Study strategies in a computer assisted study environment

Jos J. Beishuizen*, Evelien T. Stoutjesdijk

Leiden University, Department of Developmental and Educational Psychology, P.O. Box 9555, 2300 RB Leiden, The Netherlands

Abstract

A Computer Assisted Study Environment (CASE) was developed as a tool for diagnosing study problems, to be used together with other sources of information, such as learning style questionnaires and clinical interviews. Forty-one students were observed during a 1 h period of studying a text book chapter in CASE. The stages of orientation, planning, and execution were clearly separated by dividing the 1 h study session into three periods. Students could spend an unlimited part of the hour on orientation (first period). Then, not included within the hour, a plan for the task had to be made (second period), after which the remaining time could be devoted to execution of the plan or to any other form of study (third period). We administered a learning style questionnaire, measured reading speed and pretested the students on prior knowledge of the content of the study task. These data were correlated with product and process indicators collected in CASE in order to find out whether various sources of information about learning styles and study strategies provided converging evidence about potential causes of study problems. Process and product indicators of study strategies in CASE revealed differences between deep and surface learning students in orientation and planning activities. However, their actual study behaviour did not vary according to learning style. So, the differences between surface and deep learning become more apparent when we look at study activities before and after actual reading and processing information. As far as learning outcomes are concerned, students with a deep learning style obtained better results than students with a surface learning style, even for factual knowledge. Deep learning students knew more about the subject of the diagnostic study task, and developed a higher reading speed. Both student characteristics significantly determined learning outcomes. © 1999 Elsevier Science Ltd. All rights reserved.

* Author for correspondence. E-mail: beishuiz@rulfsw.leidenuniv.nl
1. Introduction

A student addresses an adviser to discuss his study problems. In a first meeting, he describes his current results and progress. His grades are insufficient to proceed to next year’s class. The adviser considers several alternatives for further diagnosis. First, she might administer a questionnaire about the student’s learning styles. Fortunately, she has several well validated instruments at her disposal. She decides to use the Dutch Inventory of Learning Styles (Vermunt & Van Rijswijk, 1988; Vermunt, 1992). This inventory makes, among other things, a distinction between deep and surface learning styles, and between preferences for internal regulation or external regulation of study tasks. However, a profile on a learning style questionnaire does not explain the causes of insufficient study results. There is no one-to-one correspondence between scores on learning styles questionnaires and study outcomes. So, the adviser has to envisage a supplementary approach. She interviews the student about his learning styles and study problems, his preferences for internal or external regulation, and his conceptions of learning. The clinical interview is an appropriate instrument to gain insight into students’ motives and considerations, as many studies according to the phenomenographic approach have shown (e.g. Marton et al., 1997a,b). However, it still does not provide conclusive answers about causes and effects. An alternative approach might be to observe the student whilst carrying out a study task, gather performance characteristics and outcome measures, and relate these figures to questionnaire and interview data. That, however, is a time-consuming endeavour. Moreover, the study task has to be carefully designed to provide reliable and valid information on inappropriate study strategies as potential causes of students’ study problems.

Designing such a diagnostic study task was the main objective of the research reported here. A computer assisted study environment (CASE) was designed to provide information on study strategies, complimentary to interview and questionnaire data. We administered a learning style questionnaire, measured reading speed and pretested the students on prior knowledge of the content of the study task. These data were correlated with product and process indicators collected in CASE in order to find out whether various sources of information about learning styles and study strategies provided converging evidence about potential causes of study problems.

In this introduction we aim at clarifying the relationships between learning styles as expressed in questionnaires and actual study strategies we expected to encounter in CASE.

According to several authors, consistency in students’ use of strategies can be attributed to their learning style (Pask, 1976; Entwistle, 1988; Schmeck, 1983; Vermunt, 1992). Schmeck (1983) defined a learning style as “a predisposition on the part of some students to adopt a particular learning strategy regardless of the specific demands of the learning task” (p. 233). Pask (1976) found two clearly distinct learning styles: the operation learning style and the comprehension learning style. The former is characterised by a pattern of activities which he called a “serialistic” or step-by-step strategy, while the latter is characterised by the use of a “holistic” or structuring strategy. Vermunt & Van Rijswijk (1987, 1988), who constructed a Dutch
learning style questionnaire, made a similar distinction between surface processing and deep processing of text. Deep processing is characterised by actively relating various parts of the text to one another and to prior knowledge, by organising separate topics into a whole to gain insight into the subject matter, and by a critical approach. Surface processing, in contrast, is characterised by step-by-step analysis of the material and reading of each part of the text thoroughly, with great attention to detail and focus on factual information, by an attempt to remember as much as possible, and by the use of study strategies like rehearsing and memorising. Deep processing and surface processing are not mutually exclusive. Both strategies can be utilised by the same student in the same study session.

A learning style is more than a habitual way of studying. Other student characteristics have been related to the learning style concept: personality characteristics and attitudes (Schmeck et al., 1991), study motives or orientations (Entwistle & Ramsden, 1983; Biggs, 1984; Vermunt, 1992), conceptions of learning, or knowledge and beliefs about learning and education (Marton & Säljö, 1997; Vermunt, 1992), and metacognitive regulation activities (Vermunt, 1992). In the view of Vermunt & Van Rijswijk (1988), a learning style “refers to a coherent whole of activities, study orientations and conceptions of learning […] that is characteristic for a certain student at a certain moment in time” (pp. 654–655). The information processing component of learning styles cannot be separated from other components of learning styles. Students with a deep learning style often prefer, for instance, to regulate their learning processes by themselves; conversely, students with a surface learning style, often prefer externally regulated learning (Vermunt, 1992). In sum, a learning style is a tendency to approach cognitive tasks with a preferred strategy or set of strategies, corresponding with a more general conception of learning.

