

ACQUIRING PROBLEM-SOLVING SKILLS IN HIGHER EDUCATION

Sequencing and Self-Regulated Learning
from Examples and Problems

Milou van Harsel



ACQUIRING PROBLEM-SOLVING SKILLS IN HIGHER EDUCATION

Sequencing and Self-Regulated Learning
from Examples and Problems

LEREN PROBLEEM-OPLOSSEN IN HET HOGER ONDERWIJS

Het sequentiëren en zelfgestuurd leren van
voorbeelden en oefenproblemen
(met een samenvatting in het Nederlands)

MILOU VAN HARSEL

Leden van de beoordelingscommissie

Prof. dr. Anique de Bruin
Prof. dr. Johan Jeuring
Prof. dr. Liesbeth Kester
Prof. dr. Sofie Loyens
Prof. dr. Alexander Renkl

All rights reserved. No part of this dissertation may be reproduced or transmitted in any form, by any means, electronic or mechanical, without the prior permission of the author, or where appropriate, of the publisher of the articles.

Cover design: www.zonderfratsen.com
Lay-out: www.zonderfratsen.com
Print: Gildeprint, The Netherlands

ISBN: 978.94.641.9260.5
DOI: 10.33540/697

© Copyright 2021 Milou van Harsel

ACQUIRING PROBLEM-SOLVING SKILLS IN HIGHER EDUCATION

Sequencing and Self-Regulated Learning
from Examples and Problems

LEREN PROBLEEM-OPLOSSEN IN HET HOGER ONDERWIJS

Het sequentiëren en zelfgestuurd leren van
voorbeelden en oefenproblemen
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op

vrijdag 1 oktober 2021 des middags te 12:15 uur

door

MILOU VAN HARSEL

geboren op 8 november 1988
te Roosendaal en Nispen

Promotor
Prof. dr. T. van Gog

Copromotoren
Dr. P.P.J.L. Verkoeijen
Dr. V. Hoogerheide

Table of contents

Chapter 1	Introduction	9
Chapter 2	Effects of different sequences of examples and problems on motivation and learning	23
Chapter 3	Examples, practice problems, or both? Effects on motivation and learning in shorter and longer sequences	53
Chapter 4	How do higher education students regulate their learning in an online environment with video modeling examples, worked examples, and practice problems?	93
Chapter 5	Instructing students on effective sequences of examples and problems: Does self-regulated learning improve from knowing what works and why?	121
Chapter 6	Summary and discussion	147
	References	163
	Supplementary materials	175
	Nederlandse samenvatting (Summary in Dutch)	201
	Curriculum vitae	215
	Dankwoord (Acknowledgements in Dutch)	221

Chapter 1

Introduction



Imagine you are having your friends over for dinner later in the day and want to impress them with homemade sushi rolls. Unfortunately, as a novice, you lack the knowledge and skills required to make the perfect maki and uramaki. *What do you do?* Option A is to buy the necessary ingredients and make the sushi rolls without any assistance – how difficult can it be? It is possible that you would be able to make presentable sushi rolls yourself. However, the process of doing so would likely be very time-consuming, effortful, and error prone (resulting in substantial loss of time and waste of food). It is also possible that you would fail, because after many unsuccessful attempts, you might become frustrated and give up in the end. In that case, your culinary adventure changes to a sad last-minute order from the local fast food joint. Option B is to consult an example, for instance by asking someone close to you who has experience with making sushi to demonstrate the steps necessary to make the perfect sushi rolls. And if you do not know of such a person in your immediate vicinity, nowadays such experts are just a click away. You open YouTube on your mobile phone and search for ‘how to make the best maki and uramaki ever?’ and you can learn from videos by experts all over the world, even real Japanese Sushi chefs! After watching one or multiple examples, you will likely be able to roll the sushi so quickly and easily, that you even have time left to make some nigiri as well. *So, which option would you choose?*

