

# The worked example and expertise reversal effect in less structured tasks: Learning to reason about legal cases

Fleurie Nievelein<sup>a</sup>, Tamara van Gog<sup>b</sup>, Gijs van Dijck<sup>c</sup>,  
Henny P. A. Boshuizen<sup>a</sup>

<sup>a</sup> Center for Learning Sciences and Technologies, Open University of the Netherlands, Heerlen, The Netherlands

<sup>b</sup> Institute of Psychology, Erasmus University Rotterdam, The Netherlands

<sup>c</sup> Faculty of Law, Tilburg University, Tilburg, The Netherlands

Preprint of

Nievelein, F., van Gog, T., van Dijck, G., & Boshuizen, H. P. A. (2013).  
*Contemporary Educational Psychology*, 38, 118–125.

<http://dx.doi.org/10.1016/j.cedpsych.2012.12.004>

## **Abstract**

The worked example effect indicates that learning by studying worked examples is more effective than learning by solving the equivalent problems. The expertise reversal effect indicates that this is only the case for novice learners; once prior knowledge of the task is available problem solving becomes more effective for learning. These effects, however, have mainly been studied using highly structured tasks. This study investigated whether they also occur on less structured tasks, in this case, learning to reason about legal cases. Less structured tasks take longer to master, and hence, examples may remain effective for a longer period of time. Novice and advanced law students received either a description of general process steps they should take, worked examples, worked examples including the process steps, or no instructional support for reasoning. Results show that worked examples were more effective for learning than problem-solving, both for novice and advanced students, even though the latter had significantly more prior knowledge. So, a worked example effect was found for both novice and advanced students, and no evidence for an expertise-reversal effect was found with these less structured tasks.

## **Keywords**

Worked examples, Expertise, Cognitive load, Law education

## 1. Introduction

A worked-example provides learners with a description not only of the initial problem state and the goal like conventional problems do, but also with a description of the entire solution procedure. The worked example effect indicates that instruction that relies more heavily on studying worked examples is more effective (i.e., higher learning outcomes) and efficient (i.e., equal/higher learning outcomes with lower/equal investment of time or effort) for novice learners than instruction consisting only of practicing with solving equivalent conventional problems (see for reviews Atkinson, Derry, Renkl, & Wortham, 2000; Renkl, 2005; Sweller, Van Merriënboer, & Paas, 1998; Van Gog & Rummel, 2010). The worked example effect used to be investigated with conventional problem solving without any instructional support or guidance as a control condition, which, according to some, was an unfair comparison condition (Koedinger & Alevan, 2007). However, recent studies have demonstrated that a heavier reliance on examples also has beneficial effects compared to tutored problem solving only (e.g., McLaren, Lim, & Koedinger, 2008; for a review, see Salden, Koedinger, Renkl, Alevan, & McLaren, 2010).

Cognitive load theory explains the worked example effect in terms of cognitive demands imposed by problem solving and example study. When confronted with a problem, novice learners can only rely on generic problem-solving strategies such as means–ends analysis, which is aimed at reducing the difference between the current problem state and the goal state (Sweller, 1988). This might be a good strategy to solve an unknown problem (i.e., performance), but it poses very high load on working memory and does not seem to contribute much to learning (Sweller, 1988; Sweller et al., 1998), that is, the capability to perform the task after the practice session as a result of practice (Salmoni, Schmidt, & Walter, 1984). By providing novices with worked examples to study, their attention can be fully devoted to learning how the problem should be solved (Sweller et al., 1998).

Worked examples are considered to be mainly effective for novices' learning; for more advanced students an 'expertise reversal effect' (see Kalyuga, 2007; Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga & Renkl, 2010) seems to occur. The expertise reversal effect constitutes "a reversal in the relative effectiveness of instructional methods as levels of learner knowledge in a domain change" (Kalyuga & Renkl, 2010, p. 209). This can either be a full reversal (e.g., the method or format that enhances novices' learning, hampers more advanced students' learning – or vice versa) or a partial reversal (e.g., the method or format that enhances novices' learning, has no effect on more advanced students' learning – or vice versa). This effect has been found with various instructional materials, including worked examples (Kalyuga, Chandler, Tuovinen, & Sweller, 2001). Kalyuga et al. demonstrated that instruction consisting of studying worked examples was effective for learners with little – if any – prior knowledge, but lost its effectiveness and even hampered learning for more advanced students who had some prior knowledge. It is assumed that the expertise reversal effect is caused by a 'redundancy effect' (Kalyuga, 2007; Kalyuga & Renkl, 2010; Kalyuga et al., 2003). The redundancy effect has mainly been studied with instructional materials other than worked examples (for a review, see Sweller, 2005; Sweller et al., 1998), and shows that cognitive load is increased and learning is hampered when redundant information has to be processed. This can be either the case when the same information is presented in two different media (e.g., text and picture), so that either one could also have been understood in isolation, or when information is enhanced or elaborated with additional information in the same

medium (e.g., a text enhanced with additional explanations) while learners do not need this additional information (Sweller, 2005). Regarding worked examples, the worked-out solution steps become redundant for more advanced students. That is, because those students have acquired a schema that can guide their problem solving, studying worked out steps is an unnecessary activity for them. Because processing redundant worked-out steps induces ineffective working memory load and also prevents students from practicing with problem solving by themselves from which more advanced students' learning might still benefit, learning is hampered for advanced students by example study relative to problem solving (Kalyuga & Renkl, 2010; Kalyuga et al., 2001).

