Differential effects of problem-solving demands on individual and collaborative learning outcomes

Femke Kirschner a,b,*, Fred Paas a,b, Paul A. Kirschner a, Jeroen Janssen c

a Centre for Learning Sciences and Technologies, Open University of The Netherlands, P.O. Box 2960, 6401 DL Heerlen, The Netherlands
b Institute of Psychology, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
c Research Centre Learning in Interaction, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, The Netherlands

Received 19 January 2010; revised 9 January 2011; accepted 12 January 2011

Abstract

The effectiveness and efficiency of individual versus collaborative learning was investigated as a function of instructional format among 140 high school students in the domain of biology. The instructional format either emphasized worked examples, which needed to be studied or the equivalent problems, which needed to be solved. Because problem solving imposes a higher cognitive load for novices than does studying worked examples it was hypothesized that learning by solving problems would lead to better learning outcomes (effectiveness) and be more efficient for collaborative learners, whereas learning by studying worked examples would lead to better learning outcomes and be more efficient for individual learners. The results supported these crossover interaction hypothesis. Consequences of the findings for the design of individual and collaborative learning environments are discussed.

© 2011 Elsevier Ltd. All rights reserved.

Keywords: Cognitive load theory; Worked-example study; Problem-solving; Collaborative learning; Learning efficiency; Learning outcomes

1. Introduction

Cognitive load theory (CLT: Kirschner, 2002; Paas, Renkl, & Sweller, 2003; Sweller, 1988; Sweller, Van Merrienboer, & Paas, 1998) focuses on learning from complex cognitive tasks based on what is known about human cognitive architecture (Sweller, 1988, 2004). This architecture consists of an unlimited long-term memory (LTM), which interacts with a working memory (WM) that is limited in both capacity (Miller, 1956) and duration (Peterson & Peterson, 1959). For new information, the processing capacity is limited to 4 ± 1 interacting information elements which is lost if not rehearsed within 30 s (Cowan, 2001). LTM stores and organises knowledge in cognitive schemas that incorporate multiple elements of information into a single element (also referred to as chunking; Chase & Simon, 1973; Miller, 1956; Simon, 1974) with a specific function (i.e., learning). If learning has occurred over a long period of time, one’s schemas may consist of huge amounts of information. Because a schema can be treated by WM as a single element or even bypass WM if it has become sufficiently automated through long and consistent practice, WM limitations will disappear for more knowledgeable learners.

According to CLT, learning task complexity is determined by the number of new (i.e., to be learned) interacting information elements; the more new interacting elements, the more complex the task. Although highly interactive information elements can be processed in isolation, they can only be understood when all of them and their interactions are processed simultaneously. Given that WM capacity is limited
with respect to new information, tasks that contain many new and interacting elements place high cognitive demands on WM. Within CLT, the load imposed on WM by the element interactivity or task complexity is called intrinsic cognitive load.

As basis for instructional design, CLT focuses on effectively dealing with individual WM limitations by creating instruction that is compatible with human cognitive architecture. Research, therefore, primarily has been on developing techniques for managing individual WM load and optimising information-processing in individual learning settings (Ayres & Paas, 2009). Kirschner, Paas, and Kirschner (2009b, 2010) have recently emphasised an alternative way of effectively dealing with individual WM limitations, namely making use of the multiple WMs of individuals in a collaborative learning setting. From their perspective, groups of collaborating learners are considered to be information-processing systems (Hinsz, Tindale, & Vollrath, 1997; Ickes & Gonzalez, 1994; Tindale & Kameda, 2000), consisting of multiple limited WM systems which can create a collective working space. Within these systems, valuable task-relevant information and knowledge held by each group member can be consciously and actively shared (i.e., retrieving and explicating information), discussed (i.e., encoding and elaborating information), and remembered (i.e., personalising and storing information) (Hinsz et al., 1997; Tindale & Kameda, 2000; Tindale & Sheffey, 2002). As long as the information is communicated between the group members and they coordinate their actions, not all group members need possess all necessary knowledge, or process all available information alone and at the same time (Johnson, Johnson, & Stanne, 2001; Langfred, 2000; Wegner, 1978, 1995). This, however, requires positive interdependence as described in Johnson and Johnson’s (1981) social interdependence theory. Positive interdependence reflects the extent to which group members must depend upon each other for effective group performance; each individual group member is responsible for the work of the group and the group as a whole is responsible for the learning of each individual group member. Group members are linked to each other such that each member cannot succeed unless the others succeed; each member’s work benefits the others and each member benefits from the others. Essential here is social cohesion and a heightened sense of ‘belonging’ to a group.

The perspective on shared memory systems in which communication plays a key role has also been investigated in the context of Wegner’s (1978) transactive memory theory. This theory is based on the idea that group members can serve as external memory aids for each other, and that they can benefit from each other’s knowledge and expertise if they develop a good, shared understanding and an awareness of ‘who knows what’ in the group. This could, for example, be achieved by training individuals on how to share their knowledge effectively (Deiglmayr & Spada, 2010; Prichard, Stratford, & Bizo, 2006). Findings indicate that transactive memory can facilitate group performance in groups whose members are aware of other group members’ knowledge and expertise as opposed to groups where the members are not aware of fellow group members’ knowledge and expertise (Michinov & Michinov, 2009). A transactive memory system enables groups to better utilise the knowledge that their members possess, and to reach higher levels of performance than they would have reached without such a system (for a review, see Moreland & Argote, 2003).

The CLT perspective, in which groups are considered to possess a shared memory system, has two conflicting consequences for individual group members. On the one hand, collaborating individuals can invest less cognitive effort compared to learners working alone, because the task’s interactive information elements with its associated cognitive load (i.e., the intrinsic cognitive load) can be divided across a larger reservoir of cognitive capacity (Kirschner, Paas, & Kirschner, 2009a; Ohtsubo, 2005; Stasser, Stewart, & Wittenbaum, 1995). This is what is called the distribution advantage. On the other hand, collaborating individuals need to invest cognitive effort in communicating information with each other and coordinating their actions, which individuals working alone do not have to exert. These, so called, transactional activities (Ciborra & Olson, 1988; Kirschner et al., 2009b; Yamane, 1996) can be beneficial for or deleterious to learning. CLT argues that, while a cognitive investment in beneficial transactional activities such as negotiating common ground should be stimulated, an investment in deleterious activities such as discussing ways to share information should be minimised.

