Opening the black box of students' text-learning processes: A process mining perspective

Amelie Rogiers[a], Emmelien Merchie[a] & Hilde van Keer[a]

[a]Department of Educational Studies, Ghent University, Ghent, Belgium

Abstract

The current study uncovers secondary school students' actual use of text-learning strategies during an individual learning task by means of a concurrent self-reported thinking aloud procedure. Think-aloud data of 51 participants with different learning strategy profiles, distinguished based on a retrospective self-report questionnaire (i.e., 15 integrated strategy users, 15 information organizers, 10 mental learners, and 11 limited strategy users), were analysed by means of educational process mining. Both the frequency of students' strategy use, as well as the temporal patterns between these strategies were studied. The process mining results clearly demonstrated differences between the strategy profiles with respect to the frequency of their applied strategies, as well as concerning the temporal sequences wherein strategies were applied throughout the course of students' text-learning process. The added value of combining both retrospective and concurrent self-report measures of students' strategies as well as conducting process mining analysis is discussed.

Process mining; learner profiles; think-aloud protocol analysis; on-line measures; off-line measures

Corresponding author: Amélie Rogiers, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium, E-mail: Amelie.Rogiers@UGent.be

1. Introduction

Recently, both educational researchers and practitioners have emphasized the importance of adjusted or personalized curricula wherein both the instructional content and methods are tailored to students' individual learning needs (Deed et al., 2014). This is also recognized by the OECD Learning Framework 2030 (2018) advocating the importance of learner-oriented teaching and learning. In view of contributing to the evidence-based design of personalized curricula, educational researchers are concerned with both measures and data analysis approaches to fully map and understand individual students' learning. Considering these *measures*, it is clear that the inclusion of both off-line and on-line instruments for measuring students' learning is preferable given their complementary properties (Veenman, 2011). While off-line measures are administered prospectively or retrospectively to performance on a learning task (e.g., self-report questionnaire data), on-line measures are gathered concurrently during task performance (e.g., think-aloud protocol or verbal self-report data). Consequently, while off-line measures enable researchers to uncover learners' perceptions of which and how often certain strategies are applied during learning, on-line measures additionally enable to map how and when these strategies are actually applied throughout the learning process (i.e., in which sequence strategies are applied or which switches occur between strategies; Merchie & Van Keer, 2014). In this respect, researchers increasingly advocate to combine both measures in view of gaining rich and detailed insight into both students' perceptions and actual strategic behaviour (Bråten & Samuelstuen, 2007; Veenman, 2005).

As to the *data analysis approaches* for gaining insight into students' learning processes, researchers call progressively for applying a more person-oriented approach, next to the rather dominant variable-oriented approach focusing primarily on analysing relationships among variables (Alexander et al., 2018; Fryer & Vermunt, 2017). Such a person-oriented approach is highly recommended as it emphasizes the study of naturally occurring clusters or profiles in students' learning (Bergman et al., 2003). Stemming from a person-oriented approach on students' text-learning strategies (i.e., strategies to select, organize, condense, and retain text information in a more memorable form; Rogiers et al., 2019a; Weinstein et al., 2011), previous research already succeeded to identify learning strategy profiles in a large sample of 1,931 secondary school students (Rogiers et al., 2019a). Four learning strategy profiles, in which students differently combine diverse strategies during text learning, were determined based on a retrospective self-report questionnaire. More particularly, *integrated strategy users* (ISU) were identified as learners with the most preferable profile, as they engaged in the strategic combination of different covert (i.e., non-observable, e.g., elaborating) and overt (i.e., observable, e.g., summarizing), cognitive (e.g., elaborating), and metacognitive (e.g., monitoring) text-learning strategies, and outperformed their peers on a subsequent performance test. The *information organizers* (IO) frequently applied text-noting strategies (i.e., highlighting, summarizing) and reported limited use of mental learning strategies. Conversely, *mental learners* (ML) restricted their repertoire to covert mental learning strategies (i.e., rereading, paraphrasing) without text-noting strategy use. Finally, *limited strategy users* (LSU) were considered as the non-strategic or less preferable profile, as they mainly focused on the frequent application of one single text-learning strategy (i.e., highlighting, rereading) and obtained the lowest performance scores afterwards. These learning strategy profiles were also identified in late elementary education (Merchie et al., 2014) and in subsequent samples of secondary school students (Rogiers et al., 2018, 2019a). Hence, the abovementioned learning strategy profiles were already corroborated several times in different age groups and independent study samples.

To date, however, there is a gap in the literature when it comes to research providing insight into the temporal sequences in which certain strategies are applied differently by learners during the course of their text-learning process. In this respect, it is seldom investigated which strategy switches unfold during this process (Cromley & Wills, 2016). Although current theories of (text) learning implicitly or explicitly state to account for what happens during this process, this matter has rarely been tested empirically with sequential analyses of real-time process data (e.g., from think-aloud protocol transcripts; Cromley & Wills, 2016). In this respect, the present study contributes to the first explicit question regarding self-report data tackled throughout the different contributions in this special issue. More particularly, it is believed that the learning process of a strategic learner can be characterised as cyclical and adaptive. First, from a self-regulated learning (SRL) perspective, students' learning process is considered as a cyclical process, consisting of different phases occurring before, during, and after learning (i.e., forethought, performance, and reflection phase; Zimmerman, 2002). These phases are not viewed as linearly structured, but considered dynamic and iterative (Panadero, 2017; Pintrich, 2000; Zimmerman, 2002). Second, next to the general comprehensive models of SRL, also more domain-specific learning strategy models (i.e., Good Strategy User Model by Pressley et al., 1987; Model of Strategic Learning by Weinstein et al., 2011; Model of Domain Learning by Alexander, 1998) point to the importance of adaptive strategy use, which encompasses engaging deliberately and flexibly in the use of various strategies. Rather than following a linear and rigid approach to text learning, strategic learners are believed to undertake learning in an adaptive way, wherein

they interactively return to prior learning activities or phases when necessary (Alexander & Jetton, 2000; McNamara et al., 2007; Simpson & Nist, 2000; Wade et al., 1990). Prior research already indicated that high achievers appear to be mostly integrated strategy users, adopting various strategies (Merchie et al., 2014; Rogiers et al., 2019a) and regulate their learning process more effectively (Stoeger et al., 2015). However, it is unclear whether these process statements as put forward in different theoretical models can be grounded empirically and whether

and how exactly specific strategy sequences unfold during the course of learners' text-learning process (Cromley & Wills, 2016). In this respect, it is often difficult to grasp the cyclical and adaptive nature of learning processes as described in the abovementioned theoretical models by means of retrospective self-report measures after learning occurred. It is therefore necessary to analyse students' learning process in a more fine-grained way (i.e., occurrence after occurrence) as it unfolds in real time during learning. Opening this black box and gaining insight into the cyclical and adaptive nature in which particular sequences and strategies unfold throughout students' learning process can offer valuable starting points for providing learner-oriented teaching and learning. If certain effective sequences between strategies come to the fore, for example, then not only strategies, but also effective sequences of applied strategies should be taught (e.g., from one learning strategy to another).

A promising and emerging technique to gain systematic insight into these sequences and analyse students' concurrent self-reports (e.g., think-aloud protocols) is Educational Process Mining (EPM). The idea behind EPM is to discover, monitor, and improve students' actual learning processes by extracting knowledge from recorded time stamps (Bannert et al., 2014). A time stamp refers to the moment wherein the learner is executing or initiating a certain learning activity (e.g., highlighting, rereading). By means of these timestamped activities derived from learners' observed learning behaviour, compact educational process models are composed (Van der Aalst, 2011). These process models provide an overview of both learners' executed activities and the paths that occur between these activities. Whereas the *activities* map the number and frequency of certain applied strategies, the *paths* represent how, and in which sequences these strategies were adopted throughout the learning process (Fluxicon, 2019; Van der Aalst, 2011). As such, EPM enables to visualise students' learning behaviour and facilitates a thorough understanding of the course of students' complex real-time learning process. In the context of SRL, for example, EPM research has shown that university students' sequences of self- or group-regulatory activities differed among successful and less successful students (e.g., Schoor & Bannert, 2012). However, the application of EPM in educational research is still in its infancy and, to date, the temporal order of students' applied strategies during task completion has been widely neglected (Bannert et al., 2014; Reimann, 2007). More in-depth EPM analyses can, therefore, yield valuable insights into students' learning process and can complement off-line measures of students' applied strategy use. In this respect, it also enables to investigate to which degree retrospective self-report measures accurately reflect students' actual strategy use that is revealed while concurrently thinking aloud. Our study adds to the literature by systematically analysing real-time think-aloud protocol (further referred to as 'TAP') data from students with different learning strategy profiles who are requested to learn an informative text and by considering strategy sequences by means of EPM. Further, this study adds to the literature by confronting the frequency of students' text-learning strategies as measured via concurrent measures on the one hand (i.e., TAP) and retrospective measures on the other hand (i.e., a task-specific self-report questionnaire) and study their overlap (Rogiers et al., 2019b).

