-management than the other conditions. This is particularly advantageous in timed-writing tasks. Students in the modelling condition performed significantly better in authenticity and source use, compared to the other conditions. Observation in this case seems to have resulted in changes in students' own perceptions about their temporal management of the writing process towards a more efficient one, and consequently, resulted in qualitatively better synthesis texts.

We interpret these findings within the framework of Bandura's Social Learning Theory (1976). Accordingly, learning of declarative and procedural knowledge, and the correlation in between, is more robust when they are enabled through vicarious learning experiences (Bandura, 1976/1986). From this perspective, modelling serves as a valuable instructional method. When students observe models, they allocate their cognitive resources to metacognitive and procedural knowledge instead of focusing predominantly on the production process (see Rijlaarsdam et al., 2005). See Table 12 for the different dimensions used in instruction across the conditions of the study.

Table 12
Differentiation of the Dimensions per Condition

	Modelling	Presentational	Control
Declarative knowledge	1	1	0
Procedural knowledge	1	1	0
**Telling	0	1	0
**Showing	1	0	0

The different types of delivery of declarative and procedural knowledge in the different conditions resulted in different effects on text quality. In both strategy conditions, we imparted students with declarative and procedural knowledge (explaining what to do and how to do it) through the TRAMPOLINE strategies mnemonic. However, in the presentational condition, this knowledge was conveyed through telling, while in the modelling condition, it was presented via showing. The mode of delivery, whether through telling or showing, when it comes to explicit strategies for synthesis writing, exerts different influences on students' writing performance. Explicit strategy instruction is effective but is most

impactful when demonstrated by a model, rather than merely explained for students to independently perform it.

4.5 *Conclusion and direction for future research*

A replication study contributes to the verification and generalization of knowledge and understanding in empirical studies (McManus, 2022). When the study is replicated by the same author, there is more confidence that the methods have been followed verbatim; however, the rate of replication by the same author, especially in education research has been found to be low compared to other domains (Makel & Plucker, 2014). In this study, we replicated the study on a larger scale to rule out sampling errors in hypothesis-testing (see Schmidt, 2017). Although in the previous study (i.e., Chapter 2) the effects of observation were already prominent, for example, for the source use aspect of text quality, in this study they are extended to also include the authenticity aspect of text quality. The different sample sizes of the two studies and the different participant profiles (coupled with the same mastery model) can explain the new findings.

It is plausible that several limitations may have influenced the results. First, as a way of channelling observers' *attention* to the modelled performance, which is the first of the four process components of observational learning (Bandura, 1986), evaluation and elaboration tasks have been successfully used in intervention studies. Since in our instructional model (i.e., SRSD) such tasks were absent, we did not include one in our study either. Rather, we presented the strategies in three subsets (i.e., TRAM, P and OLINE strategy mnemonics/videos) instead of exposing the students to an entire performance. In this way, that is, by decomposing the task in segments and by highlighting the relevant skills at the different stages, we channelled the learner's attention to the modelled activity (Bandura, 1986) while at the same time adapting the strategy to the target population. Although there are observational learning studies where no evaluation and elaboration tasks are included (see Zimmermann & Kitsantas, 2002), in experimental studies on writing (Braaksma et al., 2001) it has been shown that such tasks cultivate larger learning gains (see Braaksma et al., 2001). Thus, future studies on observation are recommended to include evaluation and elaboration tasks (among others) to ensure students' utmost engagement with the modelled behaviour.

Second, the growing interest in conceptualizing the different kinds of strategies (i.e., cognitive, metacognitive, affective) makes it important to embrace a mixed-methods approach and combine product data with process registration

data in strategy instruction interventions (Schellings & van Hout-Wolters, 2011). Interestingly however, few studies seek such an undertaking. Martínez et al. (2015) investigated the effects of strategy instruction on the distribution of activities across the writing process through individually video-recording students' writing materials during task execution as well as its effects on content learning and synthesis writing quality. Torrance et al. (2007) combined text quality data with TSSR, a process registration technique also adopted in this study. In the present study, apart from the intervention effects on writing performance, we applied a multiple process measuring method of online and offline techniques, which is recognised to enhance the validity of the results (Schellings & Van Hout-Wolters, 2011). The mixed-methods approach helped us to get further insight into the students' writing experiences from different angles, that is, their writing performance, writing processes and retained writing knowledge (1), back up assumptions drawn from one of the techniques (2), see the effects of different writing processes on writing quality (3), refine the strategy steps and the results obtained (4), as well as get an idea of the relationship between declarative and procedural knowledge (5). We believe that future studies on strategy instruction should also incorporate at least one process registration technique for the aforementioned reasons.

Preferably, the effects of instructional variables on text quality and on task processes are not determined with one and the same task when using TSSR. It could be a concern that interrupting the writing task and coding the process activity could affect text quality. In the present study, we therefore presented two tasks, one in which we determined text quality and one in which we collected self-reports on processes. This is not ideal because this way changes in processes and in text quality should not be directly linked. A study in which we could determine the effects of TSSR on writing process behaviour and on text quality could teach us more about an efficient measurement design. Such a study should also isolate the effect of TSSR training. We are well aware that the scores in the second synthesis posttest task may have been influenced by the brief TSSR video-based training. We cannot rule out that the effects on writing process behaviour may have been partly determined by that short video training, and that a potential instructional effect of the modelling instruction may have been nullified by that training, because participants from other conditions now also received some kind of training by modelling.

Third, to maximize clarity in distinguishing between the process activity categories (i.e., strategies), we conducted a thorough and intensive training for implementing TSSR. This training encompassed instructor training and subsequent student training. We avoided including overarching activity categories in the

writing logs, which had previously been used as instructional strategies, as they could overshadow the reporting of other activities, such as Thinking (the "T" in TRAMPOLINE; see Appendix A). Since our primary focus in this study was on cognitive strategies rather than metacognitive strategies, and for ease of interpreting the results, we selected activities that could be readily correlated with other factors in the various dependent variables. In retrospect, a similar approach could have been applied to the activity 'editing the target text', as it did not appear to impact text quality scores in a discernible manner. On the other hand, to prevent activity categories from overlapping, we identified two distinct reading activities within the writing logs. The first involved "reading the sources," which entailed the mere act of trying to comprehend the sources. The second, termed "working on the sources," represented a strategic approach to reading, involving attempts to identify main ideas, the important details and looking for information for source attribution. Future studies that incorporate both writing logs and activity categories alongside text quality measures might consider refining the factors within their dependent variables to establish clearer connections between them. Alternative methods of collecting process data, such as keystroke logging or think-aloud protocols, are also viable options. Such more fine-grained process data can also assist in the adaptation and refinement of strategies in instruction. However, they may not always be feasible, as was the case in the present study.

Finally, although we adopted SRSD as our instructional model and applied its stages in the design of this intervention at a micro-level, this study does not aim to test the effectiveness of SRSD as a training program, but rather, compare explicit strategy instruction to no-strategy and the two modes of instruction that are typically present in SRSD, that is, *modelling* and *presentational modes*. SRSD is a long-term strategy training program that is spread over at least a few weeks. Instructors are advised to proceed to the next stage only when the students have mastered a prior stage. The present study, on the other hand, was embedded in an actual curriculum, that is, the English preparatory program of a Turkish university, where there was no room for a longitudinal and stepwise implementation. The study was conducted in real classrooms with students trained by their main class teachers. The duration of the intervention had to match the actual time slot allocated for teaching of the synthesis task by the institution. Also, the explicit instruction we used in our study did not include strategies for self-regulation, except for the ones that were present in the partially-scripted think-aloud protocols of the peer videos, both because of time constraints, and also because we wanted to isolate the effects of cognitive strategies. Although these are also strong aspects of the study towards ecological validity, they also constitute the

reason why our results are not generalisable to studies that operationalise a form of SRSD. However, in the measurement of text quality, we used three different genres (i.e., summarising, synthesising and argumentative writing) with four different tasks, two pretests and two posttests, as well as another synthesis posttest with which we operationalised the writing processes, a forte of the study reported on here. It has been suggested that to be able to reach generalisable conclusions in L2 research with secondary school level students, there should be a total of 3 to 4 assignments per student each rated by two raters (Schoonen, 2005). Thus, we could tentatively state that the results of text quality of our study are generalisable within the strategy-focused writing intervention programs in L2 writing and provide ground for future research of similar interest.

CHAPTER 4

REACTIVITY OF DIRECTED RETROSPECTION IN INTERVENTION STUDIES: LEARNING FROM THE TEST

Hayes and Flower's state-of-the-art research (1980) introduced a new outlook on writing as a problem-solving activity and pioneered a shift of interest from the final product to writing processes. Despite the fundamental modifications introduced to its architecture by Hayes (1996/2012), the major underpinnings of the model still prevail. We now know that the written text is the product of a continuous interaction of several cognitive processes with each other, and with the task environment, including the text-written-so-far and the source materials. The temporal organisation of these cognitive processes has been the focus of interest of researchers seeking to gain an insight into the functional-dynamic nature of writing (Kellogg, 1987; Van den Bergh et al., 2016).

To observe writers' cognitive processes, a variety of methods have been adopted particularly from social sciences and psychology (Olive, 2010), others continue to be developed for the purposes of writing research. Within this methodology, a general classification is made between offline (i.e., retrospective) and online (i.e., concurrent) methods. Offline methods are executed after the writing act has been completed, and include interviews, questionnaires and stimulated recall, among several other techniques, in which writers are asked to reflect about certain aspects of their writing processes. Online methods are executed throughout the actual writing act, and include the dual/triple task technique, think-aloud protocols (TAPs) and keylogging (Rijlaarsdam et al., 2012; Vandermeulen et al., 2023). Studies using one or more of these methods either in the context of an intervention study or in a methodological study investigating method reactivity, agree that both offline and online methods provide valuable information about writing processes, albeit with several validity and reactivity pitfalls. In the context of observing the cognitive processes in writing, validity of a method indicates whether the observation completely and accurately represents the writing process; whereas reactivity of a method concerns whether the

writing performance or its outcomes are altered as a result of the observation (Ericsson & Simon, 1984/1993).

Offline methods (i.e., interviews, questionnaires, stimulated recall) entail retrospection which ensures that the technique does not interfere with the writing act itself. However, the reported activities are void of the time stamp needed to explain their temporal organisation within the writing process (Torrance et al., 1999). They rely on writers' memory and represent the writing process from the writers' own perspective (DiPardo, 1994). However, long-term memory stores the experiences transferred from short-term memory in a simplified and transmuted nature (Odell et al., 1983). This implies that the data gathered through retrospection may be incomplete because the complexity of the thought processes can be disturbed during retrieval (Van Someren et al., 1994). Considering the reporting mode, retrospective protocols are interactive and flexible (Hyland, 2016). Yet, precisely because they are interactive, the results of the study may be biased with the interviewer's selective choice of questions and/or participants' tendency to produce socially desirable answers, also known as response-bias, which can cause a measurement error (Fisher, 1993). In writing studies, for example, participants' preconceptions of a desired writing process (i.e., metacognitive knowledge) may interfere with their responses (Flower & Hayes, 1981). Thus, we can conclude that offline methods provide a viable triangulation possibility, complementing other methods. However, because of validity concerns, they are not always insightful when used on their own (Wengelin et al., 2019; Levy et al., 1996).

The typical setbacks of offline methods, that is, memory distortion and question/response bias, are circumvented in online methods, particularly, in think-aloud protocols (Green, 1998). TAPs are employed concurrently with the writing act, and writers are asked to verbalise anything that passes their mind while writing, even negligible ideas and incomplete sentences. The data is then recorded, and later transcribed and coded in categories according to the experimental goals of a study. TAPs provide rich data about specific instances of the writers' actual practices rather than their generalised statements and beliefs about what writing should be like (Van Someren et al., 1994). However, the concurrent nature of the TAPs implies that writing and thinking aloud both tap simultaneously into short-term memory which may result in a (cognitive) overload (Rijlaarsdam et al., 2012). Taking into account the demands of simultaneous decision-making and limitations of short-term memory, many important decisions are likely not to be remarked upon by writers (Odell et al., 1983), which may result in incomplete or biased protocols and pose a threat to their validity (Rijlaarsdam et al., 2012). Moreover, automated performances or immediate recognition processes are not recorded in short-term memory, hence they will not be verbalised by

writers who have frequently practised certain processes (Ericsson & Simon, 1993). Hence interpersonal and contextual variation in the quality and the quantity of verbalizations are inevitable (Barkaoui, 2011). In terms of reactivity, empirical data showed that in TAP-conditions writers produce texts at a slower rate compared to silent writing conditions (Ransdell, 1995), especially in cognitively more demanding tasks that require knowledge-transforming rather than more simple tasks such as knowledge-telling (Janssen et al., 1996). As for practical concerns, the implementation of the TAP-method and transcribing, analysing and coding of the resulting data is laborious (Green, 1998). That is why it is recommended as a research tool for use in the laboratory rather than for teaching or evaluation purposes in the classroom (Hayes & Flower, 1980). Finally, although researchers agree on a number of caveats surrounding TAPs, especially about the incompleteness of reports, the reported data is still valid and provides an insight into writing behaviour that otherwise would not be accessible with other research methods (Ericsson & Simon, 1984/1993; Hayes & Flower, 1981). It shows what happened during the process, and thus is valid, but it does not show all that happened, and thus may be incomplete.

Another online method for observing writing processes is derived from the dual task technique that originates from studies in neuropsychology (Kerr, 1973). It explores the basic principle of cognitive theory that diverse cognitive processes, when undertaken simultaneously, compete for a limited pool of attentional and executive resources (Baddeley & Hitch, 1974; Kahneman, 1973). In the dual task technique, writers are given two parallel tasks, writing as the primary task and as a secondary task, the reaction-time (RT) task that requires writers to respond to an auditory probe as quickly as possible. The time lapse between the probe and the writers' response to the probe, when compared to baseline RTs to the same stimuli as a single task, gives an insight into the amount of cognitive effort spent on the primary task. This view, however, has been challenged with the Attentional Boost Effect (ABE) hypothesis, which proposes that behaviourally relevant items presented concurrently (rather than a priori or posteriori) at the encoding stage, enhances perceptual processing at retrieval (Swallow & Jiang, 2010). As for its relevance to online writing processes, the dual task method does not specify which cognitive activity was interrupted at the time of the probe, hence it falls short in providing a complete insight into the temporality of the writing processes. Kellogg (1987/1988) therefore devised the Triple Task Technique (TTT), by including a tertiary task in order to register students' online covert writing processes, called directed retrospection (with concurrent probing) (Ericsson & Simon, 1980), or immediate introspection (Fidalgo et al., 2014). In TTT, writers are asked to concurrently (1) complete a writing task, (2) respond to

auditory probes that are played at a variable interval schedule (with a mean interval duration of 60 and 30 seconds, depending on the specific experiment) by saying "stop" as quickly as possible, and (3) choose out of the discrete writing categories (i.e., planning, translating, reviewing), based on the cognitive model of the writing process (Flower & Hayes, 1980), the activity they were engaged in at the time of the probe by pressing a key on the computer keyboard allocated to specific writing activities. A category *unrelated* is also included for all activities that do not fit in the experimenter's predefined categories. TTT enables researchers to measure the processing time along with the cognitive effort required by each writing activity, as well as their distribution and frequency throughout the writing process. It can be used in studies with different research objectives by specifying its parameters and modulating its configuration (Kellogg, 1987), which have been defined as the duration of the training phase, the interval schedule of the auditory probes and the number and type of the activity categories (Piolat et al., 1999).

Subsequent studies (Garcia-Sanchez & Fidalgo-Redondo, 2006; Fidalgo et al., 2008; Torrance et al., 1999; Torrance et al., 2007) used a variant of TTT, called Time-Sampled Self-Reporting (TSSR). In TSSR, cognitive effort is not operationalised, so the secondary RT task is precluded from its overall configuration. The function of the auditory signals, that appear at random intervals between 60-120 seconds (mean = 90 s) is to probe the students to proceed to the tertiary task of directed retrospection, which makes it the secondary task in the TSSR-configuration. The directed retrospection task in TSSR is expanded to capture more task-specific cognitive processes, hence includes a higher number of activity categories (i.e., varying between 7-12, including a category *unrelated* for activities that do not fit the given categories). Students report their writing processes using a writing log sheet by ticking the allocated category slots in paper-and-pen conditions, indicating their writing activity at the moment of the probe. The presentation of the writing log sheet, compared to the allocated keypresses on a computer keyboard in TTT, facilitated the use of TSSR in classroom settings as an online writing process registration technique.

A remarkable difference between undirected verbal responses of TAPs and directed retrospection in TTT/TSSR is that in the former, writers are instructed to report on anything while they are writing, and after the collection of the raw data, a panel of trained coders categorise the writers' undirected verbalizations in categories predefined by the experimenter. In the directed retrospection task of TTT/TSSR, the role of the coder is undertaken by the writer who has to code their thoughts when probed concurrently with composing. Of course, a disparity

between the probed categories and the students' mental representations of them can pose a threat to validity of directed retrospection data (Ericsson & Simon, 1993). Therefore, as in all data coding procedures, writers need to be trained on the directed retrospection task, to fulfil their new role as coders, and also to gain fluency in the mechanical aspects of the technique that they need to attend to concurrent to text-production. To put things into perspective, we can say that in TTT/TSSR, writers are also the panel of trained coders of the same experimental study, who have been trained before the experiment by the experimenter and are expected to code their own cognitive activities concurrent to writing. The raw self-reported data that they produce at the end of a writing task is similar to the coded self-reported data produced by the trained coders of TAPs. This aspect of TTT, makes it, in a way, an experimenter-friendly technique because it "avoids the problem of how to segment a verbal protocol after the fact". That is, instead of the experimenter, the writer is recruited for the segmentation task, which "the writer does so through the categorisation process during production itself" (Piolat et al., 2001, p.81).

Several studies tested the validity of data collected via the original TTT technique, that is, directed retrospection with four discrete writing categories. In the first study introducing TTTs, Kellogg (1987) tested the validity of the collected data. Participants (n = 12) wrote under TAP conditions before the directed retrospection training and coded their own TAPs in a delayed retrospection task (15 minutes) through stimulated recall. Substantial agreement (82%) between the categorisations of the participants and an expert judge showed that the measurement was valid (note 6, p. 266). In another study (Levy et al., 1995; as cited in Piolat et al., 2001), the researchers tested whether writers' metacognitive knowledge about the writing processes affects the kind of activities writers report when categorising by coding someone else's TAP and concluded that the collected data is valid. Piolat et al. (1999) investigated the effect of the training to see whether a three-hour group discussion on writing categories to the original 20-minute training would increase the validity of the responses and concluded that the original training is sufficient for participants to correctly categorise their own activities. It was previously shown that directed retrospection can accurately track the time course of 3-5 writing activities (Olive et. al., 2002). Therefore, the agreements reported for the original version with four abstract categories should not be generalised to studies that incorporate a higher number of activity categories, because finer-grained distinctions among processes can confuse the participants and decrease the agreement between their and the experimenter's categorisations (Kellogg, 1987). Although incorporating fewer activity categories

may go at the expense of getting a reliable picture in complex writing tasks, it has been shown to improve the validity of responses (Olive et al., 2002).

Studies looking into the reactivity of TTT ruled out a possible reactivity mostly on the theoretical account of Ericsson and Simon's (1980) protocol analysis. Although Ericsson and Simon (1980) did not make direct claims on the reactivity of the triple task technique in particular, they explained that a secondary verbalization task may not be reactive if the intermediary cognitive process that is activated by the subject to meet the verbalization requirement is in the form of *abstraction*. In subsequent reactivity discussions, it was contended that categorisations are done via abstractions, hence TTT should not be regarded as reactive, at least on theoretical grounds. Researchers also checked the reactivity of TTT via empirical studies. Kellogg (1987) compared the effects of two different dual task configurations, writing with a secondary RT task vs. with a directed retrospection task, as in the global configuration of TSSR, and did not observe any differences in their effects on students' writing quality and fluency (i.e., words per minute). Similarly, Piolat et al. (1996) tested the reactivity of TTT by comparing two conditions, composing in a single vs. triple task setting (study in French, cited in Piolat et al., 2001), and reportedly, there were no differences between the effects of the two configurations on students' fluency and revision activities. In another study, Fidalgo et al. (2014) used the original TTT in a nationwide experiment investigating the reading comprehension via summarising in L1 with 5- to 8-grade Spanish students. The reactivity of summarising (i.e. incorporation of ideas based on levels of contextual importance, irrespective of how they were expressed in students' texts) in single vs. triple-task conditions was tested on the data of a selected sample of students from the main experiment (Fidalgo et al., 2012). The authors found no reactivity effect for reading measured via writing a summary. However, more research is needed to investigate if organising and rewriting these ideas instead of only listing them may pose an additional burden to reactivity.

It is also notable that, in none of the studies that tested a possible reactivity of TTT, process measures were taken into account. However, the training session in TSSR includes some powerful instructional components which are expected to affect writing processes and foster learning. The training procedure for the directed retrospection task in TTT (Kellogg, 1987/1988) and in TSSR (Fidalgo et al., 2008; Torrance et al., 2007), are similar and proceeds as follows: (1) instruction on the activity categories (i.e., definitions, examples, counter examples) (2) practise on the activity categories by coding the TAP of a fictitious character modelling an alternate form of the same type of writing task to be reported, followed by error correction on students' categorisations, and (3) a second individual

coding practice for calculating inter-coder reliability, which is generally measured by a percentage of agreement between the experimenter's and students' categorisations. The difference between the training of TTT and TSSR is that the TAPs are presented in different modalities, that is, in written or audio format in TTT, and via peer videos in TSSR, with the exception of the study by Fidalgo et al. (2012) which is the only study that incorporated a peer video into the training session of TTT. These different modalities have traditionally been used for behavioural modelling in instructional settings for four decades (Davey, 1983), evolving due to advent of technology from single (Bridge & Winogard, 1982), to multiple modalities, video in particular (see Rijlaarsdam et al., 2008). As a matter of fact, modelling in any modality is very effective in changing students' writing process behaviours because observing provides students the opportunity to approach the writing task with both their roles as writers and learners (Rijlaarsdam et al., 2008). The effects of using both verbal and visual modalities on improving learning have also been validated, because information coded via multimodality can be organised by students into a coherent mental representation and integrated with their existing knowledge in the long-term memory (Mayer, 2009). However, empirical evidence to discount the possibility that TSSR-variant of the technique alters students' writing products and processes is missing. In the present study, we hope to contribute to writing process research by filling this gap in methodology.

In this study, we used an introductory level synthesis task (i.e., writing an integrated summary from multiple sources) that is generally used in academic writing courses in L2, to create a task representation of synthesis writing in general. Mateos and Solé (2009) showed that incorrect task representation is a major reason for failing in synthesis writing tasks, and students benefit from pedagogical interventions that use task representation strategies in a way to contribute to one of the three transformational (sub)processes of selecting, organising, and connecting (see Spivey & King, 1989; Van Ockenburg et al., 2019). Previous studies with undergraduate students also showed that when students have the underlying procedural knowledge on how to complete a writing sub-skill (i.e., revision), even a brief instruction can create, or improve a task schema at declarative level (Van Steendam et al., 2014; Wallace & Hayes; 1996).

Therefore, we propose that TSSR may create or alter students' synthesis task schema and provide cues for retrieval of students' own mental processes and task-related strategies during its successive phases of training, practice and the actual implementation. Given that the students in our study already have the underlying procedural knowledge on the several requirements of a synthesis task, a brief and direct (i.e., declarative) instruction may be sufficient to activate

their synthesis task schema, which we checked for in a presentational instruction condition in the present study. However, we also expect a difference in effects on reactivity resulting from the video instruction which features a peer modelling performance, as modelling has proved more effective over presentational instruction mode on students' writing performance in many studies (see Rijlaarsdam et al., 2008).

As a theoretical basis for the reactivity hypothesis in our study, we draw upon the concept of task schema in Hayes' (1996/2012) model of adult writing. According to Hayes (2012), task schemas, with the current plan and the motivational factors at the control level, guide the interaction of the process level (i.e., writing processes, the physical and the social task environment) and the resource level (i.e., long term memory, working memory and reading), throughout the writing task. In Hayes and Olinghouse's adaptation of the model (2015), knowledge of writing strategies is underscored as the central component of writing schemas in text construction, which is cumulatively acquired via instruction throughout schooling; hence, may be absent, incomplete, or include misconceptions at any given time. In the same vein, they can be created, altered, and/or reoriented via instruction, and can be skilfully practised with experience. Also, given the fact that task schemas are highly sensitive to years of schooling and instruction, the model of adult writing can be successfully applied to the undergraduate participant profile in the present study.

1 THE PRESENT STUDY

The main aim of this study was to test the reactivity of the TSSR-technique as an online writing process measure both in its training and data-collection phases.

Before proceeding to the research questions, we would like to clarify some terminology used throughout the study. TSSR stands for the self-reporting method used as a process measure which incorporates the secondary task of directed retrospection into the primary task of writing and its reporting on a writing log. The training (session) for TSSR is aimed at both teaching the directed retrospection categories and the concurrent reporting and writing procedure. Throughout this paper, we refer to the training using two terms interchangeably: TTT/TSSR-training when we want to emphasise the training aimed at both purposes and directed retrospection training when we want to emphasise training that is aimed at identifying activity categories. In practice they are the same and one, that is, the training session.

The research questions of the study are as follows:

RQ1: Does TSSR cause reactivity, that is, affect students' writing performance (i.e., process behaviours, text quality)?

RQ2: Does the instruction mode to prepare participants for TSSR (video vs. presentational) contribute differently to TSSR-reactivity?

RQ3: Does the TSSR-instruction mode affect the process data collected via TSSR?

We will test the following hypotheses:

HYP1: TSSR causes reactivity on process behaviours and text quality.

HYP2: TSSR instruction mode affects reactivity.

HYP3: TSSR instruction mode affects process data collected via TSSR.

2 METHOD

2.1 *Participants*

Participants were 81 1st-year undergraduate students from six intact classes studying Business Engineering and Business Administration at a Belgian university (42.5 % female). Business Engineering and Administration is a 3-year 180 ECTS academic bachelor. Included in the students' 1-year curriculum are courses in Business English and Business French. In these courses students are familiarised with economic and business terminology. Additionally, their grammar of the respective foreign languages is refreshed as they had on average four to five years of foreign language instruction (incl. Skills training) in French and English in high school. Students had an intermediate level of English after 4 years of (roughly 2 hours a week of) English-as-a-foreign language in secondary school.

Procedure and data-collection were in line with the requirements of the ethical board of the university at which the data were collected. That is, students were informed of the goal and procedure of the study prior to the study. Participation in the study was voluntary, yet encouraged one week prior to the experiment by students' respective teachers, and also at the beginning of the experiment by the researchers. All students agreed to take part in the study and signed the informed consent forms. They were also informed of the fact that their data will be removed from the study on their request at any stage of the data-collection phase.

2.2 *Design*

We tested the reactivity of TSSR on students' text quality and writing processes in a pretest – posttest between-subjects 2 by 2 factorial design with the manipulation of two factors, each with two levels: (1) Instruction mode (Video vs. Presentational Instruction), and (2) Reporting mode (TSSR vs. no-TSSR).

Half of the students from each one of the six intact classes were randomly assigned to one of the two instruction conditions (video vs. presentational). After the instruction, half of the students within each instruction condition were randomly assigned to one of the two reporting conditions for the posttest, that is, either with or without TSSR (TSSR vs. no-TSSR). Hence, we tested our hypotheses in four different conditions: Video & TSSR, Video & no-TSSR, Presentational & TSSR, and Presentational & no-TSSR. The distribution of male/female students was not significantly different across conditions. See Table 1 for the distribution of students across conditions.

Table 1
Distribution of Students Across Conditions

Random assignment of students within merged intact classes to conditions	Class A & B Class C & D Class E & F		Class A & B Class C & D Class E & F	
Instruction mode in the training	Video		Presentational	
Random assignment of students within instruction conditions to reporting conditions				
Reporting mode in the posttest: TSSR	yes	no	yes	no
N per condition	15	23	24	19
Percentage female	47%	61%	21%	42%

We ran the experiment on three days within the same week. On each one of these days, two intact classes were present at the institution. In Session 1, a baseline for text quality and writing processes for writing a synthesis text was established via a pretest. In Session 2, an intervention prepared the students for the dual task, that is, the concurrent execution of writing a synthesis task and reporting their synthesis writing activities using TSSR. The training was delivered in two instructional mode conditions: a video demonstration (i.e., the video condition) versus a teacher-led, PowerPoint presentation (i.e., the presentational condition). In Session 3, a posttest was administered under two conditions: writing under single (i.e., no-TSSR) vs. dual task (i.e., TSSR) conditions.

2.3 *Procedure*

Data was collected in a block scheduled lesson of 120 minutes. The introduction to the study, the information about participation in the study, renewing the contract with the research participants, and the instructions for the keylogging tool took 25 minutes; the pretest and posttest took 25 minutes each, and the training session took 30 minutes. We gave a 5-minute break between Session 1 and 2. The remaining 10 minutes were allocated to operational issues such as saving data, changing rooms during assignment to conditions, and possible technical mishaps.

Three research assistants led and supervised the whole procedure (i.e., baseline pretest, training and posttest sessions). They were randomly distributed across conditions, and each of the three research assistants were involved in both conditions.

In Session 1, students were in the same classroom, seated at computers on which keystroke logging software was installed. Students were unaware of the conditions they were assigned to for Sessions 2 and 3. After reading and collecting the signed consent forms, we used a PowerPoint (PPT) presentation to introduce the study and to instruct the students on how to use the keylogging tool while writing with a word processing program on their computers, as they would be using it in both in pre-and posttests. Following this introduction, we guided the students on how to download the soft copies of the pretest synthesis task and the source texts from the cloud folder. The (pretest) task sheet included a short explanation of the type of task, the topic of the task, and a two-sentence introductory prompt which students could use as the opening statement for their texts (a total of 60 words). After establishing students' baseline synthesising skills through the pretest, a random half of the students was allocated to another classroom, to form the two instruction conditions, Video Instruction and Presentational Instruction (see Table 1). We informed the students about the re-allocations of the classrooms immediately before the five-minute break that they took in between the two sessions.

In Session 2, we trained the students on how to use TSSR while completing a synthesis task. Each one of the three research assistants held the training in one of the two conditions, in four consecutive phases, which amounted to 30 minutes. In Phase 1, in both conditions, we elicited and discussed a possible list of writing activities that the students were engaged in while completing the pretest synthesis task executed in Session 1. In Phase 2, we distributed handouts to students in both conditions, which consisted of the name and the brief explanation of each writing activity category, and a visual representing that category, to

minimise the distraction that may be caused by completing the writing log while writing (Fidalgo et al., 2008). We read the contents of the handout to the students. In Phase 3, we held a 10-minute training on directed retrospection categories on the writing logs, which diverged the two conditions in instruction mode, that is, Video Instruction via a peer video and Presentational Instruction via a teacher-led PPT. In Phase 4, students in both conditions took a 10-minute demo-test which entailed writing a synthesis text based on two source texts, with concurrent probing that interrupted the primary task of writing seven times with auditory signals. These auditory signals indicated to students the instance to report their activity by coding them in the categories on the writing logs. See Table 2 for the procedure and the instructions in Session 2, p. 121.

In Session 3, we administered the posttest in two reporting mode conditions, TSSR and no-TSSR, in different computer laboratories reserved for each condition, with each condition invigilated by two of the three research assistants. Students took the posttest on the computer using a word processor program and a keylogging tool, and those in the TSSR-condition received a writing log printed on an A4 paper. Students in all conditions saw the content of the test materials including the sources and the prompt for the first time in the pretest and the posttest. The tests were not compulsory assessment components of the course, nor did they count towards students' General Point Average (GPA).

The two reporting mode conditions in the posttest was formed by dividing each of the two instruction mode conditions into two smaller groups, with random selection of names from the participants list, and cross-merging the four groups from the two instruction mode conditions under two reporting mode conditions (see Table 1).

In the TSSR-condition, students composed their texts and concurrently reported their activities on a writing log in pen and paper conditions when they were probed via audio signals (n = 22) that was played at a variable interval schedule of 60 to 90 seconds (on average every 75 seconds). The sequence of the auditory signals was indicated on the writing logs and also projected onto the board for the students to keep track. Students in the no-TSSR condition composed their texts as a single task, that is, auditory signals were not played and writing logs were not distributed. Writing processes of all students were recorded with a keylogging software as they completed the timed-posttest on a word processing program on their computers.

The type of synthesis task and instructions in the posttest were similar to the pretest. The source extracts, their level of complexity and writing task topics are discussed in the Measures section.

2.4 *Independent Variables*

We checked the reactivity of TSSR by manipulating the instruction mode (video vs. presentational) in the training phase (i.e., Phase 3 of Session 2) and the reporting mode (TSSR vs. no-TSSR) in the posttest phase (i.e., Session 3) of the experiment.

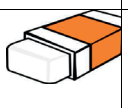
2.4.1 Instruction Mode: Video Instruction vs. Presentational Instruction

The aim of the TSSR-training is most importantly, to teach the directed retrospection categories for improving the validity of TSSR-data, and secondly, to teach the concurrent implementation of TSSR while writing.

The six activity categories in the directed retrospection task were adapted from Fidalgo et al. (2008) and validated in previous studies (Buyuktas Kara et al., 2018, 2024; Chapter 2 and 3) with minimal changes to the original. The categories represented possible writing activities that students may be engaged in when completing a synthesis task, a category 'other', and the category to signal that the process was 'finished'. They were presented in a writing log format, in slightly adapted forms across the different sessions of the intervention, that is, in the training (i.e., Phase 2 & 3, Session 2), and the practice sessions (i.e., Phase 4, Session 2). In the actual implementation at posttest (i.e., Session 3), only the TSSR conditions wrote their texts under dual task conditions, hence used the writing logs for self-reporting. Figure 1 shows the activity categories and the layout of a writing log with a representative sample of coding requirements.

