Instructional Science **32:** 233–252, 2004.

Timing of Information Presentation in Learning Statistics

LIESBETH KESTER, PAUL A. KIRSCHNER AND JEROEN J.G. VAN MERRIËNBOER
Open University of the Netherlands, Educational Technology Expertise Center, P.O. Box 2960, 6401 DL Heerlen, The Netherlands (E-mail: liesbeth.kester@ou.nl)

Received: 23 January 2003; in final form: 21 October 2003; accepted: 27 November 2003

Abstract. This study in the domain of statistics compares four information presentation formats in a 2 × 2 factorial design: timing of supportive information (before *or* during task practice) × timing of procedural information (before *or* during task practice). Seventy-two psychology and education students (7 male and 65 female; mean age 18.5 years, *SD* = 2.85) participated. The effectiveness of the learning material was measured by test performance. The instructional efficiency was measured by a combination of mental effort during practice and test performance (i.e., a high test performance combined with a low mental effort during practice denotes a high instructional efficiency). ANOVA showed a main effect for timing of supportive information: presentation *during* practice led to more *efficient* learning than presentation *before* practice. Moreover, an interaction effect was found. Simultaneous presentation of procedural information *before* and supportive information *during* practice led to the most *efficient* learning.

Keywords: Cognitive Load Theory, complex skills, information presentation, instructional design, schema acquisition, working memory

Introduction

In the 1960's and 1970's, *meaningfulness* came to be seen as a key factor in learning and remembering. According to Johnson (1975), 'meaningfulness is potentially the most powerful variable for explaining the learning of complex verbal discourse' (pp. 425–426). Learning is meaningful if learners can relate new learning tasks to their existing cognitive structures (Novak 1984; Novak and Gowin 1984; Williams and Cavallo 1995). In other words, the more that new learning can be associated with what is already known and with available cognitive schemata (i.e., existing knowledge structures such as conceptual, causal and functional models) the better it will be learned. This can be accomplished by making sure that the new information is both deeply and richly processed, since superficial similarities are not likely to lead to proper encoding.

In education, meaningful learning can be accomplished by preparing the learner for the learning tasks that are to come, for example, with the aid of advance organizers (Ausubel 1963; Ausubel and Robinson 1969; Mayer 1979). In other words, education should create a proper context within which the students can efficiently and effectively elaborate on already available cognitive structures before they start working on the learning tasks. This elaboration process is very time consuming. Therefore, in practice, a trade-off between time effectiveness and level of elaboration often has to be made.

More recent theories of learning stress the importance of instruction situated in meaningful knowledge-rich contexts, above all taking place in an interactive environment (Koschman, Myers, Feltovich and Barrows 1994; Derry and Lesgold 1996; Salomon and Perkins 1999). Within these contexts, the learner receives or actively seeks the information necessary to carry out the learning tasks at the moment the information is required. A potential problem in such contexts is that the learner must simultaneously carry out a learning task and process the necessary information, which could prove too taxing to the processing capacity of working memory. The fact that human working memory is limited is well accepted (Baddeley 1992; Miller 1956). Once the limits are reached, performance deteriorates (Chandler and Sweller 1991; Sweller 1988; Sweller, Van Merriënboer and Paas 1998). When task relevant information is presented simultaneously with the execution of the learning tasks, the risk of overloading working memory is present. The success of this approach, thus, depends on the optimal loading of the information-processing capacity of working memory. In this study, an attempt is made to design learning material with an optimal balance between the time effectiveness of the elaboration process and the level of working memory load by presenting the *right* information *just in time*.

So, which information should be presented when? To acquire a complex cognitive skill, the learner needs to master a series of (sub) skills, which are often different in nature. The performance of some skills is *consistent* across problem or task situations while the performance of other skills is highly *variable* over problem situations (Fisk and Gallini 1989). The different nature of these (sub) skills implies a difference in information necessary to master them. To design the learning material properly, a skill decomposition, a task analysis, and a knowledge analysis have to be carried out to identify these types of (sub)skills and accompanying information.

When statistics is considered, the domain of this study, a distinction can be made between skills concerned with the application of formulas (consistent skills) and skills concerned with *when* to apply *which* formulas (variable skills). To master the correct application of formulas, *procedural* information about the exact form of the formula and definitions of the elements in the formula is needed. For example, this study uses Chi-square tests and to calculate Chi-square the following formula is needed: $\chi^2 = \Sigma(o - e)^2/e$ in which χ^2 stands for Chi-square, *o* for observed frequency, and *e* for expected frequency. To master the skills that allow students to make the right choice as to whether the Chi-square test can be used, *supportive* information is needed about statistical testing in general and circumstances under which a Chi-square test is called for.

Mastering variable skills requires constructing general schemata in long term memory, which can be used in different problem situations. This construction involves a process of mindful abstractions from concrete experiences, that is, elaboration (Proctor and Reeve 1988). Mastering consistent skills requires the forming and *automation* of domain specific schemata used in similar problem situations and is accomplished through *proceduralization* (Anderson 1993, 1996). For proceduralization to occur, relevant information has to be active in working memory when the skill is practiced. The underlying cognitive processes of the variable skills and consistent skills imply that both information types require different optimal moments of presentation.