The trade-off between the advantage of dividing information-processing among group members and the disadvantage in terms of having to cognitively invest in the associated transactional activities can be an indicator for the efficiency of group learning. This so called collective working memory effect was demonstrated in a study by Kirschner et al. (2010) on the effects of low-complexity (i.e., low intrinsic load) and high-complexity (i.e., high intrinsic load) tasks on individual and group learning efficiency. Other than learning effectiveness, which is related primarily to learning outcomes (i.e., posttest performance), learning efficiency is related to the relationship between learning outcomes and the amount of mental effort learners invest to attain those outcomes; the higher the learning outcomes and the lower the effort, the higher the efficiency (Paas & Van Merriënboer, 1993; Van Gog & Paas, 2008; Tuovinen & Paas, 2004). Mental effort in combination with learning outcome can indicate the quality of learning in terms of the efficiency of cognitive schema acquisition. By giving groups and individuals low and high-complexity learning tasks and then assessing their learning efficiency on an individual posttest, Kirschner et al (2010) showed that group learning was superior to individual learning for high-complexity tasks, but inferior for low-complexity tasks.

For high-complexity tasks, Kirschner et al. (2010) argued that by sharing the task’s high intrinsic cognitive load among learners, the risk of exceeding the limits of the individual group members’ WM is reduced (i.e., distribution advantage). Although the additional cognitive load imposed by communicating information and coordinating actions (i.e., transactional activities) has to be taken into account, this load...
could be considered to be relatively low compared to the distribution advantage for complex tasks. Consequently, learning was more efficient for group members, allowing them to construct higher quality schemas in LTM than for individual learners who had to process all of the information individually (i.e., the collective working-memory effect: Kirschner et al., 2010).

For low-complexity tasks, they argued that individual learners had sufficient capacity to carry out the tasks alone and no advantage of learning together was expected. Sharing the information-processing among the group members required information communication and action coordination which for low-complexity tasks imposed a relatively high load (in relation to the benefits that will be accrued) thereby negating the distribution advantage. Consequently, learning was more efficient for individuals, allowing them to construct higher quality schemas in LTM than for group learners who had to engage in relatively high transaction activities.

In addition to the intrinsic load imposed by the complexity of the learning task, the cognitive load that learners experience is affected by a task’s instructional format. This can take two forms. When the instructional format imposes load that is not effective for learning, for example, when it provides redundant information (e.g., Chandler & Sweller, 1991), or requires students to mentally integrate spatially or temporally separated materials (e.g., Sweller, Chandler, Tierney, & Cooper, 1990), it is called extraneous cognitive load. When it is beneficial to learning, for example when it presents students with high-variability materials (Paas & Van Merriënboer, 1994), or encourages students to self-explain their actions (e.g., Renkl, 1997), it is referred to as germane cognitive load (Paas, Renkl, & Sweller, 2004; Sweller, Van Merriënboer, & Paas, 1998). Although extraneous load does not hamper learning for low in intrinsic load tasks (i.e., low-complexity), it does hamper learning for high intrinsic load tasks (i.e., high-complexity); hence, reducing extraneous load is imperative for high-complexity tasks (Van Merriënboer & Sweller, 2005). A consistent finding of CLT research is that for tasks high in intrinsic load, instruction with high problem-solving demands imposes extraneous load on novice learners (Sweller et al., 1998). In addition to Kirschner et al.’s (2009a, 2010) recent studies into the effects of different levels of intrinsic load on individual and collaborative learning efficiency, the present study kept intrinsic load constant and investigated the differential effects of extraneous load induced by the instructional format in terms of problem-solving demands. Specifically, individual and collaborative learning — by novices — from either tasks that heavily rely on problem-solving or tasks that heavily rely on example study were compared regarding their effects on learning efficiency and effectiveness.

Sweller (1988) has shown that problem-solving search when carrying out complex conventional tasks, places not only heavy intrinsic demands, but also heavy extraneous demands on WM. The strategy most commonly used by learners faced with novel problems for which they do not have previously constructed schemas — a means-ends analysis — requires them to consider the current problem state, consider the desired goal state, extract differences between the two states, and find or choose a problem-solving operator that can be used to reduce or eliminate differences between the current problem state and the desired goal state. In addition, any sub-goals that have been established need to simultaneously be kept in mind. This problem-solving search strategy imposes high extraneous cognitive load on learners and, consequently, does not leave sufficient processing capacity for them to induce the generalised solutions or schemas that are prerequisite to learning.

An effective alternative to such instruction, and one that has been supported by multiple overlapping experiments using different instructional materials in a variety of populations, is instruction with low problem-solving demands relying on example study, either by using example—problem pairs, problem—example pairs, or example study only (e.g., Carroll, 1994; Hübner, Nückles, & Renkl, 2010; Paas, 1992; Paas & Van Gog, 2006; Paas & Van Merriënboer, 1994; Stark, Kopp, & Fischer, 2010; Sweller, 1988; Sweller, 2006; Sweller & Cooper, 1985; Trafton & Reiser, 1993; Van Gog, Kester, & Paas, 2010; for an overview see Atkinson, Derry, Renkl, & Wortham, 2000). Instruction that relies heavily on studying worked examples as a substitute for instruction involving problem-solving is argued to be beneficial because it decreases extraneous load by eliminating means-ends search. In contrast to instruction with a heavy reliance on problem-solving, worked examples focus attention on problem states and associated operators (i.e., solution steps), enabling learners to induce generalised solutions or schemas. The cognitive capacity that becomes available by reducing extraneous load can be devoted to activities beneficial to learning (i.e., germane load). This reasoning has led to the counterintuitive instructional guideline that for novices learning to solve problems, studying worked examples is a better strategy than solving the equivalent problems (Rourke & Sweller, 2009; Sweller, 1988; Sweller et al., 1998).