1.2 The present study

By means of EPM, the current study investigates students' actual use of text-learning strategies when executing an independent learning task while thinking aloud. In a first step, this study aims to examine the *frequency* of students' occurred text-learning strategies depending on their learning strategy profile (RQ1). Referring to previous research using task-specific self-report questionnaires (Merchie et al., 2014; Rogiers et al., 2018, 2019a), we hypothesize more frequent verbalisations of various text-learning strategies in integrated strategy users, and less frequent and diverse strategy verbalisations in limited strategy users. In addition, we expect more frequent verbalizations of the application of overt text-noting strategies (e.g., summarizing) in information organizers and a predominant application of covert mental learning strategies (e.g., paraphrasing) in mental learners. In a second step, this study aims to explore *temporal patterns* in students' text-learning process based on the sequences in which their strategies are applied (RQ2). As the theoretical and empirical literature indicates that particularly effective learners apply diverse strategies in a flexible and systematic way (Alexander & Jetton, 2000; McNamara et al., 2007; Rogiers et al., 2019a; Simpson & Nist, 2000; Wade et al., 1990; Weinstein et al., 2011), we hypothesize a more cyclical use of text-learning strategies in integrated strategy users, including more recursive

patterns between their applied strategies. Conversely, a more linear and unidirectional text-learning process is expected in limited strategy users.

2. Methodology

2.1 Participants

A think-aloud study was carried out with 51 secondary school students (62.75% seventh and 37.25% eight graders) from 11 schools and 51 classes who were part of a large-scale study (n = 1,931, Rogiers et al., 2019a). Based on a large-scale cluster analysis, 15 integrated strategy users, 15 information organizers, 10 mental learners, and 11 limited strategy users (n = 51 participants) were identified within the sample of the think-aloud study. The sample consisted of 70.59% girls and 29.41% boys, with an overall mean age of 12.99 years (SD = .69). The majority of the students (87.23%) were native Dutch speakers, which is the language of instruction in Flanders (the Dutch speaking part of Belgium). All participants and their parents agreed to participate in the TAP administration by means of informed consent.

2.2 Instruments and procedure

The data collection procedure consisted of several steps. Figure 1 provides a visual representation of the data collection procedure.
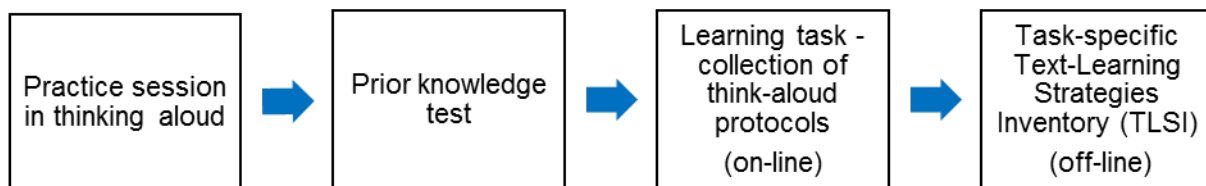

Figure 1. Chronological representation of the data collection procedure.

As can be seen in both last steps of the procedure, a combination of two self-report measures was opted for in the context of the present study, respectively an on-line and concurrent think-aloud measure on the one hand and an off-line, retrospective questionnaire on the other hand.

2.2.1 Practice session in thinking aloud

Following the recommendations of prior research (Greene et al., 2011; van Someren et al., 1994), a 20-minute practice session in thinking out loud was organised by the researcher to familiarize students with the think-aloud method. This practice session was based on prior research in a similar age group (Merchie & Van Keer, 2014; Vandevelde et al., 2005) and consisted of three phases. In a first phase, the researcher thoroughly explained the purpose and procedure of the think-aloud method. Second, the researcher modelled thinking aloud during an origami assignment. No learning task was opted for practicing thinking out loud to avoid possible training effects (Afflerbach & Johnston, 1984; Greene et al., 2011). The origami assignment provided ample opportunities for self-regulation. For instance, a step-by-step approach could be followed and there were ample opportunities for students to evaluate or adjust their approach. Third, an individual practice phase took place in which students practiced thinking out loud. During this session, students were asked to fold an origami cat while verbalizing their thoughts, feelings, and actions. During the practice session, feedback was provided on students' verbalisations in view of optimizing their thinking aloud. Accordingly, no feedback on students' approach was provided. The researcher prompted the student to continue verbalizing when (a) meaningful silences or (b) certain nonverbal behaviours took place (i.e., frowning, repeatedly turning the text page, staring; Merchie & Van Keer, 2014; Vandevelde et al., 2015), thereby thoughtfully considering the student and situation at hand to avoid the loss of meaningful information about students' behaviour (e.g., Boekaerts & Corno, 2005). As prompt, students were consistently given the instruction: "verbalize everything that you are doing or thinking" or "keep thinking aloud". In this respect, type 1 (verbal content) and type 2 (nonverbal content) verbalizations were encouraged, and type 3 verbalizations were avoided since students were not asked to explain their cognition. Consequently, researchers

were able to identify spontaneous self-regulatory learning activities (Ericsson & Simon, 1980; Vandevelde et al., 2015).

2.2.2 Prior knowledge test

As prior knowledge might influence text learning (Alexander & Jetton, 2000; Bråten & Samuelstuen, 2004), a prior knowledge test regarding the text topic was administered before the actual learning task. Students were asked to write down everything they already knew about the topic. Following prior research, the matching of students' notes to the text content was opted for to score the prior knowledge test (for more information on this procedure, see Merchie et al., 2014) The matching of students' notes to the text content revealed very limited to no prior knowledge regarding the text content (M = 5.66, SD = 2.71; Min = 0, Max = 24).

2.2.3 Learning task

Since studying in preparation for a classroom test is a regular task in secondary education, students were instructed to study an informative text in the way they would prepare for a test while thinking out loud (Fox, 2009; Samuelstuen & Bråten, 2007). For the learning task, a 442-word informative text was used of which the participants did not study the topic (i.e., chewing gum) as part of their courses. The multi-paragraph text consisted of one title (i.e., chewing gum), four sections and subtitles (i.e., history, production, advantages, and disadvantages), and three pictures. Text quality was verified in advance (see Rogiers et al., 2019a). In view of encouraging students to plan their work, they were informed to have 50 minutes time for task completion. To enable students to monitor their progress, a clock was provided, but no further time indications were given to prevent the prompted monitoring of time. In line with previous studies (Slotte et al., 2001), students were allowed, but not obligated to make notes on scratch paper while studying. During the task completion process, students were observed by the researcher and were only prompted to continue verbalizing when necessary (Greene et al., 2011).

2.2.4 Task-specific self-report inventory

Immediately after learning task execution, students completed the 'Text-Learning Strategies Inventory' (TLSI; Merchie et al., 2014). This task-specific questionnaire consists of 37 items, subdivided into nine subscales (see Appendix A) to which students respond on a five-point Likert-scale (1 = *completely disagree*, 5 = *completely agree*). In line with theoretical frameworks (Wade et al., 1990; Zimmerman, 2002), the TLSI incorporates both cognitive (e.g., paraphrasing) and metacognitive (e.g., monitoring) text-learning strategies, as well as overt (e.g., summarizing) and covert (e.g., paraphrasing) strategies. Good model fit results were obtained for this nine-factor model in prior large-scale research (Rogiers et al., 2019a). Appendix A presents the descriptive statistics and reliability coefficients for the TLSI-subscales. By means of hierarchical and K-means cluster analyses on the TLSI-subscale scores within the larger sample (n = 1,931), students learning strategy profile was determined (for a detailed description, see Rogiers et al., 2019a).

2.3 Think-aloud coding procedure of learning strategies

In a first step, think-aloud sessions were transcribed and coded. As all sessions were audio- and videotaped, both students' verbal and non-verbal behaviour (e.g., highlighting text) was transcribed to increase coding accuracy (Veenman, 2011; Young, 2005). Transcriptions were made by means of a computer program for subtitling videos (i.e., Subtitle Workshop 4). This program enables researchers to register the start and end time of each verbalisation and action. This is essential in view of conducting EPM, as the sequence of strategies is calculated based on their exact time frame. Subsequently, transcripts were segmented by one researcher into units of meaning, with one unit referring to a thematically consisted verbalization of a single text-learning activity (Scott, 2008; van Someren et al., 1994). Repeated actions were analysed as separate activities in view of considering the recurrence of different text-learning strategies. As a result, 1,015 minutes of thinking aloud, and 4,107 units of meaning were identified and coded by means of the coding scheme based on prior research of Merchie and Van Keer (2014). This coding scheme is an adapted version of the 'Text-Learning Strategy Protocol' (TLSP; see Table 1), comprising 11 subcategories referring to different text-learning strategies. In line with the self-report questionnaire, the coding scheme reflects both cognitive and metacognitive, as well as overt and covert strategies. Mean learning time was 20 minutes (SD = 3.89), with a minimum of 6 and a maximum of 34 minutes. Analysis of variance showed no statistically significant differences between the four strategy profiles in terms of their mean learning time, $F(3, 50)$ = 2.437, p = .076). Finally, two trained coders independently double-coded 27% of the protocols, resulting in high interrater reliability (Krippendorff's α = .95; Hayes & Krippendorff, 2007).