Figure 1

A Writing Log with a Representative Sample of Coding Requirements

	I am trying to understand the sources	I am paraphrasing	I am working on the sources	I am editing	I am writing my summary	I am reading my summary	Other	Finished
								
1								
2								
...								
44								

Presentational Instruction. Students in the presentational instruction condition were presented with a PowerPoint show informing them about the activity categories in the directed retrospection task of TSSR. The different activity categories (one per slide) were introduced with a brief explanation and a coloured graphical representation (See Figure 1). To familiarise students with the categories, the research assistants went through each slide by referring to the students' experience with the pretest synthesis task as a basis for discussion by asking guided questions and eliciting answers from the students. It was a teacher-led presentation and whole-class discussion based on each one of the seven writing activity categories.

Video Instruction. Students in the video instruction condition watched a training video that we developed for a previous experiment (Buyuktas Kara et al., 2018, Chapter 2). The model is a female freshman at a private university in Istanbul, who studied the English preparatory program in the previous year. The video is a compilation of nine separate footages (5 mins. 49 sec.) showing the model acting out a TAP scripted by the experimenter with fragments from a synthesis task that correspond to the directed retrospection categories. Each footage is around 15 to 90 seconds, punctuated with an auditory signal probing the students to the directed retrospection task, that is, ticking on the provided writing logs the activity category that the model was engaged in when interrupted with each auditory signal (n = 9). Following the video, the research assistants checked the answers and corrected any errors. See Table 2 for the training procedure followed in Session 2 in the two instruction mode conditions.

2.4.2 Reporting Mode: TSSR vs. no-TSSR

In Session 3, we randomised the students within each instruction condition into two reporting mode conditions (TSSR vs. no-TSSR). Students in TSSR-conditions composed their texts with concurrent probing for reporting their activities on a writing log. Students in no-TSSR-conditions were not probed with auditory signals nor given a writing log. They wrote under single task conditions (i.e., writing only).

Table 2

Training Procedure in Session 2 for the Video and Presentational Instruction Mode Conditions

Phase/Conditions	Aim	Contents	Activities and Interaction Type	Teaching Method	Materials
Phase 1 Both Conditions	Developing background knowledge on the synthesis writing activity categories	Ss reflect on the preceding synthesis task and list activities to complete the task. T corrects mistakes.	T asks Ss to share the writing activities they were engaged in. (IT: S-S). T asks: *"You just completed a synthesis task." "What are some of the writing activities that you did? "Did you include your own ideas in the text? S: Ss' answers* T writes each activity on the board, corrects mistakes. (IT: T-S-T) (7')	Guided questions/ Brainstorming/ Eliciting Joint reflection Error-correction	White board
Phase 2 Both Conditions	Familiarising the students with writing log categories	T reads the names and explanations of the activity categories	T gives a handout to the students with the name, graphic representations and explanations of the activity categories. T reads the contents of the handout out loud. (IT: T) (3')	Teacher-led self-study	Handout with the activity categories
Phase 3 Video Instruction	Training: "How to Self-Report Writing Processes" Teaching the activity categories in the directed retrospection task Training the students for writing under TSSR-conditions	Ss watch the training video to categorise the model's writing activities by ticking on the writing logs	T says: *"Now we will watch a video of a student completing a synthesis task. You will hear 9 bleep tones throughout the video. When you hear the tone, tick the student's activity on this writing log."* (IT: T-WC) Ss watch the video and complete the writing logs. T asks Ss to compare answers with their partners (IT: S-S, then share them with the class. (IT: S-WC) T corrects mistakes. (IT: T-WC) (10')	Peer-modelling Task-based evaluation Joint reflection Error-correction	Training video Writing logs

Presentational Instruction		T presents the training PPT with each activity category on a slide, with a graphic representation and its explanation.	T shows the PPT and explains each category. T: *"I am working on the sources". "Do you use all the ideas you read in the sources in your summary?" S: "No." T: "This is true. You want to find what the text is mainly about. What is it called?" S: "Main idea" T: "Yes, the main main idea of the text.* T proceeds to the next activity category. T: *"I am editing" T: "In every writing task, it is important to correct any mistakes. What kind of mistakes do we usually make?" S: "Grammar." T: "Yes, grammar, maybe spelling?* T elaborates each category. (IT: T-WC) (10')	Teacher-led presentation Guided questions/ Brainstorming/ Eliciting Joint reflection Error-correction	PowerPoint Presentation
Phase 4 Both Conditions	Simulating test conditions	Ss complete a synthesis task under TSSR-conditions	T explains how to access the task materials stored in the cloud. T gives each S a writing log, projects the video with the auditory signals and the corresponding numbers on the board. (IT: S) (10')	Individual practice	Computers; Demo-test; (soft copy) Writing logs Video

Note: IT: Interaction Type; T: Teacher; S: Student; WC: Whole Class

2.5 *Training Delivery and Intervention Fidelity*

To ensure fidelity of implementation, the three research assistants were present in their respective conditions, leading and supervising the whole procedure (i.e., baseline pretest, training and posttest sessions). To ensure comparability across conditions, they strictly adhered to the same protocol listing the different phases of the data-collection process each with an indication of timing and with verbatim instructions, as well as following a PPT projecting the verbatim instructions to the participants. After each phase one of the research assistants acting as a supervisor, time-keeper and facilitator had to tick the log, as well as log the time spent on each phase (and/or add any comments). As a second measure of fidelity, we collected the practice sheets which students used during a demo-test

simulating test conditions, which confirmed that all students actively participated in the practice session.

2.6 *Measures*

We rated text quality of pretest and posttest synthesis texts. Process behaviours were logged in the pretest and posttest, using a keylogging tool (Session 1 and 3). Time-Sampled Self-Reports (i.e., TSSR-data) were collected for the TSSR-condition only (Session 3).

2.6.1 Product Measure: Text Quality

Tasks. In the pretest and posttest, students were asked to write a synthesis text of 100 - 150 words in response to a writing prompt which was around 60 words including the instruction. The task required students to select content from three source extracts (one paragraph each), presenting complementary information, and synthesise them in a linear organisational format without incorporating their own ideas (i.e., an integrated summary from multiple sources).

Table 3
Pre-and Posttests of the Study

Synthesis	Topic	Ex-tracts	Words in Extracts	Flesch Kincaid Ease Score[1]	Words Required	Measures[2]
Pretest	"Drugging Kids on Long Flights"	3	386	54.4	100-150	TQ; KSL
Demo-test	"Academic Success"	2	218	54.4	80	None
Posttest	"Economics of Obesity"	3	382	41.5	100-150	TQ; KSL; TSSR[3]

Note. [1] Flesch Kincaid Ease Score 100 = very easy to read / 0 = very difficult to read; 54:4: Fairly difficult to read; 41,5: Difficult to read. [2] TQ: Text Quality; KSL: Keystroke logging; TSSR: Time-Sampled Self-Reporting
[3] In TSSR-condition only.

To provide a means of comparison for the difficulty level of each set of source texts that was used at each measurement occasion, we checked their Flesch Kincaid ease score. It is an automated software that calculates readability indices,

based on word and sentence length, and produces a score ranging from 0 (i.e., very difficult to read) to 100 (i.e., very easy to read) (see Flesch Kincaid Calculator)[1]. See Table 3 for the specifications of the synthesis tasks used in the present study.

Writing Performance. We defined five variables for text quality, that is, the correct selection of (1) main ideas and (2) supporting ideas, (3) text comprehensibility, (4) source comprehension, and (5) source text language use, The four subscores, except for source text language use, were merged under the composite score, content (see page 127). Hence, in the final analysis, we had two aspects for the quality of the posttest synthesis task, that is, content, and source text language use, each with an equal weighting of 50% in the final score.

Rating Procedure. An outside panel of three paid raters rated the synthesis texts. They worked independently and were blind to the conditions. All raters were fourth-year bachelor students from the English Language Teaching Department at a state university in Turkey; 22 years of age, two females and one male, with Turkish as their mother tongue. A research assistant led and supervised the rater training and the rating process. To establish an inter-rater agreement, the raters received and studied the list of idea units in the sources, and the holistic benchmarking scales and the benchmark texts, which they used to rate a sample set of six synthesis texts for each one of the four text quality aspects. They discussed the reasons for any disagreement and agreed on possible solutions.

The texts were randomly distributed among three raters. Each text was rated for (1) text comprehensibility (2) source comprehension, (3) main ideas, (4) supporting ideas and (5) source text language use. The raters examined students' texts for one aspect of text quality at a time.

For rating the main ideas and the supporting ideas, the raters counted the number of idea units that students incorporated from the source texts based on the lists that we provided to the raters and calculated them in percentage, which was used as the student's final score out of 100, for each one of the two aspects of text quality, separately.

For rating text comprehensibility and source comprehension, we used a holistic scoring procedure by employing benchmark essays, as previous research has shown that this rating technique improves rater reliability (Blok, 1986; Raedts et al., 2017; Rijlaarsdam et al., 2012; Vandermeulen et al., 2020; Van Steendam &

[1] *https://goodcalculators.com/flesch-kincaid-calculator*

Tillema, 2014). Among 81 student texts, we selected a benchmark for each of the two quality aspects that we defined as average quality. The benchmark texts were arbitrarily given a score of 100. Three raters evaluated the other synthesis texts for each quality aspect relative to the benchmark text. A text received a score of 200 if it was judged to be twice as good as the benchmark text, 50 if it was judged to be half as good, and so on. Each benchmark text was supplemented with a list of the text's weak and strong elements in relation to either text comprehensibility or source comprehension that was to be evaluated with it, so that raters could focus on the appropriate characteristics while evaluating that particular trait.

Raters assessed text comprehensibility before all aspects, without first being briefed about the task type or having read the sources, to ensure that texts were comprehensible to readers unfamiliar with the sources. For source comprehension, the raters scored the student texts considering the extent to which a reader is correctly informed about the important content in the provided source extracts, taking the sources as point of reference for comprehension. We asked them explicitly not to rate L2 linguistic quality in text comprehensibility or source comprehension, as students may have relied on source text language when incorporating source content into their texts.

Source text language use indicates the proportion of matching text between the student texts and the provided sources. Raters used WCopyFind, a free text-matching software that runs a similarity analysis and detects the matching text by comparing two documents uploaded to the system by the user.[1] The program highlights the matching text according to users' adjusted comparison parameters and reports, for each text, the calculated similarity in percentage, which we used as the source text language use score for the uploaded text. Hence, a higher score in source text language signifies fewer attempts at paraphrasing the content drawn from the sources. We opted for a chunk of four words as the minimum number of words as a parameter for matching text, in line with earlier studies (Shi, 2004; Weigle & Parker, 2012; Van Weijen et al., 2019), and excluded the references and the titles from the analysis, since activities related to source use were not included in the directed retrospection categories, or in the synthesis text quality aspects of the present study.

[1] *https://plagiarism.bloomfieldmedia.com/software/wcopyfind/*

2.6.2 Process Measures: Keylogging and Time-Sampled Self-Reporting

Keylogging. In both the pretest and the posttest of this study, we registered students' online text production processes by using Inputlog[1], a keystroke logging software specifically developed for the purposes of writing research which records the keystrokes and mouse movements of writers unobtrusively, when they are typing on their computers in real-time (Leijten & Van Waes, 2013). It provides researchers with detailed online writing process descriptors by logging and time-stamping their keystroke activities and mouse movements.

Five variables. To report students' process behaviours, we selected five process variables, based on their sensitivity for competence growth. Vandermeulen et al. (2020) set up a large-scale national assessment study on L1 synthesis writing in grades 10-12 (16-18 years old) of secondary education, and used Inputlog as the process data-collection tool. From the set of process measures that were sensitive to grade effects in their study, we selected five variables which were shown to have a statistically significant relation with synthesis writing quality: three time measures and two fluency indicators.

The three time-related measures indicate the amount or proportion of time allocated to the task (i.e., *total process time)* and how it is distributed between the two writing activities (i.e., *proportion of pause time*, *proportion of active writing time)*. Total process time is the total time spent writing. Pauses during production are periods of at least 2000 ms. in which no activity is recorded in the word document. The proportion of active writing time reflects the amount of time the author actively spent producing the text.

The two fluency process measures were the *speed of the process* (in terms of the number of strokes made in a certain episode of the writing process), and the *number of pauses.* Both variables were measured at different intervals during the writing process: for speed of the process, we divided the process into 5 equal intervals and for the number of pauses we divided the process into 6 intervals of equal length[2]. The interval-based variables allowed us to study the effect of the two experimental modes on the distribution of activities during the process, which is important because the effect of an activity on text quality depends on

---

[1] *www.inputlog.net*

[2] *The choice for 5 or 6 intervals was accidental, and otherwise has no significance or influence on the results, because the variables are not compared to each other.*

the moment of its execution in the process (see Breetvelt et al., 1994; Van den Bergh et al., 2016).

Time-Sampled Self-Reports (TSSR). We collected TSSR-data, at one measurement occasion (i.e., posttest; Session 3), in both instruction conditions. The seven activity categories in the writing logs were taken from the original (Fidalgo et al., 2008) with minor adaptations, taking into consideration the possible activities that students would be engaged in when completing a synthesis task. We also included, as in the original, a coloured graphic representation next to each activity category. We also added an activity category *finished* to identify a writing process that is cut short because of time constraints.

A randomly selected half of the student sample was assigned to the TSSR-condition, which required that they respond to auditory probes (n = 22) that were played at a variable interval schedule (mean: 75 s.) by ticking on the writing logs the writing activity they were engaged in at that moment. We projected on the board the number of the probes as they occurred, so that the students do not lose track of the sequence of activities to be reported.

2.7 *Analyses*

2.7.1 Preliminary analyses

Outliers were inspected with the SPSS Explore function. Two cases were out of the 95% range for source text language use aspect of text quality in Session 1.
To check for the relative independence of two text quality aspects, content vs. source text language use, we ran factor analyses for the scores of the five aspects of text quality, for Session 1 and 2 (i.e., pretest and posttest), separately, with maximum Likelihood Extraction and direct oblimin rotations and compared the model fit for 1 and 2 factor solutions. For Session 1, no statistical difference was observed between a one- or two factor solution, but for Session 2, a two-factor solution fitted the data best (X^2 (4) = 9.641, p = .047). While Session 2 is the session we are most interested in, given that we want to study the effect of different instruction and reporting modes on the aspects of text quality, which was measured in Session 2, we decided to divide the text quality scores in content scores, indicated by text comprehensibility, source comprehension, main ideas and supporting ideas; and in source text language use, indicated by proportion of matching text between the sources and the students' synthesis texts.

For writing process behaviour, we found max 4 cases, depending on the variable. For these cases, scores were winsored.

To check the effect of randomization, we tested the differences in baseline across conditions that were established in Session 2, with multivariate analyses. No such differences were observed for text quality (Pillai's Trace .008, $F(2,74) = 0,280$, $p = .756$). For process measures, no differences across conditions were observed in Session 1. No statistically significant differences as a result of the two factors (i.e., reporting mode and instruction mode) were found for the three *Time Allocation variables*: Total Process Time, Proportion of Pause Time and Proportion of Active Writing Time (For reporting mode, instruction mode and the interaction between the factors respectively V[1] = .025, F (3,72) = .62, p = .60; V = .013, $F(3,72) = .32$, p = .81, V = .43, F(3,72) = 1.07, p = .37). For *Fluency Indicators,* we found no statistically significant initial differences related to the two factors, neither for Strokes per Minute ($F_{\text{Reporting mode}}$ (1,78) = 2.505, p = .118, $F_{\text{Instruction mode}}$ = 2.484, p = 119, $F_{\text{Interaction}}$ (1,78) = .165, p = 658) nor for Mean Number of Pauses ($F_{\text{Reporting mode}}$ (1,78) = .273, p = .603, $F_{\text{Instruction mode}}$ = .432, p = 563, $F_{\text{Interaction}}$ (1,78) = .019, p = .891).

2.7.2 Analysis Strategy

Effects on Text Quality. We compared five nested models for the two text quality variables separately: content and source text language use. We started with an empty model (Model 0), with no intercept estimated, and session (unstructured) as a repeated variable with participants' scores nested as a random component. Subsequently, we added factors and interactions: Session (model 1 with Session for baseline and posttest), reporting mode (model 2), instruction mode (model 3), and the interaction between these factors. From the best fitting model, we report the effects on the text quality scores (B-weights) and the correlations of scores between sessions.

Effects on the Writing Process: Keylogging.
Time allocation. For the three measures indicating time allocation we compared six nested models. We started with a null model, with session (two levels) as repeated factor (unstructured) in the random component and then added effect of session (Model 1), and the effect of the two factors, each separately (reporting mode (Model 2A) and instruction mode (Model 2B); then both factors were included (Model 3), and in a final model we added their interaction.

---

[1] *V refers to Pillai's Trace index for multivariate effects in multivariate analyses.*

Interval data. For the two process behaviours with scores for different episodes or intervals during the process, we first built models to check whether the session affected the distribution of the activities across intervals. After a null model, with Session and Interval as repeated factors in the random component, we expanded the model with the effect of Session (Model 1), of Interval (Model 2A) and a model with the interaction between Session and Interval (Model 2B). That meant that we also created a variable for Interval ranging from 1 to 5 or 6 for the two interval-based Inputlog measures respectively. If Model 2B would fit best, the distribution of behaviours would (prove to) be different across sessions; if not, there would be no reason to include the interaction in subsequent models, so that we could focus on the effect of the two independent variables on the process behaviours in Session 2. In the next series of models, we added factors to the best model (2A or 2B), the effects of the two factors and their interaction, successively. Reporting mode (3A), instruction mode (3B); both factors, that is, reporting mode and instruction mode, (3C) and their interaction (3D). These models were expanded with a fourth model (4) in which we allowed the effect of the factors in 3A-3D to interact with the variable Interval.

Effects on the Writing Process: Time-Sampled Self-Reports (TSSR). For the TSSR-scores in the TSSR-condition of the reporting mode, we first built models to check whether a linear (Model 1) and a curvilinear (Model 2) component was needed to accurately model the distribution of the activities throughout the process. This is necessary to estimate if effects are linear or curvilinear, which relates to the effect of episodes (Van den Bergh & Rijlaarsdam, 1996). Then we added instruction mode as factor (Model 3), and subsequently the interaction between instruction and the linear component (Model 4), and the curvilinear component (Model 5).

3 RESULTS

We tested three hypotheses: (1) TSSR induces reactivity on students' text quality and writing processes, (2) instruction mode affects writing performance in TSSR-conditions, and (3) instruction mode affects TSSR-data.

3.1 *Effects of the Reporting Mode (i.e., TSSR vs no-TSSR) and the Instruction Mode (i.e., Video vs. Presentational) on Text Quality*

Table 4 shows the means (and standard errors) for content and source text language use aspects of text quality for four conditions.

Table 4
Text Quality Results (Content, Source Text Language Use) on Two Measurement Occasions in Four Conditions (Means, SE)

			Measurement Occasion			
			1		2	
	Reporting Mode	Instruction Mode	*M*	*SE*	*M*	*SE*
Content	no-TSSR	Presentational	76,09	6,45	70,99	6,05
		Video	74,41	5,87	75,23	5,50
	TSSR	Presentational	76,53	5,67	84,08	5,38
		Video	84,41	7,52	89,16	6,81
Source Text Language Use	no-TSSR	Presentational	22,47	3,56	28,158	5,73
		Video	21,09	3,23	33,78	5,20
	TSSR	Presentational	19,13	3,23	28,25	5,10
		Video	20,86	4,14	40,73	6,44

We tested the effect of the two independent variables, instruction mode, and reporting mode on content and source text language, by comparing the fit of nested models (see Appendix A for model comparisons). For source text language, Model 3 fitted best, indicating an effect of Instruction Mode (B = -8,56, $t(74)$ = -2,065, p = .042), indicating a higher proportion of source text language in Video Instruction than Presentational Instruction. For content, Model 2 fitted best, indicating an effect of reporting mode (B = 11,23, $t(81)$ = -2,208, p = .046), with higher content scores in TSSR-conditions.

To summarise, texts scored higher on content when written under TSSR-conditions, irrespective of the instruction mode used in the TSSR-training; and contained more source text language in the video condition irrespective of the reporting mode, that is, with or without TSSR.

3.2 *Effects of the Reporting Mode and the Instruction Mode on the Writing Process: Keylogging*

3.2.1 Time Variables: Total Process Time – Proportion of Pause Time – Proportion of Active Writing Time

We tested the effect of the two independent variables, reporting mode and instruction mode, on the process time variables, Total Process Time, Proportion of Pause Time and Proportion of Active Writing Time, by comparing the fit of nested models (see Appendix B1 for model comparisons).

For Total Process Time, a positive significant effect of the TSSR-condition was observed (F (1,76,619) = 16,037, p < .001). For Proportion of Pause Time and for Proportion of Active Writing Time, the best fitting model included the interaction between both factors: $F_{PPT\ (1,77)}$ = 9.166, p = .003; $F_{PAWT\ (1,76,934)}$ = 5,553, p = .021, with no significant main effects. Figure 1 (A-C) presents the effects of both conditions on the three variables. Figure 1A shows the effect of the TSSR-condition on Total Process Time, irrespective of the instruction mode (ES = .77): Participants spent more time on the process as a whole, due to the secondary task of directed retrospection that they had to respond to.

Figure 1B shows the interaction between the two independent factors: when in the no-TSSR-condition Video Instruction led to a larger Proportion of Pause Time than Presentational Instruction (ES = .43). In the TSSR-condition it is the other way around (ES = .75): Presentational Instruction seemed to be more sensitive to the TSSR-condition (ES = .86) regarding the Proportion of Pause Time than the Video Instruction (ES = .27).

Figure 1C presents the interaction between the two factors for Active Writing Time. Within the TSSR-condition, an effect of instruction mode in favour of Video Instruction (ES = .33) was observed for Active Writing time, whereas the effect of instruction was -38 in the no-TSSR. Students in the TSSR-condition who had been trained via Video Instruction spent more time actively writing than students who had been instructed via Presentational Instruction, in the no-TSSR-condition the effects are reversed.

Figure 1 (A-C)
Estimated Mean Scores for Instruction and Reporting Mode. Based on Best Fitting Model: (See Appendix B1) (A) Total Process Time, (B) Pause Time, and (C) Active Writing Time

Figure 1A

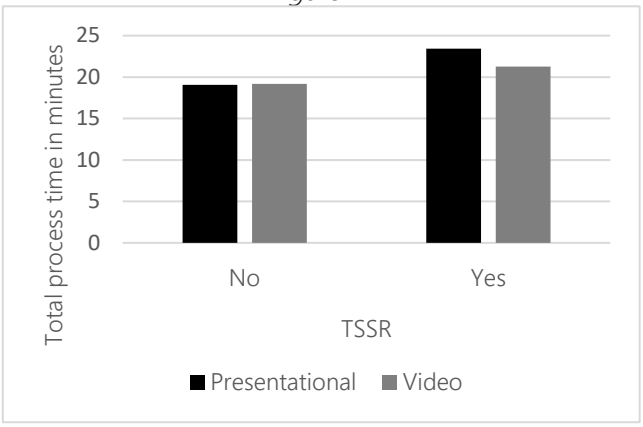


Figure 1B

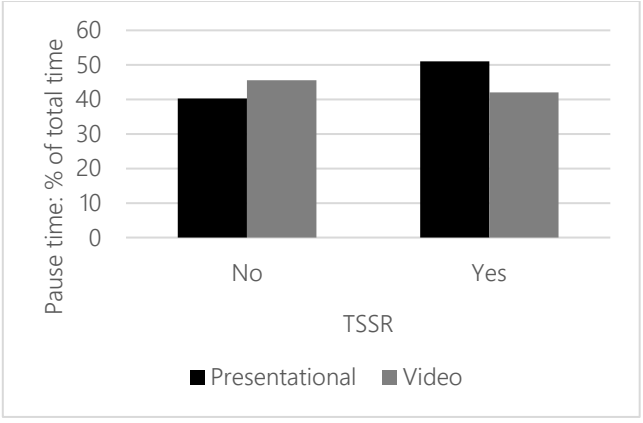
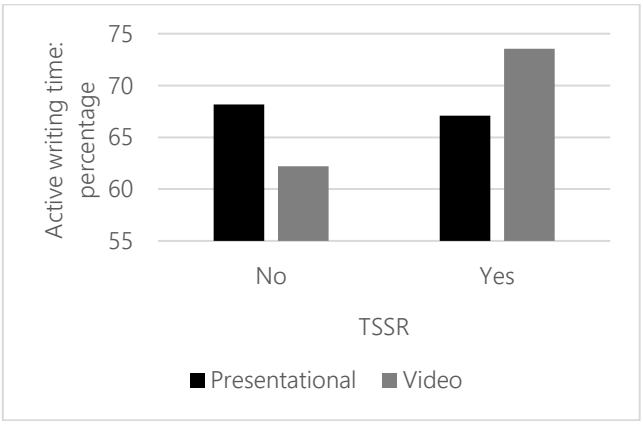


Figure 1C



3.2.2 Distribution of Writing Behaviours: Writing Speed (Number of Strokes per minute) and Number of Pauses

We tested the effect of reporting mode and instruction mode on the distribution of writing behaviours through Interval-based Inputlog measures. For both variables, main effects were observed for session and interval but no interaction between session and interval, which indicates that both effects were independent (Appendix B2).

For Speed, the model that fitted data best consisted of the two manipulated factors and their interaction, and the interactions of these three components with interval, indicating that factors affected the distribution of the speed as measured by strokes per minute (model 4D, Appendix B2). Two components explained statistically significant variation in speed: Interval (F (4;76,662) = 105,36, p < .001) and the interaction between reporting mode, instruction mode and interval (F(4;78,503) = 3,702), p = .008). Figure 2A shows the pattern. Students increasingly write faster until they reach a peak around Interval 3 to 4, roughly in the middle of the process, after which they type increasingly fewer strokes in the remaining intervals in the final stages of the writing process. In Interval 3, the effect of instruction mode is significant, with higher speed in Video Instruction than in Presentational Instruction (B = 23,349, t(69,616) = 2,354, p = .021). Figure 2A shows the pattern for the four conditions.

For the mean number of pauses, another pattern emerged. From the first set of analyses, effects of Session and Interval were observed, but no interaction. Adding the two independent variables, reporting mode and instruction mode, to the models (3A-D), showed that reporting mode explained variance (Model

3A). However, in a third round, we allowed Interval to interact with the independent variables; these more complex models proved to be a better fit, except for model 4A compared with 3A. From the four models left, Model 4B resulted in the best fit: the interaction between instruction mode and interval. In this model, the contributing elements are interval (F(5; 85,9778) = 15,343, p < .001), and the interaction between interval and instruction mode (F(5; 98,382) = 4,198, p = .002). Figure 2B shows the pattern. Interval 1 shows significantly more pauses than the other intervals (t(79,891) = 5,09, p = < .001). Next to a main effect of instruction mode in favour of Presentational Instruction (t(79,558) = 2,204, p = .03), an interaction between instruction mode and Interval 3 was observed (t(97,914) = -2,795, p = .006) in favour of Video Instruction.

Figure 2 (A-B)

Estimated Mean Scores for Instruction Mode and Reporting Mode, Based on Best Fitting Model: (2A) No. of Strokes per Minute and (2B) No. of Pauses

Figure 2A
Speed (strokes per minute)

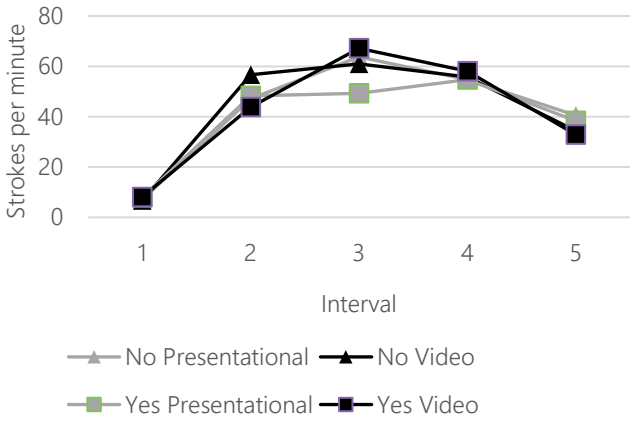
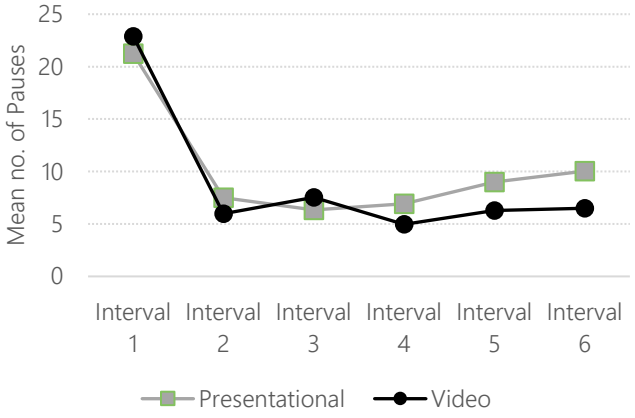


Figure 2B
Number of Pauses



3.2.3 Effects of the Reporting Mode and the Instruction Mode on the Writing Process: Time-Sampled Self-Reports

In the posttest, both TSSR-conditions generated self-reported process data. Our research question was whether the difference in training mode would affect the self-reports.

Figure 3A
Distribution of Four Writing Activities Across the Writing Process Intervals

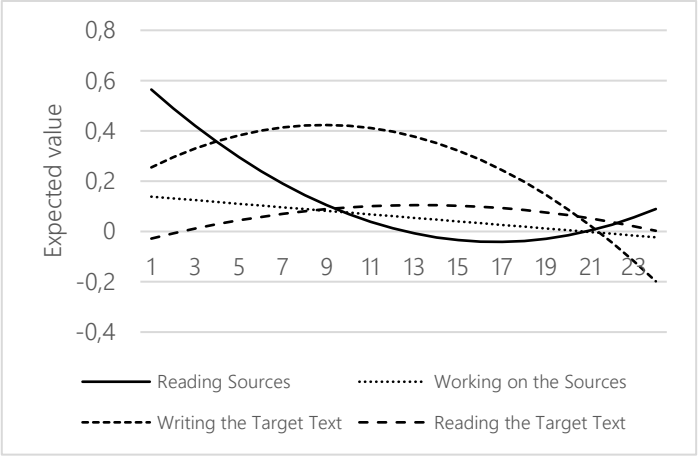
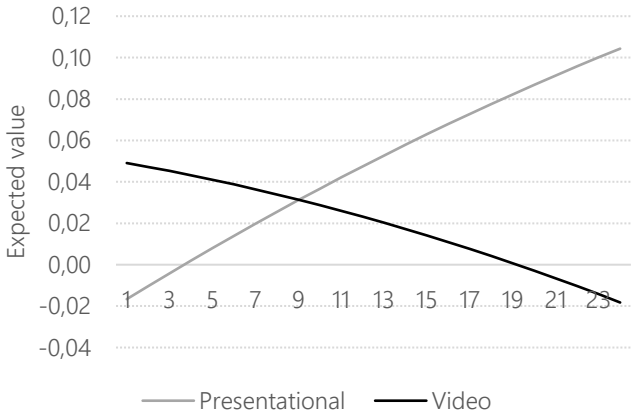


Figure 3B
The Effect of Instruction Mode on the Distribution of Paraphrasing Activity



Model comparisons (Appendix B2) showed that the curvilinear model (Model 2) was the best fit for reading the sources, editing the target text, writing the target text and reading the target text. For working on the sources, the linear model fitted best (Model 1). Addition of instruction mode did not improve these models, but it did for paraphrasing. Figure 3A presents the distribution across the writing process of reading the sources, reading the target text, writing the target text, and working on the sources.

Figure 3B shows the effect of instruction mode on the distribution of paraphrasing. In Presentational Instruction, students reported spending increasingly more time paraphrasing throughout the writing process, while in Video Instruction, students reported spending increasingly less time on paraphrasing from the beginning to the end of the process.

4 DISCUSSION

In this study, we tested the reactivity of Time-Sampled Self-Reports (TSSR), an online writing process measure. The task subject to TSSR was a basic type of synthesis that required students to include content from three source texts and write a summary without incorporating their own ideas. Data were students' synthesis texts rated on two quality aspects, content and source text language use, and process behaviours collected with keylogging and TSSR. We hypothesised that (1) TSSR causes reactivity on process behaviours and text quality, and that TSSR-instruction mode (2) affects reactivity on processes and text quality and (3)

affects process data collected via TSSR. In this section, we will explore how the results of the study align with these hypotheses, providing insights into our research outcomes. Table 5 provides an overview of the results.

4.1 *Effects of Reporting Mode on Writing Performance: TSSR vs. no-TSSR (HYP1)*

For text quality, Hypothesis 1 was partly confirmed: TSSR influenced text quality, but not on source text language use. For the writing process, students in the TSSR-condition spent more time on task compared to students in no-TSSR-condition, irrespective of the instruction mode used in the training phase. This in fact is an experimental control, given that the writing process was interrupted 22 times with auditory signals as part of the dual task, which taxed the time allocated to the writing task. However, more time on task did not imply more time in active writing.

Table 5
Overview of Statistically Significant Results
for Text Quality and Writing Behaviour Variables Across Conditions

	TSSR		no-TSSR	
	Video	Presentation	Video	Presentation
Text Quality				
Content	x	x		
Source Text Language Use	x		x	
Process - Keylogging				
Time on Task	x	x		
Active Writing Time	x			
Pause Time		x	x	
Process activities distribution				
No. of Pauses	x - Interval 3	x	x - Interval 3	x
Speed (no. of strokes)	x - Interval 3		x - Interval 3	

Note. The symbol x indicates a statistically significant result for that dimension.