To facilitate the elaboration process, supportive information should be presented *before* practice, so that the learner has the opportunity to embed new information in already available cognitive structures. Procedural information, on the contrary, should be presented *during* practice in order for proceduralization to occur (Kester, Kirschner, van Merriënboer and Bäumer 2001).

This view is supported by instructional guidelines based on Cognitive Load Theory (Sweller 1988). This theory examines the complex interactions of long term and working memory and its relationships to schema construction and schema automation (Sweller 1999). It places particular emphasis on the outcomes of this cognitive architecture for instruction and in particular the limitations of working memory and its effects on learning, which concentrates predominantly on the limitations of working memory. According to Cognitive Load Theory, different types of cognitive load are imposed on the learner by the learning material: intrinsic load (i.e., the load associated with the content of the learning material) and extraneous load (i.e., the load associated with the instructional features of the learning material).

Intrinsic cognitive load is inherent to the learning material itself and is determined by the element (i.e., anything that needs to be learned, for instance, schemata) interactivity in the material and the expertise of the learner (e.g., it is high for learning material with high element interactivity and/or learners with low expertise). Learning material with high element interactivity demands simultaneous processing of several elements in working memory while material with low element interactivity allows for serial processing of several elements (Sweller et al. 1998). For example, processing the supportive information in this study has a high element interactivity because in order to decide which statistical test is appropriate under specific circumstances, the learner has to *simultaneously* process and evaluate features of different statistical tests in working memory so as to make the right decision. Processing the procedural information, on the contrary, has a low element interactivity because in order to calculate Chi-square, for example, the learner has to *serially* process several mathematical operations in working memory. It has to be noted however that to understand the entire procedure of Chi-square testing each mathematical operation has to be related to one another which is in fact a task high in element interactivity. The element interactivity especially has implications for the timing of supportive information during the acquisition of a complex skill. The intrinsic load associated with processing the supportive information is higher than the intrinsic load associated with processing the procedural information. Therefore, it is assumed that it is better to present supportive information apart from procedural information and practice problems, so that, all working memory capacity can be allocated to processing the supportive information and thus elaboration, at that point in time.

Extraneous cognitive load is caused by the instructional features of the learning material and refers to all the processes a learner engages in, during a task, which are not directly beneficial to learning (e.g., searching for relevant information sources). Extensive research (Cerpa, Chandler and Sweller 1996; Chandler and Sweller 1991; Chandler and Sweller 1992; Chandler and Sweller 1996; Kalyuga, Chandler and Sweller 1999; Mayer and Anderson 1992; Mayer and Moreno 1998; Mayer and Sims 1994; Sweller and Chandler 1994) has shown that an effective way to minimize extraneous cognitive load is to avoid split-attention, which arises when a learner has to mentally integrate several sources of *mutually referring* information (e.g., a picture and its explanatory text) in order to understand the learning material. By integrating the different sources of information spatially or temporally, split-attention is prevented and learning is facilitated. In general, with a given intrinsic cognitive load, the extraneous cognitive load should be minimized by instructional design (Sweller et al. 1998). When this is accomplished, all cognitive capacity can be allocated to relevant learning processes, which will facilitate the mastering of variable skills (or general schema construction) and consistent skills (or domain specific schema automation).

In this study, the procedural information (e.g., the calculation procedure of expected frequencies for one sample) and the practice tasks (e.g., calculation of expected frequencies for one sample) are mutually referring. The procedural information could be applied to the practice tasks while the practice tasks rely upon the procedural information. Therefore, procedural information should be presented during practice in order to avoid *temporal* split-attention. As a result, relevant information is active in working memory when the skill is practiced which facilitates proceduralization (Van Merriënboer, Kirschner and Kester 2003). Moreover, proceduralization frees up cognitive capacity during practice that can be used to understand for example, the entire procedure of Chi-square testing or to adjust the constructed general schemata of for example, Chi-square testing based on experience.

Mayer and his colleagues (Mayer and Anderson 1991; Mayer and Anderson 1992; Mayer and Sims 1994) carried out several experiments which provide a demonstration of the *temporal* split attention effect. In the first two experiments, an instructive animation on the working of a bicycle tire pump or an automobile braking system was used to compare integrated narration and animation to: (1) successive presentation of narration and animation, (2) narration only, (3) animation only and (4) no instruction at all. It was found that integrated instruction led to better performances on relevant creative problem solving tests (Mayer and Anderson 1991; Mayer and Anderson 1992). Similar results were found in a third experiment where an instructive animation was given on the working of a bicycle tire pump or the human respiratory system. Integrated narration and animation was compared to: (1) successive presentation of narration and animation and (2) no instruction at all. It appeared that learners in the integrated condition performed better on the problem solving test than the learners in the other conditions (Mayer and Sims 1994). The present study investigates, amongst other things, whether temporally integrating two referring *visual* sources of information (i.e., text) also leads to beneficial effects on learning as described by Mayer and his colleagues (Mayer and Anderson 1991; Mayer and Anderson 1992; Mayer and Sims 1994).