Table 1

Coding scheme for analysing students' learning activities

TLSP-categories	Description	Example
Overt cognitive strategies		
Summarizing	Noting words or sentences on the scratch paper, making a summary or scheme	*'So, this part is done (student draws a line underneath his summary).'*
Highlighting	Structuring the text or one's own notes	*'I mark all these important words.'*
Covert cognitive strategies		
Memorizing	Memorizing by rereading the text or one's own notes	The student rereads his scratch paper notes.
Initial reading	Reading the text for the first time	The student reads the text for the first time.
Rereading	Rereading the source text	The student rereads the text out loud.
Rehearsing	Rehearsing the text information	*"Now I will rehearse this part again by heart."*
Paraphrasing	Retelling the text content in one's own words	*'So, in sugar-free gum, xylitol prevents tooth decay.'*
Elaborating	Activating or relating prior knowledge to the text content or providing personal remarks regarding the text content	*'Synthetic rubber is indeed petroleum, we learned that in geography!'*
Metacognitive strategies		
Planning	Exploring the text subject, detecting task demands, planning the strategic approach	*'I will first read the text, then underline important words and then try to memorise the text.'*
Monitoring progress	Checking progress during task performance, reflecting on the available time and the time schedule, monitoring the strategic approach	*'So this is what I have already done and this is what I have to do next'.'*
Monitoring comprehension	Detecting lack of comprehension or mistakes, mentioning awareness of understanding	*'Synthetic rubber… I don't know what 'synthetic' means.'*
Other[*]	Asking questions to the researcher about the overall procedure of the study, phases of silence…	*'Can I write my name on these papers?'*

Note. TLSP = Text-Learning Strategy Protocol. * In accordance to Schoor and Bannert (2012), this category was excluded from the process mining analysis, as we wanted to concentrate on task-related behaviour.

2.4 Process mining analysis on the think-aloud data

In a next step, the coded learning activities of each learner profile were analysed separately via process mining using Disco (Fluxicon, 2019). This software program enables researchers to study the course of students' actual learning processes by generating process models for each learner profile. In these process models, both (1) the *activities* performed by the learners (i.e., the executed strategies during text learning), and (2) the *paths* or connections that occurred between these activities are displayed (Fluxicon, 2019; Van der Aalst, 2011). Thus, whereas the *activities* refer to the extent in which certain text-learning strategies are adopted (i.e., boxes in Figures

2-5), the *paths* visualize the sequence of these performed activities (i.e., arrows in Figures 2-5). Above these paths, the frequency of each of these sequences is represented. Further, both unidirectional paths (→), bidirectional paths (⇆), and loops (↺) are depicted in the process models, indicating that activities have respectively been conducted in consecution, in alternation, or that the same activity was performed several times in succession.

In line with prior research in the field of SRL (Bannert et al., 2014; Schoor & Bannert, 2012) the fuzzy miner algorithm in Disco was used to perform the analysis. This algorithm relies on two metrics (i.e., significance and correlation) to calculate which activities and paths should be included in the process models and which to be excluded (Günther & van der Aalst, 2007). *Significance* refers to the relative importance of activities and paths, implying that more frequent text-learning activities are retained in the model. *Correlation* is deployed for selecting only paths of closely connected activities (Günther & van der Aalst, 2007). By means of this algorithm, Disco automatically includes strategies and paths that are often conducted by a large group of students in the process model, while less frequent activities and paths or paths that have been seldom conducted by only few students are excluded. To date, however, no specific standards are available on the amount of activities and paths that should be included in the process models. Researchers argue that the ideal number of activities and paths strongly depends on the type of research data and questions involved (Fluxicon, 2019). In general, the inclusion of as much activities and paths as possible while simultaneously avoiding too complex process models is recommended (Fluxicon, 2019). In the current study, the percentages of included activities and paths in students' process models were carefully deliberated among four experts on text learning and SRL. In this respect, 33.33% of the most frequent strategies and the 33.33% most frequent connections between these strategies were included in the analysis. As a result, initially coded categories such as paraphrasing and elaborating (see Table 1) were not included in the 33.33% process models (Figures 2-5). Although these strategies occurred commonly in the group of integrated strategy users and mental learners (Table 2), they were adopted by a rather small share of learners compared to the occurrence of the other strategies. Put differently, these activities did not belong to the 33.33% most frequent activities conducted at least once by a large group of learners.

As a final step, following Schoor and Bannert (2012) and in view of obtaining split-half-reliability for the generated process models, we repeated the EPM analyses for the five most typical individuals of each learning strategy profile. We perceived students as typical integrated strategy users (ISU) when high frequencies were found for different text-learning strategies, whereas typical limited strategy users (LSU) were characterized by the dominant application of only one strategy (e.g., highlighting, rereading). For typical information organizers (IO) and mental learners (ML), strategies with high frequency were respectively text-noting strategies (e.g., highlighting, summarizing for IO) and mental learning strategies (e.g., rereading, rehearsing for ML; Rogiers et al., 2019a). The obtained models for the five most typical individuals of each learning strategy profile were very similar to those in Figures 2-5.

3. Results

3.1 Frequency of occurrence of text-learning strategies in different learning strategy profiles' text-learning process (RQ1)

In view of the first research question, we examined which text-learning activities were executed by the different learning strategy profiles during actual text learning. Table 2 displays the frequency of occurrence of all text-learning strategies included in the process models, as well as the number of students conducting each strategy at least once. One-way analysis of variance was used to test differences between the four learning strategy profiles regarding students' use of different strategies. Additionally, Post Hoc pairwise tests with Bonferroni correction were conducted to investigate these differences in-depth. The analysis revealed significant differences between the four learning strategy profiles (see Appendix B for detailed results of the Post Hoc Pairwise comparisons and effect sizes).

As can be derived from Table 2, the results with regard to students' *cognitive strategy use* show that integrated strategy users (ISU) generally executed more diverse text-learning strategies than information organizers (IO), mental learners (ML), and limited strategy users (LSU). The frequency of occurrence of most strategies was higher in integrated strategy users, as well as the percentage of students that adopted the strategies at least once. The most occurring cognitive strategies for integrated strategy users were summarizing, highlighting, paraphrasing, and elaborating, whereas for limited strategy users highlighting, memorizing, and elaborating were

the most frequent strategies. For mental learners, summarizing and highlighting activities seldomly occurred, whereas rereading and rehearsing were frequently coded. In contrast to mental learners, summarizing and highlighting frequently occurred in the information organizers group, in addition to rereading and rehearsing. These results were reflected in the Post Hoc pairwise test results.

As can be derived from Appendix B, a statistically significant difference between the four learning strategy profiles was found for the verbalized overt text-learning strategies: summarizing ($F_{(3, 4732)}$ = 85.59, $p < .001$) particularly in favour of integrated strategy users and information organizers, and highlighting ($F_{(3, 4732)}$ = 43.88, $p < .001$) in favour of all learning strategy profiles, except for mental learners. Regarding the covert text-learning strategies, the results show that limited strategy users more frequently applied memorizing ($F_{(3, 4732)}$ = 4.54, $p = .004$) than information organizers. Further, rereading ($F_{(3, 4732)}$ = 22.89, $p < .001$) and rehearsing ($F_{(3, 4732)}$ = 144.34, $p < .001$) were most frequently executed by mental learners and information organizers and less frequent by limited strategy users, whereas paraphrasing ($F_{(3, 4732)}$ = 21.12, $p < .001$) and elaborating ($F_{(3, 4732)}$ = 5.252, $p = .001$) were less frequent performed by information organizers. No statistically significant difference between the four profiles was found regarding the execution of initial reading ($F_{(3, 4732)}$ = 2.61, $p = 0.05$).

The results with respect to students' *metacognitive strategy use* reveal no statistically significant differences between the four profiles regarding the use of comprehension monitoring activities ($F_{(3, 4732)}$ = 1.10 $p = 0.348$). In contrast, a statistically significant difference between learning strategy profiles was found with regard to planning ($F_{(3, 4732)}$ = 38.56, $p < .001$), indicating that planning activities were particularly executed by limited strategy users. In addition, a statistically significant difference with regard to progress monitoring ($F_{(3, 4732)}$ = 15.57, $p < .001$) reveals that this strategy frequently occurred in integrated and limited strategy users. However, in order to gain insight into the sequences in which these different strategies are adopted throughout students' learning process, a closer look at the process models is needed (see RQ2).

