The effects of different types of video modelling on undergraduate students' motivation and learning in an academic writing course

Mariet Raedts, Elke Van Steendam, Luc De Grez, Jef Hendrickx & Chris Masui*

University of Leuven | Belgium
*University of Hasselt | Belgium

Abstract: This study extends previous research on observational learning in writing. It was our objective to enhance students' motivation and learning in an academic writing course on research synthesis writing. Participants were 162 first-year college students who had no experience with the writing task. Based on Bandura's Social Cognitive Theory we developed two videos. In the first video a manager (prestige model) elaborated on how synthesizing information is important in professional life. In the second video a peer model demonstrated a five-step writing strategy for writing up a research synthesis. We compared two versions of this video. In the explicit-strategy-instruction-video we added visual cues to channel learners' attention to critical features of the demonstrated task using an acronym in which each letter represented a step of the model's strategy. In the implicit-strategy-instruction-video these cues were absent. The effects of the videos were tested using a 2x2 factorial between-subjects design with video of the prestige model (yes/no) and type of instructional video (implicit versus explicit strategy instruction) as factors. Four post-test measures were obtained: task value, self-efficacy beliefs, task knowledge and writing performances. Path analyses revealed that the prestige model did not affect students' task value. Peer-mediated explicit strategy instruction had no effect on self-efficacy, but a strong effect on task knowledge. Task knowledge – in turn – was found to be predictive of writing performance.

Keywords: academic writing – peer modelling – strategy instruction – observational learning – writing self-efficacy

1. Introduction

Being able to select, organize, connect and synthesize (Spivey, 1997) information from various, quite often contradictory sources, has become a crucial ability in higher education. Writing research syntheses has become a common task in different disciplines (Bridgeman & Carlson, 1984; Cooper, 2007; Graves, Hyland, & Samuels, 2010; Zhang, 2013; Zhu, 2004). However, writing a research synthesis is a cognitive and motivational challenge for students. From a cognitive point of view, students first have to identify relations and contradictions in the research findings. Next, the studies have to be evaluated with objectively defined criteria, and finally, the source articles have to be organized and integrated in a new whole (Granello, 2001; Segev-Miller, 2004). From a number of studies it emerges that successfully implementing all these different steps in writing up a research synthesis, that is, to distinguish the main information in source texts from less important information; to evaluate the sources' scientific quality; to integrate relevant research findings in a coherent text while providing a meaningful explanation for contradiction, requires higher levels of cognitive complexity, an advanced capacity for critical thought and higher-order reasoning ability (Froese, Gantz, & Henry, 1998; Granello, 2001; Jackson, 1980; Spivey, 1997).

Quite frequently, however, university freshmen lack the necessary knowledge and skills to complete an academic synthesis writing task successfully (e.g. Boscolo, Arfé, & Quarisa, 2007; Granello, 2001; Mateos & Solé, 2009). Research shows that only a minority of the students at undergraduate level construct reviews that meet academic standards (Campell, Smith, & Brooker, 1998; Makovsky, 1985). They often simplify the writing task by repeating the content of the source material sequentially in their papers, creating separate paragraphs for each study or article they review (cf. knowledge-telling strategy by Bereiter & Scardamalia, 1987) without combining and synthesizing the information (Froese, et al., 1998; Granello, 2001) as a more proficient writer would do (cf. knowledge-transforming strategy by Bereiter & Scardamalia, 1987).

Learning to write also poses motivational challenges to the students as several authors pointed out (e.g. Bruning & Horn, 2000). Too many students experience low self-efficacy beliefs or even feelings of anxiety and lack of control. Because writing requires extended periods of engagement, it is important for students to perceive writing as valuable and to believe in one's competence as a writer (Bruning & Horn, 2000, p. 28). The question then arises how we can design an instructional environment to simultaneously affect cognitive and motivational elements in the learning of synthesis writing. An instructional method which integrates motivation and cognition is observational or social learning, a core concept in Bandura's Social Cognitive Theory (1986).

2. Theoretical framework: Bandura's Social Cognitive Theory

In observational or social learning children and adults acquire a new skill or strategy that was first demonstrated by another person who acts as a model. Bandura (1986, p. 51) even posits that "Providing a model [to observe] … is one of the most effective ways to convey information about the rules for producing new behavior". Modelling can involve either live demonstrations of a skill or can be asynchronous by means of for instance a video.

Observation of a model is but a first step in learning. The learning cycle is, according to social cognitive theory (Bandura, 1997) governed by four processes: attention, retention, production and motivation processes. Attention needs to be triggered first, that is, the learning process cannot start unless learners accurately observe what has to be learned. Learners' attention can be channeled and directed by a number of techniques such as showing attractive models, including narration and/or worksheets (Bethards, 2014), contrasting modelling of good and poor performance or subdividing complex behavior in digestible chunks of knowledge (Bandura, 1986).

The second process is retention. For the noticed information to be retained, it should be stored in long-term memory but first processed in short-term memory. The latter system has a very limited processing capacity and it is important that designers of instructional videos for modelling look for load-reducing solutions. One way to do so is to use two separate channels for the processing of information. A basic assumption about the working of the human mind is that information is processed through an auditory/verbal channel and a visual/pictorial channel (Mayer & Moreno, 2003). A video clip only containing spoken words uses only one channel and the solution is to move some information to the other channel by using on-screen text. Mayer and Moreno (2003) call this the modality effect. Other techniques to enhance retention, is by presenting the information multiple times in the same (Fagundes, Chen, & Laguna, 2013) or in a different form. Models demonstrating different approaches to the same task such as weaker and stronger models are effective. This is proven in research on video modelling in presentation skills (De Grez, Valcke, & Roozen, 2009), in argumentative writing (Braaksma, Rijlaarsdam, & Van den Bergh, 2002), in academic writing (Raedts, Rijlaarsdam, Van Waes, & Daems, 2007) and in collage making and creative writing (Groenendijk, Janssen, Rijlaarsdam, & Van den Bergh, 2013a, 2013b).

The third process in observational learning is production or practice. Observers must be capable of converting information that is abstractly represented in their long-term memory into appropriate actions. To be able to do this, learners ideally should possess the necessary skill and/or (declarative, procedural, metacognitive) knowledge (Bandura, 1986).

Motivation, the fourth process, does not only play a crucial role in both attention and retention but also, and perhaps even more importantly, may determine if production actually occurs. According to Bandura (1986) learners only show what they have learned if two conditions are met. First of all, they need to be confident that they are able to perform the learned behavior. Secondly, they have to value the

consequences of that behavior. The first condition refers to learners' self-efficacy, a concept that Bandura (1997, p. 3) defines as: "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments". Learners with a positive self-efficacy are more likely to work harder, are more persistent, and attain higher achievement levels.

Bandura (1997) states that self-efficacy is enhanced by a successful performance (enactive mastery experience), by observing a successful performance (modelling) and by verbal persuasion. The first and most important source is enactive mastery experience. The second source of self-efficacy is vicarious experience or modelling. With respect to modelling, research shows that perceived similarity in competence between model and observer leads to higher self-efficacy beliefs (Bandura, 1997) and to better learning.

The second condition to be met in order for learners to show learned behavior concerns the extent to which learners value the anticipated benefits of that behavior. Task value consists of four components (Schunk, Meece, & Pintrich, 2013): attainment value (personal importance of doing well on the task), intrinsic value, utility value (how well a task relates to current and future goals) and cost (negative aspects of engaging in a task). Positive expectancies and values affect achievement behaviors beneficially. It is, as indicated, important for students to perceive writing as valuable (Bruning & Horn, 2000, p. 28). "If students believe that current educational activities are useful to them in the long run, they are more likely to be motivated to achieve" (Wigfield & Cambria, 2010, p. 53).

3. Observational learning in writing research

Also in writing instruction observational learning has proven its effectiveness either in combination with other instructional methods (in instructional packages) or as a stand-alone instructional method. First, there are the studies in which modelling is a component of a much larger, (all-encompassing) instructional package combined with subsequent collaborative and individual practice such as Self-Regulated Strategy Development or SRSD (Graham, Harris, & Mason, 2005; Graham & Perin, 2007) and Cognitive Strategy Instruction in Writing or CSIW (Englert, Raphael, Anderson, Anthony, & Stevens, 1991; Torrance, Fidalgo, & García, 2007) where it is combined with direct instruction and subsequent collaborative and individual practice. These instructional packages are based on the theoretical premise that by instructing students in the use of strategies, deliberate and effortful cognitive procedures for goal-setting and problem-solving (Alexander, Graham, & Harris, 1998) and by teaching them to self-regulate (i.e., to set themselves goals, to select strategies to manage, monitor and evaluate task execution), students can become self-regulated writers. In these strategy instruction studies modelling is usually done live by an instructor who explicitly demonstrates how to use a specific strategy to plan, structure, draft and/or revise a particular written genre. As such, the type of modelling done in SRSD and CSIW is a

quite explicit, direct type of mastery modelling in which different steps in a procedure are explicitly taught to students with the aid of mnemonics and/or graphic organizers. Mnemonics are used to reduce cognitive load and to facilitate observation, memorization and retention of the strategy modelled (Englert, Mariage, & Dunsmore, 2006). The majority of the cognitive strategy instruction studies in writing have been conducted with normally-achieving and learning-disabled writers in primary education. However, very few studies on cognitive strategy instruction in writing including modelling have been conducted at college-level (MacArthur, Philippakos, & Ianetta, 2014).

The next series of studies in which a form of modelling is used for writing instruction are studies in which modelling is the sole instructional component (Rijlaarsdam et al., 2008). These studies are based on Schunk and Zimmerman's (1997) Social Cognitive Theory of Self-Regulation according to which for the acquisition of complex skills, observation is subsequently followed by deliberate emulation or imitation, self-control and self-directed, self-regulated practice.

The type of modelling that is done in the majority of the second series of studies (on observational learning) is predominantly modelling via instructional video featuring peers verbalizing the use of specific writing processes or strategies by thinking out loud (Snowman & McCown, 2012). Students are expected to infer information about effective and less successful writing processes or strategies by evaluating (and comparing) (the) model(s') performance. Not only good models are shown to students but quite often students are asked to reflect on and evaluate both good and weak models (Braaksma et al., 2002; Groenendijk et al., 2013a; 2013b; Raedts et al., 2007).