So, to support elaboration and proceduralization during complex skill acquisition in accordance with the instructional design guidelines form Cognitive Load Theory it is assumed that supportive information is best presented *before* practice and procedural information is best presented *during* practice. To investigate this assumption, four information presentation formats were distinguished. First, the format that presents supportive information *before* practice and procedural *during* (i.e., elaboration and proceduralization are supported, the intrinsic load is properly managed and the extraneous load is minimized by avoiding temporal split attention). Second, a format that presents both information types *before* practice (i.e., only elaboration is supported, the intrinsic load is not properly managed and temporal split attention is not avoided). Third, a format that presents both information types *during* practice (i.e., only proceduralization is supported, intrinsic load is not properly managed but temporal split attention is avoided). Fourth, a format that presents supportive information *during* practice and procedural *before* (i.e., both elaboration and proceduralization are not supported, intrinsic load is properly managed but split attention is not avoided).

The effectiveness and instructional efficiency of all four information presentation formats is studied in the domain of statistics. The effectiveness of the information presentation formats is measured by test performance. Test tasks that are equivalent to the practice tasks are used to measure if the learner mastered the consistent skills, that is, is capable of using the *same* domain specific schemata the *same* way in solving equivalent test tasks. Test tasks that are very different from the practice tasks are used to measure if the learner mastered the variable skills, that is, is capable of using the *same* general schemata in a *different* way in solving transfer test tasks. The method of Paas and Van Merriënboer (1993) is used to calculate instructional efficiency scores. Instructional efficiency is determined on the basis of a combination of test performance and invested mental effort during practice. High instructional efficiency denotes higher test performance in combination with a lower investment of mental effort during practice while low instructional efficiency denotes lower test performance in combination with a higher investment of mental effort during practice.

It is hypothesized that the format that presents the supportive information *before* practice and the procedural information *during* will lead to more effective (i.e., higher test performance scores, especially for the transfer test because of richer schema construction) and efficient learning (i.e., higher test performance combined with lower investment of mental effort during practice) than the other formats. The 'supportive before, procedural during' format optimally allows for elaboration by presenting the supportive information before practice; all cognitive capacity can be allocated to process this information high in element interactivity (i.e., proper management of intrinsic load) which facilitates general schema construction and thus, mastering the variable skills. Presenting the procedural information during practice optimally allows for proceduralization; necessary procedural information is directly applied to the learning tasks hereby avoiding temporal split attention which facilitates domain specific schema automation and thus, mastering the consistent skills. Moreover, schema automation frees up cognitive resources during learning task practice that can be used to adjust and enrich the earlier constructed general schemata based on experience.

Method

Participants

Seventy-two freshman psychology and education students at the University of Gent, Belgium (7 male and 65 female; mean age 18.5 years, *SD* = 2.85) participated in this study. All of the participants spoke Dutch as their first language, the language in which the instruction was given. They voluntarily signed up for an introductory two-day statistics course to get acquainted with this subject. Only students who had four hours or less mathematics a week in secondary education could sign up for this course. In Belgian secondary schools, students who choose the social sciences typically have four hours or less mathematics a week. Students choosing engineering or natural sciences typically have seven hours or more. Secondary mathematics education in Belgium does not include any statistics, and thus it is reasonable to assume that all participants are novices in this domain. The first day of the course was meant to refresh mathematics knowledge and skills. This session was not included in the experiment. The second day encompassed the experiment and consisted of a mini course in Chi-square testing as an introduction to statistics.

Materials

Chi-square course. The Chi-square course was developed and presented in Mercator®, an electronic development and instruction environment (Valcke, Kirschner and Bos 1999). Mercator® was also used for the math course on the first day allowing the participants to get used to navigation in this program.

The goal of the Chi-square course was to teach participants when and how to use a Chi-square test for single and multiple sample situations. The participants received 24 practice tasks divided over six topics (i.e., 4 per topic), namely, frequency tables, expected frequencies for one sample, Chi-square test for one sample, crosstabulations, expected frequencies for two or more samples, and Chi-square test for two or more samples.

Each topic is composed of supportive information, procedural information and general information that was used to ‘glue’ the course together (e.g., ‘We have determined the observed frequency and the expected frequency, let’s see if we can find significant differences between these frequencies . . .’) and four practice tasks. Each topic was build up of two subsequent screens and each screen was divided in a left half and a right half. For every topic, both screens contained supportive information and/or procedural information (depending on the information presentation format) presented on the left. On screen one, this information was combined with general information presented on the right and on screen two; it was combined with four practice tasks presented on the right. The practice tasks were administered to the participants in the form of conventional problems and no feedback was given to participants about their performance. For a schematic overview of a topic in the Chi-square course see Figure 1.