1.1 Example-Based Learning

Most people would presumably choose option B, because learning from examples (i.e., *example-based learning*) is a very natural and very powerful strategy for acquiring new knowledge and skills. And this does not only apply to knowledge and skills required to perform ‘everyday’ tasks such as cooking. Decades of educational research have shown that example-based learning is a very effective and efficient way of learning new skills in educational and professional settings (for reviews, see Atkinson et al., 2000; Renkl, 2014; Sweller et al., 2011; Van Gog & Rummel, 2010; Van Gog et al., 2019). This is evidenced by studies that have shown that for novices (i.e., learners with little or no prior knowledge), example study is an effective and efficient strategy to learn new problem-solving skills relative to strategies that rely more heavily on learning by doing, such as solving practice problems. More effective refers to the finding that students attain higher scores on learning outcomes; that is, they perform better on test tasks that are isomorphic to the tasks studied or practiced in the learning phase, and sometimes also on transfer tasks, which are slightly different from the tasks encountered in the learning phase (e.g., Cooper & Sweller, 1987; Paas, 1992; Paas & Van Merriënboer, 1994). More efficient means that this (equal or) higher performance on test tasks, is achieved with less investment of time or mental effort during the learning phase, or when solving the test problems (Paas & Van Merriënboer, 1994; Van Gog & Paas, 2008).

But why do novices benefit so much from example study? When learners have not yet learned the procedures necessary for solving a problem, practice problem solving without any support forces them to resort to weak problem-solving strategies, such as trial-and-error strategies or means-ends analysis. These strategies are time-consuming, impose a high load on working memory, but barely contribute to learning (Sweller, 1988). In contrast, studying examples prevents the use of weak problem-solving strategies and allows learners to devote all available cognitive capacity to mastering the solution procedure. This helps them to develop a cognitive schema of the solution procedure and thereby to solve similar problems in the future (e.g., Sweller et al., 2011; Sweller & Cooper, 1985; Sweller et al., 1998).

The efficacy of example-based learning has been repeatedly established with different types of examples, originating from different theoretical perspectives (Renkl, 2014; Van Gog & Rummel, 2010). From a cognitive perspective (e.g., cognitive load theory; Sweller, 1988; Sweller et al., 2011), the efficacy of example-based learning has mainly been established with *worked examples*. Worked examples are text-based and consist of a problem statement and a written step-by-step explanation of a full and correct solution procedure of how to solve a problem. Social-cognitive research inspired by theories such as social learning theory (e.g., Bandura, 1977, 1986) has predominantly focused on modeling examples, in which someone else (the model) demonstrates and (possibly) explains the solution procedure step by step. Modeling examples can be presented either live (e.g., Bjerrum et al., 2013) or on video (e.g., Van Gog et al., 2014). Nowadays, there are also more ‘hybrid’ examples with characteristics of both modeling examples and worked examples (i.e., screen-recordings in which the example builds up step-by-step until the end-state is basically a worked example; e.g., McLaren et al., 2008).

With the advent of modern technology, the popularity of example-based learning has increased further in recent years (Hoogerheide & Roelle, 2020), because it has become much easier to create examples and share them (e.g., De Koning et al., 2018; Fiorella & Mayer, 2018; Van der Meij & Van der Meij, 2013). YouTube is filled with ‘how to’ videos on a variety of topics, including educational topics, many also created by teachers. Moreover, teachers are experimenting more and more with relatively novel educational concepts in which worked examples and video modeling examples play an important role. Such concepts include, for instance, having learners prepare instructional material for classes at home and apply this material during practice at school (i.e., flipping the classroom; Bergmann & Sams, 2012; Van Alten et al., 2019), combining face-to-face and technology-mediated instruction (i.e., blended learning; Stein & Graham, 2020), or creating free and easily accessible online courses aimed at unlimited participation and open access via internet (e.g., Massive Open Online Courses –MOOCs; Deng et al., 2019; McAuley et al., 2010).