The studies attest to the effectiveness of observational learning to improve the writing of secondary-school students (Braaksma et al. 2002; Couzijn, 1999; Groenendijk et al., 2013a, 2013b) compared to (more traditional) practicing methods. Also for college students, this type of modelling in which both strong and weak models were compared resulted in higher writing performance than a more traditional practising method (Raedts et al., 2007; Van Steendam, Rijlaarsdam, & Van den Bergh, 2014). In Raedts et al. (2007) first-year college students watched six videos of peer models writing up a literature synthesis of five empirical studies under think aloud conditions. The tutorial videos emphasized the importance of task orientation and planning strategies. Each video displayed two models: a good model and a weak model. The good models used effective writing strategies that led up to a cognitive complex research synthesis. These models for instance, compared the research results of the different empirical studies carefully, before they started to write. The weak models showed no task orientation and planning activities. In total nine different models appeared in the videos. Before each video, participants were instructed to pay attention to specific aspects of the models' approaches. Participants were also stimulated to take notes during their observation. After each video, the students had to identify the weak and good writer. Then, they had to elaborate on one of the model's writing strategies, linking his/her approach to the quality of the 'synthesized' text. These

metacognitive activities helped students to internalize criteria for good synthesis writing and enrich their knowledge about the text genre. Hence, observational learning not only had a positive effect on students' writing performances, but also on their knowledge about the task.

This type of modelling can be considered as a more implicit, indirect type of modelling as the different writing processes or strategies that are verbalized and demonstrated by writers are not always explicitly named nor supported by strategy steps for example. The form of modelling is thus less guided, directed and channeled than the type of modelling which can be found in strategy instruction methods where the starting point is mostly a single, specific, 'effective' strategy and teachers explicitly model and verbalize the specific steps in the strategy (e.g. the type of modelling in SRSD and CSIW).

Van Steendam, Rijlaarsdam, Sercu, and Van den Bergh (2010) tested an explicit form of modelling in which two peer models (i.e., students) via collaborative revision illustrate the use of different (subsequent) steps of an expert strategy to revise L2 texts for structure and content. The authors show that this more guided type of modelling, closely resembling the type of modelling done in SRSD and CSIW, is more effective than a more traditional practising condition for collaborative revision in a foreign language.

From the (strategy instruction and observational learning) studies discussed above hence two distinct forms of modelling can be distinguished: an implicit form of modelling from which students need to infer the necessary information with regard to successful and/or less successful writing strategies from peer models' verbalizations, and a very explicit strategy-guided form of modelling in which students are shown peers or teachers model the different steps in a (single) strategy by using mnemonics or other strategy tools. These strategy tools make writing more controllable and Bandura (1997, p.88) sees controllability as conducive to the enhancement of self-efficacy.

Very few studies to date, however, have directly compared these two forms of modelling. To the best of our knowledge, only in a single study (Fidalgo, Torrance, Rijlaarsdam, Van den Bergh, & Álvarez, 2015) a more implicit form of modelling and explicit declarative teacher instruction using mnemonics were tested on 6th-graders' writing performance. However, the study did not "permit the conclusion that modelling (combined with shared reflection) is more effective than declarative instruction" as the explicit, teacher-led strategy instruction phase followed a modelling phase nor did it offer "a direct comparison of the benefits of these two forms of instruction" (p. 48). Additionally, very little research has been conducted on the combined and separate impact of attention-catching and retention-enhancing elements of video models (Van Steendam, De Grez, Goeman, & Frawley, 2015).

4. Research goal

The objective of this study was to determine the impact of attention-, retention- and motivation-enhancing elements in an academic writing course on a new and complex writing task. Specifically, we tested the effects of two kinds of video models. The first model - a so-called 'prestige' model – was designed to enhance students' motivation (i.e., task value) for the writing task. Thus, the first research question is as follows:

- Research Question 1 (RQ1): Does a prestige model increase students' task value and, consequently, their writing performances on a new writing task?

The second model was an instructional model in two versions characterized by a varying degree of explicitness. The two versions of the model were designed to test to which degree attention- and retention-enhancing elements in the video model influence students' self-efficacy beliefs, task knowledge and writing performance. Hence, our second research question is:

- Research Question 2 (RQ2): Does modelling with explicit strategy instruction have a greater positive effect than modelling without explicit strategy instruction on students' task knowledge, self-efficacy beliefs, and writing performances?

The writing task comprised synthesizing and comparing research findings into a coherent and concise summary. "Writing a literature review presupposes complex cognitive activities such as to determine similarities and contradictions in research results and articulating them clearly in your own words" (Raedts et al, 2007, p. 226). That is why we included a measurement of students' logical thinking ability in our model to control for possible influences of this variable. Thus, our study also addresses the following research question:

- Research Question 3 (RQ3): Are the effects of explicit (respectively, implicit) strategy instruction influenced by students' logical reasoning ability?

Figure 1 is a schematic representation (path model) of the above mentioned hypothesized relationships.[1]
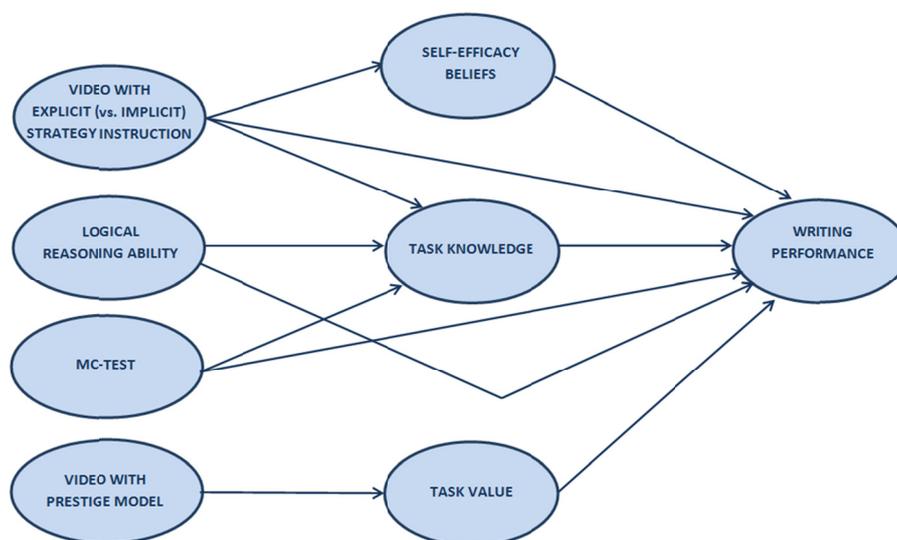
Figure 1. Path model representing the proposed relationships among the variables predicting students' writing performance on the summarizing writing task.

5. Method

5.1 Participants

Our sample consisted of 162 Dutch-speaking first-year business students at a Belgian university; 99 were male and 63 were female. All students were so-called generation students, that is, they were all in between 18 and 19 years of age. The students were enrolled in an introductory psychology course in which they did a small-group research project addressing a topic in the field of organizational psychology. The research project was spread over several months. It involved both conducting a literature review and doing qualitative research interviews. In the final stage of their project, students had to present and discuss their findings in a research report. Our writing intervention program was implemented in this latter part of the students' learning trajectory.

5.2 Research design

We conducted an experiment with a 2 (prestige model: yes/no) x 2 (explicit versus implicit strategy instruction model) between-subjects design, resulting in four experimental conditions (see Table 1). Only students in condition 1 and 3 saw a video of the prestige model. Students in condition 1 and 4 received an explicit-strategy instructional video, whereas students in condition 2 and 3 saw an instructional video with implicit strategy instruction.

Students were randomly assigned by the administration to groups (groups A to I subdivided into 44 smaller project teams of 4 to 6 students). These 44 project teams of randomly assigned generation students were randomly assigned to the four conditions (11 teams per condition). A chi-square test of independence was performed to examine the relation between gender and condition. The distribution of male and female students was not significantly different across the four conditions: $\chi^2(3, N= 162) = 1.62$, $p=.66$.

5.3 Procedure

The data were gathered on four consecutive days (one day for each experimental group). The intervention coincided with the completion of the reading stage of the literature review. As the intervention was an integrated component of the students' regular course curriculum, participation in our study was mandatory. Students could, however, notify they wanted their data to be excluded from the data analysis. One male student made use of this option.

The intervention was spread out over two sessions: a morning session (two-and-a half hours) and an afternoon session (three hours). Table 1 provides an overview of the administered pre- and post-test measurements and the students' (learning) activities during both sessions.

Table 1. Experimental design

Measurement	Condition 1 +PM ESI	Condition 2 -PM ISI	Condition 3 +PM ISI	Condition 4 -PM ESI
Morning session				
Task value (pretest)	x	x	x	x
Video prestige model	+PM	-PM	+PM	-PM
Evaluation prestige model	x	-	x	-
Self-efficacy (pretest)	x	x	x	x
Study-text on academic writing	x	x	x	x
Instructional video*	ESI	ISI	ISI	ESI
Evaluation peer model	x	x	x	x
Afternoon session				
Task value (posttest)	x	x	x	x
Task knowledge	x	x	x	x
Self-efficacy (posttest)	x	x	x	x
Writing tasks	X	x	x	x

Note. PM = prestige model (+PM = condition with prestige model vs. -PM = without prestige model) / *ESI = explicit strategy instruction / ISI = implicit strategy instruction

The pretest(s) administration and intervention itself (observation of the instructional video) were organized in a lecture hall. At the beginning of the morning session, students received a cardboard personalized folder that contained all questionnaires and documents. The various documents in the folder were printed on paper of different colours. In this way, it could be easily verified whether a student dealt with the same document as all other participants. All students were allowed the same amount of time for each activity. During the afternoon session, the posttest measures were collected in computer classes using Qualtrics 2012 survey software (Qualtrics, Provo, Utah). The first author supervised all morning and afternoon sessions. The second and/or third author were present during the afternoon sessions.

In the next paragraphs, we will outline the content of the study text on academic writing (see Table 1) and look at the writing task in more detail. Next, we describe the characteristics of the two videos in our intervention program: the video with the prestige model and the instructional video with the peer model. Finally, we give an overview of the various pre- and posttest measures.

Study-text on academic writing

Before the intervention, participants were not familiar with the rules and conventions for in-text citations in academic writing. Hence, a study-text on academic writing was given to them in the morning session. The students were told that as this text contained important information concerning the writing tasks they had to perform during the afternoon session, they were expected to read the text carefully.