2. Research Question and Hypotheses

In the context of CLT, instruction with low problem-solving demands (i.e., studying worked examples) should impose lower cognitive load on novices than instruction with high problem-solving demands (i.e., equivalent problems that need to be solved). Combining this view on instructional formats with the results of the collective working memory effect (Kirschner et al., 2010), the question arises as to whether for students learning individually, instruction emphasising worked example study would be more effective and efficient than instruction emphasising solving problems. Analogous to this is the question of whether for students learning collaboratively, instruction emphasising problem-solving would be more effective and efficient than instruction emphasising worked example study. To this end, this study tested the following hypotheses:

For learning from instruction emphasising problem-solving, the load on the limited cognitive capacity of an
individual learner should be too high for effective learning. For learners in a group, the benefits of dividing information-processing (i.e., the cognitive load) among group members would be greater than the additional cognitive investment of inter-individual communication of information and coordination of actions in the learning phase. Group members would consequently be able to devote the freed-up cognitive capacity to activities that foster schema construction and automation (i.e., germane load), resulting in higher performance (i.e., learning outcome; Hypothesis 1a) and a more favourable relationship between performance and mental effort (i.e., higher learning efficiency; Hypothesis 1b) for learners who carried out the learning tasks in groups than for those who carried out the tasks individually.

With regard to learning from instruction emphasising worked examples, learners working individually or as members of a group should have sufficient cognitive capacity to process all information themselves. Hence, the transactional cognitive activities for learners working as members of a group would be high relative to the benefits of dividing the information-processing across group members in the learning phase. Consequently, qualitative differences in constructed schemas were expected between learners learning in a group and learners learning individually, resulting in higher learning outcomes (Hypothesis 2a) as well as higher learning efficiency (Hypothesis 2b) on an individual posttest for those who learned individually than for those who learned as members of a group.

3. Method

3.1. Participants and design

The effect of instructional format on the learning outcome and efficiency of individual versus collaborative learning was investigated using learning tasks emphasizing either problem-solving (further referred to as problem-solving tasks) or worked-example study (further referred to as worked-example study tasks). In the learning phase, 140 Dutch high school sophomore students (71 boys, 69 girls) with an average age of 14.98 years (SD = 0.96), were assigned to four conditions in a worked-example study condition, learners were in addition presented with worked-out solution steps, and asked to study

3.2. Materials

The materials were in the biology domain and were concerned with heredity; specifically genotypic and phenotypic transmission of biological traits from parents to offspring (e.g., eye colour in humans, fur length in dogs, leaf shape in plants). To this end, a general introduction and instruction on solving inheritance problems, worked example study tasks, problem-solving tasks, and transfer-test tasks were designed. All materials were approved by two biology teachers as being suitable for the learners. All materials were paper based.

3.2.1. Introduction

Relevant terminology, rules, and theory underlying heredity, as well as a worked example on solving heredity problems was discussed in the introduction. This introduction gave the definition of genes, genotype and phenotype, homozygosity or heterozygosity of dominant or recessive genes, the pedigree chart, and rules concerning Punnett squares (i.e., diagrams used to predict the outcome of a particular cross or breeding experiment). The worked example demonstrated how to solve a heredity problem by combining terminology, rules, and theory. Learners were required to use the definitions, rules, and theory underlying heredity problems when carrying out the learning tasks.

3.2.2. Learning tasks

Three learning tasks that were deemed complex for novices were presented as tasks emphasising either problem-solving (i.e., problem-solving task) or worked-example study (i.e., worked-example study task). These tasks consisted of nine information elements on a biological trait in a human family (e.g., ear shape, eye colour, hair colour) and a question about the proportion of possible genotypes and phenotypes of the offspring. Each individual information element was relevant but insufficient for successfully carrying out the task. A problem could only be solved by combining all nine information elements (JIGSAW; Aronson, Blaney, Stephan, Silkes, & Snapp, 1978; Moreno, 2009; Slavin, 1990). A simplified example of such a task would be that: information element 1 is the mother’s eye colour: blue; element 2 is the father’s eye colour: brown; and element 3 is the dominance of brown over blue for eye colour. Each element gives a certain amount of information, but to determine the eye colour of the offspring, the learner must combine all three information elements. The main difference between the two instructional formats (i.e., problem-solving or worked-example study) is the degree of example elaboration. In the problem-solving condition, learners received the solution to the problem (i.e., the correct answers to the questions), and were asked to use the information elements that they had received to determine how the solution was reached. Working from the givens to the solution, learners had to rely on problem-solving techniques. The problem solution presented to the learners could be used to verify the result of their problem-solving efforts. In the worked-example study condition, learners were in addition presented with worked-out solution steps, and asked to study
the way the solution was reached (see Fig. 1 for a task emphasising problem-solving and Fig. 2 for a task emphasising worked-example study (b)). Working from the givens to the final solution learners could study 11 worked-out solution steps and sub steps.

To prevent learners from incorrectly believing that they had arrived at the correct answer, the study did not use conventional problem-solving tasks (i.e., tasks with only a description of givens and a question without the final solution; Sweller, 1988). In previous studies using the same genetic problems as conventional problem-solving tasks, learners quickly decided that they had arrived at the correct problem solution without this actually being the case (Kirschner et al., 2009b, 2010). Providing the correct solution is one way of preventing this and stimulating them to invest effort in the problem-solving steps towards the problem solution. It also provides learners with equal opportunities for learning how to correctly solve a genetics problem in both conditions (i.e., problem-solving and example study). In this way, participants become aware that they had arrived at a correct or incorrect solution, and in the latter case are stimulated to try again.