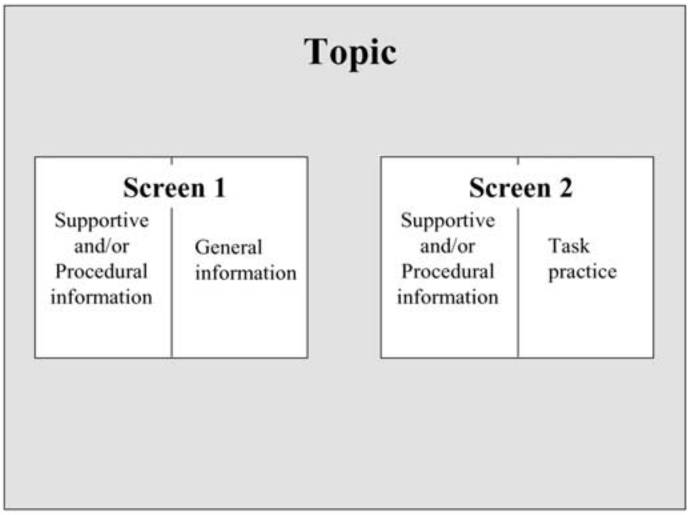

Figure 1. A schematic overview of the experimental Chi-square course.

A task analysis was carried out to determine which information in the practice tasks was supportive information and which was procedural. This was done in co-operation with, and checked by, an expert in the field of statistics. For example, the description of the research question, and examples of other research questions, are considered supportive information, and setting up a frequency table is considered procedural information (see Appendix for a detailed description of the supportive and procedural information that was determined).

Information presentation. The participants were randomly assigned to one of the four formats. In the ‘supportive before, procedural during’ format, supportive information was presented *before* practice and procedural information was presented *during* practice (n = 19). In the ‘supportive before, procedural before’ format, *both* supportive and procedural information were presented *before* the participants carried out the practice tasks (n = 16). The participants assigned to the ‘supportive during, procedural during’ format received *both* information types *during* the process of carrying out the practice tasks (n = 17). In the ‘supportive during, procedural before’ format the supportive information was presented *during* practice while the procedural information was presented *before* the participants carried out the practice tasks (n = 20). Figure 2 shows an overview of these four information presentation formats.

Test tasks. After completing the electronic Chi-square course, the participants had to carry out a paper and pencil test with equivalent and transfer tasks. The equivalent test tasks (12 items, that is, 6 topics $\times$ 2 items per topic)

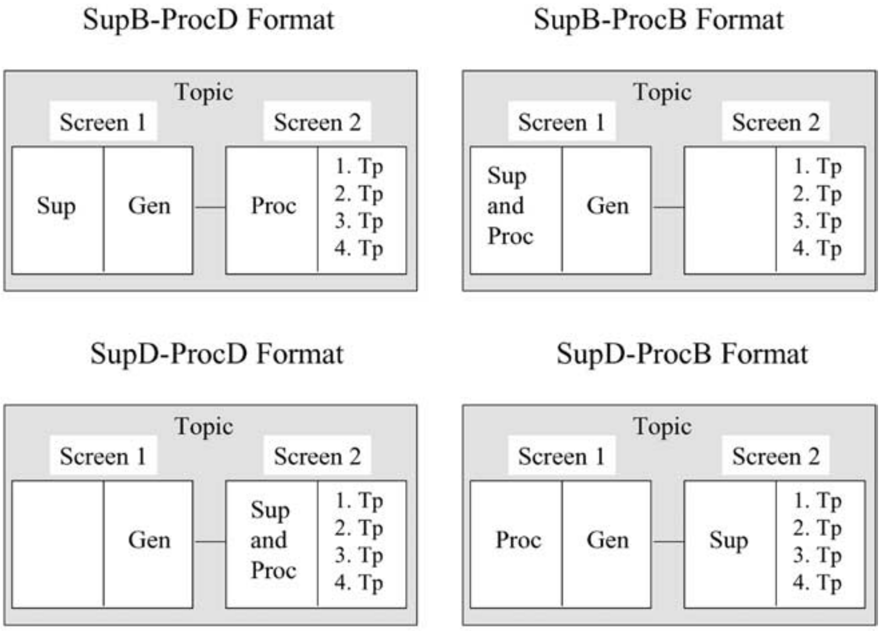

Figure 2. Overview of the four experimental information presentation formats. Sup = Supportive information, Proc = Procedural information, Gen = General information, and Tp = Task practice.

were very similar to the practice tasks. They tested whether the participants were able to independently perform the learned procedures (i.e., had formed the domain specific (automated) schemata and mastered the consistent skills). Open questions were administered in order to give the participants optimal opportunity to show they knew how to carry out the learned procedures. For example:

> Make a frequency table for the variable multiple choice question when we know that 3 participants chose answer A, 11 participants chose answer B, 27 answer C and 9 chose answer D. Calculate the percentages and enter them in the table.

The maximum score on the equivalent test tasks was 36 points, one point for each item about expected frequencies (four points in total) and four points each for the other items (32 points in total). Two raters carried out the scoring of the equivalent test tasks. The interrater reliability of the two raters was 0.93 (Intraclass Correlation Coefficient). The internal consistency of the equivalent test tasks was 0.73 (Cronbach's alpha).

The transfer test was used to measure the participant's ability to recognize situations in which a Chi-square test is called for (i.e., had constructed the general schemata and mastered the variable skills). The transfer tasks (16 items) were very dissimilar to the practice tasks. Multiple-choice (MC) questions were administered to test if participants had been able to form a good mental model of Chi-square testing. These MC questions all represented tasks for which a different statistical test could be used. The participants were asked to determine whether the specific problem called for using a Chi-square test or not. For example:

> Problem: A manufacturer of a particular drug claims that the drug is effective in 90% of all cases. In a sample of 200, the drug wasn't effective in 40 cases. Is the claim valid?
> Is the Chi-square test the right test to use to support the claim?
> 1. o Yes
> 2. o No
> 3. o I don't know.