The study-text was a four-page text (1,240 words) dealing with the following three topics: 'APA guidelines for in-text citations', 'embedding quotations into your writing' and 'making sense of contradictory research findings'. All students received the same four-page text. To make sure students read their study texts carefully, each section was tested by a multiple-choice question. Students had 30 minutes to read the study text and answer the accompanying questions (n = 5). At the end of the morning session, students had to hand in their study texts to make sure they had all spent the same amount of time on the writing course's learning content. The study text on academic writing will subsequently be referred to as MC-test.

The writing task

The writing task in this study was a mini version of "a narrative critical literature review" (Jesson & Lacey, 2006, p. 142). We defined three learning goals: (1) teaching students that "a literature review is not a list describing or summarising one article after another" (Jesson & Lacey, 2006, p. 143), but a cognitively complex text in which different publications on the same topic are juxtaposed in one paragraph (Granello, 2001; Jesson & Lacey, 2006); (2) teaching students how to incorporate quotations in their own text; and (3) teaching students to write a coherent summary with a clear introduction, body and conclusion.

The summary was based on three empirical studies. Students did not have to read these publications themselves. For each study, we summarized the following five elements on an index card: the study's research topic, the research design, the composition of the sample, a description of the variables involved in the study and the main research results. One of the index cards also contained a quote. The information was adopted from existing studies but simplified for didactic and practical reasons. The research results of one study differed substantially from the other two. Students had to incorporate valid arguments in their texts for this contradiction in the research results.

Video with prestige model

Students in both prestige-model-conditions (see Table 1) saw a two-minute video of a male adult sitting behind a desk. He was introduced to them as the managing director of Data Collection, a market research company. In the video, which was designed as an interview, the prestige model talked about various research skills he thought to be important for his employees, e.g. analyzing and synthesizing large amounts of information, skipping details, dealing with conflicting information and points of view, and giving credit to authors. During the interview, keywords of the interviewee appeared on the screen (see Figure 2).

Figure 2. Screenshots from the prestige model video with highlighted keywords.

Instructional video

For this study, we developed an instructional video in which a peer model demonstrated a five-step writing strategy for writing up a synthesis of multiple research studies. The model was an actor provided with a script and instructed to complete the writing task under think aloud conditions (cf. procedure in Raedts et al., 2007).

The video was created and produced with Camtasia. This software tool allows its users to record screen activities in real time and combine this 'movie' with external video and audio recordings. Figure 3 shows a screenshot from our video.

Figure 3. Screenshot from the instructional video.

The majority of the screen was taken by the Camtasia screen capture recordings. In the bottom-right corner we added an additional video clip showing the peer model sitting at a desk behind his laptop. Hence, the students of our intervention study could link pauses in the peer model's writing process on the computer screen to his external writing activities (e.g. rereading information on the index cards). Both recordings gave the participants insight into the model's planning, drafting and revising strategies. They saw how he dealt with the information on the index cards, how he constructed the outline for his text, where he paused during the writing process and which revisions he made. "By compiling the different types of audio–visual information in one mosaic video image, the observers did not only see how the text was constructed and which writing strategies the model used, they also heard what the model thought" (Raedts et al., 2007, pp. 229-230). Hence, information on the new writing task was presented to the students through different channels, that is, visual and auditory (cf. the dual-channel assumption of Mayer & Moreno, 2003).

The first draft of the peer model's summary contained 'errors' that are common to novice academic writing. For instance, some formulations were too informal or too

formal. Second, the model was not completely familiar with the APA guidelines for in-text citations. Third, his draft contained superfluous details. Finally, there was no concluding sentence. We added these elements in the video to make sure the model would not be too perfect. This procedure of starting from an incomplete or imperfect first draft as a starting point for revision and reflection by a peer model was also based on Van Steendam et al. (2010). In this way, we hoped that the students would identify with the model's writing approach and would perceive him as a good writer.

To facilitate retrieval of the model's writing strategy, we developed a mnemonic: the five-letter word TRACE. Each TRACE-letter represented one step in the writing process that was summarized in one sentence (e.g. Introduce the research topic). We made separate TRACE-versions for the planning stage and for the drafting and revising stage (see Table 2).

Table 2. The mnemonic TRACE for the five-step writing strategy

Planning stage		Drafting and revising stage	
T	Find the common research topic.	T	Introduce the research topic.
R	Compare the research results.	R	Present the research results.
A	Provide arguments for seemingly contradictory research results	A	Add arguments for seemingly contradictory research results.
C	Read the quotation carefully and paraphrase its content[a].	C	Embed the quotation in your text[a].
E	First, create an outline, then write your text[b].	E	Evaluate the quality of your text.

Note. [a] The Dutch word for 'quotation' is 'citaat', hence the C in step 4. [b] The Dutch word for 'first' is 'eerst', hence the E in step 5.

We constructed two versions of the video. In the implicit-strategy-instruction-video students saw the peer model tackling the writing task. The peer model thought out loud while planning, writing and revising his synthesis. He did not explicitly mention the different steps in the TRACE-strategies. The length of the video was 17 minutes and 46 seconds. Consequently, participants only observed excerpts of the model's writing process. In the explicit-strategy-instruction-video we added extra slides on which the five-step-strategy of the model was made explicit (see Figure 4). These extra slides were meant to catch and funnel viewers' attention and to enhance retention of the different steps in the strategies modelled. Due to these extra frames the explicit-strategy-instruction-video was 1 minute and 46 seconds longer than the implicit-strategy-instruction-video.
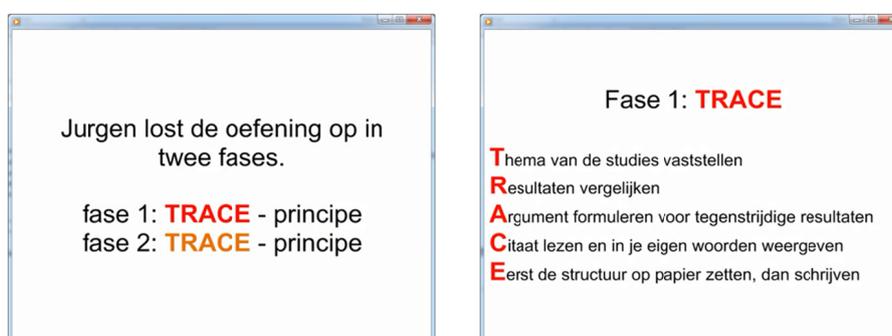
Figure 4. Screenshots from the explicit-strategy-instructional video.

Instrumentation

Writing performance. We developed three writing tasks of increasing complexity. The complexity can be characterized in terms of (1) the amount and abstractness of information on the index cards, (2) the obviousness of the arguments for the contradictory research results, (3) the possibilities for fitting the quotation into the text and (4) the abstraction level of the conclusion sentences. Students had 150 minutes to complete the three writing tasks.

We chose topics in line with students' personal interests (writing task 1: parental influences on early adolescent initiation of smoking) and students' study program, (mis)comprehension of English in standardized advertising messages (writing task 2) and the impact of negative celebrity endorser information on brand image (writing task 3). All three texts were limited to 250 words (see Appendix A for the verbatim instructions for writing task 1). Two examples of students' writing are included in Appendix B. We used three similar writing tasks and three different raters because research shows that "multiple tasks and multiple raters are necessary in order to generalize text quality scores to writing proficiency in a specific genre" (Bouwer, Béguin, Sanders, & van den Bergh, 2015, p. 93).

Students' writing tasks were scored on criteria widely considered in writing research on synthesis writing as decisive criteria to evaluate quality of the genre as becomes evident from state-of-the-art publications by Mateos, Martín, Villalón & Luna (2008), Mateos & Solé (2009) and Solé, Miras, Castells, Espino, & Minguela (2013): to be more precise, the degree of selection of (relevant) information; the correctness of the information selected; the level and quality of integration of the selected ideas (including a summarizing main idea linking the different ideas); the level of elaboration (including the quality of the paraphrasing) and textual organization. In this study these criteria were operationalized as five quality indicators or criteria: (1) the introduction of the research topic in the first sentence(s) of the text, (2) the summary and juxtaposition of the research findings; (3) the argumentation quality for the contradictory research

findings (cf. level of elaboration), (4) the implementation of the quote in their own text, and (5) the quality of the conclusion. The fourth criterion, specific to the study, partly corresponds to quality criteria for paraphrasing and quoting. The fifth criterion reflects an aspect of textual organization. As language (spelling, grammar, word choice and usage for example) and mechanics were not subject to study, they were not included as criteria.

Each criterion was rated by two of the three first authors. Coding was done following a double blind procedure: each rater rated all anonymized syntheses for a specific criterion in a random order (different for both raters) to eliminate order effects and rater biases (Rijlaarsdam et al., 2012). Both raters evaluated the text quality on a three- or four-point scale. Figure 5, for example, shows the scoring instructions for the criterion 'introduction of the research topic'.

Score	Instruction
0	The common research topic is not mentioned in the first, second or third sentence of the summary.
1	The common research topic is mentioned in the first, second or third sentence of the summary. The description of the research topic, however, is not specific enough and/or incomplete.
2	The common research topic is mentioned in the first, second or third sentence of the summary. The description of the research topic is specific enough and complete.

Figure 5. Scoring rubric for criterion 'introduction of the research topic'.

Raters used scoring rubrics which contained benchmark examples to illustrate specific scores. Research shows that raters evaluate text quality more reliably when they can compare texts "against … fixed example texts (i.e., benchmarks) that represent the range in text quality (Blok, 1986)" (as cited by Bouwer, Koster, & Van den Bergh, 2016). A translated version of the full scoring rubrics for writing task 1 is included in Appendix C. Benchmark examples have been included to illustrate the quality difference between the scale points. These scoring rubrics with benchmarks for each of the three writing tasks were developed in separate sessions.

Inter-rater reliability was good to excellent. The correlation coefficients ranged from .78 (criterion 'conclusion' of writing task 1) to .91 (criterion 'incorporation quote' of writing task 1). Scores for the five different criteria were averaged and then totalled. These totalled scores were used in the analyses.