Table 2

Frequency of occurrence of text-learning strategies for each learner profile (n = 51), including absolute frequency and number of students [a]

Coding category	ISU (n = 15)		IO (n = 15)		ML (n = 10)		LSU (n = 11)	
	Absolute frequency	*n*	Absolute frequency	*n*	Absolute frequency	*n*	Absolute frequency	*n*
Overt cognitive strategies								
Summarizing	296	15	268	14	8*	3	64*	4
Highlighting	273	15	245	12	31*	2	224	11
Covert cognitive strategies								
Memorizing	203	15	181	14	143	8	200	10
Initial reading	283	15	242	15	162	10	131	11
Rereading	31*	10	66	12	47	8	199	11
Rehearsing	45*	9	88	12	189	9	20*	6
Paraphrasing	67*	10	34*	8	83*	7	35*	7
Elaborating	38*	11	13*	5	19*	4	32*	6
Metacognitive strategies								
Planning	18	12	88	12	38	9	123	11
Monitoring progress	77	14	28*	9	9*	4	46*	6
Monitoring comprehension	21*	7	31*	8	20*	5	17	7

Note. [a] Number of students conducting the strategy at least once. * Strategies not included in the process models as they do not belong to the 33.33% most frequent activities that are conducted by a large group of students (see method section and RQ2). *ISU* = integrated strategy users, *IO* = information organizers, *ML* = mental learners, *LSU* = limited strategy users.

3.2 Temporal patterns in the different learner profiles' text-learning process

Figures 2-5 display the resulting 33.33% process models for each learning strategy profile. The direction of the arrows represents the order in which the text-learning activities were adopted throughout students' text-learning process. Strategies that took place in the beginning of students' text-learning process (e.g., planning, initial reading) are depicted at the top of the figures, while strategies that were executed at a later moment or at the end of the learning process (e.g., highlighting, summarizing, memorizing, rereading, rehearsing) are represented at respectively the centre or bottom of the figures.

When contrasting the process models of the different learning strategy profiles and focussing on the cognitive and metacognitive strategies that were included in the models, clear differences can be noticed. First, the *cognitive* activities that were included in students' process models indicate that integrated strategy users (ISU), information organizers (IO), and limited strategy users (LSU) applied a combination of both overt (e.g., summarizing and highlighting) and covert strategies (e.g., rehearsing, memorizing, rereading) during text learning, whereas mental learners (ML) exclusively adopted covert strategies. When considering the overt strategies, the process models show that integrated strategy users, information organizers, and limited strategy users frequently applied highlighting, whereas only the models of integrated strategy users and information organizers include summarizing strategies. Regarding the covert strategies, the results show that rehearsing was only included in the process models of information organizers and mental learners. Second, planning was included as a *metacognitive* strategy in all process models, while progress monitoring was only included in integrated strategy users' process model, indicating that – compared to the frequency of the other strategies – a large share of integrated strategy

users frequently tracked and controlled their progress throughout their learning process. The same applies for comprehension monitoring, which was only included in limited strategy users' process model, implying that a large group of limited strategy users actively monitored (a lack of) understanding while processing the text.

When studying the sequences in which these different strategies were adopted throughout students' learning process, differences in the phases of the text-learning process can be identified. First, when focussing on the *beginning* of students' learning process (i.e., start symbol in the process model), the results indicate that integrated and limited strategy users initiated their learning process with planning before initially reading the text. In contrast, information organizers and mental learners immediately started reading the text without planning in advance, which is indicated by the unidirectional arrows between initial reading and planning. Subsequently, they performed planning after they initially read the text.

Concerning the strategies conducted *during* actual text studying, differences between the four learning strategy profiles were found as well. In the group of limited strategy users (Figure 5), the unidirectional arrows between the different strategies indicate that planning, initial reading, highlighting, memorizing and rereading were consecutively executed. In addition, the unidirectional connection between initial reading and comprehension monitoring in limited strategy users' process model denotes that 18% of these strategy users monitored their understanding after reading the text. This strongly differs from integrated strategy users' process model (Figure 2). While highlighting is also preceded here by initial reading, bidirectional paths are found between initial reading and highlighting, as well as between initial reading and summarizing. Yet, the arrows connecting the different cognitive strategies, as well as the presence of reciprocal arrows, indicate that integrated strategy users alternately adopted these strategies before they started to memorize the text. Further, the position of progress monitoring as rather isolated from the other activities in these strategy users' process model must be noticed. This position is due to the fact that progress monitoring was applied before and after a wide variety of activities, suggesting that integrated strategy users tracked and controlled their progress throughout the entire learning process. However, since the process model only represents 33.33% of the performed activities, the wide variety of arrows were omitted by the program. When analysing the process model in detail, the large number of arrows between progress monitoring and a diverse set of text-learning strategies can be found. Furthermore, it is notable that integrated strategy users (Figure 2) alternated strategies before learning (i.e., planning) with activities during learning, which is indicated by the bidirectional arrows connected to students' planning strategy. More particularly, they considered their planning not only before reading in the pre-learning phase, but throughout the different phases in their learning process (i.e., after reading, highlighting, and memorizing).

When taking a closer look at information organizers' process model (Figure 3), many paths are visible, demonstrating that information organizers frequently switched between various strategies throughout their learning process, or frequently resumed previous strategies. This reveals that their text-learning process was rather cyclical organized. Especially the strategies 'summarizing' and 'highlighting' played a prominent role in these learners' learning process, as can be derived from the large number of incoming and outgoing arrows. For instance, after initially reading the text, the unidirectional arrows indicate that information organizers considered their planning before summarizing. After summarizing, a large share of these learners returned to reading the text or started memorizing or rehearsing the text. A clear bi-directional path is present between memorizing and highlighting activities, indicating that these activities were performed in alternation. Further, highlighting was also frequently followed by rehearsing, initial reading, and/or summarizing. Remarkable is that after conducting memorizing, summarizing, and highlighting activities, a large share of information organizers returned to initially reading the text. This could imply that these learners started to engage in different text-learning strategies without first reading or fully understanding the study text.

Although both information organizers' and mental learners' learning process was initiated by initial reading and planning, the further course of their learning process clearly differed. While a large share of information organizers started summarizing the text, a large share of mental learners (Figure 4) started memorizing the text after planning. Further, the recursive loop for memorizing, rehearsing and rereading demonstrates an alternated application of these strategies in mental learners, while the unidirectional arrows show that rehearsing was often followed by rereading and rereading by memorizing.

More fine-grained differences in the course of students' text-learning process can be detected when taking a more detailed look at the direction of the arrows in the process models for each learning strategy profile. The results show that limited strategy users followed a mainly linear structured learning process, as is indicated by the unidirectional arrows in their process model and the absence of any bidirectional paths. In contrast, the other three profiles returned more to prior activities or phases throughout their learning process, which is illustrated by the

returning arrows pointing from the bottom to the top. Hence, no strict linear, but rather a cyclical approach to learning was followed by these profiles.

At last, the arrows leading to 'stop' in the process models (i.e., stop symbol in the figures) show which strategies were conducted *lastly* by the learners. As can be derived from the models, across all learning strategy profiles, most students finished their learning process with memorizing and rehearsing. However, initial reading also occurred as a final activity in some mental learners (20%) and limited strategy users (9%), while this is not the case in the learning process of the other learning strategy profiles. In addition, some integrated strategy users (13%) finished their learning process with reflecting on their progress, while rereading also occurred as final activity in some limited strategy users (9%).
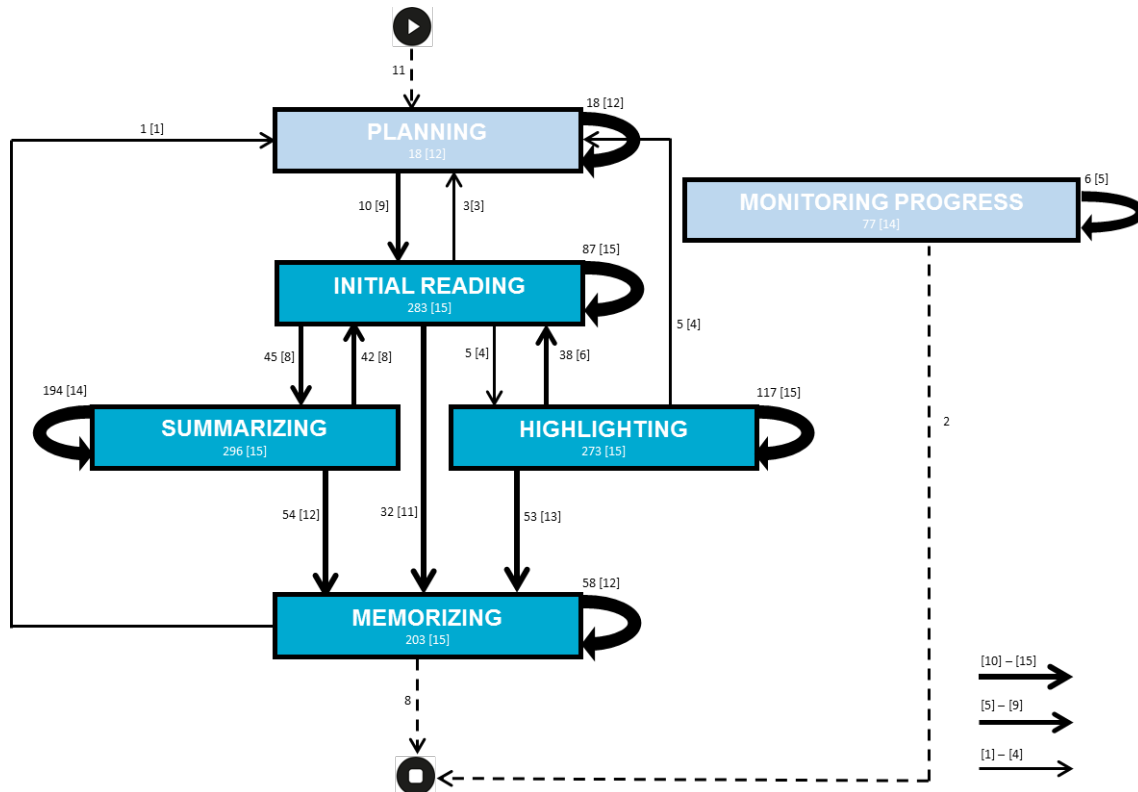