What is the relationship between the student’s learning style, identified with a questionnaire, and his or her actual study strategies in a concrete study task? Text studying can be described as a cycle of stages: orientation, planning, and execution (Kaldeway & Oost, 1990). The activities during each of these stages are meant to lead to specific (intermediate) products. Orientation, or stock-taking, entails reviewing one’s position and resources, and results into a task conception. During this first stage, the student gathers information about the study goals to be accomplished and the available means. Next, a more or less extensive plan is made by putting this information together and relating it to what the student knows or believes about study methods and their effects. The plan is carried out in the information processing stage, or execution stage, which should lead to learning results. The cycle of activities as described here is not a one-time event. In order to decide whether to stop the activities cycle or to continue, the student has to evaluate the (intermediate) learning results and the study activities undertaken so far by again taking stock of the situation. This could lead to adjustments of the plan or to the making of new plans, etc. The learning results that have been achieved when studying actually stops form the final product of the cycle of activities. This cycle appears at all levels of studying: from the way a complete course is approached, through preparation for an exam, to the study of one text passage.

From defining deep and surface processing as learning styles, it is a small step to
making stage specific predictions about study strategies. Since students with a deep learning style pay considerable attention to the organisation of a study text, they are supposed to spend more time on orientation activities, for instance by consulting the table of contents, than students with a surface learning style. In the same way, measures could be composed for the other stages and for their intermediate products. Deep orientation induces deep task conceptions and the making of plans with deep elements, while the task conceptions and plans of students with a surface learning style would contain more surface processing elements. The same would apply to predictions about execution of these plans.

With regard to post-test performance, most authors in the field of learning styles suggest that deep processing, as an important component of the meaning-directed learning style, is the superior way of learning. It should lead to better learning results than surface processing, even for factual knowledge. Students with a deep learning style build comprehensive schemata in which the facts are linked with various cues, so that they can be easily retrieved. Several researchers have indeed reported significant correlations between depth-of-processing scales and various indicators of study results for high school students (Ainley, 1993) and university students (Borduin & Van Gogh, 1990; Vermunt, 1992; Pintrich et al., 1993; Albaili, 1994; Busato et al., 1995). Positive correlations from 0.13 to 0.39 between deep processing and achievement were found in all but one study. The reported correlations between surface processing and study results are more diffuse. Negative correlations were found in two studies (Vermunt, 1992; Ainley, 1993), a small positive correlation in one (Albaili, 1994), and no significant correlations in the remaining studies.

Finally, we introduce three other factors which are responsible for considerable individual differences in learning outcomes: prior knowledge, reading speed, and test expectations. We briefly discuss relationships between learning styles and each of these three student characteristics.

Students with a deep learning style deploy more efficient storage and retrieval strategies than students with a surface learning style. Therefore, they probably can activate more and better organised prior knowledge than students with a surface learning style. By the same token, deep processing leads to faster reading than surface processing. Test expectation is known to influence study behaviour, as well as learning results (Verheul & Yang, 1989; Tang & Biggs, 1993; McDaniel et al., 1994; Thiede & Dunlosky, 1994). An important aspect of test expectation is question format: multiple choice or essay. Essay questions are perceived to demand more high-level processing than multiple choice questions. Students expecting essay questions have been found to study more intensively and to perform better on both essay and multiple choice tests than students expecting multiple choice questions (e.g. Verheul & Yang, 1989). Differences were found, for instance, in comprehension directed activities like comparing. This seems to be closely related to deep processing.

Tang & Biggs (1993) investigated the effects of learning style and criterion task expectation on preparation strategy—measured with a questionnaire—and achievement. For both a low-level test and a high-level writing assignment, they found the anticipated general congruence between learning style and exam preparation strategy.
But they also reported an interaction between style and test expectation. Students with a deep learning style were more flexible, in that they used more surface processing strategies like memorisation when they expected a test demanding low-level processing only. Other authors have also reported on the flexibility of deep processing (e.g. Vermunt, 1992). Possibly as a result of this flexibility, students with a deep learning style achieved better on both tasks.

In this study, students with deep or surface learning styles were observed during a 1 h period of preparation for a short test, covering a broad, but clearly defined, topic from a text book chapter. In one of two conditions, a test with short essay questions was announced; in the other condition, students were led to expect multiple choice questions. The stages of orientation, planning, and execution within a single study session were clearly separated by dividing the 1 h study session into three periods. Students could spend an unlimited part of the hour on orientation (first period). Then, not included within the hour, a plan for the task had to be made (second period), after which the remaining time could be devoted to execution of the plan or to any other form of study (third period). The phases of orientation and execution were monitored by means of a computer assisted study environment (CASE) in which every act was recorded. Apart from the study process, also the products of each of the three periods were registered: After orientation, students were briefly interviewed about their conception of the task, and they filled out plan forms. After completing the 1 h study session, students filled out plan forms again, this time describing the plan that would have been ideal in their view. Eventually, they answered both multiple choice and essay post-test questions, requiring comprehension or factual knowledge. Reading speed and prior knowledge had been assessed beforehand.