As such, students nowadays acquire new knowledge and skills increasingly via online learning environments (in blended or fully online courses), in which worked examples, video modeling examples, and practice problems are often embedded (e.g., Roll et al., 2011). A very popular example is Khan Academy (www.khanacademy.org), which is an online educational platform that provides many video modeling examples and additional practice opportunities for a wide range of subjects. Such online platforms seem ideal for self-regulated and self-directed learning (e.g., when doing homework or studying for a test), because they give learners the opportunity to determine themselves where (i.e., at school or at home), when, and how they want to study. As a result, personalized learning paths can be created, which has been suggested to enhance students' motivation and learning outcomes more than non-personalized instruction that is the same for all students (e.g., Niemiec et al., 1996; Schnackenberg & Sullivan, 2000).

Although modern technology provides many opportunities to enhance the use of example-based learning in education, the technological possibilities are far ahead of our understanding of how to provide learners with optimal sequences of examples and practice problems that foster their learning process (including motivational aspects of learning), and of how learners use examples and practice problems during self-regulated learning. These questions are not only theoretically relevant; they are also frequently asked by teachers and educational consultants. Therefore, the studies in this dissertation started to address those questions.

1.2 Sequencing Example Study and Practice Problem Solving

This first part of this research project was carried out at a Dutch university of applied sciences, in the context of technical and primary teacher education. The aim of the studies in the first part of this dissertation was to investigate *what sequences of examples and practice problems are most effective, efficient, and motivating for first year higher education students' learning of new mathematical problem-solving skills*. As mentioned earlier, decades of research have shown that instruction that relies heavily on example study is more effective and efficient for novices' learning of new problem-solving skills than practice problem solving only (for reviews, see Atkinson et al., 2000; Renkl, 2014; Sweller et al., 2011; Van Gog et al., 2019). However, the question is what 'heavier reliance' means exactly (Van Gog et al., 2011); examples can be presented to novice learners in many ways. For example, novices can study a sequence of examples without any problem-solving practice (i.e., *example study only*), or alternate example study and practice problem solving. In case of the latter, sequences can be created in which practice problems are solved after example study (i.e., *example-problem pairs*) or before example study (i.e., *problem-example pairs*).

What we already know from research on the worked example effect is that task sequences containing only examples (e.g., Van Gerven et al., 2002; Van Gog et al., 2006) or example-problem pairs (e.g., Carroll, 1994; Cooper & Sweller, 1987; Kalyuga et al., 2001; McLaren et al., 2008; Mwangi & Sweller, 1998; Rourke & Sweller, 2009; Sweller & Cooper, 1985) are more effective and efficient (less effortful during learning) for novices' learning than sequences of only practice problems. Research has also shown that sequences of examples only and example-problem pairs do not seem to differ in terms of the amount of mental effort required from students and the learning outcomes they attain (Van der Meij et al., 2018), even when tested one week after the learning materials have been studied (Leahy et al., 2015; Van Gog & Kester, 2012; Van Gog et al., 2015; Van Gog et al., 2011). The use of problem-example pairs, however, might be ill-advised: Despite the fact that they offer an equal number of examples as example-problem pairs, several studies have found that problem-example pairs were not more effective or efficient for learning than solving practice problems only, and less effective and efficient than studying example-problem pairs (e.g., Kant et al., 2017; Leppink et al., 2014; Van Gog et al., 2011). Although it is an open question how that finding can be explained, it has been suggested that motivational variables might play a role (Van Gog et al., 2011).

Indeed, an important open question that is relevant for theory and educational practice is whether different sequences of examples and practice problems would differentially affect not only cognitive, but also motivational aspects of learning. Sequencing research has predominantly been conducted with worked examples and against the backdrop of cognitive load theory research, which has focused on cognitive variables (i.e., learning outcomes and invested mental effort). However, student motivation has been long ignored in research with worked examples (e.g., Sweller et al., 2011, with some exceptions: e.g., Crippen et al., 2009; Crippen & Earl, 2007; Paas et al., 2005; Schnotz et al., 2009), and therefore also in sequencing research. Although there are indications that example study can influence important aspects of motivation, by fostering expectancies of one's own abilities, such as self-efficacy and perceived competence (e.g., Bandura, 1997; Crippen et al., 2009; Hoogerheide et al., 2014; Hoogerheide et al., 2018), it is an open question to what extent motivational variables underlie sequencing effects found in prior research (as suggested by Van Gog et al., 2011) and whether some sequences are more motivating than others (as suggested by Sweller & Cooper, 1985).