The effect of TSSR on the content aspect of text quality might hence be due to the mere fact that participants had to self-report their writing activities concurrent to writing. Concurrent verbal reports in general, have been shown to slow down the writing process, due to the effort of having to concurrently carry out two tasks, both physically (i.e., response latency, see Ericsson & Simon, 1993),

and cognitively, (Levy et al., 1996), without decrements in text quality (Janssen et al., 1996; Levy et al., 1996, Ransdell, 1995). On another note, the operationalisation principle of cognitive effort via the dual/triple task paradigm is based on the latency effect, that is, delays in response times to a secondary task over a baseline (Kahneman, 1973). In this line of view, the effect on the content score of TSSR-conditions could be tracked in students' writing processes as longer time spent on the writing task. Therefore, Hypothesis 1 was partly confirmed: TSSR is reactive on students' content scores and on time completing the writing task.

We explain the reactivity of TSSR on the content scores following Ericsson and Simon's view (1984/1993) that the reactivity of a secondary verbalization task is relevant to the kind of cognitive activity that is employed for producing it. In fact, this proposition has mostly been used in defence of the non-reactivity of TTT, based on the premise that abstraction is a non-reactive intermediary cognitive operation when it is employed for a secondary verbalization requirement. However, we contend that the coding requirement in TTT/TSSR cannot be met by a low level of abstraction, but rather, it requires a deep level of information processing. In the analysis of verbal protocol data, abstraction is used for creating categories (Van Someren et al., 1994), that is, by extracting similarities from particular instances to define broader concepts. In TTT/TSSR, this is done by the experimenter based on existing theoretical or empirical data (Levy & Ransdell, 1994) prior to the data-collection phase. These categories are then used, following a top-down approach, to inform the subsequent categorical membership decisions of the writer-coders (Van Someren et al., 1994), (i.e., rule-based categorisations). This is a higher-order cognitive operation that writer-coders execute via scanning and filtering of their own momentary (i.e., at the time of a probe) cognitive activities, pattern-allocation to the defined categories, and categorisation via abstraction in the case that a choice needs to be made between two closely-knit categories, which is a common occurrence with directed retrospection tasks that include a higher number of activity categories (Levy et al., 1996). Therefore, writing under TSSR-conditions is a concurrent operation of the writer-coder, but the several subprocesses of writing and coding cannot be employed simultaneously for the dual task. In fact, writing under TSSR-conditions conveniently follows a task-switching paradigm, via probed interruptions, which have been discussed to temporarily shift the control to a conscious level (Bereiter & Bird, 1985), which may promote metacognition and learning. Therefore, we believe that information processed at this level due to the requirements of the secondary verbalization task, must be reactive on writing, and the reactivity induced is likely to be a positive one (Ericsson & Simon, 1984/1993).

Abstraction is also the main cognitive operation in creating and broadening task schemas in cognition and learning (Anderson et al., 1976; Elio & Anderson, 1984). Conscious schema abstractions are more powerful because they entail making solution-oriented inferences about a problem, which subsumes related cases (Fuchs & Fuchs, 2003). In TTT/TSSR, the writer-coders are trained in using categorisation rules in the preceding training and practise sessions. Such multiple exposures to problem-solution tasks enriches schema abstractions by associating them to a wider knowledge base (Craik & Lockhart, 1972), and the level of analogy between these exposures is decisive for the success of the current solution (Fuchs & Fuchs, 2003). Of course, this is an ongoing learning process, but when it occurs at various stages of an experiment, a learning effect is highly probable. TTT/TSSR-experiments are designed in a training, practice and implementation sequence with analogical problem-solution tasks. The successive training and practise phases that immediately precede the actual implementation of TSSR can improve students' writing performance in that task, as they create opportunities for reflection and making inference-based decisions. We argue that this has augmented the positive reactivity of the directed retrospection task in the present study, which became measurable, in a robust manner, by the end of the intervention.

We also argue that the use of printed writing logs during the subsequent writing task might have contributed to the positive reactivity of TSSR. Task-related materials are part of the writer's environment, and they interact with the processes that writers employ when composing their texts (Hayes, 2012; Leijten et al., 2014). Propositional, visual and spatial input, when used as cues for encoded information can help the retrieval of knowledge from the declarative long-term memory (Olive & Passerault, 2012), and each can have different effects on the writing task, causing changes in the current plan, and the text written so far, as a result of inferencing and problem-solving (Hayes, 2012). Writing log sheets pack a set of propositional (i.e., activity categories and their definitions), visual (i.e., graphic representations) and spatial cues (i.e., the log chart). Hence, they can help to (1) alleviate the cognitive load of the simultaneous operations of working memory such as the retrieval of task schema; and to (2) partition and sequence the writing process via the log chart that divides the writing process by the expected number of probes. These cues are highly effective for learning and instruction, especially in synthesis writing instruction, because the pedagogical necessity to phase the multiple task requirements of synthesis writing has been well-documented by researchers in the field, both for effectiveness of instructional practice (Lee et al., 2018; Leijten et al., 2019; Van Ockenburg et al., 2019), and for ensuring the validity and reliability of their assessment (Bennett et

al., 2016). In fact, partitioning and sequencing the writing process of a particular writer is defined in the literature as that writer's strategic approach to writing tasks (Torrance et al. 1994), and charts and diagrams are common memory aids that have been used for strategy activation in writing (Harris & Graham, 1996; Olive & Passerault, 2012).

It should be noted here that the reporting mode in TTT is via the colour-coded and/or letter-labelled keys that are allocated to each activity on a PC keyboard. Hence, the effect of using printed writing logs as a reporting tool on the reactivity of TSSR that we discussed here is not generalisable to TTT. Moreover, in the present study the writing logs were used in both TSSR and no-TSSR-conditions in the training and practice sessions, therefore, their contribution to the positive reactivity on content quality is isolated to their use within the actual implementation phase (i.e., posttest), in which only TSSR-conditions used them. However, this does not rule out a cumulative effect resulting from its use in the training and practise sessions preceding the actual implementation of TSSR for the TSSR-conditions, and the inference-based decisions, during the course of the writing task, which can lead to the reactivity of the technique.

4.2 *Effects of Instruction Mode on Writing Performance: Video vs. Presentational (HYP2)*

Our second hypothesis that TSSR instruction mode (i.e., video vs. presentational) affects reactivity on writing products and processes was confirmed for the source text language use aspect of text quality and in terms of differences in time allocations and writing behaviours during the writing process. Results showed a main effect of Video Instruction on source text language use, indicating that students in the condition incorporated more source text language in their texts, irrespective of whether they wrote their texts in TSSR or in no-TSSR conditions. The writing processes of the Video and Presentational Instruction conditions also showed significant differences which we will discuss in association with the corresponding text quality results of the two conditions. The distribution of writing behaviours of the two instruction mode conditions was highly consistent throughout the writing task for the separate conditions, exhibiting two main instruction effects. Presentational Instruction paused significantly more often compared to Video Instruction throughout the writing process, with the exception of Interval 3, during which Video Instruction paused longer, and wrote speedier compared to Presentational Instruction. Students' process behaviours with regard to time allocation to writing processes showed that Presentational Instruction allocated more time to pausing irrespective of the reporting mode used in

the posttest (i.e., TSSR vs. no-TSSR). For the Video Instruction, time allocation to writing processes depended on the effect of the TSSR: Students in the condition spent more time actively writing when in TSSR and more time pausing when in no-TSSR (see Table 5).

A possible explanation for these results may be found in the very nature of the L2 writing task. These results are not that easy to explain but we believe that a possible explanation should take the complexity of an L2 writing task into consideration. In the type of synthesis task we used here, academic citation techniques were not expected, but students had to use indirect quotations (without using references) to incorporate content from the sources into their own texts, which entails paraphrasing of the selected content. Chenoweth and Hayes (2001), in their study with undergraduate university students, discuss the predominant role of the translator and the reviser as guiding the writing processes in L2. Moreover, when content is drawn from the sources, as in the type of task used in the present study, less cognitive effort is spent on planning processes, and more resources are directed to translation operations such as lexical retrieval and grammatical encoding (Kormos, 2011; Révész et al., 2017). As a result, what is defined as translation and revision processes in various writing tasks, should mostly refer to paraphrasing efforts in our study, particularly by changing syntax and replacing lexical items, the typical paraphrasing activities in L2 for developing writers in academic discourse (Howard, 1992). Hence, students' pausing behaviour observed in the Presentational Instruction, and a lower proportion of source text language in their texts (as compared to the Video Instruction which showed a significant effect in that aspect), can be explained by the cognitive effort spent on paraphrasing. We triangulated text quality and keylogging data with Time-Sampled Self-Report (TSSR) data, which is available only for the TSSR-condition and at one measurement occasion (i.e., posttest). TSSR-data provided complementary findings: the students in the Presentational Instruction reportedly engaged in the activity paraphrasing significantly more than the other conditions, with a stable increase from the beginning to the end, contrary to students in the Video Instruction, who reported engagement with the paraphrasing activity in a decreasing pattern throughout the writing process. See Table 6 for a crossmatch of activity categories, text quality aspects, (composite) scores and the language skill that each tap into.

Table 6
A Cross-Match of the Activity Categories, Text Quality Aspects, Composite Scores and Skills

Activity Categories	Text Quality Aspects	(Composite) Scores	Skill
Reading the source texts Working on the sources	Source Comprehension Text Comprehensibility Main Ideas Supporting Ideas	Content	Reading
Paraphrasing	Source Text Language	Source Text Language Use	Writing
Writing the target text Reading the target text Editing			

In a previous study (Buyuktas Kara et al., 2024, Chapter 3) with developing L2 learners using the same type of task, better text quality results corresponded with a global (top-down) approach to writing as reported in Learner Reports, and Time-Sampled Self-Reports (TSSR). Higher-scoring students read the texts for comprehension and idea mining before synthesising them, whereas lower-scoring students seemed to work on each single text in all its necessary dimensions, which may indicate a bottom-up approach to completing a synthesis task. Similarly, Solé et al. (2013) showed that higher-quality texts (in an integrated summary from multiple sources task) were a product of writing processes that involved a relatively more iterative engagement with sources as ongoing references throughout the writing process. Hence, the more and longer pauses observed in the Presentational Instruction may be indicative of a recursive use of reading for producing the text. These may include reading for comprehension, idea-mining, revision, checking the target text for accuracy, which we would expect to be shown as longer and more pausing. This temporal management strategy of the Presentational Instruction in the present study, of pausing more and longer, is in line with (1) the parameters of effective temporal organisation of writing activities defined in previous studies using process data (Braaksma et al., 2004), and (2) the inherent chronology of a reading-to-write task, that is, reading the sources for making associations before writing (Leijten et al., 2019), and also (3) with the empirically informed instructional recommendations, for example, to write down everything before proceeding to revision (Chenoweth & Hayes, 2001), instead of trying to manage the formulation and revision processes

together (Chanquoy, 2001), an indication of a bottom-up approach. Fayol (1999) states that such a bottom-up approach is an attempt at accommodating the extra cognitive cost that comes with the complex writing task; which is a strategic change of management of the writing process, but may result in texts that are syntactically and semantically lower in quality (Olive, 2010). More time allocated to active writing combined with a higher proportion of source text language in Video Instruction & TSSR may be an indication of a bottom-up approach to writing, and/or trying to accommodate the extra cognitive cost of the complex writing task.

We also argued in the previous section that the use of writing logs may have amplified the effects of TSSR reporting mode in general. Considering that the highest reactivity is observed in Presentational Instruction & TSSR, we argue that it may have circumvented one major shortcoming of presentational modes of instruction in general. That is, maintaining an explicit representation of a writing strategy during a subsequent writing task (Fidalgo, et al., 2015). The printed writing log sheets with the activity categories, their definitions and graphic representations may have helped as an instructional aid in the proceduralisation of the declarative knowledge on the writing activity categories in the presentational condition, and influence their writing performance in the subsequent writing task. However, irrespective of the direction of the reactivity, we can safely conclude that the mode of instruction used in the training session affects the reactivity of TSSR on students' writing performance, albeit in different directions.

Following from the results, it can be inferred that the video instruction pointed at a negative reactivity in effects on writing products and processes. One possible explanation is that the scripted TAP acted out by the peer model was not representative of an actual management of a problem-solving activity with the writing processes unfolding in a way that is comparable to the real time necessary to complete the task. The video used in the present study is a compilation of nine separate video clips (5 mins. 49 s.), ranging in duration between 10 to 80 seconds (an average of 34 s.), with quick transitions, each marked with an auditory signal that probes students to the directed retrospection task. This video was initially designed for use in the two previous studies (Buyuktas Kara et al., 2018, 2024; Chapter 2 and 3) and was aimed at teaching the writing activity categories at the declarative level (i.e., what the writing activity categories are) and the mechanical implementation of TSSR in dual task conditions at procedural level (i.e., what to do when hearing a bleep). We followed the same training procedure as in the TTT/TSSR-studies and elaborated on the activity categories during an error-correction phase that followed the video show, with the aim to match students' mental representations of the categories with that of the

experimenter. Using evaluation and elaboration tasks while observing modelled performances, is also in line with the tenets of observational learning as they help to activate metacognition and internalise the evaluation criteria (Braaksma et al. 2001). As a result, students may have taken this compact video with a rapid progression of different writing activities, shown one after the other, as comparable to an ideal performance, and hence started all processes of the complex writing task as they observed it in the video. In any case, the highly analogous pattern of this distinct change in the writing behaviours of the video instruction condition in both TSSR and no-TSSR shows the powerful instructional potential of peer videos in causing changes in students' writing processes.

4.3 *Effects of Instruction Mode on the Collected Data: Video vs. Presentational (HYP3)*

In Hypothesis 3, we proposed that when writing performance data is collected under TSSR-conditions, the mode of instruction (i.e., video vs. presentational) used in the training session affects the process data. The results indicate that, in Presentational Instruction, students allocated progressively more time to paraphrasing throughout the writing process, in stark contrast to Video Instruction, where students reported dedicating increasingly less time to paraphrasing from the beginning to the end of the writing process. From these results we presented, we can confirm Hypothesis 3 as there is a significant difference between the two instruction modes in the reported writing process activities.

The validity of the collected data was previously explained by intermediate and inferential processes that are caused by probing for specific information with a fixed set of alternative answers (Ericsson & Simon; 1984/1993), which, as well as causing reactivity on the primary task, causes production of verbal data that is not closely related to the actual thinking processes of the writers (Ericsson & Simon; 1984/1993). The reactivity and validity issue discussed here underscored the direct relationship between a technique's reactive impact on writing products and processes and the validity of the collected data. With the results of the present study, we have successfully documented that the mode of instruction used in training students for a verbalization requirement affects the data collected.

4.4 *Conclusion*

The main conclusion of this study is that each and every variable in the operationalisation of TTT/TSSR needs to be checked for reactivity within the specific experiment using it, as they are highly prone to cause reactivity. Reactivity of

TTT/TSSR may result from one or a combination of the following factors: modulations introduced to its separate components (i.e., the type and number of activity categories, etc.), different participant profiles (i.e., different levels of schooling, cognitive development), and research designs (i.e., test conditions). For example, the directed retrospection task can be modulated in the type and number of activity categories, or the secondary RT task can be modulated in the mean interval schedules of the auditory signals, which have all been shown to be sensitive to modifications when tested for the validity of the reported data. In this study, we also observed that the reporting tool in TSSR, that is, printed writing logs can increase reactivity because the propositional, visual and the spatial cues provided when encoding information can be duly used when retrieving information during the subsequent test under TSSR-conditions. That being mentioned, we argue in favour of reporting via allocated computer keypresses, as it is used in TTT, because this can circumvent the reactivity issues resulting from the printed writing log sheets that we previously discussed. Similarly, since the task schemas can be activated via the set of declarative information on the writing activity categories, incorporating a lower number of categories to select from may possibly reduce reactivity. This would, however, also mean that the data collected would provide fewer insights into the writing process. Therefore, we assert that the data collected via TTT/TSSR needs to be interpreted with care considering what the students were instructed on in the directed retrospection training.

The importance of a reactivity check in studies involving experimental manipulations may have been somewhat underestimated. Researchers have traditionally held that systematic bias due to the operationalisation of TTT/TSSR remains minimal across various experimental conditions (Piolat et al., 1999; Torrance et al., 1999). This offers a reasonable justification for employing this measurement method in empirical studies. However, it is worth noting some potential considerations. For instance, in Chapter 3, we observed that the impact of the mode of instruction, which initially appeared to be the more effective condition in writing strategy instruction, levelled out at a second posttest when TSSR was used. With the results of this study, we showed that a systematic bias in the collected data, due to the reactivity of the measurement technique, cannot be definitively excluded in studies that employ an alternate form of test-retest design (with a different topic but the same type of writing task). This can result in reactivity resulting from a learning effect in pretest-posttest experimental designs or in studies implementing two posttests to check delayed effects.

As a solution to this issue, we suggest that a reactivity analysis should be incorporated within the main experiment of each study that uses TTT/TSSR as a

process measure, which would indicate the validity of the collected data under the specific circumstances of each study. For example, Kellogg (1987) compared texts written under TTT conditions by a sample set of participants in the main experiment (n = 30) with that of a control group recruited for this purpose. Fidalgo et al. (2012) in their study followed a similar procedure and selected a sample set of data collected as part of the main experiment using TTT and compared it to data from the same participants who were required to compose their texts under single writing conditions. The results of these reactivity checks have been discussed in more detail in the earlier sections of this chapter. It should be noted, however, that from the information provided, we could not draw a timeline for the reactivity check (Fidalgo et al., 2012) for the main experiment (Fidalgo et al., 2014), that is, whether it was implemented before or after the main experiment, which seems to be an important detail given that the same students were used in both single writing versus triple task conditions. Therefore, we recommend that researchers provide detailed explanations of their reactivity checks that are preferably reported as part of the main study, to enable other researchers to comprehend the factors influencing the (non)reactivity of a technique within a particular study.

We also think that the latency effect posed by the TSSR or self-report methods on the writing performance during a dual-writing condition should not be underestimated. This is because a process cut short has the potential to induce negative reactivity on text quality in a timed-writing task, particularly on task-specific aspects that are attended to in the later stages of the writing process. In the studies of this dissertation, we showed that the source-related activities initiate the synthesis writing process, whereas target-text production activities are usually completed in the middle and the final stages of the writing process. Then we ask, could students in the TSSR-condition have incorporated a lower proportion of source text language in their texts, by paraphrasing more and better, and this way could have outperformed the no-TSSR-conditions in both aspects of text quality if they were given more time to compensate for the latency effect? The answer to this question is also relevant when interpreting how the collected data was affected by the latency effect posed by TSSR, especially considering the timed-writing conditions of the pretest and posttest in the present study. We did not opt for such a manipulation in the present study to see all possible effects of TSSR from a methodological perspective (see Barkaoui, 2016; for more information on the effect of timed-writing tasks on the writing process). Future studies aimed at producing a complete mapping of temporal distribution of writing activities, should make the necessary adaptations to their research designs to minimise these concerns.

Replicating the methodological experiment with other types of writing tasks, could also help in getting more insight into the reactivity effects discussed here, such as the latency effect. Given the ecologically valid circumstances in which we implemented our study, we used as data one type of writing task, which can be considered as a limitation of the study.

Another issue to consider is the operationalisation of source text language use with text-matching software, a technique that is recommended for its potential to provide researchers with new insight into process and product dynamics (Leijten et al., 2019). However, one limitation of assessing text quality based on computerised similarity analyses needs to be mentioned. WCopyFind indicates the overlapping parts between a student's text and the sources, hence reveals the quantity of (verbatim) source text language used in that text. The non-overlapping parts in the remainder of the text still needs to be manually coded if the aim is to assess students' texts both in terms of quantity and quality of paraphrasing skills. Researchers interested in such metrics should consider developing an analytical scoring rubric with a focus on paraphrasing quantity and quality or adding an overall quality dimension to an analytical rubric (see Yamanishi et al., 2019). Adapting the six-point holistic scale used in TOEFL IBT synthesis tasks (i.e., listening and reading integrated writing) to check student-papers' unoriginal expressions is another viable option for researchers looking for an empirically validated measure to operationalise paraphrasing in L2 synthesis writing tasks. Assessing paraphrasing quality, however, fell outside the scope of our research. We were primarily focused on quantifying the proportion of source text language taken from the sources. In this sense, WCopyFind provided us with objective, valid scores for assessing writing quality within the research objectives of the present study.

Based on the empirical evidence from this study, it becomes evident that the primary task of writing is sensitive to TSSR as a process measure, resulting in changes in students' writing processes, in the corresponding text quality results and ultimately in the quality of the process data that is collected using it. Therefore, we contend that future researchers who are interested in operationalising writing processes via either of these multi-component techniques (TTT/TSSR) should exercise caution when modulating each component in experimental studies because of their potential reactivity. In fact, we would like to conclude by stating that testing the reactivity of the technique in every study using TTT/TSSR is imperative for maintaining the quality of empirical research.

CHAPTER 5

GENERAL DISCUSSION

This dissertation aimed at providing a deeper insight into synthesis writing instruction in three empirical studies. Following the Introduction, we reported each study in a separate chapter. In this final chapter, we provide information on how the three studies tie in with each other, synthesise the main results of the studies, and present theoretical, practical, and methodological implications, as well as directions for future research. The first two studies tested the effects of instructional content (i.e., explicit strategy vs. no strategy) and instructional mode (i.e., modelling vs. presentational) on students' synthesis writing quality and writing processes. We first investigated the effects of explicit strategy instruction by comparing it to regular instructional practice (of an institution). A second research question investigated the effectiveness of two modes of instruction, namely presentational and modelling, that are typically used in explicit strategy instruction programmes. Both research questions were tested for their effectiveness on students' writing performance, which we defined in this dissertation as students' writing quality and writing processes. The second study was a replication of the first study for generalisation of results with a larger sample size and different participant profiles. In a third study, we investigated a methodological issue that was raised in the findings of Study 2, that is, a possible reactivity of the measurement method, namely the online Time-Sampled Self-Reporting (TSSR), that we used in the two studies to operationalise students' writing processes. Together the three studies provide an extensive insight on the effect of different modes of explicit strategy instruction on students' synthesis writing products and processes and into methods tapping into the writing process.

1 TWO INSTRUCTIONAL INTERVENTIONS: STUDY 1 & 2

1.1 *Instructional Method*

Our selection of the instructional method was significantly influenced by three meta-analyses. Initially, drawing from the findings of Graham and Perin's meta-analysis (2007), we opted for explicit strategy instruction. This practice, supported by evidence, is known for enhancing students' writing skills. The explicit presentation of various cognitive shortcuts makes it a time-efficient technique, particularly well-suited for instructing the synthesis task used in our context, considering the limited time available within the curriculum for this purpose.

Secondly, to decide on the mode of instruction in the delivery of the explicit strategies, we were particularly interested in Self-Regulated Strategy-Development by Harris and Graham (1996), which was shown via two meta-analyses (Graham & Harris, 2003; Graham, 2006) to be an effective strategy training program for teaching writing skills, also in reading-and-writing hybrid tasks (Mason et al., 2006; Martínez et al., 2015). Tasks of this nature are included in the studies of this dissertation. We aimed to differentiate the comparative efficacy of the two primary instructional modes in SRSD, specifically modelling and presentational, regarding their influence on the quality of students' texts and their writing processes. Additionally, given the limitations of a four-hour curriculum allocated for this specific task, we implemented SRSD at a micro-level within the lesson plans of the intervention. This approach facilitated an assessment of the effects of both instructional modes, allowing us to fine-tune and adapt them effectively to accommodate future time-constrained instructional modules.

The study by Fidalgo et al. (2015) was the only study that systematically examined the impact of the two instructional modes by isolating modelling from a direct, presentational teaching practice, and analysing their individual effects, ultimately demonstrating the benefits of the modelling instruction on students' text quality and writing processes compared to presentational instruction.

1.2 *Type of Task*

The synthesis task that we used is a source-based writing task, namely, an integrated summary of three source extracts. Source-based writing (i.e., summarising single/multiple sources) and synthesis writing are often used interchangeably in the literature, and they are expected to benefit from common instructional practices (Van Ockenburg et al., 2016). From a (cognitive) developmental learning perspective (see Bereiter & Scardamalia, 1987), the skills needed for a source-based writing task ultimately add up to three transformational stages in synthesis writing (i.e., selecting, organising, and connecting; see Spivey & King, 1989), hence facilitate their elaboration (Hirvela, 2016). This classification aligns with an expert's problem-solving behaviour when dealing with more multifaceted synthesis writing tasks such as literature reviews (Raedts et al., 2017), or argumentative writing tasks (Bennett et al., 2016). Therefore, it can be considered as a gateway to discourse synthesis in introducing the basics of source-based writing, citation use in particular. Previous studies showed that citation use needs to be phased in synthesis writing instruction both for effectiveness of instructional practice (Lee et al., 2018; Leijten et al., 2019), and for ensuring the validity and reliability of their assessment (Bennett et al., 2016). The integrated summary task

of multiple source extracts helped us phase instruction in synthesis writing (with a focus on citation use) and included the following instructional objectives:

By the end of a four-hour instruction, students are expected to have mastered all of the following skills on the citation use requirement of synthesis writing:

1. Proper source attribution using basic in-text referencing following APA guidelines.
2. Using a variety of reporting verbs to control the function of a citation (i.e., the verb *add* is used to provide complementary information; the verb *disagree* is used to present contrasting viewpoints, etc.).
3. Establishing coherence within the text via linking devices (i.e., Another reason for x is y).
4. Integrating source content in the target-text BY
   a. Identifying levels of contextual importance of the idea units in the sources (i.e., main ideas, supporting ideas and examples)
   b. Paraphrasing selected source content.

1.3 *About Creating a Strategy*

In SRSD, the teacher directly introduces explicit strategies for particular writing tasks, which are often presented in a mnemonic format and displayed in visual representations and/or charts. The mnemonics are a memory aid for students to memorise the strategies and recall the fundamental mental processes involved in using those writing strategies, as well as a tool for instructors for scaffolding strategy use, until students no longer require it (Harris et al., 2008).

Following the same practice, we created the TRAMPOLINE mnemonic, encapsulating the set of strategies for meeting the task requirements of the integrated summary task from multiple source extracts. We translated each one of the task requirements into a cognitive strategy and included metacognitive and schema-activating strategies (T of TRAMPOLINE) to support them. See Table 1.

Table 1.
The TRAMPOLINE Strategy

Think – three steps: a. Before reading the first extract: Think about the purpose why you are given different extracts on the same topic? After reading the second and the third extracts: What is the relationship of this extract to the previous one? b. What do you expect to learn from the extract?

c. What do you already know about the general topic of the extracts?

Read the extracts:
Ask - What is the main idea?
Mark the important details.
Paraphrase the main idea and the important details.
(Repeat TRAMP for each extract and OLINE for the whole summary)
Organise the paraphrased ideas.
BY
Linking the ideas with appropriate linkers,
Including APA,
Nesting reporting verbs.

Edit your summary.

1.4 *The Quality of the Writing Process*

As a parameter for effective processes in synthesis writing we adapted the quality criteria from the study by Braaksma et al. (2004). They showed for L1 argumentative writing tasks that the modelling condition when compared to a writing only condition, reported in think-aloud protocols more metacognitive activities in the initial phases and more executive activities in the later phases of the writing process. We expected these effects, also in line with the study on synthesis writing processes by Solé et al. (2013), to manifest as a more recursive use of reading (i.e., source-related activities) for text production overall, with source-related activities reported in the initial phases and more target-text-related activities in the later phases of the writing processes. The source-related activities encompassed reading the sources and working on the sources, with the former focused on comprehension, making the latter a strategic reading activity for identifying relevant content. In the target-text-related activities, the activity categories consisted of writing the target text, paraphrasing source content, reading the target text, and editing. Writing the target text was classified as a basic transcription activity, distinguishing it from more cognitively demanding and strategic activities, such as paraphrasing, reading the target text, and editing. We also expected that in their learner reports, students would report learning experiences associated with (hence, coded under) the target learning gains (i.e., strategies and/or task requirements).

1.5 *Measuring the Quality of the Writing Processes*

We measured students' writing processes by employing both online and offline methods. The online methods included the Inputlog keylogging software (Leijten & Van Waes, 2013) and Time-Sampled Self-Reporting (Torrance et al., 2007). The (retrospective) offline method was adapted from a study by Groenendijk et al. (2013), from the original Learner Report method (De Groot, 1980).

The combination of online and offline methods in our operationalisation of the writing process provided comprehensive insights into different knowledge levels and their management during the subsequent synthesis task. Low writing performance may result from incomplete declarative and procedural knowledge regarding specific task completion. Nevertheless, mere possession of this knowledge does not guarantee effective application in relevant tasks. This dormant or non-activated knowledge may exist in developing learners, but they might still struggle to perform the necessary metamemorial search to access it (Fayol, 1994). In this context, the learner reports offered valuable information about students' perceptions of their declarative and procedural knowledge of synthesis writing strategies, while TSSR allowed us to observe the actual attempts at the implementation of these strategies in students' writing processes.

Using two online process measures in the same experiment in Study 3 reported on in Chapter 4, provided a unique insight into process data collection techniques. That is, although we used keylogging in Chapter 4 to investigate the reactivity of TSSR, the TSSR data collected to that purpose provided a good triangulation possibility in the final analysis. We were able to solve some of the bottlenecks of the keylogging data when explaining students' different approaches to the temporal management of the writing process, particularly in regard to students' pausing behaviours.

1.5.1 TSSR-Training

The training session for Time-Sampled Self-Reporting (TSSR) was conducted as follows: First, we brainstormed with the students to identify potential activities they might engage in when completing an integrated summary task from multiple source extracts, which they had previously done in the pretest. We listed these activities on the board for reference. Next, we presented the students with a compilation of nine separate video clips, with each clip depicting a different writing activity category performed by a model while completing an integrated summary task. To facilitate student engagement, each video clip was punctuated with an auditory signal, prompting the students to identify the corresponding writing activity, on a short version of the writing log (that would be used in the

subsequent posttest under dual -task writing conditions), that we distributed to them. Following the video presentation, we checked students' answers as a whole class error-correction activity and provided immediate feedback. Subsequently, we created a simulated environment resembling a timed-writing exam under dual task writing conditions. We presented the students with a synthesis task along with writing logs and conducted a 10-minute demo test during which they responded to a total of nine auditory signals ('bleeps'), by reporting their activities. The entire training session was completed within one lesson hour (i.e., approximately 50 minutes).

1.6 *The Interventions*

Drawing from prior studies on observational learning (see Rijlaarsdam et al., 2008), we expected that the modelling mode of instruction would be more effective than the presentational mode of instruction for teaching explicit strategies, both on students' writing processes and also their text quality outcomes.

We tested this hypothesis in the English preparatory program of a private university in Turkey. Participants were B2 level EFL learners (according to the Common European Framework of Reference), 48 academic low-achievers in Study 1, and 155 normally performing students in Study 2. This also helped us see whether the effects hold for the two different participant profiles.

After a four-hour intervention program, three randomised conditions were evaluated: a presentational condition and a modelling condition (i.e., the explicit strategy instruction conditions), and a control condition (i.e., no-strategy condition), being the regular curriculum instruction of the university. In the strategy conditions, the instructional content was TRAMPOLINE strategies mnemonic for teaching synthesis writing. In the presentational condition, this was done via a teacher-led presentational instruction. In the modelling condition, students watched videos in which a peer modelled the use of strategy mnemonics; and in the control condition, the instruction was not strategy-focused, students inferred their own strategies from the materials and the instruction provided. Effects were tested for text quality based on three variables: content & organisation (i.e., the content composite score), source use, and authenticity (i.e., absence of verbatim source text language use); and on writing processes based on the aforementioned quality criteria for synthesis tasks, operationalised by the online TSSR-technique and the offline Learner Report method.

1.7 *Results of Study 1 and Study 2*

1.7.1 Effects on Synthesis Text Quality

The results of the two intervention studies showed that explicit strategy instruction on synthesis writing is effective when it was modelled rather than presented. In both studies, the modelling mode of instruction resulted in qualitatively better texts than the other conditions in the source use aspect of text quality. In Study 2, source use as well as authenticity (which was operationalised as less verbatim source text language use) aspects improved as a result of the modelling mode of instruction. Another finding of Study 2 was that the no-strategy condition wrote more authentic texts than the presentational condition.

1.7.2 Effects on Transfer to Another Genre

Students in all conditions in both studies took an argumentative writing exam as part of an adjunct module-end-test (MET), which served as the institutional exit test for pre-faculty students, including reading, listening (i.e., while-listening and via note-taking) and writing assessment components, introduced within a topical context. Accordingly, all other test materials belonging to other language skills (i.e., listening and reading) were collected by the examiner, but students could use their retained ideas, mental notes from the reading texts to help them support their arguments and provide counter-arguments and refutations.

Hence, unlike the synthesis tasks the argumentative tasks at the transfer test were not source-based; rather, they were developed as per the principles of a scenario-based assessment task (see Bennett et al., 2016). Therefore, source use and authenticity were not relevant in the operationalisation of text quality of the argumentative texts. Instead, text quality in the argumentative task was rated on four aspects, which were subsequently merged under two composite scores: content (a composite score of content and organisation aspects), and language use (a composite score of grammar and lexis aspects).

In Study 2, the results of this exam were used to investigate whether the conditions transferred the quality of their writing to this argumentative writing task. The results showed that the presentational condition transferred what was learned for synthesis writing to the argumentative writing task into one of the two aspects, namely, to the content aspect of text quality.