There are several reasons to question the ecological validity of the experimental study task. First, students were confined to work for just 1 h, separated in three periods. Secondly, the students had to work on their own. There were no peers around to collaborate with. Thirdly, they had to work in CASE, in which a computer displayed the study text on an A4 screen. However, CASE had been extensively tried out, evaluated and redesigned beforehand to accommodate for inconveniences in the user interface (Beishuizen et al., 1994; Verheij et al., 1996). After completing the study task, students were asked how they had appreciated the task and whether they had felt insecure about the way they carried out the task. There were no signs of embarrassment having to do with the particular conditions of CASE. Therefore, we had no reasons to expect detrimental effects of the task conditions on either students with a deep or surface learning style. We will return to this issue in Section 4.

The research questions concerned observable differences between deep and surface learning students. Would students with a deep learning style indeed show a more elaborate orientation, construct a deeper conception of the task, and plan and execute more deep activities than students with a surface learning style? And would deep learning lead to better results on the post-test than surface learning, both for comprehension and factual knowledge? A simple model was used to answer these questions: Deep and surface learning students were merely compared on study activities and learning outcome.
We pitted a more complicated model against the simple two-groups model, to find out whether learning style made an independent contribution to the explanation of learning outcomes, after test expectation, reading speed, and prior knowledge were taken into account.

2. Method

2.1. Subjects

Students were selected from 364 freshmen psychology students at Leiden University who had completed the Inventory of Learning Styles (ILS) (Vermunt & Van Rijswijk, 1987). To ensure homogeneity of general knowledge level and experience as a university student, several conditions for participation were set: having started as a university student just a few months before, having finished high school in that same year, and having participated in the first exam at the end of the introductory psychology course.

The actual selection of students with deep and surface learning styles was based on ILS scores. Students with a deep learning style were defined as students with a standardised score on the Deep Processing scale of at least 0.75, who at the same time scored below average on the Surface Processing scale. Students with a surface learning style, on the other hand, were defined as students with a standardised score on the Surface Processing scale of at least 0.75 and a below average score on the Deep Processing scale. Based on these criteria, we located 41 students who were willing to participate for a small remuneration: 21 students with a deep learning style and 20 students with a surface learning style. Students of each group were randomly assigned to two test expectation conditions. Expectations were elicited by the instruction to prepare for either a multiple choice or an essay test.

2.2. Materials

2.2.1. Learning style inventory

To measure learning styles, the Inventory of Learning Styles (Vermunt & Van Rijswijk, 1987) was used. This questionnaire consists of 120 items (in Dutch) about study habits, study motives, and learning conceptions. We used scores on two scales concerning study habits: Deep Processing and Surface Processing. Each of these scales consists of 11 statements about study activities, to be rated on a scale from 1 (Seldom or never) to 5 (Virtually always). The contents of these items are described below. Among university students, the reliabilities of the scales (Cronbach’s alpha) are 0.85 for Deep Processing and 0.78 for Surface Processing (Vermunt, 1992).

2.2.2. Computer assisted study environment (CASE)

The text to be studied was a chapter in English about Freud and his theory (c. 15 000 words) from an introductory textbook on psychology (Gleitman, 1986). An electronic version of this text was displayed on a Macintosh computer with an A4
screen. The structure of the electronic version was much like that of a book: It consisted of a table of contents, 60 short text pages, and an Index, in this order. The lay-out and the navigation facilities also looked like those of a book. Research has shown that user-friendliness of electronic texts depends on the availability of book-like cues like table of contents and page numbers. Readers perform better in reading and study tasks when the electronic text looks familiar to them (Gould & Grischkowski, 1984; Benest, 1990). Other features of CASE that resembled a book format were buttons to move back and forth one page at a time, display of the page number in the lower right corner, and the possibility of marking a text page with a “dog ear” to enable the reader to return to a marked page. Fig. 1 provides a picture of the CASE interface.

Some extra features not available in a paper version text were added to compensate for the fact that an electronic study environment does not enable the reader to get a sense of the total amount of material at a glance. First, there was an Overview Chart where all sections of the chapter and the associated pages were listed and where the user could see which text pages had been seen and which had not. This chart furthermore provided the user with the opportunity to go to a Question Box for any chosen section. Here one could answer or simply read multiple choice questions pertaining to that section. CASE checked whether the user’s answers given to these questions were correct and indicated which of the questions had already been seen. A third extra tool was an extract facility: a Selection Buffer where one could save

![CONTENTS OVERVIEW INDEX](image)

**UNCONSCIOUS CONFLICT**

Our sketch of Freud’s theory of the nature and development of human personality will concentrate on those aspects that seem to represent the highlights of a complex theoretical formulation that was continually revised and modified during the course of Freud’s long career. In this description, we will separate two aspects of Freudian theory. We will begin with the conception of the mechanisms of unconscious conflict. We will then deal with Freud’s theory of the origins of these conflicts in the individual’s life history, and of their relation to the development of sexual identity and morality.

![Fig. 1. Display of a text page in CASE.](image)
any text pages and where one could type and edit. Other extra tools were added for recording purposes. There was a clock that displayed how many minutes were left for the task, and CASE could be turned into stand-by mode when the student wanted to take a break.

The user could move through all of these pages and facilities by clicking on buttons at the top of the screen (move one page forward or backward; go directly to the Table of Contents, or to the Index, to the Overview Chart, to the Selection buffer, or to the marked page) or by clicking on the page numbers, listed in the Table of Contents, in the Index, and in the Overview Chart. Other possible actions were saving a text page in the buffer, making a new page mark, and viewing figures and footnotes mentioned in the text. All of the user’s actions were registered, together with a time-stamp.