It should be noted, however, that both the worked example effect and the expertise reversal effect have been mainly studied using well-structured cognitive tasks such as algebra (e.g., Sweller & Cooper, 1985), statistics (e.g., Paas, 1992), geometry (e.g., Paas & Van Merriënboer, 1994; Schwonke et al., 2009), mechanics (e.g., Kalyuga et al., 2001), chemistry (McLaren et al., 2008), or physics (e.g., Reisslein, Atkinson, Seeling, & Reisslein, 2006; Van Gog, Kester, & Paas, 2011; Van Gog, Paas, & Van Merriënboer, 2006). An important aspect of well-structured tasks is that they rely on the application of a limited number of concepts, rules, and principles; they have a clear goal and an algorithmic solution path. They can be contrasted with illstructured tasks, in which the problem description is not well specified, the goal is unclear, there is uncertainty about the concepts and rules to apply, and for which multiple possible solution paths or multiple possible solutions may even exist (Jonassen, 1997; Van Merriënboer, 1997). In many professions (e.g., law and medicine) people are confronted with ill-structured tasks. During training, and especially early on in training, the complexity of those ill-structured tasks is usually constrained, so as to allow learners who still need to acquire the knowledge and skills to handle these tasks to engage in relatively authentic problems, without cognitively overloading them (Van Merriënboer, 1997). Because such training tasks are not well-structured, but not fully ill-structured either, we will refer to them as 'less-structured tasks' here. Far fewer studies have addressed the effects of examples using less structured tasks such as, for example, learning to recognize designer styles (Rourke & Sweller, 2009), learning to collaboratively diagnose a patient and design a treatment plan (Rummel & Spada, 2005), or learning argumentation skills (Schworm & Renkl, 2007).

Schmidt, Loyens, Van Gog, and Paas (2007) suggested that findings concerning beneficial effects of instructional formats that provide high levels of guidance, such as worked examples, that were obtained with well-structured tasks might not necessarily apply to less structured tasks. The studies by Rourke and Sweller (2009), Rummel and Spada (2005), and Schworm and Renkl (2007) suggest, however, that worked examples may also be effective with less structured tasks. In addition, Rourke and Sweller, in their study on teaching art students to recognize the style characteristics of different different designers, showed that second year students benefitted relatively more from example study than first year students. In other words, they did not find indications for an expertise-reversal effect; example study was also beneficial for second year students. Rourke and Sweller assumed this was because second year students had developed a higher level of visual literacy, that is, the ability to interpret or make meaning from images, a skill that might be required to optimally learn from the examples of this less structured task. In sum, their findings suggest that advanced students might still benefit from examples. It is not entirely clear though, whether this was the case because the tasks they used were simply more suited for

second year students. Therefore, the present study investigated the worked example effect and the expertise reversal effect using a less structured cognitive task, that is, reasoning about legal cases, with tasks tailored to the level of novices (i.e., cases at the level of difficulty of a first-year law course).

## **2. Reasoning about cases in law education: challenges for novices**

Reasoning about legal cases plays a pivotal role in law education in both the Civil or European-Continental and the Common or Anglo-Saxon law systems. It is a complex cognitive task that requires the integration of interrelated information elements and the coordination of different cognitive processes in order for learning to occur (and the higher the number of interrelated elements a task contains, the higher the intrinsic cognitive load a task imposes; Sweller et al., 1998), and has many variable elements: Students have to read cases, formulate questions, search in information sources for applicable laws and provisions, check whether rules and provisions can be applied to the case, and finally, provide convincing, substantive argumentation to those questions, during which they also have to take into account possible counter arguments from the opposite party (Blasi, 1995; Sullivan, Colby, Welch-Wegner, Bond, & Shulman, 2007).

The number of variable elements in a task is related to structuredness: whereas highly structured tasks are constituted mainly of consistent (or recurrent) elements which rely on algorithmic, rule-based behavior after training, less structured tasks additionally contain a high amount of variable (or non-recurrent) elements which have to be performed in varying ways across problem situations (Van Merriënboer, 1997; Van Merriënboer & Kirschner, 2007). This also means that, in contrast to consistent elements, variable elements cannot be fully automated as a result of practice. Yet, individuals with higher levels of expertise can perform variable task aspects far more effectively and efficiently than novices, as a result of their experience with many task variations (see e.g., Boshuizen & Schmidt, 1992, 2008; Sweller et al., 1998; Van Merriënboer & Kirschner, 2007). Because of the need to gain experience with many task variations, however, it becomes harder or may take longer to develop expertise on a task that is less or ill-structured (Custers, Boshuizen, & Schmidt, 1998) than on tasks that are highly structured, for which problem-solving procedures can be applied similarly across a certain category of tasks (Van Merriënboer, 1997).

Given the complexity of reasoning about legal cases, it is not surprising that research has shown that many law students, especially novices, experience difficulties with reasoning about cases (Sullivan et al., 2007; Vranken, 2006). One major reason for these difficulties is that novice law students lack the necessary conceptual knowledge to interpret a case (Blasi, 1995; Deegan, 1995; Lindahl, 2004; Nievelstein, Van Gog, Boshuizen, & Prins, 2008). Without that knowledge, students are not able to understand the information in the case and to frame the problem in the correct legal context (Deegan, 1995; Lindahl, 2004; Lundeberg, 1987; Sullivan et al., 2007; Vranken, 2006).