To illustrate the reliability and the validity of the analytic scoring system by the internal raters, three additional external raters unfamiliar with the study rated a randomly chosen subsample of 135 synthesis texts (45 syntheses per writing task) with a holistic quality score for content and organization. Each rater rated (the 45 syntheses of) a single writing task.

The 135 texts were rated with a benchmarking scoring method, shown to have a high generalizability, validity and reliability (Blok, 1985, 1986; Bouwer et al., 2016; Tillema, Van den Bergh, Rijlaarsdam, & Sanders, 2012). For each writing task raters were given a benchmark essay of average text quality with regard to content and organization (one per task and per rater). These benchmarks were developed in separate sessions and considered by both a number of experts and by the raters to be a good representation of average text quality with regard to criteria for the content and organization of synthesis texts. On the benchmarks average text quality characteristics were explicitly described. Following prior research in which benchmarking is used (Schoonen, 2012; Van den Bergh, De Maeyer, Van Weijen, & Tillema, 2012), the benchmark essay was given an arbitrary score of 100. Raters had to compare all 45 student syntheses of a specific writing task to this benchmark, position and rate them relative to the benchmark synthesis text, that is, assign a higher rating than 100 if the quality of the synthesis-to-be-rated was better and a lower score than 100 if the quality of the synthesis text was considered to be lower. All possible scores could be given.

The raters were two graduate students and a highly experienced academic writing teacher at university level. The two graduate students had considerable experience rating the genre under investigation and one of them had taught academic writing to university students prior to the study. All three raters were experienced in rating synthesis texts using a benchmarking system.

The two graduate student raters were trained in two separate sessions, each lasting approximately 2 hours. During these sessions, the different synthesis tasks were explained, random subsamples of synthesis texts of differing quality were discussed, benchmarks were selected, compared and discussed and raters were trained in the specific scoring procedure and in rating until they felt confident about the rating at hand. The third rater attended the first session, participated in rating example texts of tasks 1 and 2 and then independently rated the 45 syntheses of writing task 3.

Correlations between the external raters' holistic benchmarking scoring system and the analytic scores are $r = .762, p < 0.01$ (rater 1 writing task 1), $r = .807, p < 0.01$ (rater 2 writing task 2) y and $r = .70, p < 0.01$ ('untrained' rater 3 writing task 3) respectively. These statistically significant correlations indicate that a higher score for text quality with the analytical coding system corresponds to a higher holistic quality score and is a reliable indicator of global quality of the content and organization of the genre under investigation.

We assessed students' *self-efficacy beliefs* twice: once before and once after they saw the instructional video with the peer model. The self-efficacy scale consisted of five items: 'I can draw a correct and complete picture of the studies' research findings'; 'I can deal with contradictory research findings'; 'I can leave out superfluous details and concentrate on the relevant information'; 'I can incorporate the quotation in my text' and 'I can provide a summarizing conclusion'. Students rated the strength of their beliefs on a 100-point scale, ranging from 0 (*I am not sure at all I can do this*) to 100 (*I*

am highly certain I can do this). The task-specific items were constructed following Bandura's guidelines for constructing self-efficacy scales (Bandura, 2006). We obtained Cronbach's alpha coefficients of .76 (pretest measure) and .85 (posttest measure).

Logical reasoning ability. Writing a research synthesis presupposes complex cognitive activities. Therefore, we included a test that measures students' ability to think conceptually and analytically: the Dutch translation of the AH56-L intelligence test (Minnaert, 1996). The test contains 72 multiple-choice questions of three different types (sequences, relations and analogies). It measures verbal, numerical and diagrammatic reasoning. Subscores are added up to an overall score. Minnaert and Janssen (1999) report an internal consistency coefficient of .78 (N=592). Masui, Broeckmans, Doumen, Groenen, and Molenberghs (2014) report an internal consistency coefficient of .73 (N=93). The predictive validity with respect to freshmen's overall academic performance and with respect to grades for individual courses has been established in several samples (R2=.12 to .16; e.g., Masui, 2002; Masui at all, 2014; Minnaert & Janssen, 1999). The test was organized at the beginning of the academic year (Cronbach's alpha for the 72 items was .72).

We measured students' *task value* at the beginning of the morning session (pretest) and at the beginning of the afternoon session (posttest). The items were adapted from Artino (2007). They can be divided in four dimensions: intrinsic interest (4 items), attainment value/importance (3 items), extrinsic utility value (5 items) and cost (2 items). Sample items include: 'I like(d) the subject matter of this course'; 'It is/was important for me to learn the material in this course'; 'In the long run, I will be able to use what I learn/learned in this course'; 'The work I put into this course, is/was worth the effort'. Cronbach's alpha' coefficients were .94 (pretest) and .93 (posttest measure).

Students' *task knowledge* was elicited in an indirect way. We had the students write an e-mail to a fictitious peer student who was absent during the morning session. Students were asked to write down five effective planning strategies to write up a good synthesis text and five text features their peer student specifically had to consider. The ten listed items were classified and coded by the first three authors. Disagreements among the raters were discussed until agreement was reached. The students' answers were categorized in one of the following three categories: (1) the item did not relate to one of the steps in the TRACE-strategy (e.g., 'Write an introduction'); (2) the item described one of the five steps of the TRACE-strategy, the wording however, was vague (e.g., 'Research topic'); and (3) the item clearly described one of the five steps of the TRACE-strategy (e.g., 'Start with an introduction in which you write down the research topic of the studies'). Interrater reliability was good to excellent and correlation coefficients ranged from $r = .709$ to .996 (all $p < .01$).

5.4 Data Analysis

As the theoretical model consists of several direct and indirect effects to predict writing performance, path analysis techniques are the most suitable method (Stage, Carter, & Nora, 2004). To report the model fit of the path model, we follow the guidelines by Hooper, Coughlan, and Mullen (2008). They advise to report a chi-square statistic with its degrees of freedom, the RMSEA and its confidence interval, the SRMR, the CFI and a parsimony fit index (for instance PNFI). Since parsimony fit indices do not have threshold levels and are difficult to interpret (Hooper et al., 2008), we will not report them here.

A chi-square likelihood ratio test measures the overall fit of the model and compares the likelihood value of the estimated model with a 'perfect' model. A good model should not reject the chi-square test at a 5% significance level, even though it is sensitive to sample sizes, in the sense that for large samples the test is usually always rejected. The Root Mean Squared Error of Approximation (RMSEA) is a 'badness of fit' index and one of the most popular and most reported fit indices, because it also takes into account the sample size and the number of parameters in the model. Values less than 0.07 are acceptable (Steiger, 2007). The Root Mean Square Residual (RMR) is based on the difference between observed and predicted covariances, but it is difficult to interpret, since it depends on the scales of the variables. The Standardized Root Mean Square Residual (SRMR) resolves this problem and takes values between 0 and 1. Good models have a value less than 0.05, but values less than 0.08 are acceptable. The Bentler's Comparative Fit Index (CFI) assesses the relative improvement in fit of the estimated model compared with a baseline model. If the CFI is larger than 0.95, the model is considered better than the null model. All path models were calculated using the software package AMOS (Arbuckle, 2013). Standardized path coefficients are reported, as well as the proportion of variance explained for each endogenous variable.

6. Results

Descriptive statistics for pre- and posttest measures for the four conditions can be found in Table 3. No statistically significant differences could be observed between students in the different conditions for the MC-test on academic writing ($F(3, 158)= .882$, $p = .452$, $\eta_p^2 = .016$), for logical reasoning ability ($F(3, 158)= 1.281$, $p = .283$, $\eta_p^2 = .024$), and for self-efficacy ($F(3, 158)= 1.557$, $p = .201$, $\eta_p^2 = .029$). However, for task value a statistically significant pretest difference between the different conditions was observed ($F(3, 158)= 3.471$, $p = .018$, $\eta_p^2 = .062$). Bonferroni posthoc tests revealed that Condition 2 had a higher task value than Condition 1 ($p = .016$, Cohen's $d = .69$). To control for these a priori differences we included difference scores in the statistical model. Students can perform differently on the pretest. However, it is their improvement on task value that interests us, which is measured by the difference scores. An ANOVA test showed no significant differences between students in the four

conditions for the difference scores on task value (F(3, 158)= 0.089, p = .966, η_p^2 = .002).

Table 3. Descriptive statistics (means and standard deviations) for pretests and posttests for the four conditions

Measurement	Condition 1 +PM ESI N = 40	Condition 2 -PM ISI N = 38	Condition 3 +PM ISI N =34	Condition 4 -PM ESI N = 50
MC-test[a]	3.60	3.63	3.85	3.76
	(.77)	(.75)	(.74)	(.77)
Self-efficacy pretest[b]	58.78	63.45	59.62	59.72
	(10.52)	(9.51)	(10.02)	(11.33)
Self-efficacy posttest[b]	60.84	64.87	59.62	60.63
	(10.09)	(11.52)	(12.27)	(11.18)
Self-efficacy: diff between post- and pretest	2.07	1.43	.33	.92
	(8.33)	(7.12)	(8.48)	(10.70)
Task value pretest[c]	3.75	4.40	3.95	4.17
	(.93)	(.94)	(.94)	(.97)
Task value posttest[c]	3.86	4.57	4.04	4.31
	(.88)	(.97)	(.82)	(.99)
Task value: diff between post- and pretest	.11	.17	.09	.14
	(.60)	(.61)	(.74)	(.80)
Logical reasoning ability [d]	36.78	34.37	36.12	34.30
	(7.84)	(7.51)	(6.70)	(6.32)
Task knowledge[e]	13.15	10.08	9.29	13.04
	(3.87)	(3.32)	(3.07)	(4.31)
Writing performances: writing task 1[f]	70.25	67.36	64.70	68.40
	(15.77)	(17.35)	(13.08)	(18.22)
Writing performances: writing task 2[f]	50.21	48.25	50.00	47.17
	(20.19)	16.11)	(18.12)	(16.46)
Writing performances: writing task 3[f]	42.92	40.13	41.91	38.50
	(19.57)	(15.61)	(17.47)	(16.91)

Notes. PM = prestige model (+PM = condition with prestige model vs. -PM = without prestige model) / ESI = explicit strategy instruction / ISI = implicit strategy instruction. Standard deviations appear in parentheses below means. [a]Maximum score is 5. / [b]Maximum score is 89.40. / [c]Maximum score is 6.64. / [d]Maximum score is 54/[e]Maximum score is 20/ [f]Maximum score is 100.