The tasks that had to be carried out in collaboration were structured such that there was high positive task interdependence (Johnson, 1981; Saavedra, Early, & Van Dyne, 1993). Group members had to rely on each other and interact with each other to obtain resources and to effectively carry out the task. Positive interdependence reflects the degree to which group members are dependent upon each other for effective group performance (i.e., enhanced intra-group interaction). It holds that team members are linked to each other in such a way that each team member cannot succeed unless the others succeed; each member’s work benefits the others and vice versa (Kirschner, Strijbos, Kreijns, & Beers, 2004). This was achieved by giving each group member a booklet containing only one third of the total number of information elements needed to identify or understand the solution steps. To stimulate collaboration, cognitive load distribution (i.e., the distribution advantage; Ciborra & Olson, 1988; Hinsz et al., 1997; Kirschner et al., 2009b; Yamane, 1996), and information exchange amongst the learners, there were no redundant information elements and the number of information elements was equal for all group members (i.e., three information elements per group member). In addition to the individual booklets, the whole

**TASK 1: PIET AND LUCY’S FAMILY**

**GIVEN**
- For humans, the gene for green eye color (G) is dominant over the gene for blue eyes (g).
- Sandra has green eyes.
- Sandra’s mother has green eyes.
- Sandra’s mother is homozygote for eye color.
- Sandra’s father has green eyes.
- Sandra’s father is homozygote for eye color.
- Wim has blue eyes.
- Lucy, a child of Sandra and Wim, marries Piet who has green eyes.
- Piet’s father has blue eyes.

**QUESTIONS**
What could be the genotypes and phenotypes of Piet and Lucy’s children? And what are their proportions?

**SOLUTION**
Below are the answers to the questions. However, there is no description of how the answers were arrived at. Use the given information to determine how the answers were reached.

The genotypes and their proportions are:
25% GG – 25% gg – 50% Gg

The phenotypes and their proportions are:
75% green eyes – 25% blue eyes

---

**TASK 2: PIET AND LUCY’S FAMILY**

**GIVEN**
- For humans, the gene for green eye color (G) is dominant over the gene for blue eyes (g).
- Sandra has green eyes.
- Sandra’s mother has green eyes.
- Sandra’s mother is homozygote for eye color.
- Sandra’s father has green eyes.
- Sandra’s father is homozygote for eye color.
- Wim has blue eyes.
- Lucy, a child of Sandra and Wim, marries Piet who has green eyes.
- Piet’s father has blue eyes.

**QUESTIONS**
What could be the genotypes and phenotypes of Piet and Lucy’s children? And what are their proportions?

**SOLUTION STEPS**
Below, the solution steps for answering the questions are given. You need the given information to study the solution steps and find out how the steps and final answers were reached.

**STEP 1.** Determine Piet’s genotype for eye color.
Piet’s genotype is Gg.

**STEP 2.** Determine Lucy’s genotype for eye color.
It is not possible to know Lucy’s genotype at once; first it has to be determined:

**STEP 2.1.** What is Wim’s genotype for eye color?
Wim’s genotype for eye color is gg.

**STEP 2.2.** What is Sandra’s genotype for eye color?
It is not possible to know Sandra’s genotype at once; first it has to be determined:

**STEP 2.2.1.** What is Sandra’s father’s genotype for eye color?
Sandra’s father’s genotype is GG.

**STEP 2.2.2.** What is Sandra’s mother’s genotype for eye color?
Sandra’s mother’s genotype is GG.

**STEP 2.2.3.** Make a Punnett square between the genotypes of Sandra’s mother and father.

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>GG</td>
<td>GG</td>
</tr>
<tr>
<td>g</td>
<td>Gg</td>
<td>Gg</td>
</tr>
</tbody>
</table>

**STEP 2.2.4.** Determine Sandra’s genotype.
Sandra’s genotype is GG.

**STEP 2.3.** Make a Punnett square between the genotypes of Sandra and Wim.

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>GG</td>
<td>Gg</td>
</tr>
<tr>
<td>g</td>
<td>Gg</td>
<td>gg</td>
</tr>
</tbody>
</table>

**STEP 2.4.** Determine Lucy’s genotype for eye color.
Lucy’s genotype is Gg.

**STEP 3.** Make a Punnett square between the genotypes of Piet and Lucy.

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>GG</td>
<td>Gg</td>
</tr>
<tr>
<td>g</td>
<td>Gg</td>
<td>gg</td>
</tr>
</tbody>
</table>

**STEP 4.** Determine the genotypes and their proportions of Piet and Lucy’s children.
25% GG – 25% gg – 50% Gg.

**STEP 5.** Determine the phenotype and their proportions of Piet and Lucy’s children.
75% green eyes – 25% blue eyes.

Fig. 2. A learning task emphasizing worked-example study.
group received a booklet containing either a question and its solution (i.e., problem-solving condition) or a question, the solution steps, and the solution (i.e., worked-example study condition). This booklet was available to all group members at the same time. Learners working individually received a booklet containing all nine information elements plus the question, solution, and additional solution steps depending on the experimental condition.

3.2.3. Posttest

To determine how much was learned and to see if learners could apply the knowledge and skills that they were assumed to have acquired in the learning phase to a different kind of problem, four transfer-test tasks were used. These tasks required learners to use the same basic terminology, rules, and underlying theory, but they differed from the learning tasks with respect to families and traits used, kind of information elements given, task structure, and questions asked (e.g., genealogical tree, X-chromosome linked inheritance, dihybrid crossings; see Fig. 3 for an example). The internal consistency of the posttest (Cronbach’s alpha) was 0.73.

The posttest consisted of two questions per task related to heredity characteristics of a certain trait in a family. With regard to performance, the questions could be scored on multiple elements, with 1 point for a correctly mentioned element and 0 points for an incorrectly or not mentioned element. A maximum score of 28 points could be earned for the four tasks. The minimum score for all tasks was 0. For the statistical analysis, the performance scores were transformed into proportions. In other words, a participant’s score on the four tasks was divided by the maximum score of tasks (i.e., 28).