Another difficulty for novices lies in the use of the extensive information sources that are normally used during reasoning about a case such as codes or documented jurisprudence. Nievelstein, Van Gog, Boshuizen, and Prins (2010) showed that novice, first-year students who could use the civil code during problem solving as they normally do, performed no better than novice students who had no information other than the case description at their disposal. Advanced, third-year students who had followed several courses on the topic did perform better when they

could use the code. The reason may be that effective use of such extensive information sources also relies to a large extent on conceptual knowledge to guide the search process and to be able to interpret the information that is found, and, moreover, requires knowledge of how the source is organized (e.g., Nievelein et al., 2010). Without such knowledge, students are likely to engage in ineffective search processes, which impose a high cognitive load but do not contribute much to learning (i.e., extraneous load; Sweller et al., 1998). Indeed, it has been shown that novice law students learned significantly more when the extraneous load imposed by ineffective search processes during learning was reduced by providing students with a condensed civil code that contained only those articles relevant for the cases at hand, rather than the complete code which they normally have to use (Nievelein, Van Gog, Van Dijck, & Boshuizen, 2011).

Although this condensed code provided effective instructional support for learning, there was still a substantial amount of room for improvement in the test performance scores in the study by Nievelein et al. (2011). Therefore, instructional formats offering higher degrees of instructional guidance, such as descriptions of 'process steps' or worked examples, might be even more beneficial for novice law students' performance.

Process steps provide a description of a general, systematic approach to problem solving that students should follow, but students still have to work-out each step themselves (in contrast to worked examples, where all steps are worked out for the student to study), and in law, they can be based on Toulmin's model of argument (Toulmin, Rieke, & Janik, 1984). The question is though, whether process steps would provide a sufficient degree of guidance for novices, given that the steps are not worked out. Findings by Van Gog et al. (2006) using a structured task showed that presenting students with a description of a systematic approach to problem solving did not improve novices' learning compared to problem solving without any support. The question is, however, whether this also applied to less structured tasks, and moreover, the effects of process steps as a function of expertise have not yet been explored. For advanced students, process steps might be an appropriate form of support, because they provide guidance at a general level and advanced students might have sufficient prior knowledge to be able to work out those steps themselves.

### 3. The Present Study

In sum, the present study investigated whether novice and advanced law students' reasoning about cases would improve when, during the learning phase, they were presented with worked examples to study rather than with the same cases to solve themselves, and/or with process steps derived from Toulmin's model of argument that describe a general systematic approach to solving cases. It is hypothesized that novices would learn most from studying worked examples, either with or without process steps being made explicit, because they lack the necessary knowledge to solve cases. In addition, in line with previous research on the worked example effect, it is expected that these beneficial effects on learning are obtained with less acquisition time and less investment of mental effort (see Sweller et al., 1998). When the findings concerning the expertise reversal effect would not apply to less structured tasks, advanced students would also benefit from studying worked examples; however, when they would apply, then the process steps, which provide more general guidance, or no guidance at all, would be more beneficial for advanced students' learning.

## 4. Method

### 4.1. Participants

Seventy-five first-year law students and 36 third-year law students from a Dutch university volunteered to participate in this study, after they had received information on the study in general, what was expected of them if they would participate, how their data would be handled (confidentiality), what they would receive in return for their participation, and how to contact the experimenter. None of the first-year students had taken the introductory course on private law and, accordingly, were novices on the topic. The third-year students had completed several courses on private law. For their participation, first-year students received a financial compensation of €10 and a small amount of course credit on a written exam. Third-year students received a financial compensation of €30.

This Law School sets a limit on the total number of students enrolling; at the time the study was conducted there were approximately 400 first-year students. The majority of students enter law school when they are approximately 18 or 19 years old, after completing the highest level of secondary education that gives access to university (i.e., pre-university education, which has a 6-year duration). Another possible entry route is via general secondary education (the middle tract with a 5-year duration) and 1 or 2 years of higher professional education in a relevant discipline. Finally, a small number of students enter Law School with a completed higher education degree (either professional or university education) in another area, or after several years of work experience.

### 4.2. Design

A 2 x 2 x 2 factorial design with the factors Worked Examples (Yes vs. No; i.e., Worked Examples vs. Problem Solving), Process Steps (Yes vs. No) and Student Expertise (First-year vs. Third-year) was used. As a result, at each student expertise level, there were four conditions: 'Worked Examples + Process Steps' (first-year,  $n = 19$ ; third-year,  $n = 9$ ), 'Worked Examples - No Process Steps' (first-year,  $n = 19$ ; third-year,  $n = 9$ ), 'Problem Solving + Process Steps' (first-year,  $n = 19$ ; third-year,  $n = 9$ ), and 'Problem Solving - No Process Steps' (first-year,  $n = 18$ ; third-year,  $n = 9$ ).

### 4.3. Materials

The materials used here were identical to those used by Nievelstein et al. (2011).

#### 4.3.1. Electronic experimental environment

All materials, that is, a prior knowledge test, two learning tasks, one test task, and mental effort rating scales, were presented in a web-based experimental environment. The environment logged participants' responses and time-on-task.

#### 4.3.2. Prior knowledge test

The prior knowledge test measured conceptual knowledge, by asking students

to provide the definitions of 21 concepts related to the learning tasks, for example, 'owner', 'transfer of property', or 'gift'. It served the purpose of checking whether random assignment was successful in ruling out prior knowledge differences among conditions and whether advanced students indeed had more prior knowledge than novices.