Both the wrong answer and the answer 'I don't know' were considered to indicate that the transfer had not taken place. The maximum score on the transfer test was 16 points and the test has a reliability of 0.65 (Cronbach's alpha). A closed format was chosen because it enabled the participants to compare and distinguish between problems that required different statistical tests and to identify the proper test without having to perform these tests. After all, the participants had no prior statistical knowledge and therefore did not know how to perform any statistical test other than the Chi-square test. The sequence of the test tasks was random.

Mental effort measurement. Mental effort refers to the cognitive capacity actually allocated to meet the problem requirements. The amount of mental effort invested by participants is considered to be the essence of cognitive load, therefore, mental effort is used as an index for cognitive load (Paas 1992). In instructional research and especially in research based on cognitive load theory a 9-point rating-scale is used to measure mental effort (Paas 1992; Paas, Van Merriënboer and Adam 1994). Here, for reasons of comparability, the same scale was used. Mental effort was measured during practice with a nine-point rating-scale for measuring the participants' perceived mental effort (Paas 1992; Paas, Van Merriënboer and Adam 1994). The mental effort measures ranged from very, very low mental effort to very, very high mental effort. The rating-scale was administered at fixed points during practice, namely, at the end of every topic. Participants were asked to note how much mental effort it cost them to understand the learning material and, in a second question, how much mental effort it cost them to carry out each practice task. This resulted in a total of twelve mental effort measurements. The internal consistency of these mental effort measures (Cronbach's alpha) was 0.91.

Log tool. A log tool was used to measure the time the participants spent on the electronic Chi-square course. This tool logged the window headers and every time a header changed a timestamp was generated.

Procedure

The experiment was divided into two sessions of three hours each. The morning session involved the electronic Chi-square course and the afternoon session the paper and pencil test. Before starting the experiment, participants received instructions about the general procedure of the experimental task. They were told that the Chi-square course was being used for research purposes and that they:

- had to work independently,
- were not allowed to take notes,
- had to go through the Chi-square course in the prescribed order,
- were not allowed to make any adjustments within Mercator® (during the math course the participants learned about certain personal adjustments that can be made in the program but these may not be used during the experiment), and
- were not allowed to bring anything to the PC-classroom except for a pen and a calculator.

It was made clear that they had to hand in their answer sheets when finished. Participants were allowed to ask questions before and during the experiment, but this was not encouraged. All questions about operating the electronic Chi-square course were answered. Questions about the content of the Chi-square course were not answered. In this case, participants were referred back to the information given in the course. Participants went through the electronic course at their own pace. It was possible for the participants to move back and forth both within topics and between topics.

In the afternoon session the participants were told that they had to practice what they had learned in the morning session. In fact, they received the test. It was stressed that they should not guess while answering the multiple-choice questions and that they must work independently.

Results

Test performance

Equivalent tasks. First, the scores participants obtained on the equivalent test tasks were considered. In this study, a probability level of 0.05 is used in reporting all statistical tests. ANOVA revealed no statistically significant

Table 1. Summary of the test data

Procedural information	Supportive information			
	Before		During	
	M	*SD*	*M*	*SD*
	Equivalent test tasks (max. = 36)			
Before	18.56	6.44	20.63	5.09
During	20.39	5.90	19.47	4.08
	Transfer test tasks (max. = 16)			
Before	6.37	2.00	8.45	1.32
During	6.84	2.69	6.88	2.87

Table 2. Summary of the mean mental effort data[a]

Procedural information	Supportive information			
	Before		During	
	M	*SD*	*M*	*SD*
Before	4.74	1.13	3.86	1.35
During	4.11	1.00	4.13	0.96

[a]Max = 9.

main effects or interaction effects. Equivalent test task scores are presented in Table 1.

Transfer tasks. ANOVA revealed no statistically significant main effects for the transfer test either. The results on the transfer test are summarized in Table 1.

Mental effort measures

The mean of the 12 mental effort measures was considered. ANOVA revealed no statistically significant main effects, or interaction effects regarding the mean mental effort. For an overview of these results see Table 2.

Instructional efficiency

Instructional efficiency (E) scores were calculated for mental effort and test performance (Paas and Van Merriënboer 1993). First, the mental effort measures and the performance measures per participant were transformed to *z*-scores. The grand mean is used for calculation, through which the mean *z*-score for every condition can be determined. These mean condition *z*-scores can be represented in a Cartesian coordinate system with *P*erformance *z*-scores on the horizontal axis and *M*ental effort *z*-scores on the vertical axis. The line $P = M$ through the origin of the axes indicates an efficiency of zero (slope = 45°). The relative condition efficiency is calculated as the perpendicular distance from a data point in the coordinate system to the line $P = M$ (Paas and Van Merriënboer 1993). Calculation of *E* is done, per participant, with the following formula:

$$E = \frac{P_{\text{erformance}} - M_{\text{ental Effort}}}{\sqrt{2}}$$

Equal performance (*P*) and mental effort (*M*) scores yield an instructional efficiency of zero, a neutral score. When $P > M$, the instructional material is efficient because the mental effort is lower than might be expected on the basis of observed performance. When $P < M$, the material is *not* efficient because the mental effort is higher than might be expected on the basis of the observed performance.