3.2.4. Instruction

A non-content related instruction on the procedure, rules, and regulations concerning solving problems was presented to the participants twice; once preceding the learning tasks and once preceding the posttest. The instruction preceding the learning tasks differed slightly between conditions. The individual learner had to read all information elements before solving the problem, while each group learner only had to read those information elements specifically allotted to her/him (i.e., one third of the total number of elements), but were also required to share the information elements with each other. In the problem-solving condition, participants had to find out how the solution to the problem was established, while in the worked-example study condition participants had to study the given solution steps. Because this study focuses on how instructional format influences collaborative learning, learners were not allowed to write things down during the learning phase. Using an external memory aid (i.e., writing things down) provides an external representation or visualisation of a problem and can, thus, seen as an ‘augmenting’ cognitive activity (Jonassen, Peck, & Wilson, 1999; Pea, 1993). As revealed by Duffy and Cunningham (1996) as well as Van Bruggen, Kirschner, and Jochems (2002), one impact of this is that it leads to ‘cognitive offloading’ allowing learners to devote more memory to other aspects of problem-solution. In this respect, allowing learners to use pencil and paper would skew the research since the basic premise of the research, as stated, was to determine the differential effects of instructional format (and the accompanying cognitive load) on individual and group learning when carrying out a learning task. Introducing memory aids would make it impossible to determine whether the learning formats and contexts increased or decreased cognitive load or whether the results were due to the cognitive offloading.

The instruction preceding the posttest was the same for all participants: First they had to read all information elements thoroughly, then read the questions, and finally try to answer the questions as correctly and quickly as possible using all information elements using pen or pencil and paper to write down the solution steps.

3.2.5. Cognitive load measurement

To measure learner cognitive load after each learning task and test task, the subjective 9-point mental-effort rating scale developed by Paas (1992) was used. Participants were asked to rate the level of effort required to solve a problem on a scale ranging from very, very low effort (1) to very, very high effort (9). This measure provides an overall indication of the cognitive load (i.e., the total of intrinsic, extraneous, and germane cognitive load), has been used in numerous studies dealing with cognitive load, and has proven to be non intrusive, valid, and reliable (Paas, Van Merrienboer, & Adam, 1994).

3.2.6. Efficiency measurement

The combination of performance and cognitive load measures can provide a reliable estimate of the relative efficiency of instructional methods, both in terms of the learning process and learning outcomes. Paas and Van Merrienboer’s (1993) computational approach (see Van Gog & Paas, 2008)
was used to calculate learning efficiency as a function of instructional format. The basic idea underlying this approach is that the combination of performance and mental-effort data collected during a test phase is indicative of the quality of the cognitive schemas constructed during learning. Instructional conditions with higher performance on the posttest in combination with lower or equal invested mental effort (i.e., a lower level of reported cognitive load) are more efficient than lower performance on the posttest in combination with higher or equal invested mental effort; cognitive schemas have been more efficiently acquired. Learning efficiency was calculated by standardising each learner’s scores on the posttest and the cognitive load invested working on these tests. For this purpose, the grand mean was subtracted from each score and the result was divided by the overall standard deviation, yielding z-scores for effort (R) and performance (P). Finally, each participant’s performance efficiency score, E, was computed using the formula: 

\[ E = \frac{(P - R)^2}{2} \]

High learning efficiency was indicated by relatively high posttest performance combined with relatively low mental effort. In contrast, low learning efficiency was indicated by relatively low posttest performance combined with relatively high mental effort.

### 3.3. Procedure

All participants first individually studied a paper-based general introduction to heredity-related concepts along with a worked example and returned the introduction after 15 min. During the whole experiment, the time (i.e., study and testing) was fixed and was managed by a proctor. They were then randomly assigned to one of the four experimental conditions (i.e., individual problem-solving, individual worked-example study, group problem-solving, and group worked-example study) to carry out the first series of tasks in 7 min. After each task, independent of whether in a group or in an individual condition, all participants rated the amount of invested cognitive load on the 9-point mental effort rating scale. The second and third series of learning tasks followed the same procedure. The instruction for all participants preceding the series of learning tasks consisted of advising them to read all information elements thoroughly, read the questions carefully, and finally try to identify the solution steps to the problem in the case of the problem-solving condition or understand the worked example in the case of the worked-example study, and finally try to identify the solution steps to the problem in the case of the worked-example study. After the learning phase, the test phase required participants to individually work on four transfer tasks for 5 min each. Invested mental effort was measured after each transfer-test task with the rating scale. Use of pen or pencil and paper was allowed and stimulated in this phase.

### 3.4. Data analyses

When analysing the effects of social context (individual vs. group) and instructional format (problem-solving vs. worked-example study), the data-analytical problem of nonindependence had to be taken into account (Cress, 2008; Kenny, Mannetti, Pierro, Livi, & Kashy, 2002). Because students in the group condition worked in triads, they influenced each other through their shared experiences and collaborative discussions. This might cause group members to experience similar amounts of mental effort invested in the learning tasks. Furthermore, because they discussed how to solve the learning task, their performance on the posttest may be correlated. For example, when group members collaborated effectively and found efficient solutions to the problem, they may all perform well on the posttest. On the other hand, when collaboration was not effective and they failed to find solutions to the problem, all group members may perform poorly on the posttest. This violates the assumption of nonindependence of observations of individuals, making the results of traditional analytical techniques, such as ANOVA or MANOVA unreliable (Kenny, 1995). Multilevel analysis (MLA) can cope with this nonindependence and is therefore a more appropriate technique (Bonito, 2002; Snijders & Bosker, 1999). When investigating the effects of social context and instructional format, MLA will therefore be used.

### 4. Results

The results for the learning and test phases are described separately. A significance level of 0.05 was used for all analyses. Due to registration problems there were incomplete data from 4 participants in the learning phase and an additional 5 in the test phase. Case-wise deletion of those participants was carried out in the analyses. Table 1 shows the resulting number of participants as well as the means and the resulting number of participants as well as the means and standard deviations for mental effort as a function of social context and instructional format in the learning phase, and performance, mental effort, and learning efficiency as a function of social context and instructional format in the test phase.