For instance, a motivational explanation of the finding by Van Gog et al. (2011) that sequences of practice problems only and problem-example pairs do not work so well compared to sequences of examples only or example-problem pairs, would be that starting with a practice problem and failing to solve it could cause students to lose interest in the (topic of the) task or to lose confidence in their own abilities. This in turn, could negatively affect students' willingness to work on other tasks (including examples) that follow. This would also explain why such a difference between example-problem

and problem-example pairs was not found for problems that were complex yet enjoyable or intrinsically rewarding (i.e., the puzzle problems used in Van Gog, 2011). Although this is speculative, in contrast to science or math problems, failing at puzzle problems might challenge rather than demotivate learners to study an example that follows. Sweller and Cooper (1985) used example-problem pairs in their study, based on the grounds that this would be more motivating for learners than examples only (although they did not test this assumption). Whereas example study only is assumed to be more 'passive', example-problem pairs give learners the opportunity to actively apply what they have learned and test their knowledge. If this assumption is correct, example-problem pairs might not necessarily be more effective and efficient for learning than example study only but might have a (stronger) positive effect on student motivation. Knowing whether some sequences are more motivating than others is very relevant for educational practice. Motivational aspects of learning are critical to learning and achievement, as they impact how likely a learner is to give up or push forward to reach a goal, and how much time and/or effort a learner is willing to invest in reaching that goal (e.g., Pintrich, 2003). As students often work on problem-solving tasks in online learning environments during self-study sessions (either at school or at home in preparation for class or a test), motivational aspects of task sequences likely determine whether learners decide to start, continue, or quit studying.

Another important open question in the sequencing literature is whether the established sequencing effects would change with longer task sequences. Thus far, sequencing research has been limited to rather short learning phases of two to four learning tasks. The fact that learners' knowledge gradually increases in longer learning phases may change the effectiveness and efficiency of certain sequences of examples and practice problems. For instance, once a schema of how to solve a certain type of problem has been acquired, additional example study no longer contributes to learning and students might start to benefit more from problem-solving practice (e.g., Kalyuga et al., 2001). Moreover, with longer sequences, student motivation becomes an even more important factor to consider. For instance, potential negative motivational effects of studying examples only (compared to example-problem pairs) might not be as pronounced in short sequences but might arise in longer sequences where more tasks have to be studied (or solved). Studying longer sequences is also relevant for generalizing findings to educational practice, where longer task sequences are more common.

Therefore, the empirical studies in the first part of this dissertation (**Chapters 2 and 3**), addressed the question of how different short and longer sequences of examples and practice problems affect cognitive and motivational aspects of learning.

1.3 Self-Regulated Learning with Examples and Practice Problems

The second part of this research project was carried out at a Dutch university of applied sciences, in the context of technical education. The aim of the studies in the second part of the dissertation was to examine *how and how well first year higher education students regulate their learning from examples and practice problems in an online learning environment and whether informing them about effective, efficient, and motivating instructional design principles helps to improve their task-selections, and thereby their motivation and learning outcomes*. In online learning environments, fixed task sequences can be pre-assigned to a learner by the teacher (or the computer software), but it can also be up to the learner to decide what tasks they want to work on and in what order. As most research on example-based learning has investigated effects of fixed task sequences, an important open question is what choices learners make when they can choose how to learn from examples and problems themselves (e.g., Tempelaar et al., 2020; Van Gog et al., 2019), and how well such choices align with what we know to be effective, efficient, or motivating for learning (see Table 1). When learners would make suboptimal choices, they might benefit from instructional support or advice during self-regulated learning from examples and practice problems.