In order to create assessment scales for describing patterns of study activities we attempted to stay as closely as possible to the way the concepts of deep learning and surface learning were worked out in the Inventory of Learning Styles (Vermunt & Van Rijswijk, 1987). Analysis of the contents of the items indicated that the Deep Processing scale comprises four main components: (1) trying to gain insight or deep understanding of the subject matter by actively connecting its various parts and looking for similarities and differences; (2) trying to get an overview of context, problem and materials; (3) activating prior knowledge and connect new information to existing knowledge; and (4) approaching the subject matter with a critical attitude by analysing the logic of conclusions and expert interpretations, and drawing one’s own conclusions. The Surface Processing scale contains the following components: (1) using a linear method, working step by step; (2) focusing on details, facts, definitions, and series of features; (3) working thoroughly through the subject matter, trying to study everything intensely and literally; (4) memorising by rote learning.

These components of a deep and surface learning style were used to develop a checklist for scoring the study protocols as recorded by CASE (see Appendix A). Two independent judges checked whether or not the items on the checklist were present. The criteria were mostly objective, so that in the few cases of disagreement, the judges could easily reach agreement after rechecking the data. The checklist items were combined into composite variables by computing totals for each process or product for both types of processing. The totals were converted to 10-point scales. Appendix A contains an overview of these dependent variables and the elements of which they consisted: deep and surface elements of orientation, task conception, drawing a plan and an ideal plan, and actually studying.

2.2.3. Diagnostic study task

The study assignment was stated on paper. The task was to work in CASE on studying the chapter about Freud for 1 h. The students were instructed to prepare for a test consisting of either short essay questions or multiple choice questions on Freud’s theory. The test would take place immediately after the study period. In fact, the chapter to be studied covered Freud’s theories as well as modern evidence both supportive of and contradictory to his theories, but the students were explicitly told
that the test would demand “knowledge and understanding of the theories of Freud himself, not of research and alternative explanations offered by other authors.” The task-relevant portion of the chapter took up 34 of the 60 short pages.

2.2.4. Plan forms

Standard forms were designed to guide the students in making their actual and ideal study plans. These forms consisted of a list of activities in four categories: reading, studying, note taking, and knowledge control, and one free category. The activities were identified in previous research in which students made plans without being given any suggestions. Appendix A shows, in the Planning section, the categories to which the activities in the plan form belonged. Subjects were instructed to think of what they would do first, and to write the number “1” one or more times before the activity or activities concerned. If an activity was not on the list, they could add it themselves. After this, they were asked to think of what they would do next, if anything, and to assign that activity the number “2”, etc. For the ideal plan, the same form was used, the only difference being that the student was instructed to mark the activities which would constitute the best possible plan. This was, as they were told, “the plan an ideal student would have made, given that this ideal student resembles you in domain knowledge and reading speed.” For each student, the number of deep processing and surface processing activities marked on the plan forms were separately scored.

2.2.5. Prior knowledge test

To control for prior knowledge, a pretest was administered. This prior knowledge test consisted of 12 questions about the most prominent aspects of Freudian theory. Students were asked, for instance, to place the names of the three components of Freud’s personality structure at the correct location in a scheme, to describe what fuels these components, and to categorise two examples of defense mechanisms using Freud’s concepts. After conversion to a 10-point scale, the scores on prior knowledge ranged from 0 to 10, with mean = 5.3 and S.D. = 2.4. The reliability of the test was not very high (Cronbach’s alpha = 0.52), yet satisfactory in view of the fact that a variety of knowledge types was measured (e.g. accurate memory for terminology as well as comprehension of Freud’s personality model).

2.2.6. Reading speed test

A short reading speed test was included in the experimental session. A text section from the same textbook as the text on Freud, but from another chapter (about Piaget), was presented sentence by sentence on a computer screen. The student was requested to read each sentence at a pace allowing for easy understanding. At the end of each sentence the student clicked the mouse to see the next sentence. The sentences which had already been read did not disappear from the screen, so that the text grew larger until the whole passage had been read. The test consisted of 10 sentences of various lengths, from five to 39 words, with a total of 273 words. For each student, reading speed was determined by computing the average reading time per 100 words. Separate scores for the 10 sentences showed high reliability (Cronbach’s alpha = 0.85).
The overall time per 100 words ranged from 18 to 63 s, with mean = 34.8 s and S.D. = 11.3 s.

2.2.7. Post-test

The post-test consisted of 12 multiple choice questions scored 0 or 1, and nine essay questions scored between 0 and 1. All questions regarded the relevant part of the chapter, as indicated in the assignment. The questions demanded comprehension of subject matter, knowledge of facts, and recall of terminology. Reliability analyses indicated that two of the multiple choice and one of the essay items had negative item-total correlations. A scale consisting of the 10 remaining multiple choice items had a Cronbach’s alpha reliability of 0.58; the remaining eight essay items had a Cronbach’s alpha reliability of 0.53. Multiple choice scores ranged from 1 to 10, with mean = 6.4 (S.D. = 2.0). For ease of interpretation, essay scores were converted to a 10-point scale as well. After conversion, they ranged from 0.6 to 9.2, with mean = 5.5 (S.D. = 1.8).

A factor analysis (principal-components analysis with varimax rotation) revealed three factors: (1) Terminology, or recall of terms (four multiple choice and three essay items, mean = 6.7, S.D. = 2.6, Cronbach’s alpha = 0.73); (2) Comprehension, or reasoning with and inferring from the subject matter (three multiple choice and three essay items, mean = 4.8, S.D. = 2.6, Cronbach’s alpha = 0.65); and (3) Knowledge of facts or reproduction (three multiple choice and two essay items, mean = 6.5, S.D. = 2.3, Cronbach’s alpha = 0.57).