Students had to perform three writing tasks of increasing complexity. We will present a different path model for each writing task because the writing scores of the three writing tasks only showed moderate positive correlations: $r = .409$ (correlation between the scores on writing task 1 and writing task 2), $r = .399$ (correlation between the scores on writing task 2 and writing task 3) and $r = .428$ (correlation between the scores on writing task 1 and writing task 3.

First, we present the path model for the first, least complex writing task in Figure 6. It has a non-significant chi-square value ($\chi^2 = 8.658$, $df = 11$, $p = .65$) and a Root Mean Square Error of Approximation (RMSEA) of 0.000, with [0.000, 0.068] as its 90% confidence interval. Furthermore, the Standardized Root Mean Square Residual (SRMR) is 0.0326 and the Comparative Fit Index (CFI) is 1.000. These are all excellent Goodness-of-Fit indices.

There is no significant direct effect from the video with explicit strategy instruction to writing performance, but there is a strong significant direct effect from this instructional video to task knowledge ($\beta = 0.411$) and from task knowledge to writing performance ($\beta = 0.411$), so we can conclude that there is a total effect from the video with explicit strategy instruction to the writing performance, primarily through task knowledge ($\beta = 0.099$).
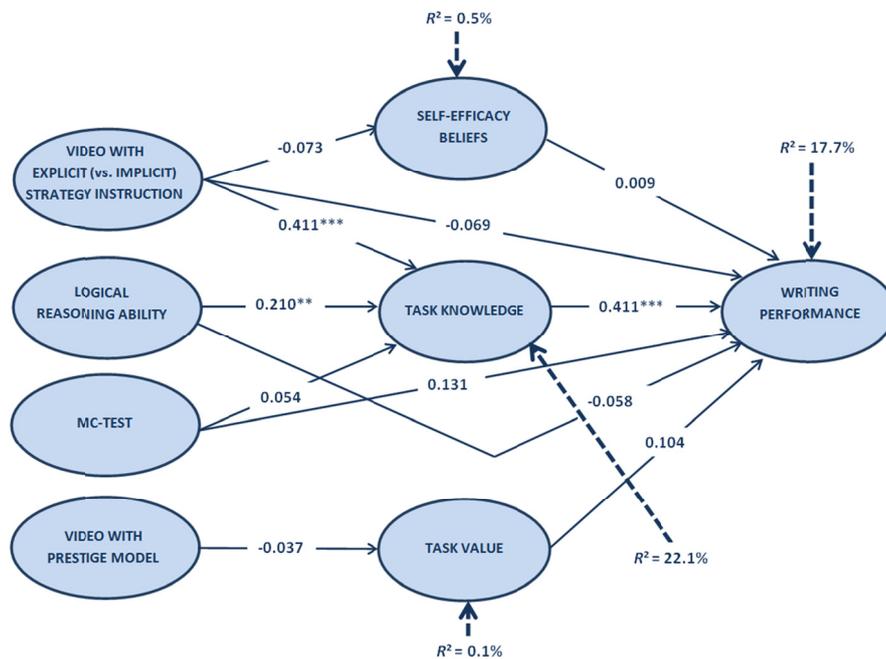
Figure 6. Final path model showing standardized path coefficients and proportion of variance explained for the first writing task. *Note.* ** $p <.01$ / *** $p <.001$.

The same applies to logical reasoning ability. There is no significant direct effect to writing performance, but there is a strong significant direct effect to task knowledge (β = 0.210), so there is a total effect to writing performance, mainly through task knowledge (β = 0.028). The other relationships are not significant, which means that we do not find a significant direct effect from the video with explicit strategy instruction to self-efficacy beliefs, from self-efficacy beliefs to writing performance, from the video with the prestige model to task value; from task value to writing performance, or from the MC-test to task knowledge or to writing performance. Overall, 17.7% of the variability in writing performance is explained by the path model. On the other hand, the video with the explicit strategy instruction and logical reasoning ability included accounts for 22.1% of the variability in task knowledge.

The model for the second writing task (Figure 7) has a non-significant chi-square value (χ^2= 9.029, df = 11, p = .62), a RMSEA of 0.000, that lies with 90% confidence between 0.000 and 0.071, a SRMR of 0.0324 and a CFI of 1.000. These results are very close to the results for the first writing task.
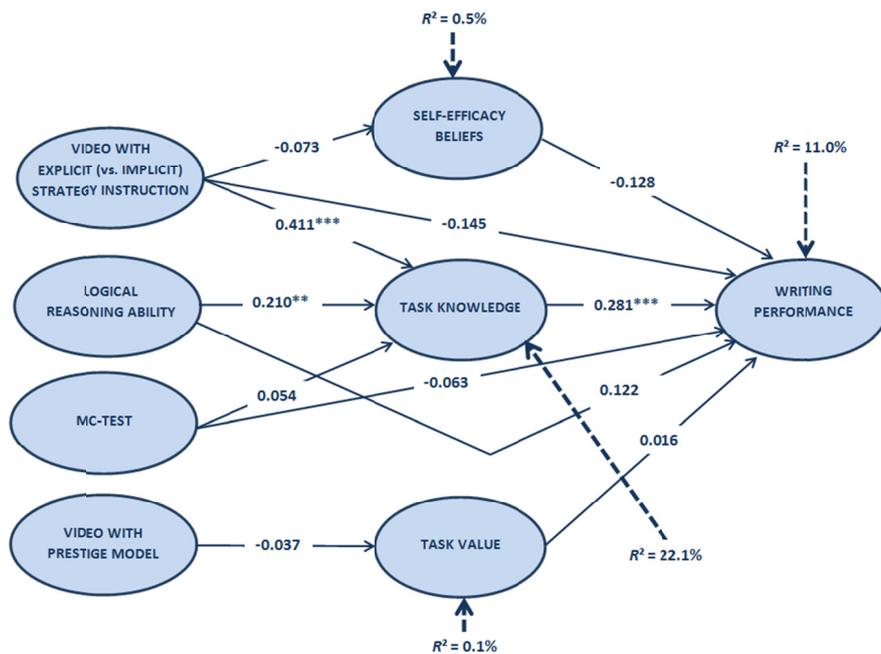
Figure 7. Final path model showing standardized path coefficients and proportion of variance explained for the second writing task. *Note. ** p <.01 / *** p <.001.*

Regarding the direct effects, again we find a strong significant direct effect to task knowledge from the video with explicit strategy instruction (β = 0.411) and from logical reasoning ability (β = 0.210), and from task knowledge to writing performance (β =

0.281). In total, 11.0% of the variability in writing performance is explained by the model.

Finally, the results for the third writing task are presented in Figure 8. The model has a non-significant chi-square value (χ^2 = 10.118, df = 11, p = .52) and a Root Mean Square Error of Approximation (RMSEA) of 0.000, with [0.000, 0.078] as its 90% confidence interval. Furthermore, the Standardized Root Mean Square Residual (SRMR) is 0.0341 and the Comparative Fit Index (CFI) is 1.000. We can draw exactly the same conclusions as in the second model.
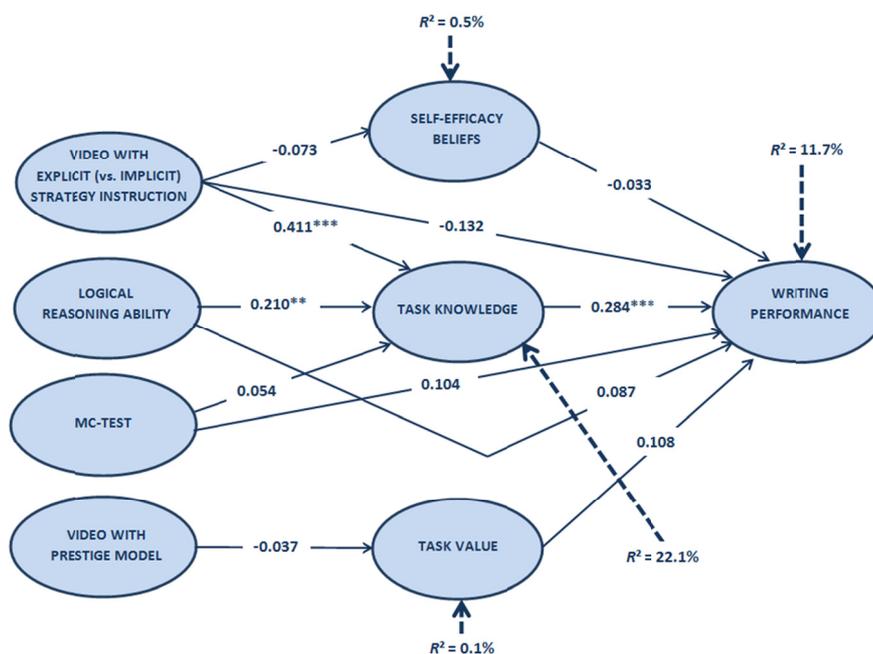
Figure 8. Final path model showing standardized path coefficients and proportion of variance explained for the third writing task. *Note*. ** p <.01 / *** p <.001.

From the different path models for the three writing tasks of increasing complexity it can be concluded that the explicit-strategy-instruction-video and logical reasoning ability have no significant direct effect on writing performance. However, they have an effect on task knowledge, which in turn has an effect on writing performance. Hence, there is an indirect effect from the video with explicit strategy instruction and logical reasoning ability to writing performance, via task knowledge. This might raise the question if logical reasoning ability has the same effect on task knowledge for students who received the video with explicit strategy compared to students with the implicit strategy video. That is why we performed a multiple linear regression analysis with task knowledge as dependent variable and as independent variables logical reasoning

ability, the video with explicit (versus implicit) strategy instruction and the product of these two to test the interaction. The results showed no significant interaction effect ($t = 1.631$, $df = 158$, $p = .105$, $R^2 = .231$), meaning that the effect of logical reasoning ability on task knowledge is not significantly different for students with the video with explicit strategy instruction compared to students with the implicit strategy instruction. Test statistics and regression coefficients are shown in Table 4.