#### 4.3.3. Learning tasks

The learning tasks consisted of two civil law cases appropriate for first-year students, with the same underlying theme, that is, whether or not someone had ownership after the transfer of property, but different contexts. The context of transfer of property in the first learning task had the sequence of rent – sell – gift. That is, person A rented a pair of skis from person X, then sold the skis to person B, who in turn gave the skis to person C as a present. The context of transfer of property in the second learning task had the sequence of rent – gift – gift. That is, person A rented a high-pressure pistol from person X, gave the pistol to person B as a present, who in turn, gave the pistol to person C as a present. The learning tasks appeared in text on the computer screen.

In the Process Steps conditions, five generic steps were provided to guide students' reasoning about the case, derived from Toulmin's model of argument (Toulmin et al., 1984): grounds, warrants, backings, rebuttals, qualifiers, and conclusion(s). First, the legally relevant facts (grounds) have to be distinguished and extract from the case information. Based on the relevant facts, applicable sources of law referred to as warrants (e.g., rules of law and statutes) have to be identified, along with possible additional information like a reference to generally accepted knowledge, norms or court judgments, which can strengthen the warrant (i.e., backings). These warrants and backings have to be compared to the grounds to test whether rules are indeed applicable to these facts. Applicable rules of law have to be placed in a specific sequence in which the more specific rules will be tested after the more general rules have proven valid. Rebuttals concern information elements from the case that are exceptions on rules, and the qualifier reflects the probability of a legally correct conclusion on the basis of the available grounds, warrants, backings and rebuttals. The final conclusion (i.e., judgment) should be drawn, based on what could be asserted from the available information. These steps to be taken by the students were listed in the instructional text above the case, and were presented in diagrammatic form beside the case. Participants received the instruction to substantiate the case according to those five steps.

In the worked examples conditions, a worked-out step-by-step argumentation according to the steps of Toulmin's model was provided of the case, with the instruction to carefully study the worked examples and to try to understand why these problem solving solutions were applied to the case. In the combined 'Worked Examples + Process Steps' condition, the process steps were explicitly mentioned in the worked examples and were presented in diagram next to the case, whereas in the examples only condition, they were only implicitly present through the worked out solutions (see also Van Gog et al., 2006; Van Gog, Paas, & Van Merriënboer, 2008).

Below the case was either a typing window in which students were required to write their argumentation about who became owner of the object after the transfer of the skis and the high-pressure pistol, respectively. There was no limitation on the number of characters that could be typed in this window. In the worked examples conditions, the worked-out solution was presented in his window and students could

not type any text themselves. All students had the Dutch civil code (a book) at their disposal.

#### 4.3.4. Test task

The test task consisted of a case that was again about ownership and transfer of property, but the context of transfer of property in this test task had the sequence of rent – gift – sell. That is, person A rented a scooter from person X, gave the scooter to person B as a present, who in turn, sold the scooter to person C. It was presented in text on the computer screen, with a typing window below the task, in which students were required to write their argumentation about who became owner of the object after the transfer of the scooter. Students could again use the civil code.

#### 4.3.5. Mental effort rating scale

Invested mental effort was measured using the 9-point subjective rating scale developed by Paas (1992). The scale ranged from very, very low mental effort (1) to very, very high mental effort (9). This rating scale is widely used in educational research (see Paas, Tuovinen, Tabbers, & Van Gerven, 2003; Van Gog & Paas, 2008). Mental effort reflects the actual cognitive load, that is, the cognitive capacity that is actually allocated to accommodate the demands imposed by the task (Paas et al., 2003).

#### 4.4. Procedure

Two-hour group sessions with maximally 20 participants per session were scheduled in a computer room at the law school. In each session, participants were randomly assigned to one of the four conditions, by having the experiment leader randomly hand out login codes that were coupled to conditions. Participants first received a short oral explanation about the procedure. Then, they were instructed to log onto the electronic learning environment and follow the directions on the screen. They could work at their own pace. All participants first completed the prior knowledge test. Then they received the two learning tasks either with or without process steps and as worked examples or problems, depending on their assigned condition. In all conditions, the test task followed immediately after the completion of the learning tasks. After each learning task and the test task, they rated their perceived amount of invested mental effort. After they had filled out the last mental effort rating scale, participants were automatically logged out of the system. Then, participants were debriefed and thanked for their participation.

#### 4.5. Data analysis

The concept definitions students provided on the prior knowledge test were rated according to their formal definitions in a Dutch legal dictionary (Algra et al., 2000). The formal definitions of the concepts consisted of one, two, or three parts. For each of these parts correctly mentioned, one point was assigned. A total of 34 points could be gained.

Performance on the test task was analyzed according to a scoring model developed by a private law professor, comparable to the models used to score examinations. The number of points to be gained for each argument, depended on the

importance of the argument to reach the correct solution. In total 100 points could be gained (see also Nievelstein et al., 2011). Two raters scored 10% of the data. The Intraclass Correlation Coefficient (ICC; Shrout & Fleiss, 1979) was .82 for the prior knowledge test and .87 for the test case. The remaining data were scored by a single rater and this rater's scores were used in the analyses.

## 5. Results

Tables 1a and 2 present the means and standard deviations of the entire sample of, respectively, first-year students' and third-year students' performance scores, mental effort ratings, and time on task. Because of the differences in participant numbers, the data for the first-year students and third-year students will be analyzed separately using 2 x 2 ANOVAs, and a comparison between first-year and third-year students will be made using non-parametric Mann–Whitney *U* tests.