2.3. Procedure

For each student, an individual 2 h session was scheduled. After a short introduction, the Prior Knowledge Test and Reading Speed Test were administered. This took about 15 min. Another 15 min were devoted to instructions for studying and operating CASE. The study assignment was handed out on paper and read aloud by the experimenter. Test expectation was manipulated by varying the assignment phrase in which the type of test questions was announced: essay or multiple choice. All functions of CASE were explained and demonstrated, and the student was given an opportunity to practise. During this introduction, not all text pages were available and the Table of Contents, the Overview Chart, and the Index were not completely visible, so that the student could not utilise this time to acquire an overview of the contents of the chapter.

The next 70 min were devoted to the study session. Before the session started, students were asked to take off their watches to make sure they used the program clock for time monitoring. The instruction for the first phase was to begin with orientation. The student was left free in deciding how much time to spend on orientation, as well as what would be done during this time. Two restrictions were made: Orientation and study time would together last 1 h, and during orientation the student was not supposed to “really study”. The student was asked to call the experimenter as soon as he or she wished to start studying. At this point the clock stopped. The student was asked what he or she had done during the period of orientation. Further-
more, questions about the student’s task conception were posed: “Do you feel that you now have an overview of the chapter?” and if so: “Can you give me a short overview?” “How many pages does this chapter have?” “With what goal will you start studying in a minute?” The experimenter wrote down the student’s answers. The student was then asked to make a plan. Then the study hour, and the clock, recommenced. The experimenter explained that it was not necessary to stick to the plan just made and left the student alone, going to an adjacent room. After a total of 60 min, CASE terminated the study session, and the experimenter came back and asked the student to make the ideal plan (c. 10 min), and then to take the post-test (c. 15 min).

2.4. Design and data-analysis

To investigate the differences between students with a deep or surface learning style (the simple model), *t*-tests were used to compare both groups on all dependent variables. Then, for each phase (orientation, planning, execution) a multivariate covariance analysis was applied to examine main and interaction effects of learning style and test expectation as the independent variables, while prior knowledge and reading speed were taken into account as control variables (the complex model).

3. Results

3.1. The simple model: comparison of students with deep and surface learning styles

Differences between students with a deep and a surface learning style were tested with *t*-tests (see Table 1). Since the expected directions were all specified, one-tailed tests might have been justifiable. Nevertheless, two-tailed significance tests were used to ensure a fair comparison with the complex analyses.

All differences between students with deep and surface learning styles were in the hypothesised direction. No significant differences were found regarding any of the variables related to surface processing. For most of the variables related to deep processing, students with a deep learning style had higher scores than students with a surface learning style. Actual orientation behaviour, and to a lesser extent the subsequent self-reports, indicated that students with a deep learning style engaged in more orientation than students with a surface learning style. Furthermore, students with a deep learning style seemed to have a deeper task conception and included more deep processing activities in their plans. Their behaviour during the execution phase of studying, however, was not different from that of students with a surface learning style: Both groups did equally well on Deep Study Activities, as well as on Surface Study Activities. From our analysis of the ideal plans, it appeared that the ideals of students with a deep learning style were a bit deeper than those of students with a surface learning style, but the two groups did not diverge as much as they did in their actual plans. Students with a deep learning style performed sig-
Table 1
Results of students with deep and surface learning style on all dependent variables

<table>
<thead>
<tr>
<th>Variables (max = 10)</th>
<th>Students with deep learning style (n = 21)</th>
<th>Students with surface learning style (n = 20)</th>
<th>Two-tailed sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Orientation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep orientation</td>
<td>5.6</td>
<td>3.2</td>
<td>3.6</td>
</tr>
<tr>
<td>activities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep orientation self-report</td>
<td>2.6</td>
<td>2.0</td>
<td>1.6</td>
</tr>
<tr>
<td>Deep task conception</td>
<td>3.0</td>
<td>1.7</td>
<td>2.1</td>
</tr>
<tr>
<td>surface conception</td>
<td>2.4</td>
<td>1.8</td>
<td>2.9</td>
</tr>
<tr>
<td>Planning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep plan</td>
<td>4.2</td>
<td>2.1</td>
<td>2.2</td>
</tr>
<tr>
<td>Deep ideal plan</td>
<td>3.9</td>
<td>1.8</td>
<td>3.0</td>
</tr>
<tr>
<td>Surface plan</td>
<td>2.0</td>
<td>1.7</td>
<td>2.6</td>
</tr>
<tr>
<td>Surface ideal plan</td>
<td>2.6</td>
<td>1.9</td>
<td>3.5</td>
</tr>
<tr>
<td>Execution</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep study activities</td>
<td>4.9</td>
<td>2.1</td>
<td>4.3</td>
</tr>
<tr>
<td>Surface study activities</td>
<td>4.1</td>
<td>1.9</td>
<td>4.7</td>
</tr>
<tr>
<td>Post-test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple choice post-test</td>
<td>7.2</td>
<td>1.5</td>
<td>5.7</td>
</tr>
<tr>
<td>Essay post-test</td>
<td>6.0</td>
<td>1.5</td>
<td>5.0</td>
</tr>
<tr>
<td>Subscore comprehension</td>
<td>5.6</td>
<td>2.4</td>
<td>4.0</td>
</tr>
<tr>
<td>Subscore knowledge of facts</td>
<td>7.5</td>
<td>1.9</td>
<td>5.4</td>
</tr>
<tr>
<td>Subscore terminology</td>
<td>7.0</td>
<td>2.2</td>
<td>6.5</td>
</tr>
</tbody>
</table>

*p < 0.05; **p < 0.001; ***p < 0.05 if tested one-tailed.

significantly better than students with a surface learning style on both the multiple choice and the essay post-tests; this was mainly attributable to a difference in factual knowledge, and to a lesser extent to a difference in comprehension. No significant differences were found on the terminology subscore.