1.7.3 Effects on Writing Processes

In both studies, the modelling condition consistently met the quality parameters for the writing process. Study 1 revealed that students in the modelling condition engaged in strategic source-related activities, particularly working on the sources, during the middle stage of the writing process, significantly more than the presentational condition. There were no particular patterns in the writing process activities of the presentational condition in this study. Study 2 provided further insights, as the strategy conditions dedicated more time to source-related activities than the no-strategy condition. This indicated a recursive use of reading to inform the production of the target text. Moreover, they displayed a greater variety in their choice of activities compared to the no-strategy condition, highlighting a more dynamic approach to the writing process.

Conversely, the writing processes of students in the no-strategy condition exhibited a sequential rather than a recursive pattern, that is, reading the sources in Phase 1 and writing the target text in Phase 3 (in Study 1). These two activities were also what we had identified as the non-strategic ones within the broader categories of source-related and target-text-related activities. In Study 2, their choice of activity was similar, that is, significant effects were observed in Phase 2 in the activity writing the target-text. This marked a distinct difference in their process behaviours compared to the strategy conditions (in Study 2), where more deliberate and planned activities were observed.

Furthermore, the modelling condition provided notable insights through the offline learner reports, revealing students' increased awareness of the procedural aspects of the writing task. In Study 1, these students shared their learning experiences related to the systematic approach to the synthesis writing task (with statements coded under abstract categories of steps/stages/subprocesses). They described their ongoing learning process of understanding and mastering the various steps involved in composing a synthesis text. In Study 2, the modelling condition expressed positive learning experiences concerning time-management, signifying a shift in their self-perception regarding their ability to efficiently manage the task requirements within the allotted time. Additionally, students in the strategy conditions shared their learning experiences related to the paraphrasing requirement of the task in Study 2. This highlights the valuable insights students gained through the instructional strategies applied.

1.8 *Discussion of the Results on Aspects of Synthesis Text Quality*

The larger-scale study, which included a more extensive sample, facilitated the generalisation of findings from the previous study, including the generalisation

of effects to an academically stronger student profile. The results attest to the results of the study by Braaksma et al. (2004) that better writing processes influences the text quality results in a positive direction. The fact that the modelling condition reported the activity working on the sources (via TSSR) and scored higher in the source use aspect of text quality in Study 1, and that they reported paraphrasing more (via TSSR) and scored better in the authenticity aspect can validate this finding.

1.8.1 Effects on Source Use

One plausible explanation for the modelling condition's success in the source use aspect is derived from previous research findings, such as those presented in the study by Braaksma et al. (2002), indicating that observation is particularly effective when applied to novel tasks. Proficiency in source use predominantly involves mastering the unique features of this new writing task. Consequently, students' relative unfamiliarity with these novel aspects of the task may have contributed to their success, particularly in the domain of source use skills, as evaluated within the context of text quality.

1.8.2 (Missing) Effects on Content

The reason the modelling condition did not lead to a substantial improvement in the content aspect of text quality can be ascribed to the instructional hierarchy in which the task requirements of this particular reading-to-write task were introduced. In this context, the emphasis was placed on citation use, as an initial phase of synthesis writing instruction. Consequently, the selection of the source extracts and the organisational requirements of the task (both contributing to the content composite score), were deliberately designed not to detract from this primary objective of prioritising citation use both in instruction, and also in testing, to ensure validity and reliability of text quality results (Bennett et al., 2016). It should be noted that the tests were in timed-writing conditions, with an allocated time of one lesson hour. In accordance with these conditions, the sources provided to the students packed the idea units that were structured hierarchically in brief extracts. This helped us prevent overloading students' L2 competence with challenging reading tasks, as slow access to these features could potentially impede performance in an L2 context (Schoonen et al., 2003), amplifying the complexities of the interplay of the reading problem and the language problem (Alderson, 1985), that are typically posed by an L2 synthesis task. In this sense, the missing effect in content indicates that the text quality measures

(i.e., synthesis writing tasks) were consistent with the approach of phasing (synthesis writing) instruction, and the collected data were ecologically valid representations of students' synthesis writing quality at the measurement occasion.

1.8.3 Effects on Authenticity

The success of the modelling condition in enhancing source use and authenticity, while not significantly impacting the content aspect, can also be explained by crossmatching the TRAMPOLINE strategies with the specific language skills they address. The content aspect of text quality primarily evaluates reading comprehension skills, represented by the first subset of strategies in the TRAMPOLINE mnemonic (i.e., TRAM: Think/Read/Ask: "What is the main idea?" and Mark the important details). Conversely, the subsequent strategies, such as Paraphrasing (P) and OLIN (Organise the ideas by using Linkers, Including APA, Nesting reporting verbs), are predominantly writing strategies and are assessed under the authenticity and source use aspects, respectively. It can be argued that modelling has a more robust impact on writing performance compared to reading comprehension. This distinction may arise because reading is essentially a receptive skill, while writing is a productive skill. The act of production in writing might lend itself more effectively to cognitive modelling of problem-solving actions, in contrast to improving reading skills, where overt problem-solving is typically less prominent due to its covert nature.

1.8.4 Transfer Effects to the Argumentative Genre

The results of transfer effects to another genre in Study 2, operationalised by the argumentative writing task at posttest showed that the presentational strategy condition produced texts that were rated higher in the content aspect of text quality than the no-Strategy condition, while the effects of the modelling condition did not reach significance. This is an important finding when the type of task is scrutinised further. First, the argumentative task featured notably greater demands in terms of content and organisation, as compared to the same quality aspects in the synthesis task (as represented by the content composite score in both). Secondly, the argumentative task was part of a scenario based-exam, that rests on the cognitive developmental perspective that a (retained) mental summary of multiple sources (from the preceding reading and listening sections of the exam) is the first of the multiple steps of an expert's problem-solving behaviour for developing successful arguments (Bennett et al., 2016).

When considered this way, the transfer effects to the argumentative genre provides stronger support for the effectiveness of strategy instruction, even when the requirements are more complex than the intended learning gains of the primary task. That is, crafting a super-proposition regarding the content (Segev-Miller, 2004), establishing a macrostructure for organising ideas (Van Ockenburg et al., 2019), and creating an intertextual connection that links all the sources (Boscolo et al., 2007), which are synthesis (or source-based) writing skills that must have been transferred to the content composite score of the scenario-based argumentative writing tasks (including the organisation aspect of text quality), because of the preceding instruction on synthesis writing.

This observation may also substantiate the above explanation of why students did not also outperform the other conditions in the content aspect of text quality in the synthesis writing tasks. We argue that if the sources were more difficult (as measured by Flesch Kincaid readability ease scores), the modelling condition would also outperform the other conditions in the content aspect. We refrained from such differentiation based on reading skills, given the ecological circumstances of the need to phase the instruction, (hence) the assessment of synthesis writing in a time-constrained program.

1.9 *Writing Processes*

In Study 2, we observed that students in the presentational mode of instruction showed similar patterns in their writing processes to the modelling mode of instruction. However, the purposeful management of their writing processes did not translate into higher scores in text quality. It seemed that students faced challenges particularly in the paraphrasing aspect of the synthesis task. We assumed this because despite exhibiting declarative knowledge on the paraphrasing requirement of the task in their learner reports, and despite also demonstrating a procedural undertaking of it with significant effects on the paraphrasing activity as measured by TSSR, students in both the modelling, and the no-strategy condition, outperformed the students in the presentational condition in the authenticity aspect of text quality.

One possible explanation is that, due to the absence of efficient proceduralization strategies, the declarative knowledge of how to complete the task may have been consuming more time. Students were instructed to read all texts before working on them. Without the benefit of guidance on this process via observation, students' time on task might have been stretched, leading to delays in paraphrasing the ideas. The modelling condition also reported significantly more

learning experiences related to the time-management of the writing process, which can attest to this explanation.

Timed-writing tasks affect students' L2 revision operations (Barkoui, 2016). In our study, revision operations can be associated with paraphrasing operations. Chenoweth and Hayes (2001) emphasised that L2 writing tasks of undergraduate university students are guided by the translator and revisor. It has also been argued that when content is taken from the sources, the translation operations are expected to be mainly in the form of lexical retrieval and grammatical encoding (Kormos, 2011; Révész et al., 2017). These are typical paraphrasing operations of developing writers in L2 engaged in academic discourse (Howard, 1992). Taken together, these converge through the paraphrasing activities in the writing processes of students and the corresponding scores in authenticity aspect of text quality. Modelling seems to be beneficial especially for tackling the time constraints of the synthesis task in the timed-writing conditions of the present study.

Additionally, we discovered that the no-strategy condition, even though they did not explicitly mention declarative knowledge of the paraphrasing requirement in their learner reports, managed to fulfil the paraphrasing requirement and demonstrated higher authenticity compared to the presentational condition. This may imply that in the absence of explicit strategies, the no-strategy group might have been employing a broader, overarching strategy to meet the task's demands. On the other hand, the presentational instruction group, while dealing with several new task requirements, may have focused on one specific requirement as a coping mechanism. This aligns with the findings of previous studies that discuss a compensatory approach, indicating an effort to focus more on the other aspects of text quality, in response to increased cognitive effort posed by the other (Rijlaarsdam & Van den Bergh, 1996; Van der Hoeven, 1997; Van Weijen et al., 2009).

2 THE METHODOLOGY OF COLLECTING WRITING PROCESS DATA: STUDY 3

One intriguing finding from Study 2 laid the foundation for the third empirical study of this dissertation. In both instructional intervention studies (i.e., Study 1 &2, Chapter 3 & 4), we conducted two posttests, one to measure the effects of instruction on text quality and the other to collect process data under dual task conditions. The text quality data from the latter was also used as an implementation check (i.e., intervention fidelity data). In the final analysis, this data from both studies pointed to contradictory results. That is, in Study 1, we observed that the effects of instruction were sustained in the second posttest. However, in

Study 2, the effects of the modelling condition on the source use aspect of text quality diminished between the two posttests. It seemed that, in Study 2, the presentational condition improved their source use skills in between the two posttests, which resulted in both the modelling and presentational conditions (i.e., the strategy conditions) outperforming the control condition in source use. We examined the factors that might have contributed to the improvement in the source use skills of the presentational condition between the two measurement occasions. In retrospect, we considered the possibility that the operationalisation of the writing processes via Time-Sampled Self-Reporting (TSSR) in the second posttest might have had a positive reactivity effect on the source use results of the presentational condition.

2.1 *Possible Causes of Reactivity*

The comprehensive training regimen and the multifaceted approach introduced various factors that could have potentially triggered reactivity. These factors were encountered by students repeatedly throughout the intervention, spanning across the training and the practice sessions, and the actual implementation of the technique in the subsequent writing tasks in dual task conditions (i.e., writing and self-reporting). In a broad sense, we postulated that one or a combination of the following factors might have activated writing strategies, thereby leading to a discernible learning effect, and caused changes in students' text quality and writing processes, and as a result, in the quality of the collected data of writing processes:

1. The explicit presentation of the activity categories and their use in the training, practise or implementation sessions
2. The mode of instruction in the training session (i.e., video-modelling)
3. The materials used: the printed writing log that was used as a reporting tool in the training, practise and implementation sessions.

2.2 *The Experiment*

To address these concerns, an experiment was designed to disentangle the impact of the training and implementation sessions, as well as the influence of the instructional mode employed during the training session. This was achieved through a 2 by 2 factorial design, creating four distinct conditions. The manipulation involved the reporting mode of the posttest (TSSR vs. no-TSSR) and the instructional mode of the TSSR-training (Video vs. Presentational). A baseline test was conducted under regular circumstances, and the various conditions

were introduced in a subsequent writing task. The study involved 81 undergraduate students from a Belgian university. The dependent variables were students' synthesis text quality, rated on content and source text language use aspects, and students' writing processes which were discreetly collected through keylogging software. We also evaluated the validity and the reliability of the writing process data collected via TSSR, based on the results of writing products and processes in the posttest.

2.3 *Reactivity Indicators in Student's Writing Performances (Text Quality and Writing Processes)*

The interpretation of the results from the keylogging data was guided by the approach we employed to analyse students' writing processes in the two previous instructional interventions of this dissertation, using the same type of synthesis task. This approach involved combining the results obtained from online and offline process measures, along with the corresponding text quality aspects. We then interpreted these findings based on the quality parameters derived from previous studies investigating students' writing processes using a similar type of synthesis task (Solé et al., 2013), instructional mode (Braaksma et al., 2004) and keylogging technique (Vandermeulen et al., 2020).

Accordingly, we assumed that students' pausing behaviours, including the time allocated to pausing versus active writing, and the duration of pauses, could serve as indicators of several process activities. First, they could indicate a recursive use of reading during the process of producing the target text, as students pause during reading to better understand source content and identify levels of contextual importance of the ideas. Secondly, pausing behaviours might signify a deliberate and strategic choice in undertaking target-text-related activities, as opposed to their lower-level or non-strategic counterparts. These target-text-related activities, as they appear on the writing logs distributed to the students, includes paraphrasing, reading the target text, and editing, as compared to the activity of writing the target text. Finally, we associated students' pausing behaviours primarily with the cognitive demands linked to paraphrasing source content, based on a line of reasoning derived from previous studies. According to these studies on undergraduate L2 writing, formulation and revision operations guide the writing process, and in synthesis tasks where content support is provided via source texts, these operations are expected to be mainly in the form of changing lexis and grammar (Kormos, 2011; Révész et al., 2017). which are associated with paraphrasing operations for the developing writers in L2 (Howard, 1992).

Considering these quality parameters, in conjunction with students' text quality results in the corresponding aspects of quality, we gained valuable and dependable insights into students' writing process behaviours. This allowed us to distinguish the effects of the two instructional modes (i.e., video versus presentational) that was used in the training session of TSSR.

2.4 *The reactivity results*

Our hypotheses were corroborated by the findings of the study. Firstly, the use of TSSR in dual writing conditions (i.e., reporting mode) was shown to enhance students' written content quality and increase the time they dedicated to the task, as compared to students in the no-TSSR-condition. Secondly, the mode of instruction used during the training phase had an impact on text quality and writing processes, regardless of the reporting mode employed in the posttest: Students in the video conditions incorporated a higher proportion of source text language into their texts and were more actively engaged in writing compared to students in the presentational condition, who paused more when in TSSR, and paused for longer durations irrespective of whether they were in TSSR or in no-TSSR-conditions. These distinct patterns in students' process behaviours (as measured via keylogging) indicated that the primary task of writing was affected by the secondary self-reporting task, specifically as a result of the mode of instruction that was used in the training phase. These findings provided evidence that the writing process data collected via TSSR is not an accurate representation of students' actual writing processes (hence, neither a valid nor reliable source of writing process data).

2.5 *Discussion of the Results of Study 3*

The combination of TSSR and presentational instruction yielded the most pronounced positive reactivity, affecting both text quality and writing processes, contrary to our initial expectation that the video used in the training phase could trigger observational learning. We ascribed this finding to the nature of the video used in the training, which was originally created and used in the two previous intervention studies, for demonstrating the mechanical aspects of implementing the technique, that is, teaching the writing activity categories at the declarative level (i.e., what the writing activity categories are) and the mechanical implementation of Time-Sampled Self-Reporting (TSSR) in dual task conditions at the procedural level (i.e., what to do when hearing a bleep).

For a think-aloud protocol (TAP) to be effective for instructional purposes, it should teach a skill rather than a body of declarative knowledge (Bereiter & Bird, 1985), which is the case with the TAP of the video used here. In line with the original purpose of creating the video, it featured a compilation of nine separate video clips with varying durations (mean: 34 s.), and a quick transition in between, punctuated with auditory signals that probe writers to self-report. As a result, the partly-scripted think-aloud protocol that was performed by the peer model in the video is not a real-time representation of the writing process that is comparable to the actual time allocations to the separate requirements of the synthesis task, nor representative of actual problem-solving operations that include self-regulatory behaviours. Students may have perceived this compact video with rapid progressions of different writing activities, presented one after the other, as an ideal performance of synthesis writing. This procedural representation may have been further internalised by students through the use of elaboration and evaluation activities that were adopted in line with the original TSSR-studies. These activities are also known to bolster the effects of observational learning from modelled performances (Braaksma et al., 2002). Consequently, students in the condition might have attempted to initiate all the processes involved in the complex writing task as they observed in the video, which is also indicative of a coping strategy that is activated for the proceduralisation (of the demands) of the complex writing task (Fayol, 1999).

The positive reactivity that was observed in the presentational condition is consistent with a previous study by Wallace and Hayes (1996), which demonstrated the effects of an eight-minute direct instruction in enhancing students' subsequent revision quality. It showed that when students have underlying procedural knowledge related to a specific task, direct instruction can lead to the improvement or alteration of their task schemas, resulting in enhanced performance in a subsequent writing task. It should also be noted however, that the effects were shown for entry level undergraduates rather than basic level students, indicating that without the underlying procedural skills, this prompt-based instruction is not sufficient for altering students task schemas. Accordingly, the researchers (Wallace & Hayes, 1996) underscored that this kind of instruction helps in re-orienting students' revision, rather than to be used as an ultimate teaching method. The fact that there were no transfer tests on long-term persistence of the learnt skills, also supports this. Further research is needed to show the long-term effects of modelling versus presentational instruction on task schemas, with manipulation of different levels of underlying procedural knowledge of task requirements. Notwithstanding these concerns, we conclude that, regardless of the direction of the reactivity that we observed, our study

demonstrated that TSSR is a reactive process registration technique that impacts text quality, writing processes and the reliability and the validity of the collected data by inducing a learning effect.

3 GENERAL ISSUES

3.1 *Methodological Implications*

3.1.1 Replication Studies

Replication studies in education play a crucial role in the generalisation of findings, particularly when conducted on a larger scale. They offer a means to ensure the validity and reliability of previous results, providing a robust foundation for extending the applicability of those findings (see Schmidt; 2017). Notably, our studies produced consistent results despite variations in participant profiles, encompassing academically low-achieving and stronger students. This consistency bolsters our confidence in making generalisation claims. Moreover, these findings provide the groundwork for proposing policy changes in preparatory schools, addressing issues that were previously unimplemented.

3.1.2 Mixed-Methods Approach

Across the three empirical studies of this dissertation, we adopted a mixed-methods approach that integrates product data with process registration data in strategy instruction interventions, which has been undertaken in relatively few studies (Schellings & van Hout-Wolters, 2011). Moreover, in each study we leveraged a combination of diverse research methods that shed light on different aspects of writing, encompassing both online and offline approaches. Our online methods involved Time-Sampled Self-Reporting (TSSR) and keylogging, while our offline methods included the Learner Report technique and the Motivated Strategies for Learning Questionnaire (MSLQ). It is worth noting that each online process registration technique presents its unique challenges during analysis. This combined approach enabled us to overcome limitations associated with each technique. We advocate for future studies to enhance the robustness of their findings by incorporating multiple techniques.

3.1.3 Writers as Process Coders

The findings from Study 3 emphasised that a particular aspect of TTT/TSSR, which was considered as the strength of the technique – solving the

segmentation problem for experimenters by having the writers act as process coders – was, in fact, the source of the reactivity issue. In the TTT/TSSR-context, writers themselves also serve as the panel of trained coders in the same experimental study. These writers have received training from the experimenter before the experiment begins. Their role is to code their own cognitive activities while writing concurrently. This approach involves writers not only in the writing task but also in the active analysis and categorisation of their thought processes during the writing process. The high-level abstraction processes such as categorical membership decisions, scanning, and filtering introduces reactivity into the primary writing task. It is noteworthy that this process of analysis is not a one-time event; it occurs during training, practice, and implementation sessions, with multiple practice opportunities. This highlights how multiple phases of the implementation of technique during the intervention fostered learning through inference-based decision-making and problem-solving processes, ultimately benefiting students in the subsequent writing task. Furthermore, the writing logs we provided, featuring propositional, visual, and spatial input, facilitated effective encoding and retrieval of information, resulting in a positive learning effect. This study shows that the reactivity of methods is a crucial factor to consider, particularly when these methods encompass multiple phases of training, practice, implementation sessions, and external materials that might inadvertently become part of the instruction. Therefore, we encourage future studies to carefully evaluate the reactivity of such techniques within the context of their specific research before proceeding with implementation.

3.2 *Timed-Writing Conditions*

In all three empirical studies within this dissertation, we conducted both pretests and posttests under timed-writing conditions. It is important to recognise that in timed-writing tasks, a prematurely truncated writing process can have detrimental effects on text quality, particularly concerning task-specific aspects typically addressed in the later stages of the writing process. We have extensively explored the potential consequences of these time constraints in the preceding sections, including students' compensatory behaviours when confronted with multiple task requirements, potential trade-offs in text quality aspects, and considerations regarding reactivity, as dual task writing conditions are often associated with longer time on task. We have also discussed how compensatory time allocations might have further influenced the results. While the use of timed-writing conditions provides strong ecological validity and aligns with the real-world context we sought to replicate, we also acknowledge the need for future

research to experiment with manipulating the time allocated for tasks. This would provide a more comprehensive understanding of how students' writing processes are affected in settings without time constraints. Therefore, researchers aiming to map the temporal distribution of writing activities should consider adjusting their research designs to address these important concerns.

3.3 *Implications for Theory on Observational Learning*

3.3.1 Self-Regulation

One contribution of our studies to the theory of observational learning is that modelling provides implicit learning opportunities for students that help them improve their self-regulation, even when the learning gains were not made explicit in strategy instruction or in the modelled performance. In Study 2, students in the modelling condition reported significantly more learning experiences relating to time-management that reportedly helped them tackle the demands of the writing task in a more time-efficient manner. This could be associated with effects on writing processes manifesting as a more time-efficient orchestration of the writing process and higher ratings in corresponding text quality aspects.

This was an unexpected finding, because time-regulation and the self-regulatory components that may have caused it, such as goal orientation, self-efficacy, and self-monitoring (Bandura, 1989; Schunk, 1989), were not explicitly included in the instructional content or in the modelled performance. Instead, these aspects were implicitly present in the think-alouds of the models. It is also worth noting that the time-management strategies demonstrated by the modelling condition in the subsequent writing task were purposeful, as students indicated these learning gains in their retrospective learner reports, indicating awareness of these strategies at declarative (knowledge) level. Of course, the timed-writing conditions of the writing exams might have influenced this outcome. However, the fact that students voluntarily attended to the self-regulation behaviours of the model and applied them due to their perceived value strengthens the self-regulation outcome (Schunk & Zimmerman, 2023), especially considering that implicit learning gains is expected to be less probable for developing L2 learners, who were exposed to the model's think-aloud in L2, as well.

3.3.2 Creating Models

Utilising technology for creating and presenting modelled performances can be highly beneficial in instruction, but it is crucial to adhere to specific quality criteria. The representation of problem-solving tasks needs to be compatible with the actual demands posed by the writing task, both in terms of the temporality, as well as the verbosity of the modelled performance, including implicit self-regulatory actions in the think-alouds of the models. In other words, an accurate task representation with real-time alignment to the expected performance in instructional modelling is of paramount importance to get the best results from modelling.

3.4 *Implications for Educational Practice at the Turkish Preparatory Schools*

The results of the three empirical studies of this dissertation provided us with extensive insight into L2 synthesis writing instruction which can inform curriculum design and assessment practices at the English Preparatory Schools in Turkey. We will discuss these implications in the following section.

3.4.1 Synthesis Writing

In the English-medium departments of universities, synthesis writing is a standard assignment in mandatory academic writing courses for first-year students. With this study, we also showed that introducing students to discourse synthesis starting at the English preparatory program is possible, as well as effective. This type of writing task is also frequently encountered in faculty departments, whether in native (L1) or non-native (L2) language contexts. Hence, this is significant not only for students in higher education, as it allows them to showcase their knowledge, but also for all language learners. Reflecting on input that holds intellectual value for the writers can also contribute to improving the L2 learning environment by providing authentic learning opportunities, which is in line with effective EFL instruction methods that are currently being followed at the preparatory programs of Turkish universities.

3.4.2 Refining Strategy Training Programs

With the instructional interventions of this dissertation, we refined Self-Regulated Strategy Development, a long-term strategy training program, both in terms of content and mode of delivery, for teaching synthesis writing strategies in relatively short L2 writing courses. In this sense, this study exemplifies how

training programs can be customised to suit the time constraints allocated for the instruction of a specific task and condenses a full training program to its most effective components.

3.4.3 Modelling Mode in Instruction

An important aspect of development for preparatory schools is incorporating modelling in their instructional practices. Implementing modelling practices in their instructional routines need not be a complex or resource-intensive task for instructors. Even with limited technological resources, instructors can engage a proficient student (a strong peer model) to showcase a skill for other students to observe and replicate. That being mentioned, our research underscores the significance of aligning the temporal demands of the writing process with students' expected performance, encompassing self-regulation components inherent in genuine problem-solving activities, in a way to meet the tenets of observational learning. This aspect is less concerning when students are spontaneously asked to model a task for their peers, as the resulting performance is a natural undertaking of the task in real-time, but becomes more critical when models are intentionally created, and the resulting material is used as instructional content.

In the case that modelling an entire process is not feasible, especially for extensive tasks like writing a source-based argumentative research paper, we recommend breaking down the task into its separate components, or task requirements, and select one specific aspect for modelling, rather than showing snippets from a whole performance, which may be counterproductive as we showed with the results of Study 3. The selection of the task requirement to model can be based on the novelty feature of that task because it was shown previously that modelling is most effective in improving novel tasks (Braaksma et al., 2004), or the novel aspects of a task, as we showed with the findings of this study.

In our studies, students responded positively to observing peer models in the instructional videos, finding the experience enjoyable and appreciating the peer support. It introduced a positive change in the learning environment. Therefore, we strongly recommend that instructors consider integrating modelling into their teaching strategies to create similar enriching experiences in their classrooms. Their experiences with an evidence-based instructional practice can contribute significantly to bridging the gap between educational theory and practical implementation in Turkey.

3.4.4 Phasing instruction and phasing assessment

The instructional recommendation to focus on one specific aspect of a task for modelling in more complex writing tasks aligns with the strategy of phasing instruction. Breaking down the instruction into more manageable segments has been shown to enhance students' performance in educational settings. In our case, we phased the synthesis writing instruction, starting with an integrated summary task that emphasised citation use. Prior research has underscored the importance of phasing in citation use during synthesis writing instruction, both for the effectiveness of teaching practices (Lee et al., 2018; Leijten et al., 2019), and the validity and reliability of their assessment (Bennett et al., 2016). It is noteworthy that this phasing approach in instruction should also extend to testing and assessment practices, particularly in the context of L2 education. Scenario-based assessment is a useful method to phase synthesis writing tasks in testing and assessment, in a way to represent the steps of an expert's problem-solving behaviour (Bennett et al., 2016). We observed in Study 2 that the presentational condition transferred the previously mastered summarising skills of multiple source extracts, in the form of creating and retaining a mental summary, to support the generation of their own argumentations in the transfer test five weeks later. This shows that building competencies via multiple contributing skills both in instruction and assessment (via phasing) can bolster learning and its transfer, especially when supported with effective instructional practices.

REFERENCES

Alderson, J. C. (1985). Reading in a foreign language. RELC Journal, 16(2), 114–116. https://doi.org/10.1177/003368828501600214

Anderson, J. R. (1976). Language, memory, and thought. Psychology Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In Psychology of learning and motivation (Vol. 8, pp. 47-89). Academic press. https://doi.org/10.1016/S0079-7421(08)60452-1

Baker, S., Gersten, R., & Graham, S. (2003). Teaching expressive writing to students with learning disabilities: Research-based applications and examples. Journal of Learning Disabilities, 36, 109 - 123. http://dx.doi.org/10.1177/002221940303600204

Bandura, A. (1977). Social learning theory. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Prentice-Hall, Inc.

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. Developmental Psychology, 25(5), 729–735. https://doi.org/10.1037/0012-1649.25.5.729

Bandura, A. (1997). Self-efficacy: The exercise of control. W H Freeman/Times Books/ Henry Holt & Co.

Barkaoui, K. (2011). Think-aloud protocols in research on essay rating: An empirical study of their veridicality and reactivity. *Language Testing*, *28*(1), 51–75. https://doi.org/10.1177/0265532210376379

Barkaoui, K. (2016). What and when second-language learners revise when responding to timed writing tasks on the computer: The roles of task type, second language proficiency, and keyboarding skills. The Modern Language Journal, 100(1), 320-340. https://doi.org/10.1111/modl.12316

Bennett, R. E., Deane, P., & van Rijn, P. W. (2016). From cognitive-domain theory to assessment practice. Educational Psychologist, 51(1), 82-107. https://doi.org/10.1080/00461520.2016.1141683

Bereiter, C., & Bird, M. (1985). Use of Thinking Aloud in Identification and Teaching of Reading Comprehension Strategies. Cognition and Instruction, 2(2), 131–156. http://www.jstor.org/stable/3233543

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition.* Hillsdale, NJ: Lawrence Erlbaum.

Berninger, V. W., & Swanson, H. L. (1994). Modification of the Hayes and Flower model to explain beginning and developing writing. In E. Butterfield (Ed.), *Advances in cognition and educational practice. Children's writing: Toward a process theory of development of skilled writing* (Vol. 2, pp. 57–82). Greenwich, CT: JAI Press.

Blok, H. (1986). Essay Rating by the Comparison Method. *Tijdschrift Voor Onderwijsresearch, 11*(4), 169–176

Bloomfield, L. (2011). WCopyFind. Plagiarism detection tool (Version 4.1.5) [Software]. Available from http://plagiarism.bloomfieldmedia.com/wordpress/software/wcopyfind/

Boscolo, P., Arfe, B., & Quarisa, M. (2007). Improving the quality of students' academic writing: An intervention study. *Studies in Higher Education, 32*(4), 419-438. https://doi.org/10.1080/03075070701476092

Bouwer, R., Béguin, A., Sanders, T., & Van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. Language Testing, 32(1), 83-100. doi: 10.1177/0265532214542994

Bouwer, R., Koster, M., & Van den Bergh, H. (2023). Benchmark rating procedure, best of both worlds? Comparing procedures to rate text quality in a reliable and valid manner. *Assessment in Education: Principles, Policy & Practice*, 30:3-4, 302-319. http://dx.doi.org/10.1080/0969594X.2023.2241656

Braaksma, M., Rijlaarsdam, G., & van den Bergh, H. (2018). Effects of hypertext writing and observational learning on content knowledge acquisition, self-efficacy, and text quality: Two experimental studies exploring aptitude treatment interactions. *Journal of Writing Research*, *9*(3), 259-300. https://doi.org/10.17239/jowr-2018.09.03.02

Braaksma, M. A., Rijlaarsdam, G., Van den Bergh, H., & van Hout-Wolters, B. H. M. (2004). Observational learning and its effects on the orchestration of writing processes. *Cognition and instruction*, *22*(1), 1-36. https://doi.org/10.1207/s1532690Xci2201_1

Braaksma, M. A., Van Den Bergh, H., Rijlaarsdam, G., & Couzijn, M. (2001). Effective learning activities in observation tasks when learning to write and read argumentative texts. *European Journal of Psychology of Education*, *16*(1), 33-48. http://dx.doi.org/10.1007/BF03172993

Braaksma, M. A. H., Rijlaarsdam, G., & Van den Bergh, H. (2002). Observational learning and the effects of model-observer similarity. *Journal of Educational Psychology*, *94*, 405-415. http://dx.doi.org/10.1037//0022-0663.94

Breetvelt, I., Van den Bergh, H., & Rijlaarsdam, G. (1994). Relations between writing processes and text quality: When and how? *Cognition and instruction*, *12*(2), 103-123. https://doi.org/10.1207/s1532690xci1202_2

Bridge, C. A., & Winograd, P. N. (1982). Readers' awareness of cohesive relationships during cloze comprehension. *Journal of Reading Behavior*, *14*(3), 299-312.

British Council (2015). *The State of English in Higher Education in Turkey. A Baseline Study*. Retrieved August, 2016, from http://www.britishcouncil.org.tr/sites/default/files/he_baseline_study_book_web_-_son.pdf

Brunstein, J. C., & Glaser, C. (2011). Testing a path-analytic mediation model of how self-regulated writing strategies improve fourth graders' composition skills: A randomized controlled trial. *Journal of Educational Psychology, 103*(4), 922–938. http://dx.doi.org/10.1037/a0024622

Buyuktas Kara, M., Van Steendam, E., Rijlaarsdam, G., & Kuru, H. (2018). The effect of two modes of instruction: Modelling vs. presentational. *International Online Journal of Education and Teaching*, *5*(2), 460-495.

Buyuktas Kara, M., Van Steendam, E., Rijlaarsdam, G., & Kuru, H. (2024). *The effect of strategy instruction and mode of instruction on both writing process and product in a foreign language: Modelling vs. presentational*. Manuscript submitted for publication.

Chanquoy, L. (2001). How to make it easier for children to revise their writing: A study of text revision from 3rd to 5th grades. *British Journal of educational psychology*, *71*(1), 15-41. https://doi.org/10.1348/000709901158370

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in Writing: Generating Text in L1 and L2. *Written Communication, 18*(1), 80–98. https://doi.org/10.1177/0741088301018001004

Clachar, A. (2000). Opposition and Accommodation: "An Examination of Turkish Teachers' Attitudes toward Western Approaches to the Teaching of Writing" *Research in the Teaching of English, 35*(1), 66-100. Retrieved from http://www.jstor.org/stable/40171507

Couzijn, M., & Rijlaarsdam, G. (1996). Learning to write by reader observation and written feedback. In G. Rijlaarsdam, H. van den Bergh & M. Couzijn (Eds.), *Effective teaching and learning of writing. Current trends in research*. (pp. 224-252). Amsterdam: Amsterdam University Press. https://hdl.handle.net/11245/1.125971

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

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behaviour*, *11*(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Davey, B. (1983). Think aloud: Modelling the cognitive processes of reading comprehension. *Journal of reading*, *27*(1), 44-47.