Figure 2

Text-learning process model of integrated strategy users (ISU; n = 15), including the frequencies of occurrence and, between brackets, the case frequencies (i.e., the number of students that conducted the activities at least once). The more frequent an activity was performed, the darker it is displayed. The more frequent a path between activities occurred, the thicker the arrow is displayed.
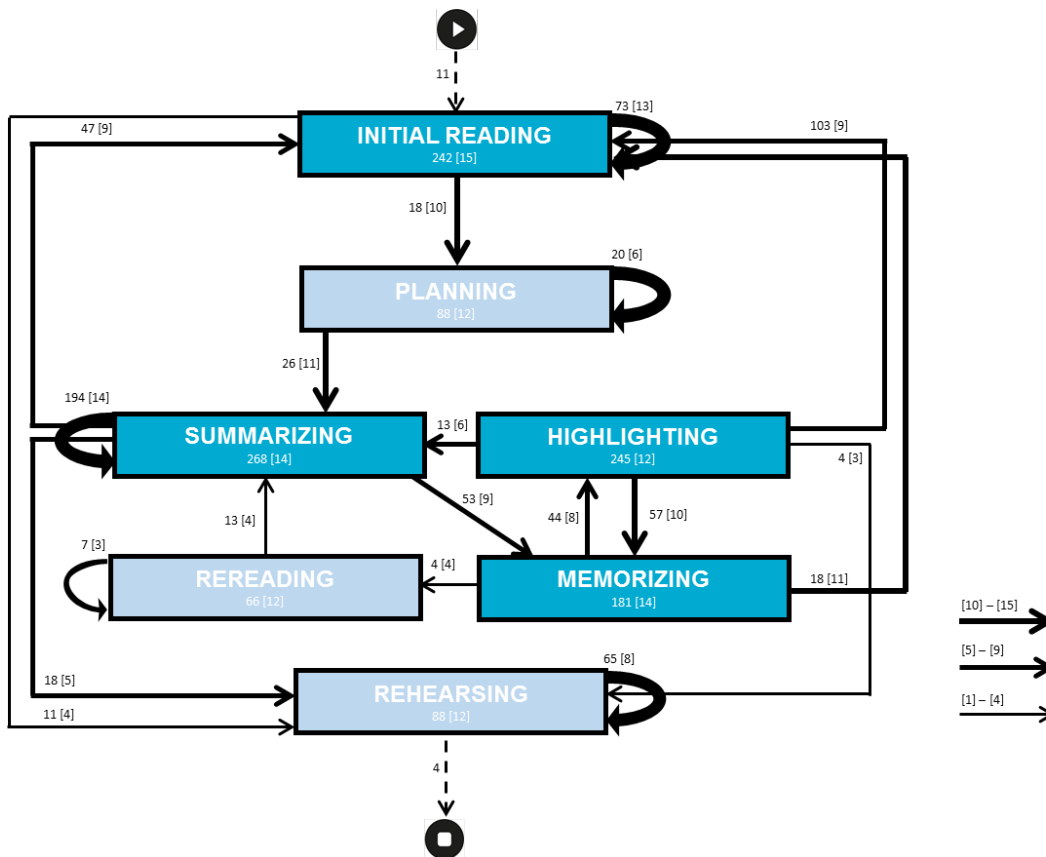
Figure 3

Text-learning process model of information organizers (IO; n = 15), including the frequencies of occurrence and, between brackets, the case frequencies (i.e., the number of students that conducted the activities at least once). The more frequent an activity was performed, the darker it is displayed. The more frequent a path between activities occurred, the thicker the arrow is displayed.

Figure 4

Text-learning process model of mental learners (ML; n = 10), including the frequencies of occurrence and, between brackets, the case frequencies (i.e., the number of students that conducted the activities at least once). The more frequent an activity was performed, the darker it is displayed. The more frequent a path between activities occurred, the thicker the arrow is displayed.
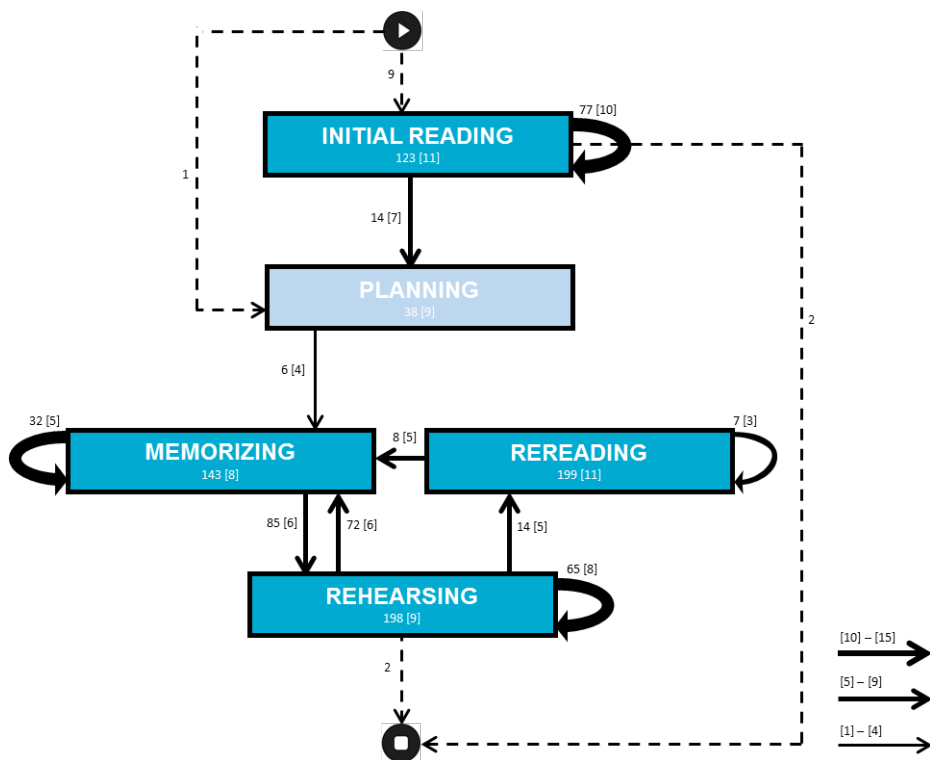
Figure 5

Text-learning process model of limited strategy users (LSU; n = 11), including the frequencies of occurrence and, between brackets, the case frequencies (i.e., the number of students that conducted the activities at least once). The more frequent an activity was performed, the darker it is displayed. The more frequent a path between activities occurred, the thicker the arrow is displayed.
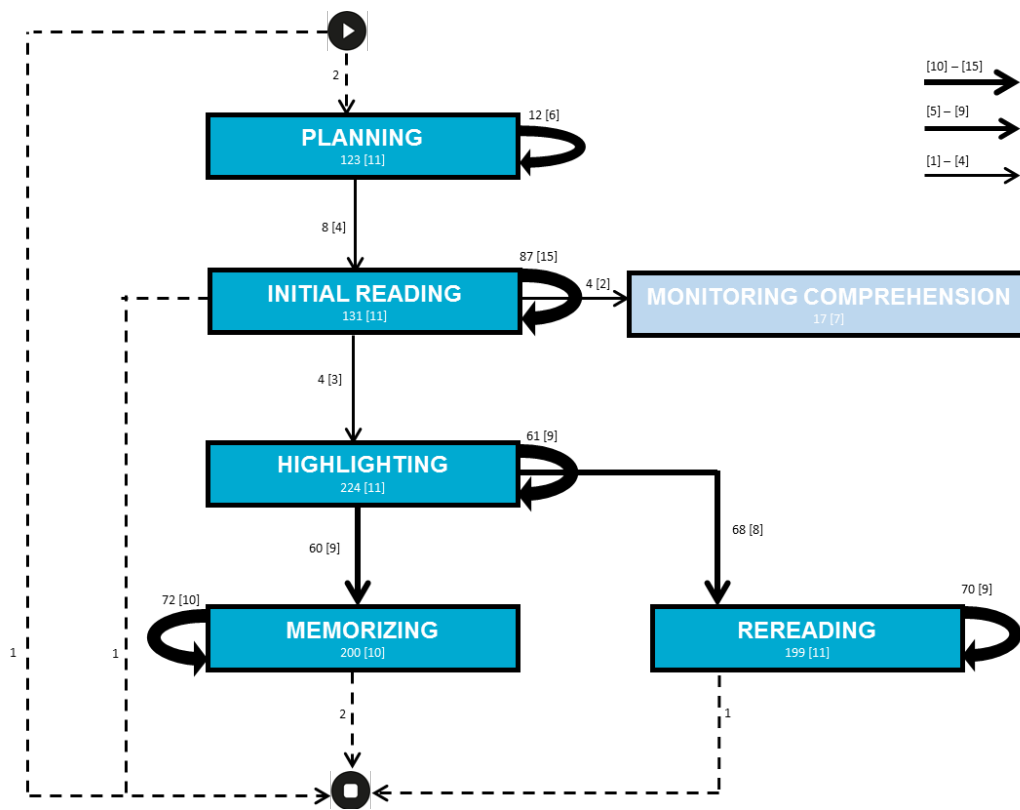
4. Discussion