For all analyses reported here, a significance level of .05 is used. For the ANOVAs Cohen's *f* is provided as a measure of effect size with .10, .25, and .40, corresponding to small, medium, and large effects, respectively (Cohen, 1988). For the Mann–Whitney *U* tests, *r* is provided as a measure of effect sized with .1, .3, and .5, corresponding to small, medium, and large effects, respectively ( $r = Z/\sqrt{N}$ ; Field, 2009).

### 5.1. Prior knowledge

Prior knowledge data from 6 first-year students and 2 third-year students were lost due to a server connection failure, so in contrast to the other analyses, the analysis of prior knowledge of the first-year students is based on: 'Worked Examples + Process Steps'  $n = 18$ ; 'Worked Examples - No Process Steps'  $n = 18$ ; 'Problem Solving + Process Steps'  $n = 16$ ; 'Problem Solving - No Process Steps'  $n = 17$ , and for the third-year students 8 rather than 9 participants were left in the 'Worked Examples + Process Steps' and 'Problem Solving + Process Steps' conditions. ANOVAs showed that there were no differences in prior knowledge among conditions in the group of first-year students,  $F(3,65) = .51, p = .675$  and in the group of third-year students,  $F(3,30) = .02, p = .998$ . As expected, a t-test showed that third-year students ( $M = 17.12, SD = 3.13$ ) had significantly more prior knowledge than first-year students ( $M = 8.86, SD = 2.85$ ),  $t(101) = -13.39, p < .001$ , Cohen's  $d = 2.75$  (approximately equal to Cohen's  $f = 1.38$ ).

### 5.2. First-year students

Prior to the analyses on the first-year students' data, outliers as identified by SPSS boxplots were removed, which were: the learning phase effort data of four participants (one in the Worked Example + Process Steps condition, two in the Worked Example condition, and one in the Problem Solving + Process Steps condition), time on task data of one participant (in the Problem Solving + Process Steps condition), test performance data of three participants (one in the Problem Solving + Process Steps condition and two in the Problem Solving condition), and test effort data of four participants (two in the Worked Example + Process Steps condition and two in the Problem Solving condition). The means and standard deviations after removal of outliers are shown in Table 1b.

The first-year students' data (test task performance, time spent on the learning

tasks, mental effort invested in the learning tasks, and mental effort invested in the test task) were analyzed using 2 x 2 ANOVAs with factors ‘Worked Examples’ and ‘Process Steps’. On test performance, there was a significant main effect of ‘Worked Examples’,  $F(1,68) = 105.221$ ,  $MSE = 426.037$ ,  $p < .001$ ,  $f = 1.24$ , indicating that students who studied worked examples during the learning phase, performed significantly better on the test task ( $M = 58.63$ ,  $SD = 27.74$ ) than students who engaged in problem solving ( $M = 8.59$ ,  $SD = 4.06$ ). There was no significant main effect of ‘Process Steps’,  $F(1,68) < 1$ , nor a significant interaction effect  $F(1,68) < 1$ . Because the assumption of homogeneity of variance was not met for the performance data (with higher variance in test performance in the worked examples groups than in the problem-solving groups), the significant main effect was also tested using ANOVA with the more robust Welch test, which showed the same result  $F(1,38.769) = 120.814$ ,  $p < .001$ .

The analysis of time on task in the learning phase showed a significant main effect of ‘Worked Examples’ (worked examples:  $M = 568.47$ ,  $SD = 227.80$ ; problem solving:  $M = 1146.15$ ,  $SD = 509.58$ ),  $F(1,70) = 58.126$ ,  $MSE = 106135.04$ ,  $p < .001$ ,  $f = 0.75$  (Welch:  $F(1,47.883) = 38.901$ ,  $p < .001$ ), a significant main effect of ‘Process Steps’,  $F(1,70) = 19.601$ ,  $MSE = 106135.04$ ,  $p < .001$ ,  $f = 0.37$  (Welch:  $F(1,55.195) = 9.405$ ,  $p = .003$ ), as well as a significant interaction,  $F(1,70) = 15.02$ ,  $MSE = 106135.04$ ,  $p < .001$ ,  $f = 0.32$ . The interaction was tested with  $t$ -tests (equal variances; Levene’s test was not significant) which indicated that the increase in learning time as a result of being given process steps was significant for students in the problem solving conditions (with process steps:  $M = 1460.72$ ,  $SD = 459.41$ , without process steps:  $M = 831.58$ ,  $SD = 337.64$ ),  $t(34) = 4.68$ ,  $p < .001$ , Cohen’s  $d = 1.56$  (approximately equal to Cohen’s  $f = 0.78$ ), whereas it was not significant for students in the worked examples conditions (with process steps:  $M = 589.37$ ,  $SD = 256.27$ , without process steps:  $M = 547.58$ ,  $SD = 200.18$ ),  $t(36) = .56$ ,  $p = .58$ .

The analysis of mental effort invested in the learning phase showed a significant main effect of ‘Worked Examples’, with students in the worked examples conditions investing less effort ( $M = 5.46$ ,  $SD = 0.79$ ) than students the problem solving conditions ( $M = 6.39$ ,  $SD = 1.39$ ),  $F(1,67) = 15.442$ ,  $MSE = 1.181$ ,  $p < .001$ ,  $f = 0.42$  (Welch:  $F(1,55.651) = 12.089$ ,  $p = .001$ ), a significant main effect of ‘Process Steps’,  $F(1,67) = 4.912$ ,  $MSE = 1.181$ ,  $p = .030$ ,  $f = .24$  (Welch:  $F(1,67.966) = 3.973$ ,  $p = .05$ ), with students in the process steps conditions investing more effort ( $M = 6.21$ ,  $SD = 1.28$ ) than students in the no process steps conditions ( $M = 5.64$ ,  $SD = 1.10$ ). The interaction effect only neared significance,  $F(1,67) = 3.513$ ,  $p = .065$ ,  $f = .20$ , but suggests that the effects of process steps on mental effort increase only occurred in the problem solving conditions and not in the worked examples conditions (see Table 1b).