Table 4. Statistics (regression coefficients, standard errors (SE), test statistics and significance value) of a multiple linear regression analysis for logical reasoning ability and explicit-/implicit-strategy instruction-video

	b	SE	t	p
Constant	9.714	.426	22.795	***p <0.001
Video with explicit (versus implicit) strategy instruction	3.358	.572	5.873	***p <0.001
Logical reasoning ability (centered)	.053	.060	.878	.381
Video × Logical reasoning ability	.132	.081	1.631	.105

Notes. Dependent variable : Task knowledge

The variable *Logical reasoning ability* was centered around its mean to avoid multicollinearity, which has no effect on the significance of the interaction term.

7. Conclusion

This study extends previous research on observational learning in writing instruction in two respects. First, we conducted a study on an academic writing task with first-year university graduates, a target audience which has not been widely studied (Raedts et al., 2007; Van Steendam et al., 2010). Secondly, to the best of our knowledge, no previous study combined the effects of several types of modelling in one experimental design: a first type of modelling directed at task value and a second type of modelling differing in degree of explicitness in instructional content.

Results show no significant effects of the first type of model (a prestige model) on task value (*Answer to RQ1*). Students who saw the video with the prestige model did not differ to a statistically significant degree in task value from the students who did not observe the prestige model. A possible explanation for the absence of a significant effect for task value could be that the time lapse between showing the video with the prestige model and the post-test measure itself was too large. It is also possible that the message from the prestige model about the value of the writing task did not affect these first-year students. First, their performances on the writing tasks did not count towards their final grade. Secondly, this group of students may not be concerned about their future professional careers or may not believe that learning to write is important for their future careers.

With regard to the instructional video, the results show that explicit strategy instruction via attention-catching and retention-enhancing elements had a positive, direct effect on students' task knowledge and an indirect effect on their writing performances compared to the video with the implicit type of modelling (*Answer to RQ2*). Apparently, the explicit form of modelling due to the attention-catching and retention-enhancing cues and the explicit strategy steps in the form of mnemonics was more successful than the implicit form of modelling in contributing to students' learning cf. the effect for task knowledge and its subsequent impact on writing performance. This finding confirms results of previous studies in which more explicit and direct forms of modelling were used (e.g. SRSD-studies discussed in Graham & Perin, 2007, Van Steendam et al., 2010). It is also a confirmation that retention is an important step towards successful production (Bandura, 1986), also in writing. That writing performance depends on task knowledge and task-specific strategies also corroborates research by Schoonen en De Glopper (1996) and Raedts et al. (2007). These findings are important for instructional designers and teachers in several ways. When using videobased instruction it is first of all worth to consider using peers as a model because they are credible coping models and can enhance knowledge and skills. It is furthermore important to keep in mind the ways in which designers and teachers can reduce cognitive overload by using instructional video. Information should be presented through different channels (Mayer & Moreno, 2003) and retention should be enhanced by providing structures or mnemonics.

Contrary to expectations, we did not find an effect of the more explicit type of modelling on students' writing self-efficacy beliefs even though a significant effect in writing performance, that is, a higher achievement level, should also correspond to significantly higher self-efficacy beliefs (Bandura, 1997). That this is not the case may be explained by the fact that students did not receive any feedback on their writing performances as could also be observed in Raedts et al. (2007). It is also possible that students had inaccurate estimations of their self-efficacy and that the observational learning experience was unable to correct these beliefs (Zimmerman & Kitsantas, 2002).

Finally, with regard to the role of logical reasoning ability, the results show that logical reasoning ability plays an important role as it has a positive effect on writing performance via task knowledge (cf. positive correlation of logical ability and task knowledge). However, this effect does not differ between the explicit and implicit form of modelling (*Answer to RQ3*), showing that students who score higher on logical reasoning ability have a higher task knowledge which results in a better writing performance regardless of the form of modelling they receive(d).

8. Future research

In this section, we briefly discuss some directions for future research. The first direction concerns the effect of prestige models on students' task value. Future studies could

incorporate first-year undergraduates' future time perspective as research from Husman (e.g. Husman & Shell, 2008) shows that it is important to consider future thinking in educational research.

Our research findings do not support previous studies examining the relationship between undergraduate students' self-efficacy beliefs and their writing performance (Prat-Sala & Redford, 2012; Pajares & Johnson, 1994; Shell, Murphy, & Bruning, 1989). This may be explained by the absence of feedback as mentioned and the fact that the task tested was a new task. Future studies could incorporate feedback on students' writing performance by a writing teacher to investigate if it makes students' self-efficacy beliefs more accurate.

The writing model only explained 10 to 15% of the variance in writing performance. Future research could include other variables such as students' reading proficiency or reading comprehension as studies show that the variables may have a significant impact on the academic research synthesis writing task (Spivey, 1997). Solé et al. (2013) for example included after a synthesis writing task a reading comprehension test of the sources students had to consult for writing their synthesis task and found a correspondence between synthesis text quality, synthesis writing processes and level of understanding of the source texts. Nevertheless, as the sources in the present study were summaries of empirical studies on index cards simplified for didactic and practical reasons, we do not expect reading proficiency in this study to play such a crucial role compared to synthesis writing studies in which the sources are not adapted or simplified. Other motivational variables from the social cognitive framework (Pintrich, 2003) such as attribution or intrinsic motivation could be included in a future study to possibly explain some of the unexplained variance. Several studies revealed a connection between attribution style and performance (e.g. Erten & Burden, 2014). To increase the model's explanatory effect, the social cognitive theoretical framework about motivation could be complemented with the self-determination theoretical framework about intrinsic motivation, as suggested by some authors (e.g. Vansteenkiste, Lens, De Witte, & Feather, 2005).

Follow-up studies could also address some of the limitations of the current study. First of all, following prior research by Raedts et al. (2007), Braaksma (2002) and Braaksma, Rijlaarsdam, van den Bergh, & Van Hout-Wolters (2004), the present study did not include a writing pretest as the genre under investigation was a new genre. In their 'Twelve recommendations for conducting high-quality writing intervention research' Graham and Harris (2014) point out the difficulty of testing or assessing "behaviors that students engage in rarely" (p. 104) such as revising content and structure for example, which may result in floor effects "preventing the researcher from precisely determining the impact of the target intervention" (p. 104). This line of reasoning also applies to the testing of writing proficiency or performance in a completely new genre students are not familiar with. Graham and Harris propose three possibilities: (1) not to pretest "students on this variable (assuming that students' performance is low and skewed" (p. 104); (2) "to develop an alternative measure to

eliminate floor effects" (p. 104) or (3) administering a pretest "acknowledging that it is flawed and tempering … conclusions accordingly" (p. 104). In this study we decided not to have students write a full synthesis task as it was a new genre but instead to administer a test measuring students' ability to think conceptually and analytically (the AH56-L intelligence test by Minnaert, 1996), considered to be important, if not crucial, for the cognitively complex activities of analyzing, integrating and synthesizing research findings of different studies into a coherent text. In addition, different other pretests were administered believed to potentially distinguish between students in the four, randomly created, conditions. Nevertheless, future studies could, next to including an intelligence test, also include a pretest measure for writing quality to control for potential a priori differences between students in different experimental conditions. How this could be achieved in studies testing the effect of an intervention on a new genre, however, remains no small feat.

Such studies could also include a higher sample size as "a sample size of at least 200 or 5 or 10 cases per parameter is recommended for path analysis or structural equation modeling (see for instance Kline, 2015). However, some recent simulation studies showed that smaller sample sizes may be sufficient (Wolf, Harrington, Clark & Miller, 2013; Sideridis, Simos, Papanicolaou & Fletcher, 2014). Even with the smaller sample size in this study some interesting statistically significant results are shown.

Another limitation or our study is that no (process) data on students' writing processes were collected. However, this was not possible for administrative and logistic reasons given that the study was organized during students' regular classes. Nevertheless, in a future study we would like to log students' writing processes with keystroke logging like Inputlog (Leijten & Van Waes, 2014), to trace if students actually apply the strategy steps they retained in the task knowledge task. Another avenue for future research would be to explore how students deal with the video and the explicit strategy steps for example by recording their eye movements with an eyetracker (e.g. Azevedo et al., 2012).

Despite its limitations, the present study is one of the first which investigated the effect of a prestige model and which compared the effects of explicit and implicit strategy instruction in an instructional video. We hope our study will inspire writing researchers to validate our results for other text genres and/or younger writers.

Notes

1. The MC-test in the model refers to a study-text on academic writing which was administered prior to the actual intervention and which could be interpreted as being part of prior task knowledge (about in-text citation) cf. Section 5.3 Procedure.

Acknowledgements

References

Alexander, P., Graham, S., & Harris, K. (1998). A perspective on strategy research: Progress and products. *Educational Psychology Review*, *10*(2), 129-154. doi: 10.1023/A:1022185502996

Arbuckle, J. L. (2013). *Amos* (Version 22.0) [Computer Program]. Chicago: SPSS.

Artino, A.R. (2007). Online military training: Using a social cognitive view of motivation and self-regulation to understand students' satisfaction, perceived learning, and choice. *Quarterly Review of Distance Education*, *8*(3), 191–202.

Azevedo, R., Landis, R., Feyzi-Behnagh, R., Duffy, M. Trevors, G., Harley, J., … Hossain, G. (2012). The effectiveness of pedagogical agents' prompting and feedback in facilitating co-adapted learning with MetaTutor. In S. A. Cerri, W. J. Clancey, G. Papadourakis & K. Panourgia (Eds*.), Intelligent Tutoring Systems (ITS) 2012: Lecture Notes in Computer Science 7315* (pp. 212-221). Heidelberg: Springer. doi: 10.1007/978-3-642-30950-2_27

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

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307-337). Greenwich, CT: Information Age Publishing.

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

Bethards, M. (2014). Applying social learning theory to the observer role in simulation. *Clinical Simulation in Nursing, 10*(2), e65-e69. doi: 10.1016/j.ecns.2013.08.002

Blok, H. (1985). Estimating the reliability, validity, and invalidity of essay ratings. *Journal of Educational Measurement, 22,* 41-52. doi: 10.1111/j.1745-3984.1985.tb01048.x

Blok, H. (1986). Essay rating by the comparison method. *Tijdschrift voor Onderwijsresearch, 11,* 169-176.

Boscolo, P., Arfé, B., & Quarisa, M. (2007). Improving the quality of students' academic writing: An intervention study. *Studies in Higher Education, 32*(4), 419-438. doi: 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. (2016). *Benchmark rating procedure, best of both worlds? Comparing procedures to rate text quality in a reliable and valid manner. Manuscript submitted for publication.*

Braaksma, M. (2002). Observational learning in argumentative writing. Unpublished dissertation. University of Amsterdam.