De Groot, A. D. (1980). Over leerervaringen en leerdoelen [About learning experiences and teaching goals] Handboek voor de onderwijspraktijk, 10 (November), B.1-B. 18. Deventer, The Netherlands: Van Loghum Slaterus.

De La Paz, S. (2005). Teaching historical reasoning and argumentative writing in culturally and academically diverse middle school classrooms. *Journal of Educational Psychology*, *97*, 139–158. http://dx.doi.org/10.1037/0022-0663.97.2.139

De La Paz, S. & Graham (2002). Explicitly teaching strategies, skills, and knowledge: Writing instruction in middle school classrooms. *Journal of Educational Psychology, 94* (4), 687 – 698. http://dx.doi.org/10.1037//0022-0663.94.4.687

De Silva, R., & Graham, S. (2015). The effects of strategy instruction on writing strategy use for students of different proficiency levels. *System*, *53*, 47-59. https://doi.org/10.1016/j.system.2015.06.009

DiPardo, A. (1994). Stimulated recall in research on writing: An antidote to "I don't know, it was fine." In P. Smagorinsky (Ed.), Speaking about writing: Reflections on research methodology (pp. 163-181). Newbury Park, CA: Sage.

Dörnyei, Z. (2009). The 2010s. *Communicative Language Teaching in the 21st Century: The "Principled Communicative Approach". Perspectives*, *36*(2), 33-43.

Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and perspectives*, *38*(1), 105-123.

Elio, R., & Anderson, J. R. (1984). The effects of information order and learning mode on schema abstraction. *Memory & cognition*, *12*(1), 20-30. https://doi.org/10.3758/BF03196994

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215–251. https://doi.org/10.1037/0033-295X.87.3.215

Ericsson K.A. & Simon H.A. (1984/93) *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Erkan, Y. D., & Saban, A. I. (2011). Writing performance relative to writing apprehension, self-efficacy in writing, and attitudes towards writing: A correlational study in Turkish tertiary-level EFL. *The Asian EFL Journal Quarterly*, *13*(1), 163-191.

Fayol, M. (1994). From declarative and procedural knowledge to the management of declarative and procedural knowledge. European journal of psychology of education, 9, 179-190.https://doi.org/10.1007/BF03172778

Fayol, M. (1999). From on-line management problems to strategies in written production. In: M. Torrance & G. Jeffery (Eds.), Cognitive demands of writing. Processing capacity and working memory effects in text production (13-24). Amsterdam: Amsterdam University Press.

Fidalgo, R., f, O. T., Olive, T., & Alves, R. A. (2012). On-line assessment of students' global reading strategies through triple task technique. In E. de Vries & K. Scheiter (Eds.), *Staging knowledge and experience: how to take advantage of representational technologies in education and training?* (pp. 70-72). Grenoble.

Fidalgo, R., Torrance, M., & Garcia, J. N. (2008). The long-term effects of strategy focused writing instruction for grade six students. *Contemporary Educational Psychology*, *33*(4), 672–693. http://dx.doi.org/10.1016/j.cedpsych.2007.09.001

Fidalgo, R., Torrance, M., Arias-Gundín, O., & Martínez-Cocó, B. (2014). Comparison of reading-writing patterns and performance of students with and without reading difficulties. *Psicothema*, *26*(4), 442-448. https://www.redalyc.org/articulo.oa?id=72732458002

Fidalgo, R., Torrance, M., Rijlaarsdam, G., van den Bergh, H., & Álvarez, M. L. (2015). Strategy-focused writing instruction: Just observing and reflecting on a model benefits 6th grade students. *Contemporary Educational Psychology*, *41*, 37-50. http://dx.doi.org/10.1016/j.cedpsych.2014.11.004

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of consumer research*, *20*(2), 303-315. https://doi.org/10.1086/209351

Flower, L.,& Hayes, J. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*, 365 - 387. http://dx.doi.org/10.2307/356600

Flower, L. S., & Hayes, J. R. (1977). Problem-solving strategies and the writing process. *College English*, *39*(4), 449-461.

Flower, L. S., & Hayes, J. R. (1980). The dynamics of composing: Making plans and juggling constraints. In Gregg, L. W and Steinberg, E. R. (Eds.), *Cognitive Processes in Writing,* Erlbaum, Hillsdale, NJ, pp. 31-50.

Fuchs, L. S., & Fuchs, D. (2003). Enhancing the mathematical problem solving of students with mathematics disabilities. In H. L. Swanson, K. R. Harris, & S. E. Graham (Eds.), *Handbook on learning disabilities* (pp. 306–322). New York: Guilford.

Gao, J., Picoral, A., Staples, S., & Macdonald, L. (2021). Citation practices of L2 writers in first-year writing courses: Form, rhetorical function, and connection with pedagogical materials. *Applied Corpus Linguistics*, *1*(2), 100005. https://doi.org/10.1016/j.acorp.2021.100005

Garcia-Sanchez, J. N., & Fidalgo-Redondo, R. (2006). Effects of two types of self-regulatory instruction programs on students with learning disabilities in writing products, processes, and self-efficacy. *Learning disability quarterly, 29*(3), 181-211. https://doi.org/10.2307/3003550

Good Calculators. (n.d.). Flesch-Kincaid Calculator. Retrieved January 22, 2023, from https://good-calculators.com/flesch-kincaid-calculator/

Grabe, W., & Stoller, F. L. (2001). *Teaching and researching reading*. London: Pearson Education. http://dx.doi.org/10.1017/S0267190504000030

Graham, S., & Harris, K. R. (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies. In L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of research on learning disabilities* (pp. 383–402). New York: Guilford.

Graham, S., & Hebert, M. (2010). *Writing to read: Evidence for how writing can improve reading (A Carnegie Corporation Time to Act Report)*. Washington, DC: Alliance for Excellent Education.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445-476. http://dx.doi.org/10.1037/0022-0663.99.3.445

Graham, S. (2006). Strategy instruction and the teaching of writing: A meta-analysis. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 187 207). New York: Guilford.

Green A (1998). *Verbal protocol analysis in language testing research: A handbook.* (Vol.5)

Groenendijk, T., Janssen, T., Rijlaarsdam, G., & Van den Bergh, H. (2008). How do secondary school students write poetry? How creative writing processes relate to final products. *L1 Educational Studies in Language and Literature*, *8*(3), 57-80.

Groenendijk, T., Janssen, T., Rijlaarsdam, G., & Van den Bergh, H. (2013). Learning to be creative. The effects of observational learning on students' design products and processes. *Learning and Instruction*, *28*, 35-47.

Guo, H., Zhang, M., Deane, P., & Bennett, R. (2020). Effects of scenario-based assessment on students' writing processes. *Journal of Educational Data Mining*, *12*(1), 19-45. https://doi.org/10.5281/zenodo.3911797

Hammann, L. A., & Stevens, R. J. (2003). Instructional approaches to improving students' writing of compare-contrast essays: An experimental study. *Journal of Literacy Research, 35*(2), 731- 756. https://doi.org/10.1207/s15548430jlr3502_3

Harris, K. R., & Graham, S. (1996). *Making the writing process work: Strategies for composition and self-regulation*. Cambridge, MA: Brookline.

Harris, K. R., Graham, S., Mason, L. H., & Friedlander, B. (2008). Powerful writing strategies for all students. *Education Review*.

Hatipoglu, C. (2016). The impact of the university entrance exam on EFL education in Turkey: Pre-service English language teachers' perspective. *Procedia-Social and Behavioral Sciences*, *232*(1), 136-144. https://doi.org/10.1016/j.sbspro.2016.10.038

Hayes, A. F. (2013).Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. Guilford Press.

Hayes, J. R., & Flower, L. (1981). *Uncovering cognitive processes in writing: An introduction to protocol analysis.* ERIC Clearinghouse.

Hayes, J. R., & Olinghouse, N. G. (2015). Can Cognitive Writing Models Inform the Design of the Common Core State Standards? *The Elementary School Journal*, *115*(4), 480–497. https://doi.org/10.1086/681909

Hayes, J. R., Flower, L. S. (1980). Identifying the organization of writing processes. In L.W., Gregg,E. R. Steinberg, (Eds.), *Cognitive processes in writing: An interdisciplinary approach* (pp. 3-30). Hillsdale, NJ: Erlbaum.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Randall (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1-27). Mahwah, NJ: Erlbaum.

Hayes, J. R. (2012). Modelling and remodelling writing. *Written communication*, *29*(3), 369-388. https://doi.org/10.1177/0741088312451

Hinkel, E. (2015). *Effective curriculum for teaching second language writing: Principles and techniques*. Routledge: New York.

Hirvela, A. (2016). Academic reading into writing. In *The Routledge Handbook of English for Academic Purposes* (pp. 127-138). Routledge. https://doi.org/10.4324/9781315657455

Howard, R. M. (1992). A plagiarism pentimento. *Journal of teaching writing*, *11*(2), 233-45. http://dx.doi.org/10.1016/j.learninstruc.2013.05.001

Hyland, K. (2016). Methods and methodologies in second language writing research. *System*, *59*, 116-125. https://doi.org/10.1016/j.system.2016.05.002

Jacobson, J. (2015). *No More "I'm Done!" Fostering independent writers in the primary grades*. Education Review.

Janssen, D., Van Waes, L., & Van den Bergh, H. (1996). *Effects of thinking aloud on writing processes*. In C. M. Levy & S. Ransdell (Eds.), The science of writing: Theories, methods, individual differences & applications (pp. 233–250). Mahwah, NJ: Lawrence Erlbaum

Kahneman, D. (1973). *Attention and effort* (Vol. 1063, pp. 218-226). Englewood Cliffs, NJ: Prentice-Hall.

Kellogg, R. T. (1987). Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. *Memory & Cognition, 15,* 256-266. https://doi.org/10.3758/BF03197724

Kellogg, R. T. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(2), 355–365. https://doi.org/10.1037/0278-7393.14.2.355

Kerr, B. (1973). Processing demands during mental operations. *Memory & Cognition*, *1*(4), 401-412. https://doi.org/10.3758/BF03208899

Kirkgoz, Y. (2005). 'English language teaching in Turkey: Challenges for the 21st century', in G. Braine (Ed.), *Teaching English to the world: History, curriculum, and practice* (pp. 159-175). Mahwah: Lawrence Erlbaum Associates.

Kirkgoz, Y. (2007). English language teaching in Turkey: Policy changes and their implementations, *Regional Language Centre Journal, 38*(2), 216-228. http://dx.doi.org/10.1177/0033688207079696

Kirkpatrick, L. C., & Klein, P. D. (2009). Planning text structure as a way to improve students' writing from sources in the compare–contrast genre. *Learning and Instruction*, *19*(4), 309-321. https://doi.org/10.1016/j.learninstruc.2008.06.001

Klein, J., & Taub, D. (2005). The effect of variations in handwriting and print on evaluation of student essays. *Assessing Writing*, *10*(2), 134-148. http://dx.doi.org/10.1016/j.asw.2005.05.002

Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, *20*(2), 148-161. https://doi.org/10.1016/j.jslw.2011.02.001

Lee, J. J., Hitchcock, C., & Casal, J. E. (2018). Citation practices of L2 university students in first-year writing: Form, function, and stance. *Journal of English for Academic Purposes*, *33*, 1-11. https://doi.org/10.1016/j.jeap.2018.01.001

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyse and visualize writing processes. *Written Communication*, *30*(3), 358–392. https://doi.org/10.1177/0741088313491692

Leijten, M., Van Waes, L., Schrijver, I., Bernolet, S., & Vangehuchten, L. (2019). Mapping Master's Students' use of external sources in source-based writing in L1 and L2. *Studies in Second Language Acquisition, 41*(3), 555-582. https://doi.org/10.1017/S0272263119000251

Leijten, M., Van Waes, L., Schriver, K., & Hayes, J. R. (2014). Writing in the workplace: Constructing documents using multiple digital sources. *Journal of writing research*, *5*(3), 285-337. https://doi.org/10.17239/jowr-2014.05.03.3

Leki, I., & Carson, J. (1997). Completely different worlds: EAP and the writing experiences of ESL students in university courses. *TESOL quarterly*, *31*(1), 39-69. http://dx.doi.org/10.2307/3587974

Levy, C. M., & Ransdell, S. (1994). Computer-aided protocol analysis of writing processes. *Behavior Research Methods, Instruments, & Computers*, *26*, 219-223. https://doi.org/10.3758/BF03204625

Levy, C.M., Marek, J.P., & Lea, J. (1996). Concurrent and retrospective protocols in writing research. In G. Rijlaarsdam, H. van den Bergh, & M. Couzjin (Eds.), Writing research: Theories, models and methodology (pp. 542–556). Amsterdam, the Netherlands: Amsterdam University Press.

Levy, M.C., Marek, J.P, & Lea, J. (1995). The reliability and validity of retrospective protocols in writing research. *Unpublished manuscript.*

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, *43*(6), 304-316. http://dx.doi.org/10.3102/0013189X14545513

Malhotra, N. K. (1991). Mnemonics in marketing: A pedagogical tool. *Journal of the Academy of Marketing Science*, *19*(2), 141-149. http://dx.doi.org/10.1007/BF02726006

Manchón, R. (Ed.). (2009). *Writing in foreign language contexts: Learning, teaching, and research* (Vol. 43). Multilingual Matters.

Mansourizadeh, K., & Ahmad, U. K. (2011). Citation practices among non-native expert and novice scientific writers. *Journal of English for Academic Purposes*, *10*(3), 152-161. http://dx.doi.org/10.1016/j.jeap.2011.03.004

Martínez, I., Mateos, M., Martín, E., & Rijlaarsdam, G. (2015). Learning history by composing synthesis texts: Effects of an instructional programme on learning, reading and writing processes, and text quality. *Journal of Writing Research*, *7*(2). http://dx.doi.org/10.17239/jowr-2015.07.02.03

Mason, L.H., Hickey Snyder, K., Sukhram, D.P., & Kedem, Y. (2006). Self-regulated strategy development for expository reading comprehension and informative writing: Effects for nine 4th grade students who struggle with learning. *Exceptional Children, 73,* 69–89. http://dx.doi.org/10.1177/001440290607300104

Mason, L.H., Reid, R., & Hagaman, J. (2012). Building comprehension in adolescents: Powerful strategies for improving reading and writing in content areas. Baltimore: Brooks Publishing Co. Inc.

Mason, L. H., Snyder, K. H., Sukhram, D. P., & Kedem, Y. (2006). TWA + PLANS Strategies for expository reading and writing: Effects for nine fourth-grade students. *Exceptional Children*, *73*(1), 69–89. https://doi.org/10.1177/001440290607300104

Mateos, M., & Solé, I. (2009). Synthesising information from various texts: A study of procedures and products at different educational levels. *European Journal of Psychology of Education*, *24*(4), 435-451. http://dx.doi.org/10.1007/BF03178760

Mateos, M., Villalón, R., De Dios, M. J., & Martín, E. (2007). Reading-and-writing tasks on different university degree courses: What do the students say they do? *Studies in Higher Education*, *32*(4), *489-510*. http://dx.doi.org/10.1080/03075070701476183

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge, England: Cambridge University Press.

McManus, K. (2022). Are replication studies infrequent because of negative attitudes?: Insights from a survey of attitudes and practices in second language research. Studies in Second Language Acquisition, 44(5), 1410-1423. https://doi:10.1017/S0272263121000838

McPherson, F. (2000). *The memory key.* New York, NY: Barnes & Noble Books.

Molfese, V. J., & Breznitz, Z. (2012). Introduction to reading and writing sections. *Reading, writing, mathematics and the developing brain: Listening to many voices*, 1. http://dx.doi.org/10.1007/978-94-007-4086-0_1

Nguyen, L. T. C., & Gu, Y. (2013). Strategy-based instruction: A learner-focused approach to developing learner autonomy. *Language Teaching Research*, *17*(1), 9-30. https://doi.org/10.1177/136216881245752

Odell, L., Goswami, D., & Herrington, A. (1983). The discourse-based interview: A procedure for exploring the tacit knowledge of writers in nonacademic settings. *Research on writing: Principles and methods*, 220, 236.

Olive, T., & Passerault, J. M. (2012). The visuospatial dimension of writing. *Written Communication*, *29*(3), 326-344. https://doi.org/10.1177/07410883124511

Olive, T., Kellogg, R. T., & Piolat, A. (2002). The triple task technique for studying the process of writing. In *Contemporary tools and techniques for studying writing* (pp. 31-59). Springer, Dordrecht. https://doi.org/10.1007/978-94-010-0468-8_3

Olive, T. (2010). Methods, techniques, and tools for the on-line study of the writing process. In N. L. Mertens (Ed.), *Writing: processes, tools and techniques (1-18).* New York: Nova Science Publishers.

Ozsevik, Z. (2010). The Use of Communicative Language Teaching (CLT): *Turkish EFL Teachers' Perceived Difficulties in Implementing CLT in Turkey.* (Doctoral dissertation). Retrieved from: http://hdl.handle.net/2142/16211

Pecorari, D., & Petrić, B. (2014). Plagiarism in second-language writing. *Language Teaching*, *47*(3), 269-302. https://doi.org/10.1017/S0261444814000056

Petrić, B. (2007). Rhetorical functions of citations in high-and low-rated master's theses. *Journal of English for Academic Purposes*, *6*(3), 238-253. https://doi.org/10.1016/j.jeap.2007.09.002

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ).* Ann Arbor: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.

Piolat, A., Kellogg, R. T., & Farioli, F. (2001). The triple task technique for studying writing processes: on which task is attention focused? *Current Psychology Letters. Brain, Behavior and Cognition, 4,* 67–83. https://doi.org/10.1002/acp.1086

Piolat, A., Olive, T., Roussey, J.-Y., Thunin, O., & Ziegler, J. C. (1999). Scriptkell: a tool for measuring cognitive effort and time processing in writing and other complex cognitive activities. *Behavior Research Methods, Instruments, & Computer, 31,* 113–12. https://doi.org/10.3758/BF03207701

Piolat, A., Roussey, J. Y., Olive, T., & Farioli, F. (1996). Charge mentale et mobilisation des processus rédactionnels: examen de la procédure de Kellogg [Mental workload and the mobilisation of writing processes: an examination of the Kellogg procedure]. *Psychologie française*, *41*, 339-354.

Plakans, L. (2009). The role of reading strategies in integrated L2 writing tasks. *Journal of English for Academic Purposes, 8*, 252–266.

Plakans, L. (2009a). The role of reading strategies in integrated L2 writing tasks. *Journal of English for academic purposes*, *8*(4), 252-266. http://dx.doi.org/10.1016/j.jeap.2009.05.001

Plakans, L. (2009b). Discourse synthesis in integrated second language writing assessment. *Language Testing*, *26*(4), 561-587. http://dx.doi.org/10.1177/0265532209340192

Raedts, M., Rijlaarsdam, G., Van Waes, L., & Daems, F. (2007). Observational learning through video-based models: impact on student's writing accuracy of self-efficacy beliefs, task knowledge and writing performances. *Studies in writing*, *20*, 219-238. http://hdl.handle.net/11245/1.270116

Raedts, M., Van Steendam, E., De Grez, L., Hendrickx, J., & Masui, C. (2017). The effect of different types of video modelling on undergraduate students' motivation and learning in an academic writing course. *Journal of Writing Research, 8*(3), p.399-435. http://dx.doi.org/10.17239/jowr-2017.08.03.01

Ransdell, S. (1995). Generating thinking-aloud protocols: Impact on the narrative writing of college students. *The American journal of psychology*, 89-98. https://doi.org/10.2307/1423102

Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 88-94.

Révész, A., Kourtali, N. E., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning*, *67*(1), 208-241. https://doi.org/10.1111/lang.12205

Rijlaarsdam, G., & Couzijn, M. (2000) Writing and learning to write: a double challenge. In R. J. Simons, J. van der Linden, & T. Duffy. (Eds.), *New Learning*. (pp. 157-189). Dordrecht: Kluwer Academic Publishers.

Rijlaarsdam, G., & Van den Bergh, H. (1996). The dynamics of composing - An agenda for research into an interactive compensatory model of writing: Many questions, some answers. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and application* (pp. 107–125). Erlbaum.

Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssen, T., Kieft, M., Broekkamp, H., & Van den Bergh, H. (2005). Psychology and the teaching of writing in 8000 and some words. In *BJEP Monograph Series II, Number 3-Pedagogy-Teaching for Learning* (Vol. 127, No. 153, pp. 127-153). British Psychological Society. https://hdl.handle.net/11245/1.254261

Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssen, T., Raedts, M., Steendam, E. V., Toorenaar A., & Bergh, H. (2008). Observation of peers in learning to write: Practice and research. *Journal of writing research*, *1*(1), 53-83. https://doi.org/10.17239/jowr-2008.01.01.3

Rijlaarsdam, G., Van den Bergh, H., & Van Steendam, E. (2015). Writing process theory: A functional Dynamic Approach. In: MacArthur C., Graham S., Fitzgerald J. (Eds.)*, Handbook of Writing Research*, Chapter 4. New York: The Guilford Press, 57-71.

Rijlaarsdam, G., Van den Bergh, H., Couzijn, M., Janssen, T., Braaksma, M. Tillema, M., Van Steendam, E., & Raedts, M. (2012). Writing. Harris, K.R., Graham, S., Urdan, T., Bus, A., Major, S., & Swanson, H. L. (Eds). *APA educational psychology handbook, Vol 3: Application to learning and teaching*, (pp. 189-227). Washington, DC, US: American Psychological Association, viii, 668 pp. http://dx.doi.org/10.1037/13275-009

Sawyer, R. J., Graham, S., & Harris, K. R. (1992). Direct teaching, strategy instruction, and strategy instruction with explicit self-regulation: Effects on the composition skills and self-efficacy of students with learning disabilities. *Journal of Educational Psychology, 84*(3), 340–352. https://doi.org/10.1037/0022-0663.84.3.340

Schellings, G., & Van Hout-Wolters, B. (2011). Measuring strategy use with self-report instruments: theoretical and empirical considerations. *Metacognition and Learning*, *6*, 83-90. http://dx.doi.org/10.1007/s11409-011-9081-9

Schmidt, S. (2017). Replication. In M. C. Makel & J. A. Plucker (Eds.), Toward a more perfect psychology: Improving trust, accuracy, and transparency in research (pp. 233–253). American Psychological Association. https://doi.org/10.1037/0000033-015

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2), 90-100. http://dx.doi.org/10.1037/a0015108

Schoonen, R., van Gelderen, A., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language learning*, *53*(1), 165-202. http://dx.doi.org/10.1111/1467-9922.00213

Schoonen, R. (2005). Generalizability of writing scores: An application of structural equation modelling. *Language Testing, 22(1),* 1–30. https://doi.org/10.1191/0265532205lt295oa

Schunk, D. H. (1989). Social cognitive theory and self-regulated learning. In B. J. Zimmerman & D. H. Schunk, Self-regulated learning and academic achievement: Theory, research and practice (pp. 83-110). New York: Springer-Verlag.

Schunk, D. H., & Zimmerman, B. J. (Eds.). (2023). Self-regulation of learning and performance: Issues and educational applications. Taylor & Francis.

Schunk, D. H, & Zimmerman, B. J. (Eds.). (1998). Self-regulated learning: *From teaching to self-reflective practice*. New York: The Guilford Press.

Schunk, D. H., & Zimmerman, B. J. (1997). Social origins of self-regulatory competence. *Educational psychologist*, *32*(4), 195-208. http://dx.doi.org/10.1207/s15326985ep3204_1

Schunk, D. H., & Zimmerman, B. J. (Eds.). (2023). *Self-regulation of learning and performance: Issues and educational applications*. Taylor & Francis.

Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational Research, 57(2),* 149–174. http://dx.doi.org/10.3102/00346543057002149

Schunk, D. H. (1996). Goal and self-evaluative influences during children's cognitive skill learning. *American Educational Research Journal, 33,* 359-382. http://dx.doi.org/10.3102/00028312033002359

Segev-Miller, R. (2004). Writing from sources: The effect of explicit instruction on college students' processes and products. *L1-Educational Studies in Language and Literature, 4*(1), 5-33. https://doi.org/10.1023/B:ESLL.0000033847.00732.af

Shi, L. (2004). Textual borrowing in second-language writing. *Written communication*, *21*(2), 171-200. https://doi.org/10.1177/0741088303262846

Solé, I., Miras, M., Castells, N., Espino, S., & Minguela, M. (2013). Integrating information: An analysis of the processes involved and the products generated in a written synthesis task. *Written Communication*, *30*(1), 63-90. https://doi.org/10.1177/0741088312466532

Sonnenschein, S., & Whitehurst, G. J. (1984). Developing referential communication: a hierarchy of skills. *Child Development, 55,* 1936e1945. http://dx.doi.org/10.2307/1129940

Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly*, *24*(1), 7–26. http://www.jstor.org/stable/748008

Spivey, N.N. (1997). *The constructivist metaphor: Reading, writing and the making of meaning*. New York: Academic Press.

Spivey, N.N. (1997). *The constructivist metaphor.* New York: Academic Press.

Swallow, K. M., & Jiang, Y. V. (2010). The attentional boost effect: Transient increases in attention to one task enhance performance in a second task. *Cognition, 115,* 118–132. https://doi.org/10.1016/j.cognition.2009.12.003

TechSmith. (2016). [Camtasia ComputerSoftware]. http://www.techsmith.com/camtasia.html

Teng, F. (2019). A comparison of text structure and self-regulated strategy instruction for elementary school students' writing. *English Teaching: Practice & Critique*, *18*(3), 281-297. https://doi.org/10.1108/ETPC-07-2018-0070

Teng, L. S., & Zhang, L. J. (2020). Empowering learners in the second/foreign language classroom: Can self-regulated learning strategies-based writing instruction make a difference? *Journal of Second Language Writing*, *48*, 1–16. https://doi.org/10.1016/j.jslw.2019.100701

Teng, L. S. (2022). Explicit strategy-based instruction in L2 writing contexts: A perspective of self-regulated learning and formative assessment. *Assessing Writing*, *53*, 100645. https://doi.org/10.1016/j.asw.2022.100645

Thompson, P., & Tribble, C. (2001). Looking at citations: using corpora in English for academic purposes. Language Learning & Technology, 5(3), 91–105.

Tillema, M., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2012). Quantifying the quality difference between L1 and L2 essays: A rating procedure with bilingual raters and L1 and L2 benchmark essays. *Language Testing, 30*(1)*,* 71-97. http://dx.doi.org/10.1177/0265532212442647

Tillema, M. (2012). *Writing in first and second language: Empirical studies on text quality and writing processes*. Utrecht: LOT Dissertation Series.

Torrance, M., Fidalgo, R., & García, J. N. (2007). The teachability and effectiveness of cognitive self-regulation in sixth-grade writers. *Learning and instruction*, *17*(3), 265-285. https://doi.org/10.1016/j.learninstruc.2007.02.003

Torrance, M., Thomas, G. V., & Robinson, E. J. (1994). The writing strategies of graduate research students in the social sciences. Higher Education, 27(3), 379–392. https://doi.org/10.1007/BF03179901

Torrance, M., Thomas, G. V., & Robinson, E. J. (1999). Individual differences in the writing behaviour of undergraduate students. *British Journal of Educational Psychology*, *69*(2), 189-199. http://dx.doi.org/10.1348/000709999157662

Torrance, M., Thomas, G. V., & Robinson, E. J. (2000). Individual differences in undergraduate essay-writing strategies: A Longitudinal study. Higher Education, 39(2), 181–200. https://doi.org/10.1023/A:1003990432398

Turkkollu, N. (1994). The error hierarchy of Turkish EFL teachers in their assessment of compositions (Master's Thesis, Bilkent University). Retrieved from: http://hdl.handle.net/11693/17571

Ulum, O. G., & Uzun, K. (2020). Critical Perspective of English Teaching and Learning in Turkey. *International Journal of Evaluation and Research in Education*, *9*(2), 456-460. https://doi.org/10.11591/ijere.v9i2.20469

Van den Bergh, H., & Eiting, M. H. (1989). A method of estimating rater reliability. *Journal of Educational Measurement*, *26*(1), 29-40 http://dx.doi.org/10.1111/j.1745-3984.1989.tb00316.x

Van den Bergh, H., & Rijlaarsdam, G. (1996). The dynamics of composing: Modelling writing process data. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 207–232). Lawrence Erlbaum Associates, Inc.

Van den Bergh, H., Rijlaarsdam, G., & Van Steendam, E. (2016). Writing process theory: A functional dynamic approach. In C.A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research*. *Second edition*. [57-71]. New York/London: The Guilford Press.

Van der Hoeven, J. (1997). "Chapter 6 Writing Ability: The Influence of Linguistic Skills on the Writing Process". In Children's Composing. Leiden, The Netherlands: Brill. https://doi.org/10.1163/9789004484689_008

Van Weijen, D., Van den Bergh, H., Rijlaarsdam, G., & Sanders, T. (2009). L1 use during L2 writing: An empirical study of a complex phenomenon. Journal of Second language writing, 18(4), 235-250.https://doi.org/10.1016/j.jslw.2009.06.003

Vandermeulen, N., De Maeyer, S., Van Steendam, E., Lesterhuis, M., van den Bergh, H., & Rijlaarsdam, G. (2020). Mapping synthesis writing in various levels of Dutch upper-secondary education: A national baseline study on text quality, writing process and students' perspectives on writing. Pedagogische Studiën, 97(3), 187-236. https://pedagogischestudien.nl/search?identifier=717755

Vandermeulen, N., Van Steendam, E., De Maeyer, S., & Rijlaarsdam, G. (2023). Writing process feedback based on keystroke logging and comparison with exemplars: Effects on the quality and process of synthesis texts. *Written Communication*, *40*(1), 90-144. https://doi.org/10.1177/0741088322112799

Van Ockenburg, L., Van Weijen, D., & Rijlaarsdam, G. (2019). Learning to Write Synthesis Texts: A review of intervention studies. *Journal of Writing Research, 10*(3)*,* 402-4028. https://doi.org/10.17239/jowr-2019.10.03.01

Van Someren, M., Barnard, Y. F., & Sandberg, J. (1994). *The think aloud method: a practical approach to modelling cognitive. London: Academic Press*, *11*, 29-41.

Van Steendam, E., Rijlaarsdam, G., Sercu, L., & Van den Bergh, H. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction*, *20*(4), 316-327. https://doi.org/10.1016/j.learninstruc.2009.08.009

Van Steendam, E., Rijlaarsdam, G.C W., Van den Bergh, H.H., Sercu, L. (2014). The mediating effect of instruction on pair composition in L2 revision and writing. *Instructional Science*, *42*(6), 905-927. https//doi.org/10.1007/s11251-014-9318-5

Van Steendam, E., Tillema, M. (2014). Assessing text quality. Invited Keynote and hands-on session on writing assessment at 'Conceptualizing of writing: urgent issues and best solutions'. Presented at the Research School preceding SIG Writing Conference on Writing Research Amsterdam & Utrecht, Utrecht University, 25 Aug 2014-26 Aug 2014.

Van Weijen, D., Rijlaarsdam, G., & van den Bergh, H. (2019). Source use and argumentation behavior in L1 and L2 writing: A within-writer comparison. *Reading and writing*, *32*, 1635-1655. https://doi.org/10.1007/s11145-018-9842-9

Veenman, M. V. J. (2011). Learning to self-monitor and self-regulate. In R. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 197–218). New York: Routledge.

Wallace, D. L., Hayes, J. R., Hatch, J. A., Miller, W., Moser, G., & Silk, C. M. (1996). Better revision in eight minutes? Prompting first-year college writers to revise globally. Journal of Educational Psychology, 88(4), 682–688. https://doi.org/10.1037/0022-0663.88.4.682

Weigle, S. C., & Parker, K. (2012). Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing*, *21*(2), 118-133. https://doi.org/10.1016/j.jslw.2012.03.004

Wengelin, Å., Frid, J., Johansson, R., & Johansson, V. (2019). Combining keystroke logging with other methods: Towards an experimental environment for writing process research. In *Observing Writing* (pp. 30–49). Brill. https://doi.org/10.1163/9789004392526_003

Wette, R. (2010). Evaluating student learning in a university-level EAP unit on writing using sources. *Journal of Second Language Writing*, *19*(3), 158-177. https://doi.org/10.1016/j.jslw.2010.06.002

Worthen, J. B., & Hunt, R. R. (2011). *Mnemonology: Mnemonics for the 21st century.* Psychology Press. https://doi.org/10.4324/9780203834107

Yamanishi, H., Ono, M., & Hijikata, Y. (2019). Developing a scoring rubric for L2 summary writing: A hybrid approach combining analytic and holistic assessment. *Language Testing in Asia*, *9*(1), 1-22.https://doi.org/10.1186/s40468-019-0087-6

Yavuz, D. & Genc, A.B. (1998). *Flexibility of setting up a writing unit at YADIM,* Unpublished Action Research Study. Çukurova University, Adana, Turkey.

Zhang, C. (2013). Effect of instruction on ESL students' synthesis writing. *Journal of Second Language Writing, 22*(1), 51-67. https://doi.org/10.1016/j.jslw.2012.12.001

Zhang, F., Schunn, C., Li, W., & Long, M. (2020). Changes in the reliability and validity of peer assessment across the college years. *Assessment & evaluation in higher education*, *45*(8), 1073-1087. https://doi.org/10.1080/02602938.2020.1724260

Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology, 94*(4), 660–668. http://dx.doi.org/doi:10.1037//0022-0663.94.4.660

Zimmerman, B. J. (1977). Modelling. In H. Hom and P. Robinson (Eds.), *Psychological processes in early education* (pp. 37-70). New York: Academic.