The differences between the two groups may have been caused by variations in reading speed or prior knowledge. The two groups did indeed differ in both reading speed and prior knowledge. On average, students with a deep learning style needed less time to read 100 words (mean = 30.7 s, S.D. = 8.8) than students with a surface learning style (mean = 39.0 s, S.D. = 12.3; p < 0.05). Students with a deep learning style obtained higher scores on the prior knowledge test (mean = 6.1, S.D. = 2.3) than students with a surface learning style (mean = 4.7, S.D. = 2.2; p < 0.05). In the complex model both learning style and test expectation were taken into account, while controlling for reading speed and prior knowledge. Results of covariance
analyses on the various study phases as well as on post-test results are presented in the next section.

3.2. The complex model: reading speed, prior knowledge, and test expectation

In Table 2, multivariate $F$-values for all analyses are reported. Only in the planning phase did the main effect of style remain evident after Reading Speed, Prior Knowledge, and Test Expectation were taken into account. The first MANCOVA concerned the orientation phase. The group of students with a deep learning style and with an essay test expectation seems to have employed and reported remarkably more activities than the three other groups. Despite a significant univariate effect of style on the amount of Deep Orientation Activities ($F(1,35) = 5.24, p < 0.05$), the multivariate $F$ for style was not significant (see Table 2). For the variables Deep Orientation Self-Report and Deep Task Conception, the univariate $F$s for style were 2.26 and 2.37, respectively, and not significant. Thus, the difference between students with deep and surface learning style on orientation activities, as revealed by the simple model, did not produce a significant effect attributed to style in the complex model. That is, Learning Style, Reading Speed and Prior Knowledge did not provide independent contributions to the variance in quantity and quality of orientation behaviour during the diagnostic study task in CASE.

In Table 3, the results of the covariance analysis on the planning variables are presented. Style had a multivariate significant effect (see Table 2), and univariate $F$ tests on each of the dependent variables were significant. These main effects of style paralleled the direction of the differences and tendencies we had found earlier, and even strengthened them. On top of the finding that students with a deep learning style scored significantly higher on Deep Plan and Deep Ideal Plan than students with a surface learning style, we found the complimentary effect that the latter stu-

<table>
<thead>
<tr>
<th>Analysis</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Planning and task conceptions</td>
<td>Plan and ideal plan</td>
<td>Execution</td>
<td>Post-test</td>
</tr>
<tr>
<td>Style (deep vs surface)</td>
<td>1.85</td>
<td>3.19*</td>
<td></td>
<td>1.16</td>
</tr>
<tr>
<td>Test expectation (multiple choice vs essay)</td>
<td>1.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Style × test expectation</td>
<td>1.33</td>
<td></td>
<td>4.27*</td>
<td></td>
</tr>
<tr>
<td>Within (reading speed, prior knowledge)</td>
<td></td>
<td>1.81</td>
<td>3.72**</td>
<td></td>
</tr>
</tbody>
</table>

*p < 0.05; **p < 0.01.

Table 2
Overview of the multivariate $F$s of covariance analyses on process and intermediate product variables of each phase. Covariates: reading speed and prior knowledge. $F$-values less than 1 are omitted
Table 3
Results of covariance analysis on planning variables. Covariates: reading speed and prior knowledge. *F*-values less than 1 are omitted

<table>
<thead>
<tr>
<th>Style (deep vs surface)</th>
<th>Deep plan</th>
<th>Surface plan</th>
<th>Deep ideal plan</th>
<th>Surface ideal plan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9.13**</td>
<td>3.95*</td>
<td>4.07*</td>
<td>7.60**</td>
</tr>
<tr>
<td>Test expectation (multiple choice vs essay)</td>
<td>1.35</td>
<td>2.14</td>
<td>2.73</td>
<td>4.51*</td>
</tr>
<tr>
<td>Style (\times) test expectation Within (reading speed, prior knowledge)</td>
<td>(5.46)</td>
<td>(3.30)</td>
<td>(3.68)</td>
<td>(4.40)</td>
</tr>
<tr>
<td>MS (\beta_{\text{Reading Speed}})</td>
<td>0.36*</td>
<td>0.15</td>
<td>-0.40*</td>
<td></td>
</tr>
<tr>
<td>Prior Knowledge (\beta_{\text{Prior Knowledge}})</td>
<td>0.16</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(p < 0.05; **p < 0.01.\)

Students scored higher on Surface Plan and Surface Ideal Plan than students with a deep learning style. No other effects on the plan variables besides those of style were found.

The third MANCOVA concerned Deep and Surface Study Activities, the dependent variables of the phase of plan execution. This phase remained unaffected by style and was also not affected by any of the other independent variables. None of the multivariate \(F\)-s even surpassed the value of 1.0 (see Table 2).

Finally, the results of the multiple choice and essay post-tests were analysed. The differences between post-test scores of students with deep and surface learning styles which came out of the analysis on basis of the simple model, did not persist in this extended analysis. Thus it appears that these differences can be accounted for by Prior Knowledge, Reading Speed, and the interaction between Style and Test Expectation (see Table 4). Students with a deep learning style performed best on multiple choice post-test questions if they expected essay questions, while students with a surface learning style did best if they expected multiple choice questions. There was no such interaction with regard to the essay post-test score. Reading Speed was negatively associated with test scores, although this relationship was only significant in the case of the multiple choice post-test. Prior Knowledge was positively associated with test scores, although this relationship was only significant in the case of the essay post-test (see \(\beta\)-s in Table 4). The more time students needed to read and understand a text section, the lower their multiple choice test scores; the more prior knowledge they had, the higher their essay scores.