Efficiency measures were calculated on the basis of the equivalent test task scores and the transfer test scores (see Table 3). For the efficiency based on the equivalent test tasks, ANOVA revealed no statistically significant effects. For the efficiency based on the transfer test tasks ANOVA revealed a statistically significant main effect for the timing of supportive information (F (1, 68) = 6.07, MSE = 0.99, $p < 0.05$; $\eta^2 = 0.082$) and a statistically significant interaction between the timing of supportive and procedural information (F (1, 68) = 6.05, MSE = 0.99, $p < 0.05$; $\eta^2 = 0.082$). Post hoc tests, using Tukey's HSD, indicated that only the 'supportive before, procedural before' format and the 'supportive during, procedural before' format differed significantly ($p < 0.05$). Participants who had received supportive information during practice (M = 0.28, SD = 0.94) had higher efficiency scores than participants who had received supportive information before practice (M = –0.30, SD = 1.10). When procedural information is presented before practice, presentation of supportive information during practice yields much higher efficiency scores (M = 0.57, SD = 0.77) than the simultaneous presentation before practice (M = –0.59, SD = 1.19). When procedural information is presented during practice, there is no difference between the presentation of supportive information during practice (M = –0.06, SD = 1.03) and before practice (M = –0.09, SD = 0.99).

Table 3. Summary of the mean efficiency measures

	Sup before				Sup during			
	Proc before		Proc during		Proc before		Proc during	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Transfer test tasks	–0.59	1.19	–0.09	0.77	0.57	0.99	–0.06	1.03
Equivalent test tasks	–0.51	1.25	0.12	0.30	0.30	1.05	–0.01	1.04

Sup = Supportive information; Proc = Procedural information.

Table 4. Mean total time (min) spent on the Chi-square course

Procedural information	Supportive information			
	Before		During	
	M	*SD*	*M*	*SD*
Before	141.27	19.77	140.92	17.15
During	139.13	26.68	146.57	17.02

Time on task

ANOVA revealed no statistically significant main effects or interaction effects regarding the total time spent on the Chi-square course. For an overview of the results see Table 4.

Discussion

In this study, contrary to what was predicted, it was not the 'supportive before, procedural during' format but rather the 'supportive during, procedural before' format that led to more efficient learning, but only in comparison with the 'supportive before, procedural before' format. The presentation of *supportive* information *during* practice in combination with the presentation of *procedural* information *before* practice yields a higher efficiency measure based on invested mental effort during practice and the *transfer* test scores than the presentation of both information types *before* practice. Moreover, presentation of supportive information during practice is superior to the presentation of this information before practice. Participants who had the supportive information available during practice exhibited a higher efficiency based on invested mental effort and transfer test scores than participants who had received supportive information before practice. No differences between formats were found in efficiency based on invested mental effort during practice and the *equivalent* test scores and no differences were found in effectiveness. The participants in all formats performed equally well on the equivalent test as well as on the transfer test.

How to explain these unexpected results? Why did the predicted beneficial effects of the ‘supportive before, procedural during’ format fail to occur? We turn first to the question of why the ‘supportive during, procedural before’ format was more efficient than the ‘supportive before, procedural before’ format. Both formats did not support proceduralization and did not avoid temporal split attention. They only differed in properly managing intrinsic load and supporting the elaboration process. Intrinsic load was properly managed in the ‘supportive during, procedural before’ format while elaboration was supported in the ‘supportive before, procedural before’ format. Based on the findings, the measures that were taken to properly manage intrinsic load *did* lead to more efficient learning but the measures that were taken to facilitate elaboration *did not*.

It seems that the students processed the supportive and procedural information differently than expected. Apparently, the ‘supportive during, procedural before’ information presentation format effectively prepared the learners for the practice tasks that were about to come by presenting procedural information before practice. This led to better appreciation of the meaningful context (i.e., availability of supportive information) during practice. Research by Carlson and colleagues (Carlson, Sullivan and Schneider 1989; Carlson, Khoo and Elliott II 1990) emphasizes the beneficial influence of a meaningful context during practice of procedural information presented earlier. They found that transfer test scores increased when learners were able to practice the procedural information in meaningful contexts.

The presentation of supportive information *during* practice provided learners with a meaningful, knowledge-rich context in which they received or could actively seek information they needed to carry out the tasks at hand. In this specific Chi-square course, the supportive information was not necessary to carry out the practice tasks, but it was necessary for a better understanding of these tasks. Apparently, when this information was presented while carrying out the practice tasks, its relevance for these practice tasks was clearer than when it was presented before the practice tasks. Presentation of supportive information during practice seemed to provide a meaningful, knowledge rich context that enabled the learners to judge this information on its own merit. They made better relations between this information and the practice tasks and therefore became more aware that they needed this information to fully understand the practice tasks, leading to more efficient learning.