To date, little is known about the sequences in which certain strategies are applied by different learners during the course of their text-learning process. In this respect, it remains unclear whether a cyclical and flexible approach to learning, as put forward as the most effective in various theoretical models, unfolds in different learning strategy profiles when learning from text. Nevertheless, in-depth insight into students' learning processes enables to be responsive to individuals' learning needs and avoid 'one-size-fits-all' approaches to learning. The purpose of this study therefore was to uncover both the frequency of students' applied strategies throughout their learning process, as well as the temporal patterns between these text-learning strategies. More particularly, the strategic behaviour of students from four different learning strategy profiles (i.e., integrated strategy users, information organizers, mental learners, and limited strategy users) based on a retrospective self-report questionnaire in a previous study (Rogiers et al., 2019a), was further depicted and compared by means of educational process mining (EPM) on their think-aloud protocol (TAP) data. In this respect, both students' concurrently and retrospectively measured strategy use was complementary taken into account.

The first research question focused merely on the quantity of students' strategy use by studying the frequency wherein text-learning strategies were executed by the different learning strategy profiles during actual text learning. The results clearly correspond to the findings of Rogiers and colleagues (2019a) who determined different learning strategy profiles based on students' retrospectively self-reported text-learning strategies. The results postulated less diverse learning strategy use for limited strategy users (LSU; e.g., highlighting and rereading), versus more varied overt and covert text-learning strategies for integrated strategy users (ISU). Similarly, the frequent use of overt text-noting strategies (i.e., highlighting, summarizing) reported by information organizers (IO) was reflected in their verbalized learning behaviour. The same applies for mental learners (ML), who both reported and actually applied the frequent use of covert mental learning strategies (e.g., memorizing, rehearsing, paraphrasing). This was also reflected in the strategies included in the different process models (RQ2). In this regard, the clusters determined based on students' retrospective self-report data were largely confirmed by their concurrent TAP data. Although some research has clearly shown discrepancies between retrospective and concurrent measures of students' strategic behaviour, our comparison overall shows that both measures enable us to uncover which strategies students do or do not use frequently. Given this convergence, the current study provides empirical support for retrospective self-report questionnaires as acceptable alternatives for more time- and labour-intensive measures such as TAP (e.g., Greene & Azevedo, 2009). It must be noticed, however, that retrospective self-reports offer a more general picture of the frequency of students' strategy use, while TAP enable a more fine-grained analysis of students' learning process, for example by exploring the temporal patterns in which it unfolds. This was particularly elaborated on in response to the second research question by applying EPM.

With respect to the second research question (i.e., studying temporal patterns in students' text-learning process based on the sequences in which their strategies are applied), the process models of the learning strategy profiles enabled a qualitative and systematic analysis based on several theoretical models in the field (see introduction section). When overviewing the results regarding the second research question, three major aspects should be noticed.

First, the findings revealed that information organizers and mental learners immediately started their learning process with reading the text before considering their planning, whereas limited and integrated strategy users initiated their learning process with planning before they started to read. In addition, planning was strongly interwoven in the different phases of integrated strategy users' learning process (i.e., before, during and after learning), which was clearly different from the other process models. The connections with planning in integrated strategy users' process model could indicate that ISU adopted a more efficient and systematic study approach (Pintrich, 2000; Zimmerman, 2002). Pressley and colleagues (1987) for instance, consider good strategy users as planful strategy users who think before they act. Their plan is not conceived as a linear sequencing of strategies, however, but rather as interacting and integrating with other strategies throughout the learning process. While at a more basic level, learners will develop a single (reading) plan for reading text materials (for the reading task) before learning, more advanced learners additionally develop a profound (action) plan for task execution and learning (Desoete, 2007; Pressley, 2000), which was the case for integrated strategy users in the current study.

Second, a remarkable difference between the learning strategy profiles regards the use of monitoring strategies. On the one hand, progress monitoring was included in integrated strategy users' process model, suggesting that a considerable share of integrated strategy users actively tracked and controlled the quality of their progress and the available time left for task execution (Meijer et al., 2006; Moos & Azevedo, 2009). By applying this progress monitoring strategy, adherence to the plan is stimulated, as well as revisions to comply with the plan (Pressley et al., 1987). In this respect, progress monitoring during learning is strongly linked to planning before

and during learning. Hence, particularly in integrated strategy users' process model, metacognitive strategies (i.e., planning and progress monitoring) and, by extension, cognitive strategies, mutually interacted. On the other hand, comprehension monitoring was included in the process model of limited strategy users. This strategy refers to control activities directed at the correctness and comprehensiveness of one's understanding (Moos & Azevedo, 2009). An indicator of applying this strategy, for example, concerns learners' noting lack of full understanding, as well as efforts to monitor their understanding after reading the text (Veenman et al., 1997). As previous research shows that limited strategy users' level of reading ability is generally lower compared to the other learning strategy profiles, limited strategy users' level of reading ability may also have played a role here (Rogiers et al., 2019a).

Finally, the results showed that limited strategy users followed a rather linear sequenced approach, whereas integrated strategy users, mental learners, and particularly information organizers adopted a more cyclical approach to learning as they often repeated or returned to prior activities. Compared to the other learning strategy profiles, limited strategy users did not seem to interact with the text as actively and recursively. Instead, they confined their study behaviour to highlighting, memorizing, and rereading. Contrary, particularly information organizers and integrated strategy users frequently switched between various strategies throughout their learning process or resumed previous strategies. According to important theoretical models concerning successful strategy use (Alexander & Jetton, 2000; Pintrich, 2000; Pressley et al., 1987; Weinstein et al., 2011; Zimmerman, 2002), also the simultaneous use of different strategies is what characterises a good strategy user. As different strategies are executed ever more efficiently in good strategy users, Pressley and colleagues (1987) state that in these learners, more short-term capacity is 'left over' to adopt other strategies simultaneously and enhance their text learning (Pressley et al., 1987). In information organizers' process model, for instance, a large share of information organizers returned to reading the text or started memorizing or rehearsing the text after summarizing. Further, memorizing and highlighting were often performed in alternation and highlighting was also frequently followed by rehearsing, initial reading and/or summarizing. Remarkable, however, is that after conducting memorizing, summarizing, and highlighting activities, a large share of information organizers returned to initially reading the text. These paths could imply that information organizers started to engage in different text-learning strategies without first reading or fully understanding the study text. Equally, this might indicate that information organizers had the tendency to interrupt their first reading with other activities (Wade et al., 1990). This could be due to the fact that they did not initiate their actual learning process by planning this process in advance. In this respect, their text-learning process seems less systematic than, for example, integrated strategy users' learning process. Rather than directly selecting important ideas in the text or starting to summarize, the findings indicate that integrated strategy users read text fragments, deliberate on the importance of the given information and then highlighted or summarized the main ideas. Subsequently, integrated strategy users applied their notes as tools to memorize. This might again indicate that these learners adopted a more strategic approach to text learning. However, it is remarkable that integrated strategy users rarely applied rereading or rehearsing strategies during their text-learning process. Since they more actively monitored their progress, it might have been the case that they did not consider it necessary or feasible to repeat or rehears the text within the given time span. Further, also the recursive loop for memorizing, rehearsing, and rereading in mental learners' process model demonstrates an alternate application of these strategies. Although we must be aware of our small sample size, the initial reading activities as final activities in the process models of mental learners and limited strategy users could imply that some students finished their learning process quite abruptly and did not implement a thoughtful text-learning approach.