Braaksma, M.A., Rijlaarsdam, G., & Van den Bergh, H. (2002). Observational learning and the effects of model-observer similarity. *Journal of Educational Psychology*, *94*(2), 405-415. doi: 10.1037/0022-0663.94.2.405

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

Bridgeman, B., & Carlson, S.B. (1984). Survey of academic writing tasks. *Written Communication*, *1*(2), 247-280. doi: 10.1177/0741088384001002004

Bruning, R., & Horn, C. (2000). Developing motivation to write. *Educational Psychologist*, *35*(1), 25-38. doi: 10.1207/S15326985EP3501_4

Campell, J., Smith, D., & Brooker, R. (1998). From conception to performance: How undergraduate students conceptualize and construct essays. *Higher Education*, *36*(4), 449–469. doi: 10.1023/A:1003451627898

Cooper, A. (2007). Writing at the graduate level: What tasks do professors actually require? *Journal of English for Academic Purposes*, *6*(3), 206-221. doi:10.1016/j.jeap.2007.09.008

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

De Grez, L., Valcke, M., & Roozen, I. (2009). The impact of an innovative instructional intervention on the acquisition of oral presentation skills in higher education. *Computers & Education, 53*(1), 112-120. doi:10.1016/j.compedu.2009.01.005

Englert, C.S., Mariage, T.V., & Dunsmore, K. (2006). Tenets of sociocultural theory in writing instruction research. In C.A MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 208-221). New York: Guilford Press.

Englert, C. S., Raphael, T. E., Anderson, L. M., Anthony, H. M., & Stevens, D. D. (1991). Making writing strategies and self-talk visible: Cognitive strategy instruction in writing in regular and special education classrooms. *American Educational Research Journal*, *28*(3), 337-372. doi: 10.1016/j.system.2014.01.006

Erten, I., & Burden, R. (2014). The relationship between academic self-concept, attributions, and L2 achievement. *System, 42,* 391-401. doi: 10.1016/j.system.2014.01.006

Fagundes, J., Chen, D.D., & Laguna, P. (2013). Self-control and frequency of model presentation: Effects on learning a ballet passé relevé. *Human Movement Science, 32*(4), 847-856. doi:10.1016/j.humov.2013.03.009

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. doi:10.1016/j.cedpsych.2014.11.004

Froese, A. D., Gantz, B. S., & Henry, A.L. (1998). Teaching students to write literature reviews: A meta-analytic model. *Teaching of Psychology*, *25*(2), 102–105. doi: 10.1207/s15328023top2502_4

Graham, S., & Harris, K.R. (2014). Conducting high-quality writing intervention research: Twelve recommendations. *Journal of Writing Research*, *6*(2), 89-123. doi: 10.17239/jowr-2014.06.02.1

Graham, S., Harris, K.R., & Mason, L. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology*, *30*(2), 207-241. doi:10.1016/j.cedpsych.2004.08.001

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

Granello, D.H. (2001). Promoting cognitive complexity in graduate written work: Using Bloom's taxonomy as a pedagogical tool to improve literature reviews. *Counselor Education & Supervision*, *40*(4), 292–307. doi: 10.1002/j.1556-6978.2001.tb01261.x

Graves, R., Hyland, T., & Samuels, B.M. (2010). Undergraduate writing assignments: An analysis of syllabi at one Canadian college. *Written Communication*, *27*(3), 293-317. doi: 10.1177/0741088310371635

Groenendijk, T., Janssen, T., Rijlaarsdam, G., & Van den Bergh, H. (2013a). Learning to be creative: The effects of observational learning on students' design products and processes. *Learning and Instruction*, *28*, 35-47. doi:10.1016/j.learninstruc.2013.05.001

Groenendijk, T., Janssen, T., Rijlaarsdam, G., & Van den Bergh, H. (2013b). The effect of observational learning on students' performance, processes, and motivation in two creative domains. *British Journal of Educational Psychology*, *83*(1), 3-28. doi: 10.1111/j.2044-8279.2011.02052.x

Hooper, D., Coughlan, J., & Mullen, M.R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods*, *6*(1), 53–60.

Husman, J., & Shell, D. (2008). Beliefs and perceptions about the future: A measurement of future time perspective. *Learning and Individual Differences, 18*(2)*,* 166-175. doi:10.1016/j.lindif.2007.08.001

Jackson, G.B. (1980). Methods for integrative reviews. *Review of Educational Research*, *50*, 438-460. doi: 10.3102/00346543050003438

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. New York: Guilford publications.

Jesson, J., & Lacey, F. (2006). How to do (or not to do) a critical literature review. *Pharmacy Education*, *6*(2), 139-148. doi: 10.1080/15602210600616218

Leijten, M., & Van Waes, L. (2012). Inputlog 4.0: Keystroke Logging in Writing Research. In M. Torrance, D. Alamargot, M. Castelló, F. Ganier, O. Kruse, A. Mangen, L. Tolchinsky & L. Van Waes (Eds.), *Studies in Writing, Volume 25: Learning to write effectively: Current trends in European research* (pp. 363-366). Bingley: Emerald Group Publishing Limited.

Makovsky,V.P. (1985). Teaching psychology in the information age. *Teaching of Psychology*, *12*(1), 23–26. doi:10.1207/s15328023top1201_7

Masui, C. (2002). *Leervaardigheid bevorderen in het hoger onderwijs: Een ontwerponderzoek bij eerstejaarsstudenten [Enhancing learning competence in higher education: A design experiment with university freshmen]*. Leuven, Belgium: Universitaire Pers Leuven.

Masui, C., Broeckmans, J., Doumen, S., Groenen, A., & Molenberghs, G. (2014). Do diligent students perform better? Complex relations between student and course characteristics, study time, and academic performance in higher education. *Studies in Higher Education, 39*(4), 621–643. doi: 10.1080/03075079.2012.721350

Mateos, M., Martín, E., Villalón, R. & Luna, M. (2008). Reading and writing to learn in secondary education: online processing activity and written products in summarizing and synthesizing tasks. *Reading and Writing: an interdisciplinary Journal*, *21*(7), 675-697. doi 10.1007/s11145-007-9086-6

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. doi: 10.1007/BF03178760

Mayer, R.E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43-52. doi:10.1207/S15326985EP3801_6

MacArthur, C.A., Philippakos, Z.A., & Ianetta, M. (2014). Self-regulated strategy instruction in college developmental writing. *Journal of Educational Psychology, 107*(3), 855-867. doi 10.1037/edu0000011

Minnaert, A. (1996). *Academic performance, cognition, metacognition and motivation. Assessing freshmen characteristics on task: A validation and replication study in higher education* (Unpublished doctoral dissertation). Katholieke Universiteit Leuven, Leuven, Belgium.

Minnaert, A., & Janssen, P. J. (1999). The additive effect of regulatory activities on top of intelligence in relation to academic performance in higher education. *Learning & Instruction, 9*, 77-91. doi: 10.1016/S0959-4752(98)00019-X

Pajares, F., & Johnson, M.J. (1994). Confidence and competence in writing: The role of writing self-efficacy, outcome expectancy, and apprehension. *Research in the Teaching of English*, 28(3), 313-331.

Pintrich, P. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, *95*(4), 667-686. doi: 10.1037/0022-0663.95.4.667

Prat-Sala, M., & Redford, P. (2012). Writing essays: does self-efficacy matter? The relationship between self-efficacy in reading and in writing and undergraduate students' performance in essay writing. *Educational Psychology*, *32*(1), 9-20. doi: 10.1080/01443410.2011.621411

Qualtrics [Computer software]. (2012).Provo, UT, USA: Qualtrics.

Raedts, M., Rijlaarsdam, G., Van Waes, L., & Daems, F. (2007). Observational learning through video-based models: Impact on students' accuracy of self-efficacy beliefs, task knowledge and writing performances. In G. Rijlaarsdam (Series Ed.), & P. Boscolo & S. Hidi (Vol Eds.), *Studies in writing, Volume 19: Writing and motivation* (pp. 219–238). Oxford: Elsevier.

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

Rijlaarsdam, G.C.W., van den Bergh, H., Couzijn, M., Janssen, T., Braaksma, M. Tillema, M., Van Steendam, E. & Raedts, M. (2012). Writing. In K.R. Harris, S. Graham & T. Urdan (Eds.), *APA educational psychology handbook. - Vol. 3: Application to learning and teaching* (APA handbooks in psychology) (pp. 189-228). Washington, DC: American Psychological Association. doi: 10.1037/13275-009

Schoonen, R. (2012). The validity and generalizability of writing scores: the effect of rater, task and language. In E. van Steendam, M. Tillema, G. Rijlaarsdam, & H. van den Bergh (Eds.), *Measuring writing: Recent insights into theory, methodology and practices (Vol. 27,* pp. 1-22). Leiden, The Netherlands: Brill. doi: 10.1163/9789004248489_002

Schoonen, R., & De Glopper, K. (1996). Writing performance and knowledge about writing. In G. Rijlaarsdam, H. Van den Bergh & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 87-107). Amsterdam: Amsterdam University Press.

Schunk, D., Meece, J., & Pintrich, P. (2013). *Motivation in education: Theory, research and applications* (4th ed). Upper Saddle River, NJ: Pearson education.

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

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*, 5-33. doi: 10.1023/B:ESLL.0000033847.00732.af

Shell, D.F., Murphy, C.C., & Bruning, R.H. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology*, *81*(1), 91–100. doi: 10.1037/0022-0663.81.1.91

Sideridis, G., Simos, P., Papanicolaou, A., & Fletcher, J. (2014). Using structural equation modeling to assess functional connectivity in the brain. Power and sample size considerations. *Educational and Psychological Measurement*, *74*(5), 733-758. doi: 10.1177/0013164414525397

Snowman, J., & McCown, R. (2012). *Ed Psych. New 1ˢᵗ Editions in Education Series*. Belmont, CA: Cengage Learning/Wadsworth.