A further exploration of post-test scores was made by analysing the subscores
Table 4
Results of covariance analysis on post-test variables: multiple choice and essay test scores, as well as univariate Fs of effects on the subscores comprehension, knowledge of facts, and terminology. Covariates: reading speed, and prior knowledge. F-values less than 1 are omitted

<table>
<thead>
<tr>
<th>Post-test Subscores</th>
<th>Comprehension</th>
<th>Facts</th>
<th>Terminology</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Multiple choice</strong></td>
<td>1.21</td>
<td>6.98*</td>
<td></td>
</tr>
<tr>
<td><strong>Essay</strong></td>
<td>1.71</td>
<td>5.02*</td>
<td></td>
</tr>
<tr>
<td><strong>Style (deep vs surface)</strong></td>
<td>6.91*</td>
<td>5.00*</td>
<td>1.07</td>
</tr>
<tr>
<td><strong>Test expectation (multiple choice vs essay)</strong></td>
<td>6.03**</td>
<td>4.57*</td>
<td>3.78*</td>
</tr>
<tr>
<td><strong>Style × test expectation</strong></td>
<td>(2.60)</td>
<td>(2.72)</td>
<td>(5.07)</td>
</tr>
<tr>
<td><strong>Within (reading speed, prior knowledge)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Within (error MS)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{RT}$</td>
<td>-0.41*</td>
<td>-0.19</td>
<td>-0.31*</td>
</tr>
<tr>
<td>$\beta_{FK}$</td>
<td>0.25</td>
<td>0.39*</td>
<td>0.24</td>
</tr>
</tbody>
</table>

*p < 0.05; **p < 0.01; ***p < 0.001.

for different knowledge types. Comprehension was significantly affected by Test Expectation ($F(1,35) = 5.02, p < 0.05$): Students who expected an essay test did better on comprehension questions than students who expected a multiple choice test. In addition, the control variables had a significant effect ($F(2,35) = 3.78, p < 0.05$), mainly due to Reading Speed ($\beta = -0.31, p < 0.05$). Style and the interaction between Style and Test Expectation showed no effects at all ($Fs < 1.0$). The pattern of effects on Knowledge of Facts turned out to be contrary to the pattern found for comprehension. Both style and the interaction between Style and Test Expectation were significant factors, while Test Expectation by itself and the control variables were not. Students with a deep learning style did generally better on knowledge questions than students with an surface learning style ($F(1,35) = 6.98, p < 0.05$). Furthermore, an essay test expectation fostered the acquisition of factual knowledge for students with a deep learning style, while a multiple choice test expectation did so for students with a surface learning style ($F(1,35) = 5.00, p < 0.05$). On the third type of question, requiring knowledge of exact Terminology, only the control variables showed significant effects ($F(2,35) = 4.22, p < 0.05$), in particular Prior Knowledge, which contributed positively to Terminology subscores ($\beta = 0.33, p < 0.05$).

4. Discussion

The diagnostic study task, completed in a computer assisted study environment (CASE), revealed differences between deep and surface learning students in orien-
tation and planning activities. However, their actual study behaviour did not vary according to learning style. As far as learning outcomes are concerned, students with a deep learning style obtained better results than students with a surface learning style, even for factual knowledge. However, learning style was found to contribute to (multiple choice) test scores only in interaction with test expectation: A match resulted in better scores than a mismatch, whereby the essay expectation matches deep learning and the multiple choice expectation matches surface learning.

If we use the learning style questionnaire as an external criterion to assess the ecological validity of CASE, then the fact that deep and surface learning students deployed various orientation and planning activities in accordance with their style of learning can be considered as supporting evidence. However, the lack of typical study activities of students with deep and surface learning styles puts the value of CASE into question. Similar results have been obtained in earlier research in which study tasks were conducted in an electronic learning environment. Samarapungavan & Beishuizen (1994) found that domain experts and beginners deployed different activities when studying a text in hypertext format, although both groups read the text pages in the same linear order when the text was presented in an ordinary linear format. Verheij et al. (1996) found clear differences between students with deep and surface learning styles when a particular piece of information had to be found in a text in hypertext format. Deep learning students more often used a concept map as an orientation and starting point for further explorations. Surface learning students did not utilise this facility and followed text relations connecting one text page to another in search of the requested information. Strategy differences did not show up when both groups of students were asked to study a text in hypertext format to prepare for a test. All students chose a linear reading order. These findings confirm the results of this study: deep and surface learning styles do not necessarily lead to different approaches to actual learning (i.e. reading and processing information), but cause different ways of orienting, planning, and organising information which has been collected by studying. So, the differences between surface and deep learning become more apparent when we look at study activities before and after actual reading and processing information.

What are the implications of these findings for the influence of test expectations on study outcomes? When students are supposed to comprehend subject matter, essay tests should be given. Essay tests invoke an essay test expectation, which apparently stimulates both deep and surface learning students to direct their studying towards comprehension. When the goal of instruction is the acquisition of different types of knowledge, including factual knowledge, students with a deep learning style again appear to benefit most from an essay test expectation, while students with a surface learning style benefit more from a multiple choice test expectation. Such a differentiation would not be convenient for educational practice. An alternative would be to reduce the expectation effects of question format by directly generating expectations about the required type of knowledge. To accomplish this, teachers should construct questions which, irrespective of their format, actually test the type of knowledge they want students to acquire, and they should prepare students on the expected level of processing. If students do not have specific information on the required type of
knowledge, they will infer from the expected question format and from prior experience what type of knowledge they are supposed to acquire. Explicit instructions on the required type of knowledge, however, would reduce these effects of anticipation on a particular test format (Verheul & Yang, 1989).