While providing learners with control over their task-selections might have motivational benefits for learning, which in turn may increase learning outcomes (e.g., Zimmerman, 2002), one could expect that they would struggle selecting the right task according to their learning needs. Self-regulated learning of problem-solving tasks is notoriously difficult, because learners not only need to assess their performance on the current task, but also use this information to select the next task with the right level of task complexity and support (e.g., De Bruin & Van Gog, 2012). Hence, it may not come as a surprise that research has shown that learners often experience problems with accurately assessing their own knowledge gaps and determining which (order of) tasks will help them improve (e.g., Corbalan et al., 2008; Kostons et al., 2012). Learners, especially novices, tend to focus more on irrelevant task aspects when selecting their own tasks such as the cover story of the task than task aspects that contribute to learning (i.e., level of support and complexity; e.g., Corbalan et al., 2009).

These findings raise the question of how and how well students regulate their learning when acquiring problem-solving skills. A recent study conducted by Foster and colleagues (2018) suggested that students made rather suboptimal decisions when learning from examples and problems (i.e., compared to what we know to be effective sequences). For instance, we know that for novices, studying examples, and particularly at the start of the learning phase, is effective and efficient, but their findings showed that on average, students opted more often for (completion) problems than examples and rarely started the learning phase with example study. However, given the paucity of studies, more research on self-regulated learning with examples and practice problems is needed, and especially in the context of actual study programs with content that is relevant to students.

Another important question is how we can support students' self-regulated learning from examples and problems. One approach that might work well and is quite easy to implement in practice is to explicitly inform learners about effective and efficient principles derived from instructional design research (see Table 1). It has been suggested that explicitly informing learners might help them to become more aware of the value of certain strategies and to increase their metacognitive knowledge (i.e., knowledge about why and which strategies are beneficial for learning and which strategies are not). Consequently, this can increase the likelihood that strategies are applied (e.g., Tullis et al., 2003; Yan et al., 2014). Indeed, explicitly informing or instructing students about effective (meta)cognitive learning strategies has shown to be successful for increasing learners' metacognitive beliefs or knowledge e.g., Endres et al., 2021; Lineweaver et al., 2011; McCabe, 2011; Yan et al., 2016) and their use of those learning strategies (e.g., Biwer et al., 2019). Ariel and Karpicke (2017) found that explicitly informing students about learning strategies (i.e., retrieval practice) also improved their learning outcomes regarding word-pair learning. When learning from examples and problems, it could mean that explicitly informing students about findings from sequencing studies, would help learners to select the right tasks (according to their level of expertise) faster and more often. However, it is an open question whether this approach would generalize to problem-solving skills, because self-regulated learning of problem-solving skills requires much more complex regulation decisions (e.g., De Bruin & Van Gog, 2012) than self-regulated learning of verbal learning tasks or text comprehension (as often used in research on learning strategies).

Therefore, the studies in the second part of this dissertation (**Chapters 4 and 5**) address the questions of 1) how and how well learners regulate their learning from examples and problems, and 2) whether informing learners about effective, efficient, and motivating instructional design principles helps to improve their task-selections, and thereby their motivation and learning outcomes.

Table 1.
Effective, Efficient, and Motivating Principles Derived from Instructional Design Research on Learning New Problem-Solving Skills.