#### 4.1. Learning phase

The data from the learning phase were analysed using a random intercept multilevel model (Snijders & Bosker, 1999) that included three predictor variables: social context (dummy coded with 0 = individual and 1 = group), instructional format (dummy coded with 0 = problem-solving and 1 = worked-example study), and the interaction between social context and instructional format. The dependent variable was the perceived amount of mental effort invested in studying. The random intercept regression equation is given in Eq. (1), where \( \gamma_0 \) is the intercept, \( \gamma_{10} \) is the regression coefficient of the group level variable social context, \( \gamma_{20} \) is the regression coefficient for the group level variable instructional format, \( \gamma_{30} \) is the regression coefficient for the interaction...
Table 1
Means and standard deviations of the dependent variables in the learning and test phase as a function of social context and instructional format.

<table>
<thead>
<tr>
<th>Instructional format</th>
<th>Social context</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Individual</td>
</tr>
<tr>
<td></td>
<td>M</td>
</tr>
<tr>
<td>Learning Phase&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Mental Effort (1–9)</td>
</tr>
<tr>
<td></td>
<td>Example study</td>
</tr>
<tr>
<td>Test Phase&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Performance (0–1)&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Example study</td>
</tr>
<tr>
<td></td>
<td>Mental effort (1–9)</td>
</tr>
<tr>
<td></td>
<td>Example study</td>
</tr>
<tr>
<td></td>
<td>Learning efficiency&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Example study</td>
</tr>
</tbody>
</table>

<sup>a</sup> n = 34 for each condition.
<sup>b</sup> n = 32 for all individual conditions, n = 34 for the − group-problem-solving conditions, n = 33 for the group − worked-example study conditions.
<sup>c</sup> Performance is the proportion of correct answers on the posttest.
<sup>d</sup> Based on z-scores of mental effort and performance in the test phase.

between social context and instructional format, $U_{ij}$ is group level variance, and $R_{ij}$ is individual level variance.<sup>1</sup>

Mental effort<sub>ij</sub> = $\gamma_{00} + \gamma_{01}\text{Context}_j + \gamma_{02}\text{Format}_j + \gamma_{03}\text{Context}_j \times \text{Format}_j + U_{ij} + R_{ij}$  

(1)

This model was estimated with MLwiN version 2.18. Table 2 shows the estimated model parameters. The MLA shows a significant effect of social context on perceived mental effort, $p < .01$. The positive sign of $\gamma_{01}$ shows that group members rated the mean mental effort higher than individuals, independent of instructional format. No effect of instructional format was found, nor was there an interaction effect between social context and instructional format in the learning phase. The deviance reported in Table 2 can be used as a test for the goodness-of-fit of the multilevel model. The deviance value shown in Table 2 to the deviance of a multilevel model without predictor variables (i.e., empty model, labelled Model 1 in Table 2), one can test whether the estimated model fits the data better than the empty model. In this case, the decrease in deviance was significant, $\chi^2(3) = 7.444, p < .05$. As seen in Table 2, the model including social context and instructional format was able to explain 12% of the group level variance and 8% of the individual level variance.

4.2. Test phase

The data from the test phase were analysed using a similar multilevel model as the one in Eq. (1). Three different multilevel models were estimated with posttest performance, mental effort, and learning efficiency as dependent variables.

For posttest performance, the MLA results in Table 3 show a significant effect of social context, $p < .01$ and instructional format, $p < .01$. These main effects were qualified by a crossover interaction effect of social context and instructional format, $p < .01$. Inspection of regression coefficient $\gamma_{03}$ shows that participants who had learned individually performed better on the posttest when they learned from solving problems than when they learned from studying worked examples. The opposite effect was found for participants who

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{00} = \text{Intercept}$</td>
<td>4.09</td>
</tr>
<tr>
<td>$\delta_1 = \text{Social context}$</td>
<td>0.821**</td>
</tr>
<tr>
<td>$\beta_1 = \text{Instructional format}$</td>
<td>0.11</td>
</tr>
<tr>
<td>$\beta_1 = \text{Social context} \times \text{Instructional format}$</td>
<td>−0.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Group level</td>
<td>0.71</td>
<td>0.20</td>
</tr>
<tr>
<td>Individual level</td>
<td>0.60</td>
<td>0.13</td>
</tr>
</tbody>
</table>

| Deviance | 385.91 |
| Decrease in deviance | 7.45* |

* $p < .05$ ** $p < .01$. 

1 Note: In our study we have a multilevel structure with students nested in groups only as a subset of our sample. For students who worked individually (I-condition), there is no hierarchical clustering in groups, whereas there was clustering for students who worked in triads (T-condition). The research design, thus, can be considered to be a partially nested design (cf., Bauer, Sterba & Hallfors, 2008). The analyses deal with this by considering students in the I-condition to be members of different 1-person groups. This means that in the ML analyses a nested structure was forced on students in both the I- and the T-condition. We are aware that this is not fully consistent with our partially nested design. This for example means that for students in both the I- and T-conditions variance in the dependent variable is decomposed into individual level variance as well as group level variance, while there is actually no group level variance in the I-condition. Bauer et al. proposed a method that better matches such a partially nested dataset where students in the I-condition are still considered to be sole members of a group. However, by using a fixed intercept random slopes multilevel model, variance is decomposed into group and individual level variance only in the T-condition, which is consistent with the partially nested design. When using Bauer et al.’s approach, highly similar results were obtained. This leads to the conclusion that the initial MLA approach was appropriate. Accordingly, the results presented hereafter refer to ML analyses that do not take the partial nesting into account.
had learned in groups. They performed better on the posttest when they learned from studying worked examples than when they learned from solving problems (see Fig. 4). These results confirm Hypotheses 1a and 2a. Because of this interaction effect, it is important that the effects of social context and instructional format on posttest performance are not considered separately, but rather that they are interpreted together. The goodness-of-fit of the model was adequate as indicated by a significant decrease in deviance compared to the empty model, $\chi^2(3) = 12.272$, $p < .01$. Furthermore, the model explained 22% of the variance of the group level variance and 12% of the individual level variance.