On mental effort invested in the test task no significant effects were found (main effects, both  $F(1,67) < 1$ ; interaction effect,  $F(1,67) = 2.44$ ,  $p = .12$ ).

### 5.3. Third-year students

The same 2 x 2 ANOVAs were performed for the third-year students. On test performance, there was a significant main effect of ‘Worked Examples’,  $F(1,32) = 37.03$ ,  $MSE = 497.06$ ,  $p < .001$ ,  $f = 1.02$ , indicating that students who studied worked examples during the learning phase, performed significantly better on the test task ( $M = 82.28$ ,  $SD = 18.11$ ) than students who engaged in problem solving ( $M = 37.06$ ,  $SD = 26.71$ ). There was no significant main effect of ‘Process Steps’,  $F(1,32) < 1$ , nor a significant interaction effect,  $F(1,32) = 3.06$ ,  $MSE = 497.06$ ,  $p = .09$ . Because

performance data were not normally distributed in all conditions, we also conducted non-parametric Mann–Whitney  $U$  tests, which showed the same results: a significant effect of ‘Worked Examples’  $Z = 4.13, p < .001, r = .688$ , but no significant effect of ‘Process Steps’  $Z = .127, p = .899$ .

The analysis of time on task in the learning phase showed a significant main effect of ‘Worked Examples’ (worked examples:  $M = 427.36, SD = 162.21$ ; problem solving:  $M = 975.58, SD = 363.38$ ),  $F(1,32) = 36.03, MSE = 75078.72, p < .001, f = 1.00$ . There was no significant main effect of ‘Process Steps’,  $F(1,32) = 2.43, p = .13$ , nor a significant interaction effect,  $F(1,32) = 1.43, p = .24$ .

The analysis of mental effort invested during the learning phase showed no significant main effects (‘Worked Examples’:  $F(1,32) = 2.22, p = .15$ ; ‘Process Steps’:  $F(1,32) < 1$ ) or interaction effect ( $F(1,32) = 1.10, p = .30$ ).

Neither did mental effort invested in the test task show significant main effects (Worked Examples:  $F(1,32) < 1$ ; ‘Process Steps’:  $F(1,32) = 1.19, p = .28$ ) or interaction effect ( $F(1,32) < 1$ ).

#### 5.4. First-year vs. third-year students

Mann–Whitney  $U$  tests showed that third-year students performed significantly higher on the test case ( $M = 59.67, SD = 32.12$ ) than first-year students ( $M = 34.76, SD = 31.64$ ),  $Z = 3.71, p < .001, r = .35$ , had to invest less effort in the learning tasks ( $M = 4.46, SD = 1.52$ ; mean rank: 37.51) than first-year students ( $M = 5.75, SD = 1.41$ ; mean rank: 64.87),  $Z = -4.22, p < .001, r = .40$ , and had to invest less effort ( $M = 4.61, SD = 1.50$ ; mean rank: 39.58) in performing the test task than first-year students ( $M = 5.88, SD = 1.47$ ; mean rank: 63.88),  $Z = 3.81, p < .001, r = .36$ . Time spent on the learning tasks did not differ significantly between third-year ( $M = 701.47, SD = 392.68$ ; mean rank: 49.14) and first-year students ( $M = 882.87, SD = 561.81$ ; mean rank: 59.29),  $Z = -1.56, p = .120$ .

## 6. Discussion

This study investigated whether providing first- and third-year students with instructional support consisting of worked examples, process steps, or both, during learning, improved their learning to reason about this type of cases as measured by performance on a test task. In line with our hypothesis, a worked example effect was found for novice students: First-year students performed best on the test case when they had studied worked examples during the learning phase. Moreover, despite their higher level of expertise as corroborated by their higher prior knowledge scores, higher test performance, and lower mental effort ratings, the same applied to third-year students. That is, an expertise reversal effect did not occur.

The finding that studying worked examples leads to better test performance with less time on task, and less mental effort invested during learning (at least for first-year students; the difference in invested effort failed to reach significance for third-year students, possibly due to the smaller sample size), replicates that of other studies (e.g., Paas & Van Merriënboer, 1994; Van Gog et al., 2006). In contrast to some other studies (e.g., Paas, 1992; Paas & Van Merriënboer, 1994), however, we found no differences among conditions in mental effort invested in the test task. Potentially, this is a consequence of differences between the current study and prior studies in the number of examples provided; studies using less structured tasks often provide more than two examples (which is possible because the tasks are much

shorter), which would offer more opportunities for schema construction or schema elaboration. Nonetheless, these findings do indicate that studying worked examples was more efficient in terms of the learning process, that is, higher test performance was reached with lower investment of time and effort during learning, and in terms of the quality of learning outcomes, that is, higher test performance was reached with equal investment of mental effort in the test (see Van Gog & Paas, 2008).