Zimmerman, B. J. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 202-231). New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511527692.009

Zimmerman, B. J. (2013). From Cognitive modelling to self-regulation: A social cognitive career path. *Educational Psychologist*, *48*(3), 135-147. https://doi.org/10.1080/00461520.2013.794676

AUTHOR INDEX

APPENDICES

TABLE OF CONTENTS

CHAPTER 2

APPENDIX A

WRITING PROMPTS

1 PRETEST 2: SUMMARY TASK

Write ONE paragraph summary of the text below. Use no more than 150 words.

FORTUNE-TELLING[1]

What will the future bring: Adventure, an unexpected marriage, or a long journey? Methods of predicting the future through fortune telling have existed for centuries. Although many people dismiss these practices as superstitious, others accept them and fortune telling remains popular in many cultures. From analysing the formation of clouds to opening fortune cookies at a Chinese restaurant, people do numerous things to make predictions about the future.

Reading the palms has long been a favourite method of fortune telling. The shape of fingers, the appearance of fingernails, and in particular the lines of the palm are important features for interpretation. Some say that there are three main lines in the hand, the lifeline, heart line and head line, revealing intellectual and emotional aspects of the person.

A method of fortune telling using coffee is practiced throughout the world, especially in the Middle East. After the coffee is drunk, the fortune-teller turns the cup several times and then looks at the coffee grounds, the tiny pieces of coffee beans left in the cup. She analyses the shapes of the coffee grounds, which supposedly, have various meanings. A heart could forecast love, and a ring suggests a marriage in the future.

---

[1] *Adapted from Ready to Read More Student Book (2006).*

Dreams are also regarded as indicators of the future. Colours, images and events occurring in dreams supposedly have special significance; some believe they reflect subconscious fears, concerns, or desires. Others believe that dreams can predict future events. For example, in Chinese legends, a dream of a rainbow predicts eventual good fortune.

Whether true or not, quite a few people enjoy fortune telling as forms of entertainment all around the world.

2 POSTTEST 1: SYNTHESIS TASK

You are writing about *basic memory processes*. Expand the following main idea by summarising the most relevant information from the three extracts below. You should NOT use quotations, but you are welcome to use the underlined terms.

Write ONE paragraph of 150- 200 words using correct in-text references. The word limit does NOT include the text below.

There are several stages of memory construction.

_____
_____
_____

Extract 1

First, information must be put into memory, a step that requires *encoding*. Just as the incoming sensory information must be coded so that it can be communicated to the brain, information to be remembered must be put in a form that the memory system can accept and use. In the memory system, sensory information is put into various memory codes, which are mental representations of physical stimuli[1]. Imagine that you see a billboard that reads "Huey's Going out of Business Sale," and you want to remember it so you can take advantage of the sale later. If you encode the sounds of the words as if they had been spoken you are using acoustic encoding, and the information is represented in your memory as a sequence of sounds. If you encode the image of the letters as

[1] *Stimuli: things that cause growth or activity*

they were arranged on the sign, you are using visual encoding, and the information is represented in your memory as a picture. Finally, if you encode the fact that you saw an ad for Huey's you are using semantic encoding, and the information is represented in your memory by its general meaning. The type of encoding used can influence what is remembered. For example, semantic encoding might allow you to remember that a car was parked in your neighbors' driveway just before their house was robbed. If there was little or no encoding, you might not be able to remember the model or the color of the car.
Source: Petersen, M. (2000). Psychology: Themes and Perspectives, 5th Ed. London: Collins.

Extract 2
The second basic memory process is ***storage***, which refers to the maintenance of information over time, often over a very long time. When you find it possible to recall a vacation from many years ago, you are depending on the storage capacity of your memory.
Source: Bernstein, D. (2005). Memory and our life. *Psychology today*, 204(22), 386-390.

Extract 3
The third process, ***retrieval***, occurs when you locate information stored in memory and bring it to consciousness. Retrieving stored information such as your address or telephone number is usually so fast and effortless that it seems automatic. Only when you try to retrieve other kinds of information—such as the answer to a quiz question that you know but cannot quite recall—do you become aware of the searching process. Retrieval processes include both recall and recognition. To recall information, as on an essay test, you have to retrieve it from memory without much help. Recognition is retrieval aided by clues such as alternatives given in a multiple-choice test item. Accordingly, recognition is easier than recall.
Source: Clark-Stewart, A., & Penner, L. A. (2009). *Myths about memory recognition.* London: Longman.

3 DELAYED POSTTEST: SYNTHESIS TASK

You are writing about academic success. Expand the following main idea by summarising the most relevant information from the three extracts below. You should NOT use quotations.

Write ONE paragraph of 150- 200 words using correct in-text references. The word limit does NOT include the text below.

There are several factors that play a role in academic success.

_____

_____

_____

_____

_____

Being present in the lessons is necessary for academic success. This sounds obvious but isn't always as simple and straightforward as it might seem. This is important for several reasons. For example, it may help the professors get to know the student's name and personality. Students who attend class, and who use the opportunity to engage in discussion with the professor – during, before or after class – will begin to build a relationship with the professor. This means that when it comes to grading time, the professor will be able to put a face and a name together. Students can also make connections with other students in the class. This can be helpful in forming study groups, sharing class notes and helping each other with assignments.
Calfee, R. C., & Valencia, R. R. (1991). *Your network is your net worth*. Washington, DC: Best Dissertations.

It is very important to encourage students to stay involved in the rich life on campus. Many of the most successful students on campus are also those students who are most involved in various activities on campus. As with so many people that we know in life, those people who are doing the most are often the most successful. Students who are involved in their college's life are often more motivated and focused to do well. Athletes, student leaders, campus activists, club organizers, residence assistants, are all often some of the great students as well.
Plain, S. (2000). *The unabridged journals*. K. V. Kukil (Ed.). New York, NY: Anchor.

College provides students with independence – and the responsibility for using that independence wisely. Many classes meet only a couple of times a week – or even once a week. Many assignments are long term assignments, with the

assignment mentioned on the student calendar and then due weeks later. Students are expected to plan accordingly. Chapters may be assigned, but no time spent in class following up until several weeks later. Successful students learn to plan ahead. They learn to use their free time wisely. They break assignments down into manageable parts. They plan realistically for how much time it will take to read a chapter, write a paper, work with a group on a project, study for a test. Successful students balance coursework, work time, social time. They set deadlines and priorities for themselves. It is not wrong to say that success in university is all about good time management.

Wiener, P. (1973). *Dictionary of the history of ideas* (Vols. 1-4). New York, NY

CHAPTER 2

APPENDIX B

LEARNER REPORT[1]

What did you learn during "summary from multiple sources" lessons? It can be anything. Maybe you were surprised about certain things. Write this down in full sentences. Try to make as many sentences as possible about what you learned or what surprised you. Write down only what you <u>really</u> learned.

Examples of how to start a sentence:

General
I learned that… (Something is like)
_____
_____

I learned to… (How to do something)
_____
_____

I learned that… (Something works like…)
_____
_____

I learned how… (to do something)
_____
_____

I learned…

_____

[1] *Adapted from Groenendijk et al. (2013). Learning to be creative. The effects of observational learning on students' design products and processes Learning and Instruction. 28, 35-47.*

_____
_____

Surprises/exceptions

I learned that it is not always like…

_____
_____

What did you learn about yourself during the lesson?

I noticed that I…

_____
_____

I liked to…

_____
_____

I learnt that I….

_____
_____

What did you NOT like about the lessons:

_____
_____

Anything else you would like to add:

_____
_____

CHAPTER 3

APPENDIX A

TRAP IDEAS AND TRAMPOLINE STRATEGIES

1 TRAP IDEAS READING-AND-WRITING STRATEGIES FOR SUMMARISING

Think before reading
Read the paragraph
Ask: "What is the paragraph mostly about?"
Paraphrase the important information

Identify important details to support the main idea
Delete trivial details
Eliminate redundant details
Add a term for a list of words or concepts
Summarise

2 TRAMPOLINE STRATEGIES FOR WRITING A SYNTHESIS TEXT

Think – Three steps

Step 1: Before reading the first extract: Think about the purpose why you are given different extracts on the same topic?
After reading the second and the third extracts: What is the relationship of this extract to the previous one?
Step 2: What do you expect to learn from the extract?
Step 3: What do you already know about the general topic of the extracts?

Read the extracts
Ask - What is the main idea?
Mark the important details
Paraphrase the main idea and the important details

(Repeat TRAMP for each extract and OLINE for the whole summary
Organise the paraphrased ideas BY
Linking the ideas with appropriate linkers
Including APA
Nesting reporting verbs
Edit your summary

CHAPTER 3

APPENDIX B

TRAINING PROCEDURE ACROSS THE THREE CONDITIONS

Sessions	Conditions		
	Presentational	Modelling	Control
1	Session 1 started with the instructor initiating discussion about a controversial topic to create the context for the theme of the material at hand: *"Medicating kids on long flights"*. Students read two brief extracts taken from different sources on the topic and one strong and one weak sample of student syntheses of these two extracts. The instructor led a discussion during which the students talked about the strong and the weak points of the syntheses, and subsequently discussed the possible process steps that the writer of the good sample may have followed in order to write their text (i.e., elicitation of the task requirements). Then the	Session 1 was the same as in the presentational condition.	Session 1 was the same as in the strategy conditions (i.e., modelling & presentational) except for the introduction to TRAMPOLINE strategies. Students received the same good/weak student synthesis samples and studied them in groups to discover the requirements of the task (procedures). Teacher asked the main differences in these samples, elicited procedures from students, and asked them to complete a one-item exercise about these procedures and a number of sentence-level paraphrasing practises. On the PPT, the only declarative instructions were: "underline

	instructor introduced the TRAMPOLINE strategies mnemonic for writing an effective synthesis, with each step followed by one or two practise items to reinforce understanding. Finally, to enable memorisation, the strategies were reviewed in the form of a whole class drill.		the key points" and "paraphrase". APA in-text referencing, reporting verbs and linking words were the focus of the previous weeks, so students were also able to recognise these items on the samples and got a chance to review their previous knowledge on the introductory PPT, but instruction was not explicit and followed student discovery.
2	In Session 2, students received another synthesis task titled: *"The reasons for the increase in divorce rate",* with three brief extracts, each taken from different sources on the topic. After looking up the unknown vocabulary, students read the sources. Then, the instructor presented the TRAM strategy, the first of the three sets of strategies of the TRAMPOLINE strategy series. After the presentation by the instructor, participants practised the strategies individually (finding the main idea and the important details) for extracts 2 and 3. Finally, the instructor projected the answer key on the board and held a whole class feedback session. During the last part of Session 2, the instructor presented the sub-skills of effective paraphrasing (i.e., step P in the TRAMPOLINE mnemonics) and also the	In Session 2, students in all conditions received the same synthesis task titled: *"The reasons for the increase in divorce rate".* They were given time to read the three extracts and study the unknown vocabulary. Then, they watched the first of the three videos, where a peer modelled an actual account of a student thinking aloud while working on the first extract using the TRAM strategies. After observing the model in the video, participants emulated the strategies individually (finding the main idea and the important details) for extracts 2 and 3. Finally, the instructor projected the answer key on the board and held a whole class feedback session. During the last part of Session 2, students watched a 17-minute	In sessions 2 and 3, students worked on the same synthesis task as the students in the strategy conditions. Taking the weak and strong synthesis samples as reference tasks, they worked on the first source of the new synthesis task. The instructor asked guided questions to elicit task requirements, that is, finding the main idea, supporting ideas and the formalities of effective paraphrasing in the second session and more surface-level concerns such as reporting verbs, linkers, APA conventions in the third session. Thus, the sequence of content coverage was similar to that of the strategy conditions. Some guided questions were: "What would you include in your synthesis?" "Why did you choose that sentence?", and so on. The talk was partially improvised

	paraphrased versions of the main idea and the important details in extract 1.	video of the model paraphrasing (i.e., step P in the TRAMPOLINE mnemonic) the main idea and the important details in extract 1. In this video, the sub-skills of effective paraphrasing were shown.	with the instructor asking the next question according to the responses of the students to the previous question, so instruction was student-centred. Students read the second and the third sources and underlined the key points individually. After checking the answers, the participants paraphrased the underlined ideas in groups of four (collaborative practice). The instructor supervised the activity and showed sample paraphrased sentences, which were reported using various reporting verbs, APA conventions, combined with linkers and required some editing, which students were expected to discover and point out.
3	Session 3 started with students in groups of four, paraphrasing the main ideas and the important details in extracts 2 and 3 (subsequent collaborative practice). The instructor supervised the activity, and scaffolded paraphrasing by projecting samples of paraphrased sentences on the board. Towards the end of the session, the instructor presented the OLINE strategies.	In Session 3, in groups of four, students paraphrased the main ideas and the important details in extracts 2 and 3 (subsequent collaborative practice). The instructor supervised the activity, scaffolded paraphrasing by projecting samples of paraphrased sentences on the board. Towards the end of the session, students watched a 12-minute video of the model completing the OLINE strategies.	

4	In Session 4, students completed a new synthesis task individually with minimal support to ensure consolidation of the learning content. Students were free to refer to the TRAMPOLINE mnemonic handout and could ask for minor assistance from the instructor.	Session 4 was the same as in the presentational mode condition. The videos were not available to the students after the screening, but they were free to refer to the TRAMPOLINE mnemonic handout and could ask for minor assistance from the instructor.	Session 4 was the same as in the strategy conditions. Students were free to use their notes and in-class materials.

CHAPTER 3

APPENDIX C

MOTIVATED STRATEGIES FOR LEARNING QUESTIONNAIRE[1]

The following questions ask about your motivation for and attitudes about this class. Remember there are no right or wrong answers; just answer as accurately as possible. Use the scale below to answer the questions. If you think the statement is very true of you, circle 7; if a statement is not at all true of you, circle 1. If the statement is more or less true of you, find the number between 1 and 7 that best describes you.

1 2 3 4 5 6 7

Not at all Very true of me

1. I think I will be able to use "summary from multiple sources" in other courses in the faculty.

1 2 3 4 5 6 7

2. I believe I will successfully summarise multiple sources in future circumstances.

1 2 3 4 5 6 7

3. I'm certain I can understand the most difficult reading texts given to me for summarising.

1 2 3 4 5 6 7

4. It is important for me to learn the "summary from multiple sources" task.

1 2 3 4 5 6 7

5. I'm confident that I can learn the basic concepts taught in the writing lessons.

1 2 3 4 5 6 7

6. I'm confident that I can understand the most complex material presented by the instructor in the writing lessons.

---

[1] *Adapted from Pintrich et al. (1991)*

1 2 3 4 5 6 7

7. In a class like this, I prefer course material that is interesting even if it is difficult to learn.

1 2 3 4 5 6 7

8. I am very interested in summarisation.

1 2 3 4 5 6 7

9. I'm confident that I can do very well in the timed-writings and the midterm writing sections.

1 2 3 4 5 6

10. I expect to do well in the writing lesson.

1 2 3 4 5 6 7

11. The most satisfying thing for me in this course is trying to understand the content (the sources) as thoroughly as possible.

1 2 3 4 5 6

12. I think a summary from multiple sources is useful for me to learn.

1 2 3 4 5 6 7

13. Even if they don't guarantee a good grade, I am happy that I learnt the "summary from multiple sources" task.

1 2 3 4 5 6 7

14. I like writing.

1 2 3 4 5 6 7

15. Understanding how to write effectively is very important to me.

1 2 3 4 5 6 7

16. I'm certain I can be good at writing skills being taught in this class.

1 2 3 4 5 6 7

17. Considering the difficulty of this course, the teacher, and my skills, I think I will do well in this class.

1 2 3 4 5 6 7

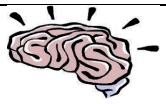
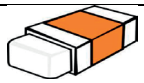
CHAPTER 3

APPENDIX D

TIME-SAMPLED SELF-REPORTING (TSSR)

You will hear 45 bleeps during your timed-writing exam. Please put a tick (✓) into the box for the activity you are engaged in at the moment when you hear the bleep sound.

If you are not sure about the number of the bleep sound, check the number that appears on the board when you hear the bleep sound.

	I am trying to understand the sources	I am para-phrasing	I am working on the sources	I am editing	I am writing my sum-mary	I am reading my sum-mary	Other	Fin-ished
								
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
...								
...								
44								

CHAPTER 3

APPENDIX E

LEARNER REPORT[1]

What did you learn during "summary from multiple sources" lessons? It can be anything. Maybe you were surprised about certain things. Write this down in full sentences. Try to make as many sentences as possible about what you learned or what surprised you. Write down only what you <u>really</u> learned.

Examples of how to start a sentence:

General

I learned that… (Something is like)

_____
_____

I learned to… (How to do something)

_____
_____

I learned that… (Something works like..)

_____
_____

I learned how… (to do something)

_____
_____

I learned…

_____

[1] *Adapted from Groenendijk et al. (2013). Learning to be creative. The effects of observational learning on students' design products and processes Learning and Instruction. 28, 35-47.*

_____

Surprises/exceptions
I learned that it is not always like…

_____
_____

What did you learn about yourself during the lesson?

I noticed that I…

_____
_____

I liked to…

_____
_____

I learnt that I….

_____
_____

What did you NOT like about the lessons:

_____
_____

Anything else you would like to add:

_____
_____

CHAPTER 3

APPENDIX F

WRITING PROMPTS

1 PRETEST SUMMARY TASK (SUM)

Write ONE paragraph summary of the text below. Use no more than 150 words.

FORTUNE-TELLING[1]

What will the future bring: Adventure, an unexpected marriage, or a long journey? Methods of predicting the future through fortune telling have existed for centuries. Although many people dismiss these practices as superstitious, others accept them and fortune telling remains popular in many cultures. From analysing the formation of clouds to opening fortune cookies at a Chinese restaurant, people do numerous things to make predictions about the future.

Reading the palms has long been a favourite method of fortune telling. The shape of fingers, the appearance of fingernails, and in particular the lines of the palm are important features for interpretation. Some say that there are three main lines in the hand, the lifeline, heart line and head line, revealing intellectual and emotional aspects of the person.

A method of fortune telling using coffee is practised throughout the world, especially in the Middle East. After the coffee is drunk, the fortune-teller turns the cup several times and then looks at the coffee grounds, the tiny pieces of coffee beans left in the cup. She analyses the shapes of the coffee grounds, which supposedly, have various meanings. A heart could forecast love, and a ring suggests a marriage in the future.

Dreams are also regarded as indicators of the future. Colours, images and events occurring in dreams supposedly have special significance; some believe they reflect subconscious fears, concerns, or desires. Others believe that dreams can predict future events. For example, in Chinese legends, a dream of a rainbow predicts eventual good fortune.

---

[1] *Adapted from Ready To Read More Student Book (2006)*

Whether true or not, quite a few people enjoy fortune telling as forms of entertainment all around the world.

2 PRETEST ARGUMENTATIVE WRITING TASK (ARG_1)

Write a well-developed essay of 300+ words answering the following question.

People tend to consume and go shopping more than ever before. What are the reasons that there is a growing increase in the rate of consumption?

3 POSTTEST ARGUMENTATIVE WRITING TASK (ARG_2)

"Our addiction to technology is causing us to become less intelligent people."

How much do you agree with this statement?

Write an argumentative essay.
You must provide at least one counterargument and refutation.
Give reasons for your answer and include any relevant examples from your own knowledge and/or experience.
Write at least 300 words.

4 POSTTEST 1. SYNTHESIS TASK (SYN_1)

You are writing about lying. Expand the following main idea by summarising the most relevant information from the three extracts below. You should NOT use quotations.
Write ONE paragraph of 100-150 words using correct in-text references. The word limit does NOT include the text below.

There are good reasons why we need to make a conscious decision not to lie._____

Extract 1
The idea of not lying is strangely controversial. Most people seem to feel lying in some circumstances is not only acceptable but desirable. If you have to lie, for example, to save someone's life or if lying turns out to be ultimately more compassionate than telling the truth, it is probably the correct course of action

to take. But most of the time most of us do not lie out of a spirit of compassion. Even when we do, we usually make the assumption that people are essentially weak and have egos that will likely collapse or at the very least be injured if they hear unpleasant feedback.
Source: Frazier, R. (2012). Psychology: Themes and Perspectives, 5th Ed. London: Collins.

Extract 2
There is a wonderful, yet not very obvious, benefit of aiming for honesty in as many circumstances as we can: it motivates us to try to become the good people lying helps us pretend we are. By facing the reasons for lying, we have the opportunity to change. Living with this intention helps us avoid doing anything we feel the need to cover up and this leads to a remarkably stress-reduced life.
Source: Miller, J. (2014). Pathological Liars. *Psychology today*,204(22), 386-390.

Extract 3
Imagine developing a reputation for honesty upon which others know they can always rely. What an invaluable resource you'd become! Most people say they want to hear the truth but in reality, they are more interested in being praised. When they do hear the truth, they learn either not to ask you for your views or that the value of hearing the truth, no matter how painful, is greater than protecting their egos because it affords them the opportunity to reflect and self-improve. Others often have a far more accurate perspective on our character flaws than we do. If we are genuinely interested in improving ourselves or our work, what we need is the truth, not flattery, even in matters that appear at first glance unimportant.
Source: Bernstein, D., & Cohen, R. A. (2009). *Myths about lying.* London: Longman.

5 POSTTEST 2. SYNTHESIS TASK (SYN_2)

You are writing about alcoholism. Expand the following main idea by summarising the most relevant information from the three extracts below. You should NOT use quotations.
Write ONE paragraph of 100 – 150 words using correct in-text references. The word limit does NOT include the text below.

There are several factors that cause alcoholism.

_____

Extract 1

According to the latest research, certain psychological factors may contribute to the development of alcoholism. These factors include: high stress and/or anxiety levels, emotional pain, low self-esteem and depression. Drinking under these circumstances is often called "self-medicating," because the person is abusing alcohol to "treat" one or more emotional and/or psychological problems. Having these psychological issues makes a person more likely to become an alcoholic but does not necessarily mean that the person is definitely going to become an alcoholic.

Brighton, A. H., & Loraine, T. F. (1990). *From monotony to monopoly*. Washington, DC: Best Dissertations.

Extract 2

The complete impact of genetics on addiction remains unknown, but scientists know enough to recommend that children, grandchildren, and siblings of alcoholics need to be careful because they are at higher risk of developing alcoholism themselves. Some studies have indicated that family members of an alcoholic are three- to four-fold more likely to develop the disease than people with no alcoholic family members. One thing behind the connection between alcoholism and heredity may be a tendency to enjoy alcohol more than other people. People who are closely related to an alcoholic, including brothers, sisters, and children, have a more positive reaction to alcoholic drinks. This may be because they have a gene that makes alcohol use more pleasurable, which could make addiction more likely.

Rice, J. (2003). *It is in my genes.* T. V. Pupil (Ed.). New York, NY: Anchor.

Extract 3

With so many genetic factors involved, it can be tempting to think of alcoholism as a purely genetic disease. However, some people with none of the known genetic factors and no family history of alcoholism still become addicted. This is because the social and environmental factors can be strong enough in some cases to cause alcoholism even though the person is at lower risk genetically. Social pressure, in the form of friends and colleagues who drink heavily, is a big factor in the development of alcoholism. A tendency to view alcohol as a form

of stress relief or as a way to become less nervous in social situations can also make alcoholism more likely, even if the genetic influences are minimal.

Fried, P. (1984). *Addiction redefined.* (Vols. 1-4). New York, NY: Scribner's.

CHAPTER 3

APPENDIX G1

THE ASSESSMENT OF THE ARGUMENTATION: STRENGTH OF THE ARGUMENTATIVE TEXTS

With regard to argumentation strength, the text below is an average text of the collection texts that you are going to assess. The score of this text is 100. The other texts should be scored in comparison with this text. For instance, if you consider a text twice as good as this one, you give it a score of 200. And if you consider the text half as good as this text, you give it a score of 50.

As the nature of world, some people have less than others. Even if someones claim that their poorness is because of their unadequate work, unemployment, corruption and unequality of chance in education are the actual reasons of poorness of majority people.

To begin with, in our country, unemployment is increasing year-by-year. It does not seen is statistic, since they work in some insufficient jobs for live even if they hate. Long work hours and salaries which are not enough to live are making people tired, and some of them quit their jobs. Also the government and rich people who in relation with government do not care about people life standarts. They just think about the "unemployment statistics". In addition, in capitalist world, some jobs are disappeared, because of the machines and neglecting. It is another reason of unemployment and poornes.

Secondly, corruption is important fact that should be seen about poorness in our country. For example, few years before, workers of Tekel are left alone with their fate. Government said, if you want you can work for me in worse conditions and very lower salary, or you can quit. It is the favour of governments, if you close your minds. In spite of protests while few months nothing is changed. On the other hand, in choosing new colleagues, there is no justice in my country. If you are not familiar of boss or manager. your chance is close to imposible for job in local compa

nies. It also happens in big companies but thanks to system
it has a title and it is not guilt: reference. Those
corruptions lead to increase tho poorness of people.

Opposing of my view, somebody can say that the
people who have not income for live are taking they
deserved, because if they graduate high school, they
will not be poor. However, it should be seen the unequ
al financial resources for education. For graduate a high school
in different city is high cost, and everyone can
not pay this money. Also there are few city that
have remarkable universities. So, poorness is main reason
of poorness of poor people.

Consequently, insufficient possible jobs, injustices and
inaduquate possibility of education are the reasons of
poorness, not less labour, and we have the chance
for change this situation.

When assessing the texts, you focus on the following: *does the argumentation in the text convince the reader?* With regard to that, the text above has the following strong and weak points.

Strong points:

- The writer wants to convince the reader that unemployment, corruption, and unequal educational opportunities are causes of poverty. Therefore, he gives at least two arguments per cause.
- The writer gives a clear example to convince the reader that corruption is a cause of poverty in his home country. He discusses the consequences of the corrupt acts of the government.
- The writer accounts for possible counter arguments that the reader might have. He mentions one counter argument: *People should blame themselves for their poorness. The reason for their low salary is their low educational level.* The writer refutes that counter argument with an explanation why we may assume that people cannot be blamed for their low educational level.
- The reader would probably agree with the writer that corruption and unequal educational opportunities are causes of poverty. The writer supports that opinion with relevant and adequate arguments.
- Weak points:

- The writer does not agree with the issue that people are poor because they do not work enough. However, he does not give arguments to support that opinion. He just gives arguments for his opinion that there are three different causes for poverty. Thus, he does not address the issue about which they had to write an argumentative text, but another one.
- After reading the argumentative text, the reader probably has not been convinced that unemployment causes poverty. The writer states that the figures prove the contrary. Consequently, he tries to prove that the figures are incorrect. Taking into consideration the arguments that he uses therefore, it seems that the writer does not exactly know what unemployment is. Poor people seem to work a lot instead of not at all. If they therefore resign, they are responsible for their unemployment; the government is not.
- The reader probably does not understand why the fact that the government does not worry about the life conditions of the citizens proves that unemployment causes poverty. It is rather another cause of the poverty problem: the government does not take steps to solve the poverty problem.
- The reader probably does not agree with some of the writer's conclusions. Is capitalism really the cause of industrialization? (last sentence paragraph 2) And is it really natural that one person has less than another? (first sentence paragraph 1).

CHAPTER 3

APPENDIX G2

ARGUMENTATION BENCHMARK SCALING

	ORGANISATION	CONTENT	GRAMMAR & PUNCTUATION	LEXIS
10	Introduction begins with a hook or general statement that grabs attention. Introduction successfully narrows down to the thesis. Introduction ends with an explicit thesis statement with a clear stance. Body paragraphs contain clear topic sentences, elaborate on the thesis, are appropriate length, and are well-connected with transition words. Conclusion summarises the main points in the body or restates the thesis, and finishes with a concluding remark	Fully addresses the question at hand. All main points are elaborated and explained thoroughly with sufficient supporting details that provide full reasoning and exemplification. Paragraphs are very clear, coherent, and unified.	Skilful command of language with almost no grammatical errors. Level appropriate and varied sentence structure. Almost impeccable use of punctuation and capitalization	Sophisticated range of level appropriate vocabulary. Almost no word formation errors and almost impeccable spelling.

8	Introduction has a hook or general statement but may not successfully connect to the thesis statement. Introduction somewhat successfully narrows down to the thesis. Introduction has a clear thesis statement, previewing supporting ideas. Body paragraphs have satisfactory topic sentences and elaborate on the thesis statement with sufficient use of transitional signals. Conclusion summarises the main points, but might have repeated the thesis word-for-word	Sufficiently addresses the question at hand. Presents a developed and sufficient argument. Main points are supported with information that provide adequate reasoning and exemplification. Paragraphs are clear, coherent, and unified.	Good command of language with minor grammatical errors that do not impede understanding. Level appropriate sentence structure and adequate range. Good use of punctuation and capitalization.	Sufficient range of level appropriate vocabulary Few word formation errors with mostly accurate spelling.
6	The hook or general statements do not lead to the thesis statement / narrowing down not successful / may start too general or too specific. Attempt to create a thesis statement but may be unclear or may not pose a stance. Topic sentences are unclear/weak or not well connected to the thesis. Body paragraphs are too short or not divided proportionately or not well-	Somewhat responds to the question at hand. There may be more than one central argument / some supporting ideas may be irrelevant. Content may have inadequate or excessive information or examples. Repetition of ideas either in the same paragraph or other paragraphs. Some effort may have been made to write coherently and clearly.	Some structures are accurate but sentence structure errors predominate / many minor errors that at times confuse. Only a limited range of level appropriate sentence structure/ attempts to use level-appropriate sentence structures with some mistakes. Limited command of punctuation and capitalization.	Somewhat sufficient range of level appropriate vocabulary. Some major word formation errors and spelling errors that do not impede understanding.

	connected with insufficient use of transition signals. The arguments are somewhat reviewed in the conclusion / a new idea might be introduced.			
4	There is an introduction but there is no hook or general statement, or general statements leading to thesis are irrelevant or non-existent. There is a thesis statement, but it is vague, or weak. Topic sentences are non-existent, or they are contradictory to the thesis and /or they do not correspond to the thesis. Disproportionate paragraphs and use of only simple transition signals. The conclusion lacks a summary of the body / the arguments are not reviewed.	Barely responds to the question at hand. Main argument may be too vague, weak, or underdeveloped/ Several arguments may have been made, but no central idea is in focus. Presents inadequate information with little or no supporting details. Limited clarity, coherence, or unity	Weak command of language with many grammatical errors so much as to hinder comprehension Sentence structures below level expectations / only simple sentences Use of punctuation and capitalization below level expectations.	Limited range of level appropriate vocabulary. Frequent errors of word forms that confuse meaning with many spelling errors.
2	Produces a simple written text (not in essay form) that lacks cohesion. Inappropriate paragraphing, no thesis statement, no conclusion.	Fails to respond to the question at hand. Produces a simple written text that shows minimal coverage of the assignment/task. No consistency, no unity. Not enough ideas or information to support ideas.	So many grammatical errors that comprehension is impossible.	Range and accuracy of lexis fall significantly short with too many errors in word formation and spelling.
0	NOT ENOUGH OF A SAMPLE TO GRADE			

CHAPTER 3

APPENDIX G3

THE ASSESSMENT OF THE GRAMMAR IN THE ARGUMENTATIVE TEXTS

With regard to grammar, the text below is an average text of the collection texts that you are going to assess. The score of this text is 100. The other texts should be scored in comparison with this text. For instance, if you consider a text twice as good as this one, you give it a score of 200. And if you consider the text half as good as this text, you give it a score of 50.

Nowadays, many countries have poverty problem, except big countries. If countries which have poverty problem want to be development nations, they should solve this problem. In my country, people have some poverty problem because poor people do not work sufficient.

First of all, if poor people do not want do live in poverty, they souhld work hard However, in Turkey, many people want do be rich but they do not work enough. They always meet friends, go for shopping, visit another countries. Therefore, this kind of Turkish people still poor. For the sake of example, in my country many teacher who graduated University now, they do not want to be assistant teacher they want to directly be teacher. These people, firstly, should be assistant teacher and work hard, after this situation, they want to be teacher. On the other hand, in my country, poor do not work suffic ient moreover, they blame to goverment for poverty. Many poor people who are unemployer, they do not like another job, all of them want to quality job and high maste , therefore, they are still unemployer. If they choose another job and work hard, they will have a comfortable and quality life standart. Moreover, in Turkey, according to survey(1997) , Turkey have a big poverty problem but today this situation is change. Since the rate of poor people is decrease in last decate. Therefore, people shouldn't blame to goverment for poverty. It is only related to poor people characteristics at the work time. In brief, Turkey has some poor people because poor people do not work hard.

To conclusion, in my home country people are poor do not work enough, thus, they are poor. However, in last decate, Turkish people life standart İncrease and number of poor people decrease. In my Perspective, poor people must work hard and enough For being a rich person.

When you score the text for grammar, you take into account the complexity of the grammatical constructions that are used and the correctness of the grammar. With regard to that, the text above has the following strong and weak points.