In several studies, relationships between learning style and intelligence were found (Riding & Mathias, 1991; Ainley, 1993; Busato et al., 1995). These studies also reported unique contributions of learning styles to achievement. We did not find such independent effects of learning style on post-test results, when reading speed and prior knowledge were taken into account. These two student characteristics may be more completely determined by domain related knowledge than intelligence. Both reading speed and prior knowledge are, to a certain extent, the result of a particular learning style. In our sample, the correlation between deep learning style and reading speed was $-0.3510$ ($p < 0.05$) and the correlation between deep learning style and prior knowledge was $0.3683$ ($p < 0.05$). These moderate but significant correlations reflect the relatedness of learning style and both student characteristics. Deep learning students knew more about the subject of the diagnostic study task, and developed a higher reading speed.

This study aimed at providing a more direct means of observing and assessing study behaviour, to be used together with other instruments for advising students how to improve their study methods. The data confirm the ecological validity of CASE and point towards the importance of orienting and planning as preliminary study strategies. A CASE based assessment may be used to fuel reflections on adequate orienting and planning strategies and on the influence of learning styles and test expectations on test performance. Moreover, the role of reading speed and prior knowledge can be illustrated by comparing student characteristics with group norms. These reflections and analyses may be initiated by an adviser, but should become part of the student’s repertoire of metacognitive knowledge and skills. Analysing study problems in terms of both questionnaire data and performance on a diagnostic study task may be a more convincing approach to strategy and attitude change than just advocating deep learning per se.

Acknowledgements

Netherlands Organization for Scientific Research (NWO) is gratefully acknowledged for funding this project. This research was conducted while E. T. Stoutjesdijk was supported by a grant of the Foundation for Behavioural and Educational Sciences of this organization (575-90-702) awarded to Dr J. J. Beishuizen.

Appendix A

Description of Dependent Variables in Terms of Behavioural Characteristics: Process Indicators and Interview Categories, for Students with a Deep and Surface Learning Style

<table>
<thead>
<tr>
<th>Variable</th>
<th>Behavioural characteristics</th>
</tr>
</thead>
</table>


Orientation
Deep Orientation Activities
(recorded by CASE)
- Total orientation time more than 90 s
- Total orientation time more than 300 s
- Time for Table of Contents or Index more than 30 s
- Time for Table of Contents or Index more than 120 s
- Number of text pages chosen from Table of Contents more than 0
- Number of text pages chosen from Table of Contents more than 1
- Total time for text scanning more than 30 s
- Total time for text scanning more than 120 s

Deep Orientation Self-Report
(interview data)
- View structure
- Look for known/unknown
- View contents
- View contents selectively
- Select text for task
- Watch length with contents

Deep Task Conception
(interview data)
- Can give outline
- Goal: overview/understand
- Goal: global, line of argument
- Goal: select towards task
- Goal: focus on main points

Surface Task Conception
(interview data)
- Number of pages correct, but no outline given
- Goal: remember/reproduce
- Goal: thorough, detailed method or knowledge
- Goal: do or know everything

Planning
Deep (Ideal) Plan
(recorded in interactive dialogue by CASE)
- General approach: selective
- First reading: global
- Repeated readings: increasing intensity
- Selective reading at least once
- Read test questions to control knowledge
Reflect on terms in index to control
After knowledge control more processing
Write down own questions
Notes: scheme(s)

Surface (Ideal) Plan
(recorded in interactive dialogue by CASE)

General approach: read everything
Go through the whole chapter once
Detailed reading at least once
Rote learning activities
Notes: terms and definitions list
Notes: extensive summary on paper
Notes: extensive summary in computer

Execution
Deep Study Activities
(recorded by CASE)

Time for Table of Contents or Index more than 120 s
Time for Table of Contents or Index more than 240 s
Time for Table of Contents or Index more than 360 s
Number of pages chosen from Table of Contents more than 1
Number of pages chosen from Table of Contents more than 3
Number of pages turned back more than 1
Number of pages turned back more than 3
Reading time differences between pages in favour of important pages more than 0.05 s per word
Reading time differences between pages in favour of important pages more than 0.10 s per word
Reading time differences between pages in favour of task relevant pages more than 0.05 s per word
Reading time differences between pages in favour of task relevant pages more than 0.10 s per word
Number of test questions read without answering more than 1
Number of test questions read without
Surface Study Activities (recorded by CASE)

- Linear text movements more than 65%
- Linear text movements more than 75%
- Linear text movements more than 85%
- Thorough linear reading more than 10 min
- Thorough linear reading more than 20 min
- Thorough linear reading more than 30 min
- Thorough linear reading more than 40 min
- Watch the clock at least every 10 min
- Watch the clock at least every 5 min

Notes: scheme(s)

Linear text movements more than 65%
Linear text movements more than 75%
Linear text movements more than 85%
Thorough linear reading more than 10 min
Thorough linear reading more than 20 min
Thorough linear reading more than 30 min
Thorough linear reading more than 40 min
Watch the clock at least every 10 min
Watch the clock at least every 5 min
Notes: list of terms/definitions on paper
Notes: extensive summary on paper
Notes: summary in computer

Note. If a characteristic is applicable, then one point is added to the value of the pertinent variable.

References