Before a final conclusion is reached, we should try to explain why the predicted beneficial effects of the ’supportive before, procedural during’ format failed to occur. This format aimed at properly managing intrinsic cognitive load and avoiding temporal split-attention, so that the learners could devote all their cognitive capacity to elaboration and proceduralization. However, when the mental effort results are considered, low mental effort scores for all formats were found. In every format, the mean mental effort scores never exceeded the rating ‘not low, not high’. Since there is no indication that the Chi-square course exceeded working memory capacity, managing intrinsic load and prevention of temporal split attention will not have facilitating effects on the learning processes. Therefore, superiority of one of the formats over the others is not to be expected. This was true for most measures in this study namely the effectiveness (i.e., equivalent an transfer test performance), and for the instructional efficiency based on the equivalent test scores.

Further research is needed to find out under which circumstances and in which domains the different information presentation formats are successful. In this study optimal moments of information presentation were defined based on supporting elaboration and proceduralization, properly managing intrinsic load based on the element interactivity of the information and avoiding temporal split attention. However, the element interactivity of the information was not objectively measured. Future studies should address this issue. Furthermore, in statistics it was rather difficult to describe the task in terms of independent pieces of knowledge. It is probably better to choose a task in which that is not such a problem. In this study, an attempt was made to present a well rounded statistical practice task, but every topic in statistics elaborates on other topics and therefore it is very difficult to find strictly limited practice tasks. Moreover, in hindsight, because of the obvious interdependence of the practice tasks and the procedural information, the learners may have gotten the false notion that the supportive information was not of much relevance for the task while in fact it was meant as input for a deeper understanding of the learning material. This may have interfered with the learning process in general. Limited, well-rounded, practice tasks, in which both the importance of supportive information and procedural information is clear, might be better found in technical domains, such as engineering or mechanics, or scientific domains, such as physics.

Bearing this study in mind, it can be concluded that the ‘supportive during, procedural before’ information presentation format leads to more efficient learning (i.e., based on the transfer test scores) than information presentation according to the ‘supportive before, procedural before’ format. So, presenting all information at the same time is a sub optimal option. This is important to note because conventional classroom practices often provide students a textbook that contains all necessary information which they have to study before they start to practice. Based on the results of this study, these practices should be adjusted. Instead of presenting all information at the same time in a textbook, the necessary information should be presented just in time before and during practice. However, no differences were found in effectiveness and efficiency based on the equivalent test scores. Evidently, more research is needed to determine just in time presentation of appropriate information to help learners master complex skills.

Acknowledgements

The authors gratefully acknowledge the comments, support and assistance of Hannelore Dekeyser during the setting up and carrying out of this experiment.

Appendix

Summary of supportive and procedural information

Supportive information	Procedural information
Task cluster 1: Frequency tables	
• Description of a research question (with several examples).	• Definition of analysis units (with different examples).
• Explanation of variable analysis units, operationalization of variables (with examples).	• Definition of a variable, a measurement value and classes (with an example).
• Description of a measurement instrument and the process of data collection.	• Definition of raw data.
• Introduction of the structioning technique, 'frequency table'.	• Definition of a frequency table (with an example). Definition of relative frequencies and the calculation procedure of relative frequencies (with an example). Description of setting up a frequency table inclusive relative frequencies (with an example).
Task cluster 2: Expected frequencies for one sample	
• Description of the assumptions a researcher can have regarding a research question (with an example).	• Definition of expected frequencies and the calculation procedure of expected frequencies for one sample (with an example).

Supportive information	Procedural information
Task cluster 3: Chi square test for one sample	
• Introduction in which the question is raised: 'How to decide whether differences between observations and expectations are significant or not?' A description of statistical testing, differences between parametrical and non-parametrical tests and the Chi-square test for one sample (with examples).	• Definition of the quantity Chi-square.
• Description of the interpretation of the value of Chi-square.	• Calculation procedure of Chi-square (with an example). Introduction Chi-square table and an explanation of how to use it (with an example). Definition of the calculation procedure of the degrees of freedom, the critical values and how to interpret them (with a example).
Task cluster 4: Cross tables	
• Introduction of the structuring technique, cross tables (with an example).	• Definition of a cross table.
• Description of the relations that can exist between variables (with an example). Introduction of the relative frequency.	• Calculation procedure of the relative frequency. Definition independent and dependent variables.
• Explanation of the relevance of relative frequencies (with an example).	
Task cluster 5: Expected frequencies for two or more samples	
• Connotation in which is stressed that the expected frequencies for two or more samples are no percentages.	• Definition and calculation procedure of expected frequencies for two or more samples (with an example).
Task cluster 6: Chi-square test for two or more samples	
• Introduction in which the question is raised: 'How to decide whether two variables are related or not?'. A description of the Chi-square test for two or more samples. An explanation of how to interpret the value of Chi-square.	• The formula of Chi-square for two or more samples (with an example). Description of how to use the Chi-square table (with an example). Definition of the calculation procedure of the degrees of freedom, the critical values an how to interpret them (with an example).

References

Anderson, J.R. (1993) *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum.

Anderson, J.R. (1996) ACT: A simple theory of complex cognition. *American Psychologist* 51: 355–365.

Ausubel, D.P. (1963) *The Psychology of Meaningful Verbal Learning*. New York: Grune & Stratton.