Strong points:
- The different word phrases are positioned correctly in most sentences. The inflected verb has generally been positioned at the second place in the sentence. It is preceded by a subject and followed by the objects.
- The writer has used a few single sentences, mainly multiple sentences. These multiple sentences are either two coordinate clauses or two subordinate clauses.
- Sometimes a clause has been embedded in a main clause to further specify the subject of the main clause: *many teacher who graduated University* en *Many poor people who are unemployer*.
- Sometimes the writer has used a gerund.
- The writer has varied in the structure of the sentences.
- The author has usually inflected the verbs correctly.
- The writer has used frequently and adequately modal verbs to indicate how certain he is about something or how desirable he considers it.
- When the writer wanted to deny something, he has done it in the right way: with a form of the verb *to do + not + an uninflected verb*, or with the modal verb *should + not + an uninflected verb*.
- The writer has correctly used collocations that frequently occur in English texts, for example, *want to*.
- The author has usually applied the English punctuation rules correctly. At the beginning of the sentence, he has used a capital letter. At the end of the sentence, he has produced a dot. Main and subordinate clauses have usually been separated with a comma. When a sentence is preceded by an adverbial expression, a comma has been produced before the sentence has been written.

Weak Points
- The writer has not used infinite clauses.
- When a clause has been embedded in the main clause to further specify the subject, the writer has not started the sentence with an inflected verb, but with a repetition of the subject of the main clause. In the following sentence, the writer should not have used *they*: *in my country many*

teacher who graduated University now, they do not want to be assistant teacher.

- The writer has not always used a gerund in the right cases. In line 12, a gerund could be used (when *for* is deleted). On the other hand, the writer could better have used an infinite clause with *to be* than the expression *for being* in line 40.
- The writer has never used a verb in the progressive tense, although this seems to have been necessary in some cases. For example, in line 38, it should have been *changing* and *decreasing*.
- The writer has not used the possessive suffix. For instance, it should have been *poor people's characteristics* in line 31 instead of *poor people characteristics.*
- The writer often has used an adjective to modify a verb instead of an adverb. He has often used expressions like: *work sufficient* and *work hard*.
- Sometimes there is no agreement between the subject and the main verb, for example in line 37 and 38.
- The writer has not used articles correctly. In line 1 and 2, they are missing. In line 12, it should have been *other* instead of *another*, because it is a plural noun.
- The writer has not correctly used some English collocations. The verb *blame* should not have been followed by the preposition *to*, and *for* should not have been used in the expression *go for shopping*.
- The author has made some punctuation errors. After some sentences, no dot has been produced. A point instead of a comma separates some main and subordinate clauses. Sometimes a comma is missing after an adverbial expression.

CHAPTER 3

APPENDIX G4

THE ASSESSMENT OF THE LEXIS IN THE ARGUMENTATIVE TEXTS

With regard to lexis, the text below is an average text of the collection texts that you are going to assess. The score of this text is 100. The other texts should be scored in comparison with this text. For instance, if you consider a text twice as good as this one, you give it a score of 200. And if you consider the text half as good as this text, you give it a score of 50.

Poverty

All around the world, there are a lot of poor people It is demonstrated that they don't have job to earn money enoughly. However, it is decreasing in the developed country Also, there are a lot of poor in every countries because there is no helping programs for them. Job opportunities and helping programs are the leading reasons for poverty.

One of the most significant reason for poverty is Job opportunities. The most relevant point to the issue is unemployment when the rate of unemployment is peak of the top, there are a lot of poor people in that country since the people can't find a job to earn money. Some surveys show that if there are many poor people in a country, there are some unemployment issues and poverty problems in that country. Moreover, equally related point to the issue is that people can't find a job easily. Imagine that a country has some problems that there is no job enoughly. This is a government problem because the government must create job opportunities for people so this is a crisis both people and that country Therefore, the rate of poverty can increase in that country. ıt is believed that there are poverty in every countries but the rate of poverty shouldn't increase. To sum up, the government should create some job opportunities for people and the government should promote poor people.

Second of the most important reason for poverty is helping problems. The most relevant point to the issue is individual helping problems. The people should be aware of their responsibility for the poor people. To illustrate the people can give small job or helping materials like bread, post every day. If people promote to the poor people, it can be solved by people. It is believed that this is a individual helping way to solve to poverty. They can provide them a small job to work by this way. Furthermore, the equally related point to the issue is that the developed countries and big companies should create some helping programs. For instance, they can provide job opportunities or helping programs. Some researchers, suggest that the big companies can help to have job for poor people. It can be a good way. In brief, the big companies should create some helping programs.

All in all, there a lot of poor people in every country, they don't work enoughly becaues of some problems. As far as I am concerted, I agree with this idea because poor people can't work enoughly because of unemployment, job opportunities and some government problems.

When assessing the lexis in the texts, you should take into account the complexity of the words that are used in the text, the variation in word choice and the extent to which the used words enhance the comprehension of the text. With regard to that, the text above has the following strong and weak points.

Strong points:
- The word choice enhances the comprehension of the text. The reader gets to know which message the writer wants to express, since he has chosen the right words to express it.
- Almost all words are correctly spelled. This enhances the text comprehension. The reader can immediately try to determine the meaning of the used words. He does not need to determine first which word was actually meant.
- Nouns are often correctly inflected. Therefore, the reader knows whether one or two of the identified things are meant.
- The writer does not only use English words for concrete things, but also for abstract things like *unemployment, poverty, responsibility* and *government*.
- The writer correctly uses some common English expressions, like *a lot of* and *all in all*.
- The writer correctly uses a few collocations, like *be aware of* and *related point to*.

Weak points:
- Sometimes, it is difficult to determine what the writer means, since he has not chosen the right word to express his message. This can be a consequence of his wish to use sophisticated words. In line 25 and 31, he could have better used the less sophisticated word *help* instead of *promote*. In addition, the expression *peak of the top* is inappropriate in line 11/12.
- Some parts of the text are difficult to understand, because the writer has forgotten some words. This is often a consequence of insufficient grammatical knowledge. For instance, the writer does not seem to know that an article should precede singular nouns. In some cases, the writer has forgotten a word, which is crucial for the expression of the intended meaning. For example, the word *people* is missing in line 6 and *are* in line 42.
- Although the writer uses English words for abstract things, he does not seem to know many English terms with an abstract meaning. He uses words most frequently with a concrete meaning and repeats the abstract words he knows very often.

- The writer does not vary much in the words he uses. He does not only repeat some words very often, but also some expressions, like *The most relevant point to the issue* en *equally related point to the issue*.

CHAPTER 3

APPENDIX G5

THE ASSESSMENT OF THE ORGANISATION OF THE ARGUMENTATIVE TEXTS

With regard to organisation, the text below is an average text of the collection texts that you are going to assess. The score of this text is 100. The other texts should be scored in comparison with this text. For instance, if you consider a text twice as good as this one, you give it a score of 200. And if you consider the text half as good as this text, you give it a score of 50.

Nowadays poverty is the most significant
İssiu around the world. It has lots of
reasons to country. In my home country
there are many poor people and some of
them do not work enough some of them
are working hardly but still live in poverty.

My first point is that in my home
country, there are lots of work
opportunities but many people poor due to
the fact that people don't want
each job. They want always best job. To
illustrate, a man not studying University
want to work like people graduated
University. He just want to sit in front of
the computer and answer phones. Hence,
he can not find a job and become poor.
Moreover, some man in my home country
earn low money and it is not enough
for their family. Maybe, if the men let
their wife to work, they can have
better life quality but some men don't let.
To illustrate, when two people work hardly in
a family, their cost will be double and
they won't live in poverty.

On the other hand, some people in my
home country are working hardly but they
still poor. There are two main reason for this
situation, first one is having many family
members and second one is living big
city. To illustrate, some people in my home country
work hardly and earn lots of money; however,
when they have lots of children, aspecially
young children, this money is not enough
for this family. The other example is
that some people work hardly but they are
living a big city. They earn money, so
their money is not enough to don't live
in poverty.

As a result, in my home country
there are hardworking poor people and
not working poor people.

When assessing the texts, you focus on the following: *is the argumentation in this text presented in a logical order to the reader?* With regard to that, the text above has the following strong and weak points.

Strong points:

- The reader recognises the different parts of the argumentative text, since the writer has split the text in paragraphs: an introductory paragraph, two body paragraphs and a concluding paragraph.
- In the beginning of the text, the writer creates a common ground with the reader. He starts with mentioning something everybody presumably knows: poverty is a major problem in the world. Subsequently, he takes that common ground as a starting point for his argumentative text in which he presents information the reader was probably ignorant of.
- The introduction narrows down to the thesis statement, on which is elaborated in the body paragraphs. In the thesis statement, the main topics of both body paragraphs are mentioned. After reading the thesis statement, the reader consequently knows what is discussed in the rest of the text.
- In the main part of the text, the reader is logically informed why the writer, on the one hand, thinks that the poverty problem can be solved when people work harder, and why he, on the other hand, thinks that the problem cannot be solved in that way. In every body paragraph, he elaborates on one of two sides of his opinion.
- The reader can understand the argumentation in the body paragraphs, since it is presented in a logical order. In the beginning of the paragraphs, general statements are made. These are substantiated with specific examples in the remainder of the paragraphs.
- The main idea of a body paragraph is supported with several arguments. The reader is correctly informed how the arguments are related to each other with connectives.
- In the second body paragraph, two main arguments are given to support the idea that the poorness of some people is not a consequence of their laziness. Both arguments are already introduced in the beginning of the paragraph. The reader consequently knows what is discussed in the rest of the paragraph.
- At the end of the text, the reader gets to know which conclusion can be drawn from the argumentation in the main part of the text. The conclusion is a reformulation of the thesis statement.

Weak points:

- The reader presumably would not characterise this text as an argumentative essay. The reader would rather think that the writer wants to give an explanation of the poverty problem in his text. In the introductory as well as the concluding paragraph, the writer presents the arguments for his opinion as explanations for the poverty problem. A clear thesis statement is missing.
- The introduction just grabs a reader's attention, when the reader is very interested in the poverty problem and its causes.
- The first sentence of the second paragraph does not function as a topic sentence of the whole paragraph. The sentence introduces in general terms the ideas that are presented in the first part of the paragraph; it does not introduce the ideas of the second part of the paragraph.
- The writer uses many connectives to draw relationships between the information in the body paragraphs. Sometimes the wrong connective has been chosen. This could confuse the reader. For instance, in the last sentence of paragraph two, a subordinate argument is given for the argument in the preceding sentence, rather than an illustration.
- It would have been reader friendly to repeat the main ideas of the body in the concluding paragraph. Then, the reader can decide whether he thinks that the writer's conclusion can be drawn based on the information in the text.

CHAPTER 3

APPENDIX G6

THE ASSESSMENT OF THE COMPREHENSIBILITY OF THE SUMMARIES

With regard to comprehensibility, the text below is an average text of the collection of texts that you are going to assess. The score of this text is 100. The other texts should be scored in comparison with this text. For instance, if you consider a text twice as good as this one, you give it a score of 200. And if you consider the text half as good as this text, you give it a score of 50.

First, some people dismiss the fortune-telling predictions but many people do this prediction numerous to see the future. There are some way to prediction for the future. The first way is reading the palms. There are main lines in the hand. The can predict when they look at your palm. The other method is using the coffee. This method is favorite in the Middle East. After the coffee is drunk, the fortune teller try to predict looking your coffee glass about your future. The other method is dreams. This is also favorite in China. Colors, images and events are so important to comment about your future. In conclusion, the fortune-teller is so common in some areas to predict about the future or to enjoy

When assessing the texts, you focus on the following: *does the reader know and understand the traditions of fortune telling and people's attitude towards it after reading the text?* With regard to that, the text above has the following strong and weak points.

Strong points:

- The reader learns what the three main ways of fortune telling are: the analysis of the palms of one's hand, the analysis of the residues of coffee in a coffee cup and the analysis of colours, images and events in dreams.
- After reading the text, the reader knows roughly what the different methods of fortune telling entail.
- The reader realizes how the methods of future prediction relate to each other and to the subject of the text: Even now, future predictions are done

quite often. There are several methods of fortune telling. Three of them are listed in the text.
- The reader gets to know how a part of the world population perceives fortune telling: as bad.
- The reader learns that the future is frequently predicted in some parts of the world.

Weak points:
- The reader has to read the description of a method of fortune telling completely in order to know what the method entails. The sentences in which a method of fortune telling is introduced are so short that a reader cannot determine what a method of fortune telling entails based on these sentences alone. For instance, the reader knows that one method has something to do with dreams, after having read the following sentence: *The other method is dreams*. Yet, it is unclear what these dreams can reveal about the future, and how they can reveal that.
- The reader does not get to know how the future can be predicted when studying the lines in one's hand palm, when looking at the coffee cup from which one has been drinking or when analysing one's dreams. The reader does not know which meaning is assigned to these things and how that is done.
- The reader may assume that a part of the world population perceives fortune telling positively. Otherwise it would not have been done that frequently to get to know the future or to have fun. However, in the text, it is not explicitly stated that some people perceive fortune telling positively.
- After having read the summary, the reader does not draw the conclusion that is presented in the final sentence. That sentence contains new information. Consequently, the writer should have used another connective than 'in conclusion'.
- The reader has to have read the text entirely, before he knows what the main topic and main aim of the text is: informing about the fortune telling traditions and people's attitude towards it. It would have been easier to understand the text, when the main topic was introduced in the opening sentence and was maybe even repeated in the closing sentence.

CHAPTER 3

APPENDIX G7

THE ASSESSMENT OF CONTENT IN THE SUMMARY TEXTS: "FORTUNE-TELLING"

FORTUNE-TELLING

What will the future bring: Adventure, an unexpected marriage, or a long journey? Methods of predicting the future through fortune telling have existed for centuries. Although many people dismiss these practices as superstitious, others accept them and fortune telling remains popular in many cultures. From analysing the formation of clouds to opening fortune cookies at a Chinese restaurant, people do numerous things to make predictions about the future.

Reading the palms has long been a favorite method of fortune telling. The shape of fingers, the appearance of fingernails, and in particular the lines of the palm *[at least one of these should be mentioned]* are important features for interpretation. Some say that there are three main lines in the hand, the lifeline, heart line and head line, revealing intellectual and emotional aspects of the person.

A method of fortune telling using coffee is practiced throughout the world, especially in the Middle East. After the coffee is drunk, the fortune-teller turns the cup several times and then looks at the coffee grounds, the tiny pieces of coffee beans left in the cup. She analyses the shapes of the coffee grounds, which supposedly, have various meanings. A heart could forecast love, and a ring suggests a marriage in the future

Dreams are also regarded as indicators of the future. Colors, images and events *[at least one of these should be mentioned]* occurring in dreams supposedly have special significance; some believe they reflect subconscious fears, concerns, or desires. Others believe that dreams can predict future events. For example, in Chinese legends, a dream of a rainbow predicts eventual good fortune.

Whether true or not, quite a few people enjoy fortune telling as forms of entertainment all around the world.

Main ideas:

- Methods of predicting the future through fortune telling have existed for centuries.
- Although many people dismiss these practices as superstitious, others accept them
- Fortune telling remains popular in many cultures.
- Quite a few people enjoy fortune telling as forms of entertainment all around the world.

Supporting ideas:

- Reading the palms has … been a favorite method of fortune telling.
- A method of fortune telling using the coffee is practiced throughout the world,
- She analyses the shapes of the coffee grounds, which supposedly, have various meanings.
- Dreams are also regarded as indicators of the future.
- Colours, images and events *[at least one of these should be mentioned]* occurring in dreams supposedly have special significance;

Examples:

- The shape of fingers, the appearance of fingernails, and in particular the lines of the palm [at least one of these should be mentioned] are important features for interpretation.
- After the coffee is drunk, the fortune-teller turns the cup several times and then looks at the coffee grounds,
- Some believe they reflect subconscious fears, concerns, or desires.
- Others believe that dreams can predict future events.

CHAPTER 3

APPENDIX G8

THE ASSESSMENT OF CONTENT IN THE SYNTHESIS TEXTS: "BASIC MEMORY PROCESSES"

STUDY 1, TEST 1

Extract 1

First, information must be put into memory, a step that requires *encoding*. Just as the incoming sensory information must be coded so that it can be communicated to the brain, information to be remembered must be put in a form that the memory system can accept and use. In the memory system, sensory information is put into various memory codes, which are mental representations of physical stimuli. Imagine that you see a billboard that reads "Huey's Going out of Business Sale," and you want to remember it so you can take advantage of the sale later. If you encode the sounds of the words as if they had been spoken you are using acoustic encoding, and the information is represented in your memory as a sequence of sounds. If you encode the image of the letters as they were arranged on the sign, you are using visual encoding, and the information is represented in your memory as a picture. Finally, if you encode the fact that you saw an ad for Huey's you are using semantic encoding, and the information is represented in your memory by its general meaning. The type of encoding used can influence what is remembered. For example, semantic encoding might allow you to remember that a car was parked in your neighbors' driveway just before their house was robbed. If there was little or no encoding, you might not be able to remember the model or the color of the car.

Source: Petersen, M. (2000). Psychology: Themes and Perspectives, 5th Ed. London: Collins.

Main ideas:

- Information must be put into memory
- a step that requires *encoding*... information to be remembered must be put in a form that the memory system can accept and use.

Supporting ideas:

- sensory information is put into various memory codes, which are mental representations of physical stimuli.
- as if they had been spoken you are using acoustic encoding, and the information is represented in your memory as a sequence of sounds
- the letters as they were arranged on the sign, you are using visual encoding ... is represented in your memory as a picture.

- semantic encoding, and the information is represented in your memory by its general meaning.

Examples:

- The type of encoding used can influence what is remembered.

Extract 2

The second basic memory process is *storage*, which refers to the maintenance of information over time,

often over a very long time. When you find it possible to recall a vacation from many years ago, you are depending on the storage capacity of your memory.

Source: Bernstein, D. (2005). Memory and our life. *Psychology today*,204(22), 386-390.

Main ideas:

- The second basic memory process is *storage*, which refers to the maintenance of information over time, often over a very long time.

Supporting ideas:

- When you find it possible to recall a vacation from many years ago, you are depending on the storage capacity of your memory.

Extract 3

The third process, *retrieval*, occurs when you locate information stored in memory and bring it to consciousness. Retrieving stored information such as your address or telephone number is usually so fast and effortless that it seems automatic. Only when you try to retrieve other kinds of information—such as the answer to a quiz question that you know but cannot quite recall—do you become aware of the searching process. Retrieval processes include both recall and recognition. To recall information, as on an essay test, you have to retrieve it from memory without much help. Recognition is retrieval aided by clues- such as alternatives given in a multiple-choice test item. Accordingly, recognition is easier than recall.

Source: Clark-Stewart, A., & Penner, L. A. (2009). *Myths about memory recognition*. London: Longman.

Main ideas:

- The third process, *retrieval*, occurs when you locate information stored in memory and bring it to consciousness.

Supporting ideas:

- Retrieval processes include both recall and recognition.
- To recall … retrieve it from memory without much help.
- Recognition is retrieval aided by clues
- Accordingly, recognition is easier than recall.

Examples

- your address or telephone number … automatic
- essay test
- a multiple-choice test item

CHAPTER 3

APPENDIX G9

THE ASSESSMENT OF CONTENT IN THE SYNTHESIS TEXTS: "LYING"

STUDY 1, SECOND TEST

Extract 1

The idea of not lying is strangely controversial. Most people seem to feel lying in some circumstances is not only acceptable but desirable. If you have to lie, for example, to save someone's life or if lying turns out to be ultimately more compassionate than telling the truth, it is probably the correct course of action to take. But most of the time most of us do not lie out of a spirit of compassion. Even when we do, we usually make the assumption that people are essentially weak and have egos that will likely collapse or at the very least be injured if they hear unpleasant feedback.

Source: Frazier, R. (2012). Psychology: Themes and Perspectives, 5th Ed. London: Collins.

Main ideas:

- Most of the time, people do not lie out of a spirit of compassion.

Supporting ideas:

- Lying is probably the correct course of action to take to save someone's life, or if lying turns out to be ultimately more compassionate than telling the truth.
- (People are) weak and have egos that will likely collapse or at the very least be injured if they hear unpleasant feedback.

Extract 2

There is a wonderful, yet not very obvious, benefit of aiming for honesty in as many circumstances as we can: it motivates us to try to become the good people lying helps us pretend we are. By facing the reasons for lying, we have the opportunity to change. Living with this intention helps us avoid doing anything we feel the need to cover up and this leads to a remarkably stress-reduced life.

Source: Miller, J. (2014). Pathological Liars. *Psychology today*,204(22), 386-390.

Main ideas:

- Honesty motivates us to try to become the good people lying helps us pretend we are.

Supporting ideas:

- The opportunity to change.
- The need to cover up.
- Stress-reduced life.

Extract 3

Imagine developing a reputation for honesty upon which others know they can always rely. What an invaluable resource you'd become! Most people say they want to hear the truth but in reality they are more interested in being praised. When they do hear the truth, they learn either not to ask you for your views or that the value of hearing the truth, no matter how painful, is greater than protecting their egos because it affords them the opportunity to reflect and self-improve. Others often have a far more accurate perspective on our character flaws than we do. If we are genuinely interested in improving ourselves or our work, what we need is the truth, not flattery, even in matters that appear at first glance unimportant.

Source: Bernstein, D., & Cohen, R. A. (2009). *Myths about lying*. London: Longman.

Main ideas:

- If we are genuinely interested in improving ourselves or our work, what we need is the truth, not flattery,

Supporting ideas:

- Most people say they want to hear the truth but in reality, they are more interested in being praised.
- it affords them the opportunity to reflect and self-improve.
- Others often have a far more accurate perspective on our character flaws than we do.

CHAPTER 3

APPENDIX G10

THE ASSESSMENT OF CONTENT IN THE SYNTHESIS TEXTS: "FACTORS THAT PLAY A ROLE IN ACADEMIC SUCCESS"

STUDY 2, TEST 1

Extract 1

Being present in the lessons is necessary for academic success. This sounds obvious, but isn't always as simple and straightforward as it might seem. This is important for several reasons. For example, it may help the professors get to know the student's name and personality. Students who attend class, and who use the opportunity to engage in discussion with the professor – during, before or after class – will begin to build a relationship with the professor. This means that when it comes to grading time, the professor will be able to put a face and a name together. Students can also make connections with other students in the class. This can be helpful in forming study groups, sharing class notes and helping each other with assignments.

Calfee, R. C., & Valencia, R. R. (1991). *Your network is your net worth*. Washington, DC: Best Dissertations.

Main ideas:
- Being present in the lessons is necessary for academic success.

Supporting ideas:
- build a relationship with the professor
- make connections with other students in the class.

Examples:
- grading time
- forming study groups, sharing class notes and helping each other with assignments *[at least one of these should be mentioned]*

Extract 2

It is very important to encourage students to stay involved in the rich life on campus. Many of the most successful students on campus are also those students who are most involved in various activities on campus. As with so many people that we know in life, those people who are doing the most are often the most successful. Students who are involved in their college's life are often more motivated and focused to do well. Athletes, student leaders, campus activists, club organizers, residence assistants, are all often some of the great students as well.

Plain, S. (2000). *The unabridged journals.* K. V. Kukil (Ed.). New York, NY: Anchor.

Main ideas:
- It is very important to encourage students to stay involved in the rich life on campus.

Supporting ideas:
- People who are doing the most are often the most successful.
- Students who are involved in their college's life are often more motivated and focused to do well.

Examples:
- Athletes, student leaders, campus activists, club organizers, residence assistants

Extract 3

College provides students with independence – and the responsibility for using that independence wisely. Many classes meet only a couple of times a week – or even once a week. Many assignments are long term assignments, with the assignment mentioned on the student calendar and then due weeks later. Students are expected to plan accordingly. Chapters may be assigned, but no time spent in class following up until several weeks later. Successful students learn to plan ahead. They learn to use their free time wisely. They break assignments down into manageable parts. They plan realistically for how much time it will take to read a chapter, write a paper, work with a group on a project, study for a test. Successful students balance coursework, work time, social time. They set deadlines and priorities for themselves. It is not wrong to say that success in university is all about good time management.

Wiener, P. (1973). *Dictionary of the history of ideas* (Vols. 1-4). New York, NY: Scribner's.

Main ideas:
- It is not wrong to say that success in university is all about good time management.

Supporting ideas:
- College provides students with independence
- – and the responsibility for using that independence wisely.
- Successful students learn to plan ahead.
- They learn to use their free time wisely.

Examples:
- Many assignments are long term assignments, with the assignment mentioned on the student calendar and then due weeks later.
- break assignments down into manageable parts.
- plan realistically for how much time it will take to read a chapter, write a paper, work with a group on a project, study for a test

CHAPTER 3

APPENDIX G11

THE ASSESSMENT OF CONTENT IN THE SYNTHESIS TEXTS: "ALCOHOLISM"

STUDY 2, TEST 2

Extract 1

According to the latest research, certain psychological factors may contribute to the development of alcoholism. These factors include: high stress and/or anxiety levels, emotional pain, low self-esteem and depression. Drinking under these circumstances is often called "self-medicating," because the person is abusing alcohol to "treat" one or more emotional and/or psychological problems. Having these psychological issues makes a person more likely to become an alcoholic, but does not necessarily mean that the person is definitely going to become an alcoholic.

Brighton, A. H., & Loraine, T. F. (1990). *From monotony to monopoly*. Washington, DC: Best Dissertations.

Main ideas:

- Certain psychological factors may contribute to the development of alcoholism.

Supporting ideas:

- High stress and/or anxiety levels, emotional pain, low self-esteem and depression.
- "self-medicating", because the person is abusing alcohol to "treat' one or more emotional and/or psychological problems.

Extract 2

The complete impact of genetics on addiction remains unknown, but scientists know enough to recommend that children, grandchildren, and siblings of alcoholics need to be careful because they are at higher risk of developing alcoholism themselves. Some studies have indicated that family members of an alcoholic are three- to four-fold more likely to develop the disease than people with no alcoholic family members. One thing behind the connection between alcoholism and heredity may be a tendency to enjoy alcohol more than other people. People who are closely related to an alcoholic, including brothers, sisters, and children, have a more positive reaction to alcoholic drinks. This may be because they have a gene that makes alcohol use more pleasurable, which could make addiction more likely.

Rice, J. (2003). *It is in my genes.* T. V. Pupil (Ed.). New York, NY: Anchor.

Main ideas:

- Because of genetics, children, grandchildren, and siblings need to be careful because they are at higher risk of developing alcoholism themselves.

Supporting ideas:

- a tendency to enjoy alcohol more than other people.
- a more positive reaction to alcoholic drinks.
- They have a gene that makes alcohol use more pleasurable, which could make addiction more likely.

Extract 3

With so many genetic factors involved, it can be tempting to think of alcoholism as a purely genetic disease. However, some people with none of the known genetic factors and no family history of alcoholism still become addicted. This is because the social and environmental factors can be strong enough in some cases to cause alcoholism even though the person is at lower risk genetically. Social pressure, in the form of friends and colleagues who drink heavily, is a big factor in the development of alcoholism. A tendency to view alcohol as a form of stress relief or as a way to become less nervous in social situations can also make alcoholism more likely, even if the genetic influences are minimal.

Fried, P. (1984). *Addiction redefined.* (Vols. 1-4). New York, NY: Scribner's.

Main ideas:

- social and environmental factors can be strong enough in some cases to cause alcoholism.

Supporting ideas:

- Social pressure, in the form of friends and colleagues
- a form of stress relief
- a way to become less nervous in social situations.
- even if the genetic influences are minimal.

CHAPTER 3

APPENDIX G12

THE ASSESSMENT OF AUTHENTICITY[1] IN THE SYNTHESIS TEXTS

You have to assess to which extent the students have plagiarized the source texts. You have to score the texts between 0 and 100, depending on how much of the texts has been copied from the sources. Texts with the scores 0, 50 and 100 are presented below, in order to give you an idea how the texts should be scored.

An example of a text with the score 0

First information must be put into memory, a step that requires encoding. In the memory system, sensory information is put into various memory codes, which are mental representations of physical stimuli. If you encode the sounds of the words as if they had been spoken you are using acoustic encoding, and the information is represented in your memory as a sequence of sounds. The type of encoding used can influence what is remembered. The second basic memory process is storage. It is for over a very long time. So, we are depending on the storage capacity of our memory. The third basic memory process is retrieval, which occurs when you locate information stored in memory and bring it to consciousness. Retrieval processes include both recall and recognition. Recognition is retrieval aided by clues such as alternatives given in a multiple-choice test item. As a consequence, recognition is easier than recall.

Justification for the score: Almost 100% of the text is copied from the sources. The student has just replaced a few words by a synonym, or added a few words to promote the coherence of the text.

An example of a text with the score 50

There are several stages of memory construction. The first way of memory construction is encoding. The information communicate to the brain owing to codes. Sensory

[1] *The absence of plagiarism.*

information is put into several memory codes which are logical representation of physical stimuli. Petersen. M (2000) According to Dentstain. D. the second stage of memory construction is storage. To storage provides to maintenance of information over time, you are subject to the storage capacity of your memory when you find it possible to remember a work from many years ago. (2005). Finally. Clark-Stewart A. and Penner believe that the way of memory construction is retrieval. This process happens when you locate information stored in memory and bring it to awareness. This retrieving stored information is usually so fast and easyness that it seems automatic. Only when you try to retrieve other kinds of information do you become conscious of the searching process(2009)

Justification for the score: The student wrote approximately 50% of the text and copied about 50% of the text from the sources. In 40% to 50% of the sentences, no trace of plagiarism can be observed. Those phrases are completely formulated by the student. The students had permission to use the terms encoding, retrieval and storage. The other 40% to 50% of the sentences are almost completely copied from the original sources. Often, only a phrase is replaced by a paraphrase or a few words by a synonym.

An example of a text with the score 100

To begin with. The first stage to construct the memory is encoding which has a few steps this encoding means all the information that the person recieve must be coded so it will be connected with the brain and the remember information should be put in the some type that can be used easily by the memory system the first step is saving the sounds of the informations as they have been pronounced that the brain can easy remember them the second step of encoding is saving the image of the letters as they were written that this information will be appeared as a pictures in the brain the last step of encoding is using your main memory. that makes you get the information correctly even if it is wrong. Moreover, the next stage for constructing memory is strong which is related to the offerial or certain information throughout the time of after a very old time also it defunds on the storage amounts of the memory. Furthermore, the last stage is retrieval. Which usually happenes when the person won to remember some information so

he can remember it spontainsly without processing
or Retrieval process has two ways which are
remembering and catching up for example
the person can remember the information by
some helps. If he is in an exam catch
the answer through some hents.

Justification for the score: Almost 0% of the text is copied from the sources. With permission, the student has used the same technical terms as the sources, like encoding, retrieval and storage. In addition, the writer has copied some words from the sources, which have no to a very few synonyms.

CHAPTER 3

APPENDIX G13

THE ASSESSMENT OF THE COMPREHENSION OF THE SOURCES OF THE SYNTHESIS TEXTS

With regard to source comprehension, the text below is an average text of the collection of texts that you are going to assess. The score of this text is 100. The other texts should be scored in comparison with this text. For instance, if you consider a text twice as good as this one, you give it a score of 200. And if you consider the text half as good as this text, you give it a score of 50.

The first level of memory construction, as Petersen point
out that encoding is on significant step to locate information
in mind. The memory system includes sensory information in
many types of codes. Also, visual encoding, sounds of the
words encoding and semantic encoding are three examples
of representing in formation in mind.
of level of memory. Another stage in memory construction,
according to Bernstein, is storage makes arrangement of
information from past a long. You require storage to recall
a too earlier holiday if you want. The third stage of
memory construction , as Stewart explain that retrieval
needs to you put information on bring it to consciousness.
For instance, retrieving is so quick and effortless automaticly.
Behind to this, retrieval level in mind have not only
recall but also recognition. For example, on an essay
exam, retrieving it from memory with much help is
obligation so as to recall information. And also, as an example
of retrieval help, there are clue such as multiple-choice
test item to recognise which is easier than recall.

When assessing the texts, you focus on the following: *is the reader correctly informed about the process of knowledge construction?* With regard to that, the text above has the following strong and weak points.

Strong points

- The reader learns that three processes play a role in the process of knowledge construction: encoding, storage and retrieval.
- The reader realizes that these three processes occur in a certain order during the process of knowledge construction.

- The reader can determine what the process of retrieval is. Therefore, he needs to reread the sentence in which it is described several times.
- The reader learns what the function of encoding is.
- The reader is correctly informed about how information is stored in our memory: *The memory system includes sensory information in many types of codes.*
- The reader knows what the three types of encoding are.
- The reader learns what the function of storage is.
- A correct example of 'retrieval help' is given.

Weak points:

- It can be doubted that the reader can correctly determine what the process of encoding is, based on the information in this text. While reading the text, the reader has to make several inferences in order to get a complete idea of the process of encoding. It is possible that the reader thinks that encoding is a method to locate information in our memory, after having read this text.
- After having read this text, the reader probably does not exactly know what the process of storage entails. It is explained what the function of storage is; it is not explained what it actually is.
- The reader does not get to know that recall and recognition are sub-processes of retrieval.
- The reader does not get to know correctly what recall entails: it is a retrieval process without much help.
- The reader can hardly determine what recognition entails.

CHAPTER 3

APPENDIX G14

THE ASSESSMENT OF SOURCE USE IN THE SYNTHESIS TEXTS

With regard to source use, the text below is an average text of the collection of texts that you are going to assess. The score of this text is 100. The other texts should be scored in comparison with this text. For instance, if you consider a text twice as good as this one, you give it a score of 200. And if you consider the text half as good as this text, you give it a score of 50.

There are several way to remember information, which
is your mind that one of these is encoding. It mean
is install of meaning. According to author Peterson,
M.(2000) if you put information into your mind
with a sound of the worlds, you will remember
easily because it will come automaticly. And this
state can improve your memory. Second way is
storage. Which is mean keeping information or picture.
Bernstein, D (2005) believes that if you remember
what you want words you should keep on your
mind that it is possible. You have to rely
on the storage limit of your memory.
However according to Clark-Stewart, A
and Penner, L.A (2009) recognition is more
necessery than remember. İf you store your
memory, you can easily remember and
that is way you can spend less effort.