Providing students only with process steps to be taken (i.e., without being combined with worked examples), resulted in higher time on task for the first-year students, but did not improve their learning, so this does not seem to be an effective form of instructional guidance. Possibly, students are not able to effectively use the steps to guide their problem solving, because they need to find out for themselves what they have to do at each of those steps and why, which does not reduce cognitive load imposed by (ineffective) search processes and strategies.

There are several (potential) limitations to this study. First, a limitation of this study was the small sample of third year students, due to which the power of the analyses of the third-year students data was lower (.64 according to post hoc power analysis using G\*Power 3.1.3, assuming large effects; Faul, Erdfelder, Buchner, & Lang, 2009) than the power of the analyses of first-year students' data (.93 according to post hoc power analysis using G\*Power 3.1.3, assuming large effects; Faul et al., 2009). So it is possible that some effects that neared significance, would have been significant with larger sample sizes. Second, a potential limitation of this study is that we cannot rule out that the activation of conceptual knowledge (required for the prior knowledge test) may have had a differential effect for first and third year students during the learning phase. However, even if the activation of prior knowledge would have boosted third year students' performance more than first year students' performance, it is unlikely that this would be the sole explanation for the performance differences between first and third year students on the test case, because the available conceptual knowledge would likely also have been activated by the cases in the learning phase. More importantly, because prior knowledge test performance did not differ between conditions for either the first or the third year students, the instructional effects are not due to differences in prior knowledge. Third, another potential limitation is that we did not use a transfer test case, on which different laws would apply. As a consequence, it is not entirely clear whether the beneficial effects of worked examples on performance on the test case result from an improvement in knowledge about the laws that applied in the cases used here, or from an improvement in the process of reasoning itself, or both. Future studies might analyze performance on novel cases where other laws apply to get an indication of the extent to which worked examples contribute to enhanced reasoning processes (i.e., transfer in terms of preparation for future learning; Bransford & Schwartz, 1999).

Nevertheless, the findings from this study clearly show that worked examples are also effective for learning less structured tasks as Rourke and Sweller (2009) suggested, despite the suggestion by Schmidt et al. (2007) that findings on instructional support on highly structured tasks might not necessarily apply to less structured tasks. These findings suggest that implementing a heavier reliance on worked examples in curricula may be an interesting option for law educators. Of course, because this study was highly controlled and of short duration, future research should establish whether the beneficial effects of worked examples could also be found in more ecologically valid settings over a longer period of time.

Moreover, the findings from this study suggest that an expertise reversal effect may not occur on less structured tasks, even though the tasks were at a level

appropriate for novices, and even though the advanced students had a much more prior knowledge and had a much higher level of performance on those tasks than the novices – as becomes clear when comparing the scores from the problem solving conditions (see Tables 1a, 1b and 2). What is not entirely clear, however, and should be addressed in future research, is what this means for the expertise reversal effect. As mentioned above, based on our data it is not possible to distinguish whether the beneficial effects of worked examples on performance on the test case result from an improvement in knowledge about the laws that applied in the cases used here, or from an improvement in the process of reasoning itself, or both, and so it might be the case that the reason why worked examples were effective are different for first year and third year students. It is possible that third-year students in the problem-solving conditions performed better on the tasks because they had more conceptual knowledge and were more familiar with the organization of the civil code. This combination would result in more effective and efficient search processes for those students, while first-year students experience many difficulties with those aspects of the task (Nivelstein et al., 2011). However, subsequently using the information found in that search process to develop a good line of argument in relation to the case, is a highly variable task component that might still be difficult for third-year students (see e.g., Voss, 2006), which might explain why they still benefitted from the examples. In other words, it is possible that third-year students did not need the support provided by worked examples on certain aspects of the task on which first-year students did benefit from it, but that third-year students still benefitted from the support the examples provided on other aspects of the task. Unfortunately, based on our data, it is not possible to distinguish on which components they did and did not benefit from the examples; future studies could address this question by using process-tracing techniques such as verbal reporting (Ericsson & Simon, 1993), eye tracking (Holmqvist et al., 2011) or a combination of both (e.g., Van Gog, Paas, Van Merriënboer, & Witte, 2005), to study which parts of examples are studied most deeply by advanced students and how this affects performance on a test case.

Another option to address this question would be to look at completion problems (i.e., partly worked examples with blank steps for the student to complete; Paas, 1992) that vary in which aspects are worked out and which aspects the student has to complete. This would also shed light on what our finding regarding the expertise reversal effect would entail for a completion or fading strategy (see Renkl & Atkinson, 2003) with less structured tasks. Research with well-structured tasks has shown that instructional support needs to be high initially and should then be gradually decreased, or faded out, when students' expertise on the task increases. This is done by providing students with fully worked examples initially, followed by completion problems with an increasing number of blanks for the learner to complete, ending with solving problems without any support. The findings of our study suggest that the dynamics of this fading principle may be different with less structured tasks, and that in deciding which steps to delete, it should be considered which task aspects can be faded earlier and which aspects should only be faded relatively late in a task sequence.

#### Acknowledgments

We would like to thank LL.M. Helma Severeijns-Wijenbergh for developing the materials, the Faculty of Law at Tilburg University for the facilitation of this study, and Prof. Dr. Huub Spoormans for his helpful suggestions concerning previous drafts of this article.