Ausubel, D.P. & Robinson, F.G. (1969) *School Learning: An Introduction to Educational Psychology*. London: Holt, Rinehart & Winston.

Baddeley, A.D. (1992) Working memory. *Science* 255: 556–559.

Carlson, R.A., Sullivan, M.A. & Schneider, W. (1989) Component fluency in a problem solving context. *Human Factors* 31(5): 489–502.

Carlson, R.A., Khoo, B.H. & Elliott II, R.G. (1990) Component practice and exposure to a problem-solving context. *Human Factors* 32(3): 267–286.

Cerpa, N., Chandler, P. & Sweller, J. (1996) Some conditions under which integrated computer-based training software can facilitate learning. *Journal of Educational Computing Research* 15(4): 345–367.

Chandler, P. & Sweller, J. (1991) Cognitive load theory and the format of instruction. *Cognition and Instruction* 8(4): 293–332.

Chandler, P. & Sweller, J. (1992) The split attention effect as a factor in the design of instruction. *British Journal of Educational Psychology* 62: 233–246.

Chandler, P. & Sweller, J. (1996) Cognitive load while learning to use a computer program. *Applied Cognitive Psychology* 10: 151–170.

Derry, S. & Lesgold, A. (1996) Toward a situated social practice model for instructional design. In David C. Berliner & Robert C. Calfee (eds), *Handbook of Educational Psychology*, pp. 787–806. New York, London et al.: Macmillan; Simon & Schuster.

Fisk, A.D. & Gallini, J.K. (1989) Training consistent components of tasks: Developing an instructional system based on automatic-controlled processing principles. *Human Factors* 31: 453–463.

Johnson, R.E. (1975) Meaning in complex learning. *Review of Educational Research* 45(3): 425–454.

Kalyuga, S., Chandler, P. & Sweller, J. (1999) Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology* 13(4): 351–371.

Kester, L., Kirschner, P.A., Van Merriënboer, J.J.G. & Bäumer, A. (2001) Just-in-time information presentation and the acquisition of complex cognitive skills. *Computers in Human Behavior* 17: 373–391.

Koschman, T.D., Myers A.C., Feltovich, P.J. & Barrows, H.S. (1994) Using technology to assist in realizing effective learning and instruction: A principled approach to the use of computers in collaborative learning. *Journal of the Learning Sciences* 3(3): 227–264.

Mayer, R.E. (1979) Can advance organizers influence meaningful learning? *Review of Educational Research* 49: 371–383.

Mayer, R.E. & Anderson, R.B. (1991) Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology* 83: 484–490.

Mayer, R.E. & Anderson, R.B. (1992) The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology* 84(4): 444–452.

Mayer, R.E. & Moreno, R. (1998) Split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology* 90(2): 312–320.

Mayer, R.E. & Sims, V.K. (1994) For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology* 86(3): 389–401.

Miller, G.A. (1956) The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63: 81–97.

Novak, J.D. (1984) Application of advances in learning theory and philosophy of science to the improvement of chemistry teaching. *Journal of Chemical Education* 61(7): 607–612.

Novak, J.D. & Gowin, D.B. (1984) *Learning How to Learn*. Cambridge: Cambridge University Press.

Paas, F. (1992) Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology* 84(4): 429–434.

Paas, F. & Van Merriënboer, J.J.G. (1993) The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors* 35(4): 737–743.

Paas, F., Van Merriënboer, J.J.G. & Adam, J.J. (1994) Measurement of cognitive load in instructional research. *Perceptual and Motor Skills* 79: 419–430.

Proctor, R.W. & Reeve, T.G. (1988) The acquisition of procedural productions and modification of declarative representations in spatial precueing tasks. *Journal of Experimental Psychology: General* 117: 293–307.

Salomon, G. & Perkins, D.N. (1998) Individual and social aspects of learning. *Review of Research in Education* 23: 1–24.

Sweller, J. (1988) Cognitive load during problem solving: Effects on learning. *Cognitive Science* 12: 257–285.

Sweller, J. (1999) *Instructional Design in Technical Areas*. Camberwell, Australia: ACER Press.

Sweller, J. & Chandler, P. (1994) Why some material is difficult to learn. *Cognition and Instruction* 12(3): 185–233.

Sweller, J., Van Merriënboer, J.J.G. & Paas, F. (1998) Cognitive architecture and instructional design. *Educational Psychology Review* 10(3): 251–296.

Valcke, M.P., Kirschner, P.A. & Bos, E. (1999) Enabling technologies to design, produce, and exploit flexible, electronic learning materials. In: J. van den Akker, R.M. Branch, K. Gustafson, N. Nieveen & Tj Plomp (eds), *Design Approaches and Tools in Education and Training*. Dordrecht: Kluwer.

Van Merriënboer, J.J.G., Kirschner, P.A. & Kester, L. (2003) Taking the load of a learner's mind. Instructional design for complex learning. *Educational Psychologist* 38(1): 5–13.

Williams, K.A. & Cavallo, A.M. (1995) Relationships between reasoning ability, meaningful learning and students' understanding of physics concepts. *Journal of College Science Teaching* 24(5): 311–314.