Principle	Explanation	References
Example-based-learning-principle	Replacing all or a substantial number of practice problems with examples helps novices to learn more (i.e., is more effective) with less time and effort investment (i.e., is more efficient) than solving practice problems without any instructional support	e.g., Sweller et al. (2011), Van Gog et al. (2019)
Example-study-first-principle	Novices should start the learning phase with an example instead of a practice problem, as this was found to be more efficient, than starting with problem-solving only	e.g., Van Gog et al. (2011)
Lowest-level-first-principle	Novices should start with a task at the lowest level of complexity	Van Merriënboer (1997), Van Merriënboer & Kirschner (2013)
Simple-to-complex-principle	Novices should gradually increase the level of task complexity as their knowledge increases	Van Merriënboer (1997), Van Merriënboer & Kirschner (2013)
Start-each-level-with-example-principle	Novices should receive a high level of instructional support (like an example) at the start of each new complexity level	Van Merriënboer (1997), Van Merriënboer & Kirschner (2013)

1.4 Overview of the Studies in this Dissertation

This dissertation is divided into two parts. Part I, *Sequencing Example Study and Practice Problem Solving*, consists of two chapters (**Chapters 2 and 3**), each presenting two experimental studies in which it was investigated what sequences of examples and practice problems are most effective, efficient, and motivating for first year higher education students' learning of new mathematical problem-solving skills. Part II, *Self-Regulated Learning with Examples and Practice Problems*, contains two chapters, each presenting a study in which it was examined how and how well first year higher education students regulate their learning from examples and practice problems in an online learning environment (**Chapters 4 and 5**), and whether informing them about effective, efficient, and motivating instructional design principles helps to improve their task-selections, and thereby their motivation and learning outcomes (**Chapter 5**).

PART I: Sequencing Example Study and Practice Problem Solving

Chapter 2 contains two experiments investigating whether different short sequences of examples and practice problems (i.e., 4 learning tasks) differ in effectiveness, efficiency, and how they affect motivational aspects of learning. In Experiment 1 ($N = 124$), technical higher education students learned how to approximate the region under a graph using the trapezoidal rule (a math task) by means of example study only, example-problem pairs, problem-example pairs, or problem solving only. Subsequently, we conducted a second experiment in order to examine whether results would replicate with a different sample (i.e., students with a non-technical background). Therefore, Experiment 2 ($N = 81$) used the same materials and design as Experiment 1, but with a sample of primary teacher training students. Effectiveness was measured by assessing performance on the isomorphic test tasks, procedural transfer task, and conceptual transfer task. Efficiency was measured by logging time-on-task and rating invested mental effort after each task in the learning and posttest phase. Motivation was measured by means of short self-efficacy, perceived competence, and topic interest questionnaires, provided to learners before and after the learning phase.

In **Chapter 3**, two experiments are described examining whether different short (i.e., 4 tasks) and longer (i.e., 8 tasks) sequences of example study only, example-problem pairs, problem-example pairs, and practice problem-solving practice show differences in effectiveness, efficiency, and motivation. In Experiment 1 ($N = 157$), it was examined whether the results of Experiment 1 described in **Chapter 2** would replicate with a conceptual pretest instead of a procedural pretest and whether the effects remain stable on a delayed test one week later. Moreover, it was investigated how self-efficacy *during* learning is affected by the order of examples and practice problems. Technical higher education students learned how to approximate the region under a graph using the trapezoidal rule in a short learning phase consisting of four learning tasks. In Experiment 2 ($N = 105$), it was investigated whether the results found with short

sequences would be different with longer sequences. Experiment 2 used the same materials and design as Experiment 1; however, students studied or solved eight learning tasks. Effectiveness was measured by assessing performance on the isomorphic test tasks, procedural transfer task, and conceptual transfer task. Efficiency was measured by logging time-on-task and rating invested mental effort after each task in the learning and (delayed) posttest phase. Motivation was measured by means of short self-efficacy, perceived competence, and topic interest questionnaires, provided to learners before and after the learning phase and at the start of the delayed posttest. Self-efficacy was also measured after each task in the learning phase.