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. doi 10.1177/0741088312466532

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

Stage, F. K., Carter, H. C., & Nora, A. (2004). Path analysis: An introduction and analysis of a decade of research. *The Journal of Educational Research, 98*(1), 5-13. doi: 10.3200/JOER.98.1.5-13

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42*(5), 893-898. doi:10.1016/j.paid.2006.09.017

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,* 71-97. doi:10.1177/0265532212442647

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. doi:10.1016/j.learninstruc.2007.02.003

Van den Bergh, H., De Maeyer, S., Van Weijen, D., & Tillema, M. (2012). Generalizability of text quality scores. In E. van Steendam, M. Tillema, G. Rijlaarsdam, & H. van den Bergh (Eds.), *Measuring writing: Recent insights into theory, methodology and practices (Vol. 27)* (pp. 23-32). Leiden, The Netherlands: Brill. doi: 10.1163/9789004248489_003

Van Steendam, E., De Grez, L., Goeman, K., & Frawley, M. (2015, August). *Interactive instructional video for L2 writing in higher education: The impact on student learning*. Paper presented at the 16th biennial EARLI Conference, Limassol, Cyprus.

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. doi:10.1016/j.learninstruc.2009.08.009

Van Steendam, E., Rijlaarsdam, G., & Van den Bergh, H. (2014, August). *The interplay between instruction, group composition and interaction in EFL collaborative revision*. Paper presented at the 14th biennial Earli SIG Writing Conference, Amsterdam and Utrecht, the Netherlands.

Vansteenkiste, M., Lens, W., De Witte, H., & Feather, N. (2005). Understanding unemployed people's job search behaviour, unemployment experience and well-being: A comparison of expectancy-value theory and self-determination theory. *British Journal of Social Psychology, 44*(2), 269-287. doi: 10.1348/014466604X17641

Wigfield, A., & Cambria, J. (2010). Expectancy-value theory: Retrospective and prospective. In S. Karabenick & T. Urdan (Eds.), *Advances in motivation and achievement* (Vol. 16A, pp. 35-70). Bingley, UK: Emerald Group.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models. An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, *73*(6), 913-934. doi: 10.1177/0013164413495237

Zhang, C. (2013). Effect of instruction on ESL students' synthesis writing. *Journal of Second Language Writing*, *22*, 51-67. doi: 10.1016/j.jslw.2012.12.001

Zhu, W. (2004). Writing in business courses: An analysis of assignment types, their characteristics, and required skills. *English for Specific Purposes*, *23*(2), 111-35. doi: 10.1016/S0889-4906(02)00046-7

Zimmerman, B., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology, 94*(4), 660-668. doi:10.1037/0022-0663.94.4.660

Appendix A: Writing Task 1

Please summarize and synthesize the results of the three studies below in a text of 200 to 250 words. The text is part of a larger literature review on smoking behaviour of young people.

Study 1

Author	Ellen Goddard
Year of publication	2009
Objective of study	To explore parental influence on the probability of adolescent smoking initiation
Methodology	Surveys of 3,694 American youngsters (12 to 13 years old at the first data-collection phase (1st cohort at baseline) 3 cohorts: 2006 – 2007 – 2008
Results	Parental influence: large ▪ If one or both parents smoke, the odds that the adolescent starts smoking, are larger ▪ Both parents smoke: 19% of the youngsters started smoking between 2006 and 2008 ▪ One of the parents smokes 17% of the adolescents started smoking ▪ None of the parents smoke: 13% started smoking

Study 2

Author(s)	John Pierce
Year of publication	2005
Study objective	To investigate parental influence on adolescent smoking initiation
Methodology	Survey of 4,502 American 15- to 17-year old adolescents
Results	Parental influence: small Parental quitting of smoking up to 5 years prior to adolescent smoking initiation: very small impact on smoking initiation of son or daughter The timing of parental smoking cessation[1] appears to be important. Parental quitting[2] is most effective[3] in reducing initiation if it occurs[4] before the child reaches 9 years of age. (p. 217)

[1] cessation = to quit smoking
[2] quitting = to stop
[3] effective = achieving its objective, successful
[4] occur = to happen

Study 3

Author(s)	Carine Vereecken and Lea Maes
Year of publication	2010
Study objective	To investigate parental influence on adolescent smoking initiation
Methodology	Survey with 1,219 Flemish children (12 to 13 years old)
Results	Parental influence: high ▪ Both parents smoke: 22.2% of the adolescents smokes at least one cigarette a day ▪ 1 of the parents smokes: 15.7% of adolescents smokes at least 1 cigarette a day ▪ Both parents do not smoke: 6.2% smokes at least 1 cigarette a week

Appendix B: Student sample 1 (low achievement level)

In a series of studies researchers have investigated the influence of parental smoking on smoking. On the one hand, this was done by Ellen Godard (2009) (and John Pierce (2005) as well as Carine Vereecken and Lea Maes (2010)). They have conduted [sic] this research on American youngsters and Flemish children.

The parents have to give their child a good education before they turn 9. This is explained in the following quote. The quote is as follows: The timing of parental smoking cessation appears to be important Parental quitting is most effective in reducing initiation if it occurs before the child reaches 9 years of age.

We can thus conclude that parental education is very important in about [sic] if children start smoking or not.

Word count: 125 words (Dutch text: 123 words)

Text criteria	Score	Maximum score
T: theme	1	2
R: results	0	2
A: arguments	0	2
C: embedding quote	1	2
E: concluding sentence(s)	1	2
Total score	**3**	**10**

Student sample 2 (high achievement level)

A series of studies have been conducted into the influence parents have on the likelihood that their child starts smoking.

Both Goddard (2009) and Vereecken and Maes (2010) have found that parental influence is big.

The results of the study by Pierce (2005) indicate that there is very little influence. The difference between both studies can be explained by the fact that Pierce's study (2010) has a different objective. To be more precise, Pierce studies what the impact is of parent cessation is going back 5 years prior to the their children's smoking initiation. The two other studies, on the other hand, studied the likelihood of their children's smoking initiation if one or both parents smoked.

From a quote by Pierce (2010) it follows that this influence is not high if the child is older than 9 years and the parents then quit smoking.

The timing of parental smoking cessation appears to be important. Parental quitting is most effective in reducing initiation if it occurs before the child reches [sic] 9 years of age (pg 217).

In short, if parents smoke, the odds that their child also starts smoking are high. If the parents quit smoking five years prior to their child's smoking initiation and the child is older than 9, the influence is not big and there is a chance that the child will start to smoke.

Word count: 228 words (Dutch text: 225 words)

Text criteria	Score	Maximum score
T: theme	2	2
R: results	2	2
A: arguments	2	2
C: embedding quote	2	2
E: concluding sentence(s)	2	2
Total score	**10**	**10**

Appendix C: Scoring Rubrics Writing Task 1

Category 1: Introduction of the research topic

0	The common research topic is not mentioned in the first, second or third sentence of the summary.
1	The common research topic is mentioned in the first, second or third sentence of the summary. The description of the research topic, however, is not specific enough and/or incomplete.
2	The common research topic is mentioned in the first, second or third sentence of the summary. The description of the research topic is specific enough and complete. Benchmark example 1: A specific, summarizing research topic linking and joining together the three studies such as "The influence of parents on the odds that children smoke' or 'parental influence on early adolescent initiation of smoking" (Benchmark C1.2.1)

Category 2: Summarizing results

0	There is no synthesis or summary of research findings or results of the studies discussed in the sources: the three studies are juxtaposed without integration or clear links. Each paragraph deals with a separate study or each study is separately discussed following the structure: '*A first study….; In a second study…; A third study'*. (Benchmark C2.1.0)
1	There is an attempt at summarization of the research findings e.g. by using connectives to link the research results of the studies, and/or by combining either the two studies or three studies (but without a clear indication of similarity and contradiction and without a clear focus on the main findings first). '"*Two studies show a significant influence… ; one study does not.*" (Benchmark C2.1.1) "*From a first study (Goddard, 2009) it becomes clear that … Pierce concludes that…. In a final study also becomes clear that parental influence is significant." … contains a lot of detail about the study's findings*." (Benchmark C2.1.2)
2	The research findings are synthesized and the synthesis is good: the two studies with similar results are discussed jointly in a single sentence or by using connectives and contrasted with the third study with contradictory results. '"*Contrary to Pierce (2005), Goddard (2009) and Vereecken and Maes (2010) claim that parental influence on children's smoking is high*" (Benchmark C2.2.1) "*Goddard (2009) and Vereecken and Maes (2010) showed that parental influence on children's smoking is high. They showed that the probability that children start smoking is higher when both parents smoke rather than if only a single parent or none of the parents smoke. Pierce (2005) showed that parental influence on children's smoking is rather low*" (Benchmark C2.2.2)

Category 3: Providing arguments and explanations for contradictory research findings

0	No argument is provided.
1	Only a single argument, that is, the difference in age groups in the different studies, is provided. This argument is the least feasible and least convincing explanation.
2	The argument and explanation that there is a difference in research focus in the different studies is provided, either in addition to the 'age group argumentation' or separately.

Category 4: Inserting the quote in the text

0	The quote is not integrated. *"Goddard and Vereecken and Maes are convinced that parents exert a significant influence on the odds that their children initiate smoking, whereas Pierce considers the influence to be negligible. Pierce concludes: "The timing…"* (Benchmark C4.0: very general, not integrated nor linked to the information preceding)
1	The quote is embedded but not very successfully integrated (the introductory clause is general; the quote is 'dangling'). *"However, Pierce (2005) concluded that parents have little or no influence on children who start smoking. He concluded the following (Pierce, p. 217):* *'The timing of parental smoking cessation appears to be important. …'"* ((Benchmark C4.1.1) *"Pierce (2005) investigated the influence by looking at the parents as well. … He explained a low influence as follows: "The timing of parental smoking cessation.."*" (Benchmark C4.1.2)
2	The quote is successfully integrated. *"Pierce disagrees as parental influence is small. He states that it is the period during which the parent quits smoking that is decisive: "The timing…. "'* (Benchmark C4.2.1) *"The researcher concludes that the specific moment at which parents quit smoking is of significant importance: "The timing of parental smoking…""* (Benchmark C4.2.2)

Category 5: Concluding the synthesis

0	There is no conclusion. The text is not complete.
1	There is a conclusion and the text is complete e.g. by synthesizing the results of the different studies at the end of the text.
3	There is a conclusion and the text is complete. Additionally, the conclusion is of a higher, abstract level also drawing inferences, interpreting the impact of the studies or pointing out the significance and relevance of the studies.