## References

- Algra, N. E., Barmantlo, D. G., Beelen, S. T. M., Van den Bergh, A., Boel, F. C., De Boer, A. H., et al. (2000). *Juridisch woordenboek* [Legal dictionary]. Nijmegen. The Netherlands: Ars Aequi.
- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. W. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*, 181–214.
- Blasi, G. (1995). What lawyers know: Lawyering expertise, cognitive science, and the functions of theory. *Journal of Legal Education, 45*, 313–397.
- Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates, and novices. *Cognitive Science, 16*, 153–184.
- Boshuizen, H. P. A., & Schmidt, H. G. (2008). The development of clinical reasoning expertise. In J. Higgs, M. A. Jones, S. Loftus, & N. Christensen (Eds.), *Clinical reasoning in the health professions* (3rd ed., pp. 113–121). London: Butterworth-Heinemann.
- Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*, 61–100.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Custers, E. J. F. M., Boshuizen, H. P. A., & Schmidt, H. G. (1998). The role of illness scripts in the development of medical diagnostic expertise: Results from an interview study. *Cognition and Instruction, 16*, 367–398.
- Deegan, D. H. (1995). Exploring individual differences among novices reading in a specific domain: The case of law. *Reading Research Quarterly, 30*, 154–170.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev.ed.). Cambridge, MA: MIT Press.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.
- Field, A. (2009). *Discovering statistics using SPSS*. London: Sage Publications.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking a comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.
- Jonassen, D. H. (1997). Instructional design models for well-structured and illstructured problem-solving learning outcomes. *Educational Technology Research and Development, 45*(1), 65–94.
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509–539.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23–31.
- Kalyuga, S., Chandler, P., Tuovinen, J. E., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*, 579–588.
- Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science, 38*, 209–215.
- Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*, 239–264.
- Lindahl, L. (2004). Deduction and justification in the law: The role of legal terms and

- concepts. *Ratio Juris*, 17, 182–202.
- Lundeberg, M. A. (1987). Metacognitive aspects of reading comprehension: Studying understanding in case analysis. *Reading Research Quarterly*, 22, 407–432.
- McLaren, B. M., Lim, S., & Koedinger, K. R. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 2176–2181). Austin, TX: Cognitive Science Society.
- Nievelstein, F., Van Gog, T., Boshuizen, H. P. A., & Prins, F. J. (2008). Expertise-related differences in ontological and conceptual knowledge in the legal domain. *European Journal of Cognitive Psychology*, 20, 1043–1064.
- Nievelstein, F., Van Gog, T., Boshuizen, H. P. A., & Prins, F. J. (2010). Effects of conceptual knowledge and availability of information sources on law students' legal reasoning. *Instructional Science*, 38, 23–35.
- Nievelstein, F., Van Gog, T., Van Dijck, G., & Boshuizen, H. P. A. (2011). Instructional support for novice law students: Reducing search processes and explaining concepts in cases. *Applied Cognitive Psychology*, 25, 408–413.
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology*, 84, 429–434.
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38, 63–71.
- Paas, F., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive load approach. *Journal of Educational Psychology*, 86, 122–133.
- Reisslein, J., Atkinson, R. K., Seeling, P., & Reisslein, M. (2006). Encountering the expertise reversal effect with a computer-based environment on electrical circuit analysis. *Learning and Instruction*, 16, 92–103.
- Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skills acquisition: A cognitive load perspective. *Educational Psychologist*, 38, 15–22.
- Renkl, A. (2005). The worked examples principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 229–244). New York: Cambridge University Press.
- Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognise designers' styles. *Learning and Instruction*, 19, 185–199.
- Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *Journal of the Learning Sciences*, 14, 201–241.
- Salden, R. J. C. M., Koedinger, K. R., Renkl, A., Alevin, V., & McLaren, B. M. (2010). Accounting for beneficial effects of worked examples in tutored problem solving. *Educational Psychology Review*, 22, 379–392.
- Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin*, 95, 355–386.
- Schmidt, H. G., Loyens, S. M. M., Van Gog, T., & Paas, F. (2007). Problem-based learning is compatible with human cognitive architecture: Commentary on Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42, 91–97.

- Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. (2009). The worked-example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior, 25*, 258–266.
- Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology, 99*, 285–296.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.
- Sullivan, M. W., Colby, A., Welch-Wegner, J., Bond, L., & Shulman, L. S. (2007). *Educating lawyers*. San Francisco: Jossey-Bass.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.
- Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*, 59–89.
- Sweller, J. (2005). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 159–167). New York: Cambridge University Press.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296.
- Toulmin, S. E., Rieke, R., & Janik, A. (1984). *An introduction to reasoning*. New York: Macmillan Publishing Company.
- Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology, 36*, 212–218.
- Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*, 16–26.
- Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2006). Effects of process-oriented worked examples on troubleshooting transfer performance. *Learning and Instruction, 16*, 154–164.
- Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2008). Effects of studying sequences of process-oriented and product oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction, 18*, 211–222.
- Van Gog, T., Paas, F., Van Merriënboer, J. J. G., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied, 11*, 237–244.
- Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22*, 155–174.
- Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.
- Van Merriënboer, J. J. G., & Kirschner, P. A. (2007). *Ten steps to complex learning: A systematic approach to four-component instructional design*. Mahwah, NJ: Erlbaum.
- Voss, J. F. (2006). Toulmin's model and the solving of ill-structured problems. In D. Hitchcock & B. Verheij (Eds.), *Arguing on the Toulmin model: New essays in argument analysis and evaluation* (pp. 303–311). Dordrecht, The Netherlands: Springer.
- Vranken, J. B. M. (2006). *Exploring the jurist's frame of mind*. Deventer, The

Netherlands: Kluwer.