The results of the MLA of the effect of social context and instructional format on the perceived mental effort invested in solving test-problems are shown in Table 4. As can be seen, the effects of social context and instructional format, as well as the interaction between social context and instructional format were not significant. The goodness-of-fit of the model was poor, $\chi^2(3) = 0.480$, $ns$. Compared to the empty model (Model 1 in Table 4), the model including social context and instructional format explained only a small part of the variance at the group- and individual level variance (2% and 1% respectively).

Finally, Table 5 shows the results of MLA on the effect of social context and instructional format on learning efficiency. A significant effect was found for social context, $p < .05$, and instructional format, $p < .05$. These main effects were qualified by the crossover interaction effect of social context and instructional format, $p < .05$. The negative sign of $\gamma_3$ should be interpreted as follows: participants who had learned individually carried out the posttest more efficiently — as indicated by a more favourable relationship between posttest effort and posttest performance — when they learned from solving problems than when they learned from studying worked examples. The opposite effect was found for participants who had learned in groups; they carried out the posttest more efficiently when they learned from studying worked examples than when they learned from solving problems. This resulted in an interaction pattern which was similar to the pattern for the posttest performance. Therefore, Hypotheses 1b and 2b were both confirmed. Inspection of the decrease in deviance compared to the empty model revealed that the goodness-of-fit of the model was only marginally significant, $\chi^2(3) = 5.572$, $p = .05$, and that the model explained 5% of the group level and 8% of the individual level variance.

5. Discussion

This study examined the effects of instructional format and social context along with their interaction on learning outcomes, cognitive load, and learning efficiency. From the performance interaction graph (Fig. 4) it is apparent that

![Fig. 4. Performance scores in the test phase as a function of social context and instructional format.](image)

### Table 3
Estimates for random intercept model for performance in the test phase.

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$\beta$</th>
<th>$SE$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{00}$ = Intercept</td>
<td>0.58</td>
<td>0.03</td>
</tr>
<tr>
<td>$\beta_1 = $Social context</td>
<td>0.180**</td>
<td>0.05</td>
</tr>
<tr>
<td>$\beta_2 = $Instructional format</td>
<td>0.12**</td>
<td>0.05</td>
</tr>
<tr>
<td>$\beta_3 =$Social context $\times$ Instructional format</td>
<td>-0.22**</td>
<td>0.07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Group level</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Individual level</td>
<td>0.03</td>
<td></td>
</tr>
</tbody>
</table>

Deviance: 63.17  
Decrease in deviance: 12.27**

### Table 4
Estimates for random intercept model for perceived amount of mental effort in the test phase.

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$\beta$</th>
<th>$SE$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{00}$ = Intercept</td>
<td>4.91</td>
<td>0.24</td>
</tr>
<tr>
<td>$\beta_1 = $Social context</td>
<td>-0.03</td>
<td>0.36</td>
</tr>
<tr>
<td>$\beta_2 = $Instructional format</td>
<td>-0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>$\beta_3 =$Social Context $\times$ Instructional format</td>
<td>0.24</td>
<td>0.51</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Group level</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>Individual level</td>
<td>1.64</td>
<td></td>
</tr>
</tbody>
</table>

Deviance: 451.72  
Decrease in deviance: 0.48

### Table 5

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$\beta$</th>
<th>$SE$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{00}$ = Intercept</td>
<td>-0.37</td>
<td>0.20</td>
</tr>
<tr>
<td>$\beta_1 = $Social context</td>
<td>0.64*</td>
<td>0.31</td>
</tr>
<tr>
<td>$\beta_2 = $Instructional format</td>
<td>0.59*</td>
<td>0.29</td>
</tr>
<tr>
<td>$\beta_3 =$Social context $\times$ Instructional format</td>
<td>-0.88*</td>
<td>0.44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Group level</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>Individual level</td>
<td>1.08</td>
<td></td>
</tr>
</tbody>
</table>

Deviance: 404.87  
Decrease in deviance: 6.57*
although learning from solving problems in a group led to higher learning outcomes than learning from studying worked examples, the reverse relationship between learning from solving problems and studying worked examples was found for individual learners. As there were no differences in the amount of mental effort invested during the posttest, the similar interaction between context and format that was found for learning efficiency seems to be caused by learning outcomes only. These test phase results provide an affirmative and clear answer to the research question, namely that for students learning individually, instruction emphasising worked example study is more effective and efficient than instruction emphasising solving problems while for students learning collaboratively, instruction emphasising problem-solving is more effective and efficient than instruction emphasising worked example study.

Learning outcome and learning efficiency were determined by the crossover interaction between learning individually or in a group from tasks imposing high or low total cognitive load. For problem-solving tasks, though the total cognitive load was high for the individual learner, within groups it could be distributed across group members. Whereas individual learners can be expected to have difficulties processing all information and constructing schemas, as indicated by the lower learning outcome and learning efficiency scores, group members could devote freed-up capacity to activities that fostered schema construction and learning, as indicated by the higher learning outcome and (Hypothesis 1a) learning efficiency scores (Hypothesis 1b). Those results make clear that the cognitive investment of inter-individual integration and coordination of information were lower within groups than the benefits of dividing the processing of information across individuals. With regard to studying equivalent worked examples, it can be argued that the decrease in extraneous load left individual learners with enough processing capacity to successfully deal with the information, negating the distribution advantage of groups. Although both individual and group learners could successfully process the information, only group learners had to invest in transactional activities. This disadvantage for group learners resulted in group learners constructing lower quality schemas, as indicated by a lower learning outcome (Hypothesis 2a) and efficiency (Hypothesis 2b) scores than individual learners.

These results were supported by the differential effect of social context on mental effort investment in the learning phase where learners in the group condition invested more mental effort than learners in the individual condition, regardless of instructional format. For learning by studying worked examples, this result supports our explanation that group members, as opposed to individual learners, need to invest additional mental effort for inter-individual integration and coordination of information and that this additional effort is not directly effective for learning. While group members invested more mental effort than individual learners, this did not result in the construction of cognitive schemas; they learned less efficiently than individual learners. For learning by solving problems, the results support our explanation that the freed-up WM capacity of group members which resulted from the distribution advantage was devoted to activities that foster learning. The greater invested mental effort of group members in the learning phase resulted in higher performance as well as a more favourable relationship between mental effort and performance on the posttest. This explanation should be further investigated by, for example, observing and analysing group communication and coordination processes to determine what the precise nature of the invested mental effort was.