4.1 Limitations and implications

The present study is associated with some strengths and concerns regarding both the measure and data analysis approach used. First, we must be aware of the fact that various self-report measures reflect learning strategy conceptualizations in a different way. Retrospective self-report data has shown to be valuable in prior research to provide insight into the frequency and variety of applied overt, covert, cognitive, and metacognitive learning strategies during a learning task (e.g., Rogiers et al., 2019a; Merchie et al., 2014). However, this particular data provides us with less information on the cyclical and adaptive nature of these processes, characteristics that have been identified in various theoretical models as being essential in strategic learning. TAP can be regarded as concurrent self-report measures and are recognised as useful data sources to provide additional insight (Dinsmore, 2018; Veenman, 2011). More particularly, by the unique combination of concurrent self-report think-aloud data and educational process mining in this study, we were able to shed light on not only the diversity of applied learning strategies, both also on their cyclical and adaptive nature. In this way, EPM on students' concurrent TAP really opened the black box and provided in-depth insight into the course of different learners' actual text-learning process (Cromley & Wills, 2016; Veenman, 2005). In this respect, this study illustrates the complementarity of both retrospective (i.e., task-specific self-report questionnaires) and concurrent self-reports (i.e., TAP). This

reflection touches upon the first explicit question regarding self-report data tackled throughout the different contributions in this special issue. However, limitations of this study should equally be recognised. A first risk inherent to thinking out loud concerns the incompleteness as automated or unconscious behaviour is not explicitly verbalized (Boekaerts & Corno, 2005). It is possible that students' actions and thoughts might have sometimes remained covert, making them difficult to record in the TAP. Second, students were instructed to report both verbal and nonverbal processes during thinking out loud. To prevent that students' verbalisations did interfere with their learning process (Greene et al., 2011), they were not asked to explain these processes. Therefore, TAP gave no insight into students' underlying motives for their executed activities and sequences that occurred in the process models. Including retrospective stimulated interviews (Schellings & Broekkamp, 2011) based on students' TAP could be useful in future research to learn more about the underlying motives of students' behaviour.

As to the data analysis approach, no specific EPM guidelines are currently available with respect to the number of activities and paths to be included in the process models. More process mining research in educational settings, as well as exploring EPM techniques that rely on different algorithms could therefore contribute to a better understanding of students' learning process on the one hand and to more evidence-based guidelines for conducting EPM on the other hand (Bannert et al., 2014). Related to this, it is to be recommended as well to engage in more fine-grained coding of the think aloud data in future research in view of considering valences of specific SRL processes during students' learning. Positive judgment of learning (e.g., "I am getting this"), for example, can elicit a distinct subsequent SRL process than a negative judgment of learning (e.g., "I am so confused with this paragraph"). Within the scope of the current study, however, this fine-grained coding was not applied (e.g., both 'detecting lack of comprehension' and 'mentioning awareness of understanding' were more generally coded as 'monitoring comprehension'). We therefore make a plea for more fine-grained coding of the distinct subprocesses of particular learning strategies, such as for instance 'comprehension monitoring', to enable the study of more detailed subprocesses and their temporal nature.

Further, prior studies have shown that students adapt their strategies and the effort they spend on studying according to the learning task, their prior domain knowledge, and their learning goals (e.g., Boekaerts & Niemivirta, 2000; Broekkamp & Van Hout-Wolters, 2007). As a result, students may decide to select from their available strategies these strategies that are most appropriate given the assigned learning task, their prior knowledge, and/or the learning goal they set for themselves. In this respect, they might opt, for example, to systematically reread the study text instead of engaging in summarizing and paraphrasing the text (Broekkamp & Van Hout-Wolters, 2007). It will therefore be interesting in future research to study students' strategy use across more and varied learning tasks as well as to investigate the impact of their prior knowledge and their personal learning goals. In this respect, it is to be recommended to also consider other types of prior knowledge tests, such as open questions, multiple choice tests, cloze tests, completion tests, and recognition tests, which also provide valid means of assessment (Dochy et al., 1999).

With regard to the implications for research, this study extends earlier work by including new possibilities for analysing learning processes by means of EPM. This study must be considered as a first important investigation in unravelling patterns in secondary school students' text-learning processes. Educational research is encouraged to fine-tune this type of analysis by the suggestions mentioned above. Further, our results encourage data triangulation in future research, preferably combining both off-line (e.g., self-report questionnaires) and on-line measures (e.g., TAP) to gain a more accurate portrayal of students' learning process (Rogiers et al., 2019b; Veenman, 2005). In view of implications for theory, this study showed that EPM can be used to test the cyclical and adaptive use of learning strategies as put forward in different theoretical models. The present study confirmed differences between four learning strategy profiles in secondary school students. These findings carry important implications for educational practice to help and support students to also evolve towards the more adaptive and cyclical use of strategies. The proposed process models provide a detailed picture on students' text-learning process and could be used as starting points for supporting learner-oriented teaching and learning.

Keypoints

- Integrated strategy users execute more diverse text-learning strategies than the other learner profiles
- Limited and integrated strategy users initiate their learning process with planning this process, while planning is strongly interwoven in the different phases of integrated strategy users' learning process
- Progress monitoring is included in integrated strategy users' process model, comprehension monitoring is included in limited strategy users' process model
- Limited strategy users follow a rather linear sequenced approach to learning whereas integrated strategy users, mental learners, and particularly information organizers adopt a more cyclical approach to learning

References

Alexander, P. A. (1998). The nature of disciplinary and domain learning: The knowledge, interest, and strategic dimensions of learning from subject matter text. In C. R. Hynd (Ed.), *Learning from texts across conceptual domains* (pp. 263-287). New York, NY: Routledge.

Alexander, P. A., & Jetton, T. L. (2000). Learning from text: A multidimensional and developmental perspective. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. III, pp. 285-310). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Alexander, P. A., Grossnickle, E. M., Dumas, D., & Hattan, C. (2018). A retrospective and prospective examination of cognitive strategies and academic development: Where have we come in twenty-five years? In A. O'Donnell (Ed.), *Oxford Handbook of Educational Psychology.* Oxford, UK: Oxford University Press.

Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning, 9*(2), 161–185. https://doi.org/10.1007/s11409-013-9107-6

Bergman, L. (2001). A person approach to adolescence: Some methodological challenges. *Journal of Adolescent Research, 16*(1), 28–53. https://doi.org/10.1177/0743558401161004

Boekaerts, M., & Corno, L. (2005). Self regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology, 54*(2), 199–231. https://doi.org/10.1111/j.1464-0597.2005.00205.x

Boekaerts, M., & Niemivirta, M. (2000). Self-regulated learning: Finding a balance between learning goals and ego-protective goals. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation*. San Diego, CA: Academic Press.

Bråten, I., & Samuelstuen, M. S. (2004). Does the influence of reading purpose on reports of strategic text processing depend on students' topic knowledge? *Journal of Educational Psychology, 96*(2), 324-336. https://doi.org/10.1037/0022-0663.96.2.324

Båten, I., & Samuelstuen, M. S. (2007). Measuring strategic processing: comparing task-specific self-reports to traces. *Metacognition and Learning, 2*(1), 1-20. https://doi.org/10.1007/s11409-007-9004-y

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.

Cromley, J. G., & Wills, T. W. (2016). Flexible strategy use by students who learn much versus little from text: transitions within think-aloud protocols. *Journal of Research in Reading, 39*(1), 50-71. https://doi.org/10.1111/1467-9817.12026

Deed, C., Lesko, T. M., & Lovejoy, V. (2014). Teacher adaptation to personalized learning spaces. *Teacher Development, 18*(3), 369–383. doi:10.1080/13664530.2014.919345

Dinsmore, D. L. (2018). *Strategic Processing in Education*. New York, NY: Routledge.

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research, 69*(2), 145-186. https://doi.org/10.3102/00346543069002145

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

Fluxicon. (2019). *Disco user's guide*. Retrieved from https://fluxicon.com/disco/files/Disco-UserGuide.pdf

Fox, E. (2009). The role of reader characteristics in processing and learning from informational text. *Review of Educational Research, 79*(1), 197-261. https://doi.org/10.3102/0034654308324654

Fryer, L. K., & Vermunt, J. D. (2017). Regulating approaches to learning: Testing learning strategy convergences across a year at university. *British Journal of Educational Psychology*, *88*(1), 21-41. https://doi.org/10.1111/bjep.12169

Günther, C., & Van der Aalst, W. (2007). Fuzzy mining: Adaptive process simplification based on multi-perspective metrics. In G. Alonso, P. Dadam, & M. Rosemann (Eds.), *International conference on business process management (BPM 2007)* (pp. 328–343). Berlin, Germany: Springer.

Greene, J. A., Robertson, J., & Croker Costa, L.-J. (2011). Assessing self-regulated learning using think-aloud methods. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 313–328). New York, NY: Routledge.

McNamara, D. S., Ozuru, Y., Best, R., & O'Reilly, T. (2007). The 4-pronged comprehension strategy framework. In D. S. McNamara (Ed.), *Reading Comprehension Strategies: Theories, Interventions, and Technologies* (pp. 465-496): New York, NY: Erlbaum

Merchie, E., & Van Keer, H. (2014). Learning from Text in Late Elementary Education. Comparing Think-aloud Protocols with Self-reports. *Procedia - Social and Behavioral Sciences, 112*(2013), 489–496. https://doi.org/10.1016/j.sbspro.2014.01.1193

Merchie, E., Van Keer, H., & Vandevelde, S. (2014). Development of the Text-Learning Strategies Inventory: Assessing and profiling learning from texts in fifth and sixth grade. *Journal of Psychoeducational Assessment, 32*(6), 1-15. https://doi.org/10.1177/0734282914525155

Organisation for Economic Cooperation, Development [OECD]. (2006). Personalizing Education. Paris: OECD Publishing.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*(422), 1–28. https://doi.org/10.3389/fpsyg.2017.00422

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, *16*(4), 385–407. https://doi.org/10.1007/s10648-004-0006-x

Pressley, M., Borkowski, J. G., & Schneider, W. (1987). Cognitive strategies: Good strategy users coordinate metacognition and knowledge. *Annals of Child Development*, *4*, 89-129.