Justification for the score:
Strong points:

- The reader knows what the sources of the information are: they are all mentioned.
- The reader knows how he can find the information sources, thanks to his acquaintance with the APA style. In accordance with the APA style, the last name of the authors has been given, likewise as the year of publication.
- The reader gets to know the information sources while reading. The writer refers to the sources in the running text, not at the end of a sentence between brackets.

- The reader is correctly informed about the manner in which Petersen and Clark-Stewart & Penner present the paraphrased information in their texts. The writer uses the expression 'according to'.
- Before the reader reads the discussion of the (sub-processes of the) retrieval process, he already knows what the source of the presented information is.

Weak points:
- The writer does not refer entirely correctly to the sources. The writer has misspelt Petersen's name, and has also given the initial letters of the authors.
- The writer does not vary a lot in the manner in which he presents the authors of the sources. He often uses the expression 'according to'
- The reader is not correctly informed about the manner in which Bernstein presents the paraphrased information in his text. He just presents a belief, which many people share.
- In the middle of the discussion of the process of encoding and storage, the reader gets to know what the sources of the presented information are.
- The reader does not know what the source is of the information that is given in the first sentence.

CHAPTER 3

APPENDIX G15

THE ASSESSMENT OF TEXT COMPREHENSIBILITY OF THE SYNTHESIS TEXTS

With regard to comprehensibility, the text below is an average text of the collection of texts that you are going to assess. The score of this text is 100. The other texts should be scored in comparison with this text. For instance, if you consider a text twice as good as this one, you give it a score of 200. And if you consider the text half as good as this text, you give it a score of 50.

The first process of memory construction, Petersen(2000)states
that you should memory informations so the first step is requires
encoding. When you encode the sounds of the some words as
you can say and use acoustic encoding, the information which
is showed in your mind as a sequence of sounds. If you
use encoding, it can be effective what you wanted to remember.
The second step of memory construction, Bernstein(2005)
demonstrates that you can remember eaisly what you did
many years ago with storage. If you want to remember very significant
and valid a vacation in past for you, you can do because you
develope in the storage capability of your memory. The third
process, According to Clark-Stewart and Penner(2009), retrieval.
You put into information in memory with it. Retrieving stored
information meaning shortly is that you can say fastly your telephone
number or addres and it can seen automatic but retrieval
steps have recall and recognition. When you take an essay test, you must
retrieve it from memory yourself to recall information so recall is
harder than recognition.

When assessing the texts, you focus on the following: *does the reader know and understand the process of knowledge construction after reading the text?* With regard to that, the text above has the following strong and weak points.

Strong points:

- The reader learns that three processes play a role in the process of knowledge construction: encoding, storage and retrieval.
- The reader gets a very general idea of what the different processes entail.

- The reader realizes that these three processes occur in a certain order during the process of knowledge construction: After information is encoded, it is stored in memory. Subsequently, the stored information can be retrieved.
- The reader learns that in the process of *encoding* and *retrieval* several subprocesses can be discerned: *acoustic encoding* in the process of *encoding* and *recall* and *recognition* in the process of *retrieval*.

Weak points:

- The reader has to read the text several times in order to grasp the abovementioned facets of the knowledge construction process.
- The reader does not get to know what the different processes in knowledge construction entail, when he just reads the text literally. He has to interpret the information and examples in the texts in order to determine that.
- Sometimes an incorrect word choice hampers the understanding of the text. For example, the word 'but' in the penultimate sentence was wrongly chosen.
- The reader has to have read the text entirely, before he knows what the main topic and main aim of the text is: informing about the process of knowledge construction. It would have been easier to understand the text, when the main topic was introduced in the opening sentence, and was maybe even repeated in the closing sentence.

CHAPTER 4

APPENDIX A

MODEL COMPARISONS: TEXT QUALITY SCORES

Table A1. Model comparisons for Text Quality: Source Dependent Language Use

Text Quality: Source dependent Language Use							
	Models			Comparisons			
		X^2	df	Models	X^2	df	p
Null model	0	1457,568	3				
Session	<.0011	1365,409	5	0 vs 1	92,159	2	<.001
Bleep	2	1365,109	6	1 vs 2	0,3	1	0,584
Bleep & Instruction	3	1360,957	7	2 vs 3	4,152	1	0,042
Bleep, Instruction, & Bleep* Instruction	4	1360,795	8	3 vs 4	0,162	1	0,687

Table A2. Model comparisons for Text Quality: Content

Text Quality: Content							
	Models			Comparisons			
		X^2	df	Models	X^2	df	p
Null model	0	1738,445	3				
Session	1	1524,038	5	0 vs 1	214,407	2	<.001
Bleep	2	1520,075	6	0 vs 1	3,963	1	0,047
Bleep & Instruction	3	1519,621	7	1 vs 2	0,454	1	0,500
Bleep, Instruction, & Bleep* Instruction	4	1519,594	8	2 vs 3	0,027	1	0,869

CHAPTER 4

APPENDIX B1

MODEL COMPARISONS: EFFECTS ON WRITING BEHAVIOURS

Model comparisons for three Writing Behaviours: Total Process Time, Proportion of Pause Time and Proportion Active Writing Time

Total Process Time winsored

		Models		Comparison			
		X^2	df	Models	X^2	df	p
Null model	0	2380,398	3				
Session	1	2075,215	5	0 vs 1	305,183	2	<.001
Bleep	2A	2060,858	6	1 vs 2	14,357	1	<.001
Instruction	2B	2073,407	6	1 vs 2	1,808	1	0,179
Bleep & Instruction	3	2060,07	7	2A vs 3	0,788	1	0,375
Bleep, Instruction, & Bleep*Instruction	4	2059,093	8	3 vs 4	0,977	1	0,323
Proportion Pause Time							
Null model	0	1405,517	3				
Session	1	1165,471	5	0 vs 1	240,046	2	<. 001
Bleep	2A	1161,636	6	1 vs 2	3,835	1	0,050
Instruction	2B	1164,316	6	1 vs 2	1,155	1	0,283
Bleep & Instruction	3	1160,959	7	2A vs 3	0,677	1	0,411
Bleep, Instruction, & Bleep* Instruction	4	1152,301	8	3 vs 4	8,658	1	0,003
Proportion Active Writing Time Winsored							
Null model	0	2227,752	3				
Session	1	1997,872	5	0 vs 1	229,88	2	<. 001
Bleep	2A	1995,861	6	1 vs 2	2,011	1	0,156
Instruction	2B	1997,703	6	1 vs 2	0,169	1	0,681
Bleep & Instruction	3	1995,465	7	2A vs 3	0,396	1	0,529
Bleep, Instruction, & Bleep* Instruction	4	1990,149	8	3 vs 4	5,316	1	0,021

CHAPTER 4

APPENDIX B2

MODEL COMPARISONS: EFFECTS ON THE DISTRIBUTION OF WRITING BEHAVIOURS

Model comparisons for two Writing Behaviours. Effect on the distribution of behaviours across the process: Process Speed and mean number of pauses

Speed: Strokes per minute in five intervals (winsored)

	Models			Comparison			
		X^2	df	Models	X^2	df	p
Null model	0	6591,621	55				
Session	1	6567,686	57	0 vs 1	23,935	2	<. 001
Interval	2A	6402,376	61	1 vs 2	165,31	4	<. 001
Session * Interval	2B	6397,578	65	1 vs 2	4,798	8	0,779
Effect of Factors and Interval in session 2							
Dual Task and Interval	3A	6402,284	62	3A vs 2A	0,092	1	0,762
Instruction and Interval	3B	6402,344	62	3B vs 2A	0,032	1	0,858
Dual Task & Instruction & Interval	3C	6402,271	63	3C vs 3B	0,073	1	0,787
Dual Task* Instruction & Interval	3D	6402,27	64	3D vs 3C	0,001	1	0,975
Effect of Factors on distribution							
Dual Task*Interval	4A	6398,781	66	4A vs 3A	3,503	4	0,477
Instruction*Interval	4B	6398,871	66	4B vs 3B	3,473	4	0,482
Dual Task*interval; Instruction*Interval	4C	6394,861	72	4C vs 3C	7,41	9	0,595
Dual Task* Instruction*Interval	4D	6382,129	76	4D vs 3D	20,141	12	0,064
Final Model Comparison							
				4B vs 3A	7,423	4	0,115
				4C vs 4B	4,01	6	0,675
				4D vs 4C	12,732	4	0,013
				4D vs 4B	16,742	10	0,080

Arithmetric Mean of Pauses in six intervals (winsored)							
Null model	0	6409,541	78				
Session	1	6273,233	80	0 vs 1	136,308	2	< .001
Interval	2A	6214,252	85	1 vs 2	58,981	5	< .001
Session*Interval	2B	6205,9	90	1 vs 2	8,352	10	0,594
Effect of Factors and interval in session 2							
Dual Task and Interval Session 2	3A	6208,752	86	3A vs 2A	5,5	1	0,019
Instruction and Interval Session 2	3B	6213,224	86	3B vs 2A	1,028	1	0,311
Dual Task & Instruction & Interval Session 2	3C	6208,229	87	3C vs 3A	0,523	1	0,470
Dual Task* Instruction & Interval Session 2	3D	6206,402	88	3D vs 3C	1,827	1	0,176
Effect of Factors on distribution							
Dual Task*Interval	4A	6202,276	91	4A vs 3A	6,476	5	0,263
Instruction*Interval	4B	6195,676	91	4B vs 3B	17,548	5	0,004
Dual Task*Interval; Instruction*Interval	4C	6184,989	98	4C vs 3C	23,24	11	0,016
Dual Task* Instruction*Interval	4D	6176,636	103	4D vs 3D	29,766	15	0,013
Final Model Comparison							
				4B vs 3A	13,076	5	0,023
				4C vs 4B	10,687	7	0,153
				4D vs 4C	8,353	5	0,138
				4D vs 4B	19,04	12	0,088

CHAPTER 4

APPENDIX C

MODEL COMPARISONS: EFFECTS ON SELF-REPORTED PROCESS ACTIVITIES

Model Comparison for Self-reported process activities

Paraphrasing

		Models		Comparison			
		X^2	df	Models	X^2	df	p
Null model	Model 0	-178,366	3				
Timepoint (linear)	Model 1	-209,975	4	0 vs 1	31,609	1	<. 001
Timepoint (curvilinear)	Model 2	-210,062	5	1 vs 2	0,087	1	0,768
Instruction	Model 3	-215,774	6	2 vs 3	5,712	1	0,017
Instruction *linear	Model 4	-220,399	7	3 vs 4	4,625	1	0,032
Instruction * curvilinear	Model 5	-220,7	8	4 vs 5	0,301	1	0,583
Reading Sources							
Null model	Model 0	549,727	3				
Timepoint (linear)	Model 1	347,895	4	0 vs 1	201,832	1	<. 001
Timepoint (curvilinear)	Model 2	216,779	5	1 vs 2	131,116	1	<. 001
Instruction	Model 3	216,778	6	2 vs 3	0,001	1	0,975
Instruction * linear	Model 4	216,613	7	3 vs 4	0,165	1	0,685
Instruction * curvilinear	Model 5	215,265	8	4 vs 5	1,348	1	0,246
Working in Sources							
Null model	Model 0	-80,421	3				
Timepoint (linear)	Model 1	-123,828	4	0 vs 1	43,407	1	<. 001
Timepoint (curvilinear)	Model 2	-124,397	5	1 vs 2	0,569	1	0,451
Instruction	Model 3	-124,463	6	2 vs 3	0,066	1	0,797
Instruction * linear	Model 4	-124,469	7	3 vs 4	0,006	1	0,938
Instruction * curvilinear	Model 5	-124,954	8	4 vs 5	0,485	1	0,486
Editing							
Null model	Model 0	-212,801	3				
Effect Timepoint (linear)	Model 1	-212,801	4	0 vs 1	0	1	1,000
Timepoint (curvilinear)	Model 2	-235,555	5	1 vs 2	22,754	1	<. 001
Instruction	Model 3	-236,937	6	2 vs 3	1,382	1	0,240
Instruction * linear	Model 4	-236,99	7	3 vs 4	0,053	1	0,818
Instruction * curvilinear	Model 5	-238,831	8	4 vs 5	1,841	1	0,175

Source Writing							
Null model	Model 0	1107,142	3				
Timepoint (linear)	Model 1	1010,436	4	0 vs 1	**96,706**	1	<. 001
Timepoint (curvilinear)	Model 2	933,095	5	1 vs 2	**77,341**	1	<. 001
Instruction	Model 3	933,06	6	2 vs 3	**0,035**	1	0,852
Instruction * linear	Model 4	932,396	7	3 vs 4	**0,664**	1	0,415
Instruction * curvilinear	Model 5	932,396	8	4 vs 5	**0**	1	1,000
Reading Synthesis							
Null model	Model 0	-6,799	3				
Timepoint (linear)	Model 1	-8,228	4	0 vs 1	1,429	1	0,232
Timepoint (curvilinear)	Model 2	-31,569	5	1 vs 2	23,341	1	<. 001
Instruction	Model 3	-32,441	6	2 vs 3	0,872	1	0,350
Instruction * linear	Model 4	-32,522	7	3 vs 4	0,081	1	0,776
Instruction * curvilinear	Model 5	-32,615	8	4 vs 5	0,093	1	0,760

SUMMARY

CHAPTER 1: INTRODUCTION

The premise of this dissertation was an attempt at consolidating the writing instruction of the English preparatory school at a private university in Turkey and the writing skills requirements of the departmental courses, an issue which was raised as a result of a needs analysis meeting. The issue was a general one, as was shown by previous needs analysis studies conducted at different universities, that is, a gap in the writing curricula of the preparatory schools, namely instruction on synthesis writing. Students needed to be introduced to discourse synthesis, so that they would be familiarised with academic writing conventions, which they could improve via academic writing courses in their first year at the faculty.

This new practice necessitated a shift in instructional methods from those traditionally employed in EFL instruction. In English preparatory schools in Turkey, EFL instruction typically follows a guided-discovery approach, promoting meaningful learning opportunities resembling an authentic language learning environment. Rule-based instruction takes a secondary role, emerging implicitly in instructional materials and becoming explicit over consecutive sessions. Instructors undergo regular in-service training and performance evaluations to maintain these practices. This is a legacy of the grammar-translation method, which has had a lasting impact on L2 instruction, and is still followed by primary and secondary state schools in Turkey (Hatipoglu, 2016; Ulum & Uzun, 2020). According to this method, practical language use is secondary with a greater emphasis placed on explicit linguistic rules and translation between native and target languages. Therefore, transitioning from the methods used in EFL instruction to academic writing instruction was essential, and required delving into effective academic writing practices including in L1, which we undertook via the intervention studies of this dissertation.

CHAPTER 2: THE EFFECT OF TWO MODES OF INSTRUCTION: MODELLING VS. PRESENTATIONAL

In this study, conducted at the EFL preparatory program of a Turkish university, we aimed to investigate the impact of explicit strategy instruction in two different

modes: one involving peer observation (i.e., modelling mode) and another involving teacher presentation (i.e., presentational mode). These modes were compared with the regular practice of the institution of a more implicit nature (i.e., control condition), in a four-hour instructional intervention for teaching an introductory level synthesis task. We involved 48 upper-intermediate level EFL learners, who were academically low-achievers, and randomly assigned them to one of three conditions: modelling, presentational, and no-strategy conditions.

Our hypotheses were that (HYP1) the conditions involving strategy-focused instruction, would positively impact students' synthesis writing quality and writing processes compared to the no-strategy condition, and that (HYP2) the modelling mode of instruction would yield the best results.

The first hypothesis was rejected: the strategy conditions did not outperform the no-strategy condition in effects on writing performance. We confirmed the second hypothesis: the modelling mode of instruction resulted in qualitatively better texts than the presentational and the no-strategy conditions in the source use aspect of text quality, which was maintained in the second posttest four weeks later. The modelling condition also reported, via the offline learner reports, more learning experiences related to the procedural aspects of the writing task (i.e., steps/stages/subprocesses). They expressed that they were in the process of learning how to navigate the various steps involved in writing a synthesis text and often attributed their success to this acquired process knowledge. The online process results, as measured by Time-Sampled Self-Reporting, showed distinct patterns in the writing processes of the modelling condition, which could be associated with their superior text quality results in source use, and also with their reported learning experiences regarding the procedural knowledge on the stepwise handling of the writing process. To specify the particular effects, in the middle phase of the writing process, students in the condition reported the activity working on the sources significantly more than the presentational condition, which indicated the use of strategies (i.e., OLIN strategies of the TRAMPOLINE mnemonic) associated with the source use aspect of text quality (i.e., source attribution using APA principles, using a variety of reporting verbs, establishing coherence within the text via linking devices). The no-strategy condition, on the other hand, reported reading the sources (for comprehension) in the first phase and writing the target text (transcription) in the last phase of the writing process, significantly more than the strategy conditions. Based on our quality parameters, the sequential (rather than a recursive) temporal management of the writing process, as well as the choice of nonstrategic (rather than cognitively more demanding) activities within the broader source-related and target-text-related

categories, indicated engagement with more simple and general activities, resulting in a less effective writing process.

CHAPTER 3: THE EFFECT OF STRATEGY INSTRUCTION AND MODE OF INSTRUCTION ON BOTH WRITING PROCESS AND PRODUCT IN A FOREIGN LANGUAGE: MODELLING VS. PRESENTATIONAL

This study is a larger-scale replication of the previous study, including a different student profile. In this study, conducted at the EFL preparatory program of a Turkish university, we aimed to investigate the impact of explicit strategy instruction in two different modes: one involving peer observation (i.e., modelling mode) and another involving teacher presentation (i.e., presentational mode). These modes were compared with the regular practice of the institution of a more implicit nature (i.e., control condition), in a four-hour instructional intervention for teaching an introductory level synthesis task. We involved 155 upper-intermediate level normally-performing EFL learners, and randomly assigned them to one of three conditions: modelling, presentational, and no-strategy conditions. Our hypotheses were that the conditions involving strategy-focused instruction, would positively impact students' synthesis writing quality, effects on transfer to another genre, and writing processes compared to the no-strategy condition, and that the modelling mode of instruction would yield the best results.

The results of this study largely corroborated the findings of the previous study. Explicit strategy instruction proved to be effective when it was modelled rather than presented, this time with additional effects on text quality: students in the modelling condition produced qualitatively better texts than the presentational condition, in both source use and authenticity aspects of text quality. Another finding that was new to this study (when compared to the previous one), is that the no-strategy condition wrote more authentic texts compared to the presentational condition. As for the effects on transfer to the argumentative task, presentational strategy condition was more effective compared to the no-strategy condition; however no differential effects between the modelling condition and the presentational condition were found in effects on transfer. As for the writing processes, participants in the strategy conditions (i.e., modelling and presentational) reported via online reporting methods several common patterns in their process behaviours, which pointed towards a more recursive approach to the writing process: they allocated more time to source-related activities throughout the writing process, and also showed more variety in their choice of activities, when compared to the no-strategy condition. The results of the offline

process measure, Learner Reports, indicated that the explicit strategy conditions reported more learning experiences in the paraphrasing requirement of the synthesis tasks, and the modelling condition mentioned more learning experiences overall and also more experiences regarding time-management. Thus, the results of this replication study, encompassing a more extensive sample, allowed for the generalisation of findings from the previous study to a student profile with stronger academic capabilities.

CHAPTER 4: REACTIVITY OF DIRECTED RETROSPECTION IN INTERVENTION STUDIES: LEARNING FROM THE TEST

This study investigated a possible reactivity effect of Time-Sampled Self-Reporting (TSSR) in a pretest-posttest experimental design with 81 undergraduate students at a Belgian university. To address the distinct phases of the TSSR technique, that is, the training and the actual implementation phases, and explore their effects on the validity and the reliability of the collected data, we formulated three hypotheses. Firstly, we examined the reactivity of TSSR on students' process behaviours and text quality in the subsequent test under TSSR conditions (HYP1). Next, we investigated whether the mode of instruction used in the training session (video vs. presentational) had varying effects on the reactivity imposed on students' text quality and writing processes in the subsequent posttest under TSSR conditions (HYP2). Lastly, we explored how the mode of instruction (video vs. presentational) used in the training session affected the validity and the reliability of the process data collected under TSSR conditions (HYP3).

To test these hypotheses, we employed a 2x2 between-subjects factorial design, manipulating the reporting mode in the posttest (TSSR vs. no-TSRR) to examine the first hypothesis and the instruction mode in the TSSR training (video vs. presentation) to evaluate the second hypothesis. This design allowed us to assess the impact of the intervention in four conditions: (1) Video Instruction & TSSR, (2) Video & no-TSSR, (3) Presentational & TSSR, and (4) Presentational & no-TSSR. We evaluated students' synthesis texts in two quality aspects: content and the proportion of source text language. We also collected and analysed students' process behaviours through keylogging software, focusing on time allocation and distribution based on empirically validated variables from previous research (Vandermeulen et al., 2023). Additionally, students' self-reported writing logs (TSSR) were collected in the posttest phase.

We confirmed Hypothesis 1: TSSR improved the content quality of students' synthesis texts, which corresponded with the process indicator of more time spent on the task. Hypothesis 2 was also confirmed, but in a different direction

than we had initially expected. Students in the presentational condition used less source text language compared to students in the video condition. Presentational Instruction & TSSR corresponded with the highest reactivity effect, in a positive direction, indicating a learning effect resulting from the TSSR technique. We argued that the explicit and declarative instruction on the activity categories in the training phase helped in activating/reorienting synthesis task schema and the writing logs with the activity categories, their descriptions and a graphic representation may have aided with the retrieval of the encoded information. The multiple encounters with these factors in the consecutive training, practice and implementation sessions, may have resulted in inference-based decisions about the solution of the current (writing) problem. Moreover, the short video featuring snippets from a writing task in a fragmentary format may have created a wrong task-representation in terms of the temporal management of the writing process, thereby resulting in a negative reactivity effect. The observed variations in students' writing process activities, influenced by the instructional mode used in the training session, and their subsequent impact on students' text quality, raise concerns regarding the validity and reliability of the process data collected through TSSR, thereby proving Hypothesis 3.

CHAPTER 5: GENERAL DISCUSSION

In this final chapter, we presented a comprehensive overview of each study, discussing and synthesising their results. Our research demonstrates the valuable contributions to the field of observational learning, methodology in writing studies, and potential implications for policy changes in English preparatory schools at Turkish universities.

Our instructional intervention studies revealed the remarkable efficacy of modelling in explicit strategy instruction, rendering the inclusion of additional presentational components unnecessary to achieve the desired effects. Moreover, the large-scale replication study allowed us to generalise these findings across a diverse range of participant profiles, encompassing both academically low-achieving and high-achieving students.

Furthermore, our results highlighted the enduring positive effects of this strategy over time. In Chapter 2, we found that improvements were still evident in a second posttest conducted four weeks later (Study 1). In Chapter 3, we demonstrated the successful transfer of learning to the argumentative genre, underscoring the adaptability of the skills acquired in one context to various writing tasks. Additionally, our studies unveiled an unanticipated impact of modelling on students' self-regulation, particularly in time-management. This aspect

was not explicitly included in the strategy instruction or the content of the modelled performance. Thus, modelling transcended its conventional definition as an instructional mode, creating a distinct learning environment that significantly contributed to learning gains.

Our research used a variety of measures to assess text quality, writing processes (both online and offline), learning experiences, taking into consideration possible confounding effects such as motivation and initial levels of writing proficiency. Data collected via these measures, when analysed in conjunction with one another, provided a comprehensive understanding of the results from multiple dimensions. We also evaluated a possible reactivity issue of TSSR, utilised for operationalising students' online writing processes, by employing another process measure through keylogging software in a third study. This provided further insight into the methodology of instruction and assessment of writing products and processes, as well as how different instructional modes and their respective materials (e.g., writing logs, video) contributed to their reactivity and the validity of the collected process data.

Finally, we discussed how our studies possess the potential to make substantial contributions to the theory of observational learning and the methodology employed in writing studies. These contributions lay a solid foundation for recommending policy changes in English preparatory programs at Turkish universities, addressing issues that were previously left unaddressed, which we presented as implications for educational practice in the final section of the chapter.

ABOUT THE AUTHOR

Mujgan Buyuktas Kara was born in 1984, in Eskisehir, Turkey. She pursued her education in Western Languages and Literatures at Bogazici University, followed by English Language and Literature studies at Istanbul Aydın University. She completed her PhD at the University of Amsterdam, Research Institute of Child Development and Education, focusing her dissertation on L2 writing skills.

Her career commenced as a research assistant in English Language Teaching (ELT), subsequently serving as an English language instructor in the preparatory schools of various foundation universities for nearly a decade. During this time, she also took on additional responsibilities such as test writing, teacher training and researching within these institutions.

She holds a Cambridge ESOL Certificate for English Language Teaching to Adults (CELTA) and is halfway through completing the Diploma degree of the same program (DELTA). Additionally, she possesses a TEFL International TESOL certificate. Throughout her academic pursuits, she received numerous grants and scholarships and has presented at various international conferences.

Commencing in February 2024, Mujgan Buyuktas Kara will be instructing second and third-grade students in Literature, Linguistics, ELT, and Educational Psychology courses within the ELT department of a foundation university in Istanbul.

ACKNOWLEDGEMENTS

My PhD journey began as a personal pursuit for professional development; however, it evolved into something far more profound as it progressed. It was a transformative journey intertwined with the tapestry of life, fostering my personal growth along the way. I am deeply grateful for the collective efforts and collaborative spirit that underpinned this achievement, and I extend my sincere acknowledgment to all involved.

I owe my deepest gratitude to my supervisors, Prof. Dr. Gert Rijlaarsdam, and Prof. Dr. Elke Van Steendam, who formed an exceptional team for this PhD endeavour. From the outset, Gert Rijlaarsdam warmly welcomed me and provided unwavering support, demonstrating a comforting presence through every challenge. His demeanour will always stay with me in my future career endeavours, guiding me to overlook mistakes and encourage my students in a way to bring out the best in them. Elke Van Steendam, my co-supervisor, played a pivotal role in my transformation from a graduate of English Literature to an adept academic writer. She introduced me to the nuanced academic standards, consistently upholding these benchmarks and steadfastly illuminating the path forward. Her graceful guidance was instrumental in shaping me into the researcher I am today, and for this, I am profoundly thankful.

I am grateful to Dr. Ton Koet who encouraged me to pursue my PhD at UvA shortly after submitting my Master's thesis. His assistance in introducing me to the UvA research team was invaluable. Special thanks to Hüseyin Eren Kuru for conducting the majority of the statistical analyses in this dissertation and to Dr. Margot Chauliac for working on the keylogging data. Their collaboration was indispensable for the completion of this work. I also extend my gratitude to Melanie Hof for meticulously planning the rating procedure and to Noortje Hemmen for organising the rating sessions. Their constant support and willingness to address my numerous questions were instrumental throughout this process.

I am deeply grateful to the administration of Istanbul Sehir University English Preparatory School, particularly Dr. İlke Büyükduman, for her support, as well as to my dear colleagues who dedicated their valuable time and effort to the intervention studies. Their professional cooperation and friendly encouragement facilitated a seamless training and data collection process when conducting the studies of this dissertation. I also express my thanks to Ahsen Keskin, the peer

model in the training videos, whose remarkable wisdom beyond her years and exceptional acting skills greatly enriched the project.

A special acknowledgement goes to the students of İstanbul Şehir University and KU Leuven, who participated in the studies. Their enthusiasm, support, and kindness towards me were invaluable. They were not just participants but the key contributors and beneficiaries of this dissertation.

I'm deeply thankful to my extended family for their encouragement, particularly my mother, a dreamer herself, and my sister Diler, who, amidst the challenges, revealed herself as my kindred spirit. Your belief in me has been invaluable.

I am immensely thankful to my friends, Anıl and Özge, who accompanied me on every trip to Amsterdam, adjusting their holiday schedules so that we could explore together. The memories we created on these travels will remain cherished throughout our lives. My best friends forever, Aslı, Erinç, Hilal, Mehtap, and Sevtap, who stood by me through several life adventures, witnessing and supporting all my aspirations, I am deeply grateful! Aslı, I eagerly anticipate celebrating your upcoming PhD ceremony together!

My dear children Deniz and Leyla, you are my little muses of inspiration. Your constant reminders of the fun things we would do once my 'homework' was finished kept us going. Emre, the love of my life, thank you. You have witnessed me pursue my countless dreams, watched me let go of some and pick up more! I evolved a thousand times during the two decades we have been together. Thank you for always making me feel appreciated and loved. Your support means the world to me.

Finally, to my angels here and above. My pets Pippa, Kahraman, Juju, and the dearly departed yet never forgotten Şermin, thank you for being a source of comfort during challenging times. To my dear father, Nuri, your enduring light transcends time and space, and the joy of feeling it always prevails.

DISSERTATION SERIES INNOVATIVE LANGUAGE EDUCATION LAB

Jannemieke van de Gein	1991	The sense of sentences. Effects of grammar skills on aspects of writing profiency.
Rob Schoonen	1991	Constructie en validering van gestructureerde stelopdrachten voor zesdeklassers [Construction and validation of structured writing tasks for eight graders.]
Amos van Gelderen	1992	Fonetisch en linguïstische validatie van oordelen over spreekprestaties [Phonetic and linguistic validation of rating of oral performances]
Michel Couzijn	1995	Observation of Writing and Reading Activities: Effects on Learning and Transfer
Marianne Overmaat	1996	Empirische studies naar effectieve schrijfdidactiek. [Empirical Studies in Written Composition Methodology].
Tanja Janssen	1998	Literatuuronderwijs bij benadering [Approaches in Literature Education]
Martine A.H. Braaksma	2002	Observational learning in argumentative writing.
Hein Broekkamp	2003	Task demands and test expectations. Theory and empirical research on students' preparation for a teacher-made test.
Isabelle de Ridder	2003	Hyperlinks & Vocabulary Learning from texts in L2.
Rita Rymenans & Sven de Maeyer	2004	Onderzoek naar kenmerken van effectieve scholen. Kritische factoren in een onderzoek naar schooleffectiviteit in het technisch en beroepssecundair onderwijs in Vlaanderen. [Investigating characteristics of effective schools. Critical factors in a study of school effectiveness in technical and vocational secondary education in Flanders.]
Lap Quoc Trinh	2005	Stimulating learner autonomy in English language education. A curriculum innovation study in a Vietnamese context.
Marleen Kieft	2006	The effects of adapting writing instruction to students' writing strategies.
Rita Rigo-Toth	2007	A study of the influence of home and preschool factors on young children's literacy achievement
Ton Koet	2007	Polder English in Dutch ears. Empirical studies on the evaluation of the pronunciation of English as a foreign language.
Theo Witte	2008	Het oog van de meester.
Mariet Raedts	2008	De invloed van zelfeffectiviteitsverwachtingen, taakkennis en observerend leren bij een nieuwe en complexe schrijftaak. [The effect of self-efficacy expectations, task knowledge and observational learning in a new and complex writing task].
Elke Van Steendam	2008	Effective instructional strategies in collaborative revision in ESL. The effects of Two empirical studies.
Daphne van Weijen	2009	Writing processes, text quality, and tasks effects Empirical studies in first and second language writing

Lieve Verheyden	2010	Achter de lijn. Vier empirische studies over ontluikende stelvaardigheid. [The story behind the line. Four empirical studies on writing by third and fourth graders of primary school.
Tatiana Berden-Antonenko	2010	Stimulating intercultural intellectual capabilities in intercultural communication. Testing an innovative course design.
Talita Groenendijk	2012	Observe and Explore. Empirical studies on Learning in creative writing and visual arts.
Marion Tillema	2012	Writing in first and second language. Empirical studies on text quality and writing processes
Phuong Nam Thi Nguyen	2012	Second language writing and literary reading in university: Three empirical studies.
Dick van Dijk	2013	De Taal van oplossingen. Een empirisch begrippenkader voor oplossingsgerichte interactie. [The Language of Solutions. An empirical conceptual framework for solution-focused interaction]
Milan Kríz	2014	Teaching Dutch as a foreign language in Eastern Europe
Marco Kragten	2015	Comprehending process diagrams in biology education
Marrit Hoeks- van de Guchte	2015	Focus on Form in Task-based Language Teaching
Marie-Thérèse van de Kamp	2017	Re-Imagine, Re-Design and Transform. Enhancing Generation and Exploration in Creative Problem Finding Processes in Visual Arts Education.
Marloes Schrijvers	2019	The story, the self, the other Developing insight into human nature in the literature classroom.
Klaske Elving	2019	Effectieve Leeractiviteiten voor het Schrijfonderwijs in HAVO 4 [Effective Learning Activities for the Teaching of Writing in General secondary education, grade 10].
Suzanne Luger	2020	Lost in Latin Translation. Teaching students to produce coherent target texts
Nina Vandermeulen	2020	Synthesis Writing in upper-secondary Education
Magdalena Flores-Ferrés	2021	Teaching Writing in Chile. An evidence base for policies on writing instruction for Grades 7-12 of Chilean public schools
Martijn Koek	2022	Think Twice: Literature Lesson that Matter.
Liselore van Ockenburg	2022	Synthesis writing. Teaching high school students how to read, plan, draft, and revise
Edith Alkema	2022	Schrijven met kennis van zaken. Een didactiek voor de zaakvakken. [Knowledgeable writing. Writing instruction for social sciences]
Chelsea O'Brien	2023	Dealing with setbacks. Exploring academic buoyancy in Latin students who are cognitively gifted
Lieke Holdinga	2023	Disciplinary Writing. Four empirical studies on historical and philosophical literacy
Anouk ten Peze	2024	Creative Writing. Thinking beyond the standard text