PART II: Self-Regulated Learning with Examples and Practice Problems

The first aim of the study presented in **Chapter 4** was to explore what choices students make (and why) when they can regulate their own learning from different examples and practice problems (at different complexity levels). A second aim was to explore to what extent their task-selection decisions match with principles for effective, efficient, and motivating learning sequences derived from instructional design research. Finally, this study examined whether there is a relation between the extent to which students' choices match with these instructional design principles and their scores on cognitive (isomorphic test tasks, procedural transfer task, and conceptual questions, mental effort, and time-on-task) and motivational variables (self-efficacy, perceived competence, and topic interest). Technical higher education students ($N = 147$) learned how to solve problems on the trapezoidal rule in an online learning environment by selecting six learning tasks from a database with 45 tasks that varied in task format (video examples, worked examples, and practice problems), complexity level (three levels), and cover story. Effectiveness was measured by assessing performance on the isomorphic test tasks, procedural transfer task, and conceptual questions. Efficiency was measured by logging time-on-task and rating invested mental effort after each task in the learning and posttest phase. Motivation was measured by means of short self-efficacy, perceived competence, and topic interest questionnaires before and after the learning phase. Self-efficacy was also measured after each task in the learning phase.

In **Chapter 5**, a study is described in which the question was first addressed how students regulate their learning when they can decide for themselves how to sequence examples and practice problems (i.e., to replicate the study in **Chapter 4**). Second, it was investigated whether self-regulated learning would be as effective, efficient, and motivating as a fixed task sequence based on the principles derived from instructional design research. Third, this study examined whether explicitly informing learners about the principles derived from instructional design research would enhance their self-regulated learning of examples and problems (at different complexity levels), performance, and motivation compared to self-regulated learning without such information, and whether this would be as effective for learning as engaging in a fixed

task sequence. Technical higher education students ($N = 150$) learned how to use the trapezoidal rule by engaging in a fixed task sequences condition, a self-regulated learning condition, or an 'informed' self-regulated learning condition. In the self-regulated learning conditions, students selected six learning tasks from the task database (cf. Chapter 4). Before selecting their own learning tasks, students in the 'informed' condition received a video instruction explaining effective and efficient (and motivating) principles from instructional design research. Effectiveness was measured by assessing performance on the isomorphic test tasks, procedural transfer task, and conceptual questions. Efficiency was measured by logging time-on-task and rating invested mental effort after each task in the learning and posttest phase. Motivation was measured by means of short self-efficacy and perceived competence questionnaires before and after the learning phase. Self-efficacy was also measured after each task in the learning phase.

The final chapter (**Chapter 6**) presents a summary of the main findings and a discussion of the theoretical and practical implications.

Chapter 2

Effects of different sequences of examples and problems on motivation and learning



This chapter was published as:

Van Harsel, M., Hoogerheide, V., Verkoeijen, P. P. J. L., & Van Gog, T. (2019). Effects of different sequences of examples and problems on motivation and learning. *Contemporary Educational Psychology, 58*, 260–275. <https://doi.org/10.1016/j.cedpsych.2019.03.005>

Acknowledgement of author contributions:

MvH, PV, and TvG designed the study, MvH recruited participants and collected the data, MvH analyzed the data, VH checked the data package, MvH drafted the manuscript, all authors contributed to critical revision of the manuscript, VH, PV, and TvG supervised the study.

Abstract

Recent research has shown that example study only (EE) and example-problem pairs (EP) were more effective (i.e., higher test performance) and efficient (i.e., attained with less effort invested in learning and/or test tasks) than problem-example pairs (PE) and problem solving only (PP). We conducted two experiments to investigate how different example and problem-solving sequences would affect motivational (i.e., self-efficacy, perceived competence, and topic interest) and cognitive (i.e., effectiveness and efficiency) aspects of learning. In Experiment 1, 124 technical students learned a mathematical task with the help of EEEE, EPEP, PEPE, or PPPP and then completed a posttest. Students in the EEEE Condition showed higher posttest performance, self-efficacy, and perceived competence, attained with less effort investment, than students in the EPEP and PPPP Condition. Surprisingly, there were no differences between the EPEP and PEPE Condition on any of the outcome measures. We hypothesized that, because the tasks were relevant for technical students, starting with a problem might not have negatively affected their motivation. Therefore, we replicated the experiment with a different sample of 81 teacher training students. Experiment 2 showed an efficiency benefit of EEEE over EPEP, PEPE, and PPPP. However, only EEEE resulted in greater posttest performance, self-efficacy, and perceived competence than PPPP. We again did not find any differences between the EPEP and PEPE Condition. These results suggest that, at least when short training phases are used, studying examples (only) is more preferable than problem solving only for learning. Moreover, this study showed that example study (only) also enhances motivational aspects of learning whereas problem solving only does not positively affect students' motivation at all.