Reimann, P. (2007). Time is precious: Why process analysis is essential for CSCL (and can also help to bridge between experimental and descriptive methods). In C. Chinn, G. Erkens & S. Puntambekar (Eds.), *Mice, minds, and society. Proceedings of the computer-supported collaborative learning conference (CSCL 2007)* (pp. 598–607). New Brunswick, NJ: International Society of the Learning Sciences.

Rogiers, A., Merchie, E., & Van Keer, H. (2018). *Fostering text-learning strategies in secondary education through explicit strategy-instruction.* Paper presented at the International Conference of the EARLI SIG 2, Freiburg, Germany, 27-29 august, 2018.

Rogiers, A., Merchie, E., & Van Keer, H. (2019a). Learner profiles in secondary education: Occurrence and relationship with performance and student characteristics. *The Journal of Educational Research, 112*(3), 385-396. https://doi.org/10.1080/00220671.2018.1538093

Rogiers, A., Merchie, E., & Van Keer, H. (2019b). What they say is what they do? Comparing task-specific self-reports, think-aloud protocols, and study traces for measuring secondary school students' text-learning strategies. *European Journal of Psychology of Education*, 1-18. https://doi.org/10.1007/s10212-019-00429-5

Samuelstuen, M. S., & Bråten, I. (2007). Examining the validity of self-reports on scales measuring students' strategic processing. *British Journal of Educational Psychology, 77*(2), 351-378. https://doi.org/10.1348/000709906X106147

Schellings, G. L. M., & Broekkamp, H. (2011). Signaling task awareness in think-aloud protocols from students selecting relevant information from text. *Metacognition and Learning, 6,* 65–82. https://doi.org/10.1007/s11409-010-9067-z

Slotte, V., Lonka, K., & Lindblom-Ylanne, S. (2001). Study-strategy use in learning from text. Does gender make any difference? *Instructional Science, 29*(3), 255-272. https://doi.org/10.1023/a:1017574300304

Schoor, C., & Bannert, M. (2012). Exploring regulatory processes during a computer-supported collaborative learning task using process mining. *Computers in Human Behavior, 28*(4), 1321–1331. https://doi.org/10.1016/j.chb.2012.02.016

Scott, D. B. (2008). Assessing text processing: A comparison of four methods. *Journal of Literacy Research, 40*, 290-316. doi:10.1080/10862960802502162

Simpson, M. L., & Nist, S. L. (2000). An update on strategic learning: It's more than textbook reading strategies. *Journal of Adolescent & Adult Literacy*, *43*(6), 528-541.

Stoeger, H., Fleischmann, S., Obergriesser, S. (2015). Self-regulated learning (SRL) and the gifted learner in primary school: The theoretical basis and empirical findings on a research program dedicated to ensuring that all students learn to regulate their own learning. *Asia Pacific Education Review,* 16, 257-267. https://doi.org/ 10.1007/s12564-015-9376-7

Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In C. Artett & B. Moschner (Eds.), *Ledrnstrategien und metakognition. Implikationen für forschung und praxis* (pp. 77-99). Münster: Waxmann.

Veenman, M. V. J. (2011). Alternative assessment of strategy use with self-report instruments: A discussion. *Metacognition and Learning, 6*(2), 205–211. https://doi.org/10.1007/s11409-011-9080-x

Veenman, M. V., Elshout, J. J., & Meijer, J. (1997). The generality vs. domain-specificity of metacognitive skills in novice learning across domains. *Learning and Instruction, 7*(2)*,* 187–209. https://doi.org/10.1016/S0959-4752(96)00025-4

Wade, S. E., Trathen, W., & Schraw, G. (1990). An analysis of spontaneous study strategies. *Reading Research Quarterly, 25*(2), 147-166. https://doi.org/10.2307/747599

Weinstein, C. E., Jung, J., & Acee, T.W. (2011). Learning strategies In V. G. Aukrust (Ed.), *Learning and Cognition in Education* (pp. 137-143). Oxford, UK: Elsevier Limited.

Van der Aalst. (2011). *Process mining: Discovery, conformance and enhancement of business processes.* New York, NY: Springer.

van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method. A practical guide to modelling cognitive processes.* London, UK: Academic Press.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice, 41*(2), 64–71.

Young, K. A. (2005). Direct from the source: the value of 'think-aloud' data in understanding learning. *Journal of Educational Enquiry, 6*(1), 19-33.

Appendices

Appendix A

Descriptive statistics and reliability coefficients of the different Text-Learning Strategies Inventory subscales

TLSI-subscales	N items	Example item	M (SD)	Cronbach's α
Summarizing and schematizing (SS)	7	*I repeated the text with my summary or graphic organizer on my scratch paper*	3.14 (1.30)	.92
Highlighting (HL)	1	*I marked the most important things*	4.31 (1.31)	/
Rereading (RR)	3	*To learn the text, I read the text a lot of times*	3.14 (1.07)	.72
Paraphrasing (PAR)	7	*I covered up the text information and tried to recall it*	3.02 (0.82)	.71
Linking with prior knowledge (LPK)	3	*Before learning, I thought about what I already knew*	3.16 (1.09)	.75
Studying titles and pictures (TP)	3	*I looked at the titles to understand the text*	2.81 (1.07)	.70
Planful approach (PA)	3	*First, I read the whole text and then I started learning*	3.86 (1.08)	.65
Self-evaluation (SE)	5	*While learning, I checked what I had already done and how much I still had to do*	4.03 (0.63)	.70
Monitoring (MON)	5	*I managed to learn the text in a good way*	3.21 (0.88)	.63

Note. TLSI = Text-Learning Strategies Inventory. Cronbach's α is based on the total sample of 1,931 students wherein learner profiles were determined.

Appendix B

Results of the Post Hoc Pairwise comparisons between the four learning profiles on the different coding categories

Coding category		IO		ML		LSU		
Overt cognitive strategies		**p**	**Cohen's d**	**p**	**Cohen's d**	**p**	**Cohen's d**	**F**
Summarizing	ISU	.128	.075	.000***	.700	.000***	.403	$F(3, 4732) = 85.59, p < .001$
	IO			.000***	.631	.000***	.329	
	ML					.000***	-.351	
Highlighting	ISU	.291	.077	.000***	.511	.951	.000	$F(3, 4732) = 43.88, p < .001$
	IO			.000***	.435	.288	-.076	
	ML					.000***	-.510	
Covert cognitive strategies		**p**	**Cohen's d**	**p**	**Cohen's d**	**p**	**Cohen's d**	**F**
Memorizing	ISU	.720	.058	.987	-.028	.183	-.081	$F(3, 4732) = 4.54, p = .004$
	IO			.373	-.086	.002**	-.139	
	ML					.893	-.053	
Rereading	ISU	.001***	-.164	.000***	-.208	.006**	.199	$F(3, 4732) = 22.89, p < .001$
	IO			.996	-.046	.000***	.335	
	ML					.000***	.372	
Rehearsing	ISU	.013*	-.141	.000***	-.617	.835	.063	$F(3, 4732) = 144.34, p < .001$
	IO			.000***	-.494	.000***	.205	
	ML					.000***	.671	
Paraphrasing	ISU	.023*	.160	.000***	-.193	.045*	.083	$F(3, 4732) = 21.12, p < .001$
	IO			.000***	-.356	.000***	-.233	

		p	Cohen's d	p	Cohen's d	p	Cohen's d	F
	ML					.156	.107	
Elaborating	ISU	.004**	.184	.978	.064	.991		$F_{(3, 4732)} = 5.252$, $p = .001$
	IO			.226	-.080	.004**	-.146	
	ML					.874	-.063	
Metacognitive strategies		**p**	**Cohen's d**	**p**	**Cohen's d**	**p**	**Cohen's d**	**F**
Planning	ISU	.000***	-.264	.010*	-.180	.000***	-.423	$F_{(3, 4732)} = 38.56$, $p < .001$
	IO			.438	.089	.000***	-.178	
	ML					.000***	-.264	
Monitoring progress	ISU	.000***	.210	.000***	.280	.232	.093	$F_{(3, 4732)} = 15.57$, $p < .001$
	IO			.995	.082	.016*	-.116	
	ML					.001***	-.189	

Note. *p* ≤ .05, ** *p* ≤ .01, ****p* ≤ .001. To interpret effect sizes, Cohen's benchmarks for the social sciences apply (i.e., small effect size: *d* = 0.2, medium effect size: *d* = 0.5, and large effect size: *d* = 0.8; Cohen, 1977). *ISU* = integrated strategy users, *IO* = information organizers, *ML* = mental learners, *LSU* = limited strategy users.