A closer look at the absolute scores of invested mental effort in the learning phase reveals fairly low average scores (i.e., group $M = 4.83$; individual $M = 4.19$), and subtle but significant differences between the experimental conditions. These results could be explained by the consistently found effect that, although the scale ranges from 1 to 9, participants tend to stay in the middle range of the scale (Paas, Tuovinen, Tabbers, & Van Gerven, 2003), and that the subtle differences are reliably indicative for substantial differences in experienced cognitive load (Van Gog & Paas, 2008).

Although the cognitive load experienced by individual learners learning from tasks emphasising problem-solving was not high enough for it to be considered cognitive overload (i.e., group $M = 4.83$; individual $M = 4.19$), their posttest performance was relatively low. An alternative explanation for this might be that these learners were not able to profit from the discussion, argumentation, and reflection in the group, rather than from not having the advantage of an expanded cognitive capacity. In line with this, the finding that individuals had relatively high posttest performance when learning from tasks emphasising worked-example study could be explained by the absence of having to engage in these collaboration processes. The provided solution steps in these learning tasks may have made elaborate problem-solving activities unnecessary. Consequently, individual learners may have performed better than group learners because the latter could not profit from positive interdependence, but more likely suffered from working together with other group members (i.e., collaborative inhibition; Weldon & Bellinger, 1997). To investigate this, combined qualitative process-oriented research and quantitative cognitive load theory research, would lead to better understanding of the coordinative and communicative processes that contribute to the cognitive investment made by learners in collaborative learning (see Janssen, Kirschner, Erkens, Kirschner, & Paas, 2010).

The crossover interaction found for learning individually or collaboratively by solving problems or studying worked examples is similar to the collective working memory effect found by Kirschner et al. (2010) for learning individually or collaboratively on low and high-complexity problems. Combining the findings of both studies, one can argue that the efficiency of individual versus collaborative learning is determined by the complexity of the learning task (i.e., intrinsic load) and the instructional format of the learning tasks (i.e., extraneous and germane load). At a more general level, one can argue that the total cognitive load imposed by the learning tasks, that is, the sum of intrinsic, extraneous, and germane cognitive load (Paas, Tuovinen et al., 2003),...
determines the efficiency of individual versus collaborative learning. Learning from tasks that impose high cognitive load can be expected to be more efficient for groups than for individuals, while the opposite can be expected for learning from tasks that impose low cognitive load. An interesting topic for future research would be to investigate the level of cognitive load at which it becomes more effective and/or efficient to learn in a group as opposed to learning individually.

The positive interdependence that was stimulated in this experiment by dividing the information elements over the group members might have influenced the effects of social context. Each of the group members received three information elements which had to be exchanged to solve the problem or to understand the problem solution. Individual learners, in contrast, were confronted with all nine information elements which might have initiated qualitatively different mental representations for group and individual learners. To determine the magnitude of this effect, future studies should include an experimental condition where, similar to the individual learning condition, each learner in a group is confronted with all information elements from the outset.

It should be noted that the learning conditions in this study were designed to optimise collaboration. This creates tension between ecological validity and experimental validity. The learning environment differed from settings encountered in ‘real’ education in that all collaborating participants received only part of the unique information elements and, consequently, were required to exchange information to solve the problems or study the worked examples. Also, participants were not allowed to offload their WMs by using pencil and/or pen and paper while learning, which also might have stimulated them to collaborate. Finally, the learning setting was highly structured and scripted causing a minimal cognitive investment with respect to transactional activities. In this sense, it is not clear to what extent the results obtained in this study can be generalised to real classroom settings. It can be assumed that when all individuals have access to all information, where offloading WM is possible, and where there are normal transactional activities, different results might be obtained. Future research should investigate the contribution of the ‘artificial’ aspects of the study to the effects of the load imposed by learning tasks on the effectiveness and efficiency of learning in a group.

Finally, a theoretical implication of the results is that the limited processing capacity of an individual learner can be expanded by learning in collaboration with other learners. From this perspective, collaborative learning can be considered an instructional technique for managing individual working memory load. When groups of collaborating learners are considered as information-processing systems (see e.g., Hinsz et al., 1997), the information necessary for carrying out a learning task and its associated cognitive load can be distributed across multiple collaborating working memories. The freed-up cognitive capacity of group members can consequently be devoted to activities that foster learning. In other words, through good instructional design, a collective working memory effect can be achieved. This view has implications for CLT, one of which is that it seems that the functional properties of the learner’s cognitive architecture change when collaborating with others. The expanded limited processing capacity can only be used effectively by individual information communication and coordination processes. It is clear that these processes not only lead to a cognitive investment, but also to affective investments. Until now, CLT has focused on the alignment of instruction with cognitive processes, without recognising the role of affective ones. This research on group learning might stimulate cognitive load theorists to address affective issues in their research.

Practically, the results suggest that wholesale adoption of collaborative learning is not a sensible educational practice. The challenges that a learning task poses to the learner’s cognitive capacity and/or the amount of cognitive load a task imposes, should be determining factors when deciding whether to employ a learning model or environment based upon an individual or a collaborative learning paradigm. The higher the load imposed by the learning tasks, the more likely that collaborative learning will lead to better learning outcomes, in terms of effectiveness, efficiency, or both. This means that when choosing collaborative learning as an educational approach, educational designers — most often the teachers themselves — must assure themselves that the learning tasks given to the groups (e.g., problems, projects, et cetera) are complex enough that they cannot be carried out easily by an individual. This also suggests that it would be better for practitioners not to make an exclusive choice for individual or collaborative learning, but rather vary their approach depending on the complexity of the learning tasks to be carried out and the goals of the instruction.

References