Keywords: example-based learning, worked examples, problem solving, motivation, mental effort, instructional design

2.1 Introduction

Example-based learning is an effective and efficient instructional strategy for novices to acquire new problem-solving skills. Research has repeatedly shown that instruction that relies more heavily on example study, yields better learning outcomes than engaging in practice problem solving only (for reviews, see Atkinson et al., 2000; Renkl, 2014; Sweller et al., 2011; Sweller et al., 1998; Van Gog & Rummel, 2010). This is known as the worked example effect. Notwithstanding several decades of research, an important open question in research on example-based learning is whether and when example study should be alternated with practice problem solving to be effective and efficient for learning.

2.1.1. Different Sequences of Examples and Problems

Historically, most studies on the worked example effect have used *example-problem pairs* (e.g., Carroll, 1994; Cooper & Sweller, 1987; Kalyuga et al., 2001; McLaren et al., 2008; Mwangi & Sweller, 1998; Rourke & Sweller, 2009; Sweller & Cooper, 1985). Others used *example study only* (e.g., Van Gerven et al., 2002; Van Gog et al., 2006). Both approaches were found to be more effective and efficient for learning and transfer than *problem solving only*. Another means of implementing examples and problems is to use *problem-example pairs* (e.g., Hausmann et al., 2008; Reisslein et al., 2006; Stark et al., 2000). In a direct comparison of all four approaches of Van Gog and colleagues (2011), students were randomly assigned to learn how to troubleshoot electrical circuits (in four training tasks) by means of example study only (EEEE), example-problem pairs (EPEP), problem-example pairs (PEPE), or practice problem solving only (PPPP). Time-on-task was kept constant. Results showed no differences in test performance or mental effort investment in the training phase between the EEEE condition and EPEP condition and between the PEPE condition and PPPP condition. The EEEE condition and EPEP condition were, however, more effective (i.e., attained significantly higher test performance; medium to large effect) and more efficient (i.e., attained significantly higher test performance with less invested mental effort in the training phase; medium to large effect) compared to the PEPE condition and PPPP condition (for a discussion of efficiency, see Hoffman & Schraw, 2010; Van Gog & Paas, 2008).

2.1.2. Example Study Only versus Example-Problem Pairs

Several studies have by now replicated the finding that learning outcomes after EEEE and EPEP do not differ significantly (Van der Meij et al., 2018), even on a delayed posttest (Leahy et al., 2015; Van Gog & Kester, 2012; Van Gog et al., 2015). Note though, that these studies did not include motivational variables (which have largely been ignored in worked example research; Renkl, 2014; Sweller et al., 2011; Van Gog & Rummel, 2010). It has been suggested –though not yet tested– that solving a (similar)

problem immediately after studying an example may be more motivating for students than passively studying examples only, because practice problem solving requires learners to actively apply what they have learned (e.g., Sweller & Cooper, 1985; Trafton & Reiser, 1993). If EPEP was found to be more motivating than EEEE whilst yielding comparable levels of learning outcomes, it would be highly relevant for educational practice. Outside a laboratory research setting, motivational variables might affect learning outcomes via persistence. That is, in (online) learning environments, students can decide for themselves whether they continue to work on a task (sequence) or not, so how motivating a task sequence is becomes important.

2.1.3. Example-Problem Pairs versus Problem-Example Pairs

Another noteworthy finding in the Van Gog et al. (2011) study in which motivational aspects of learning might have played a role, was that EPEP was more effective and efficient for learning than PEPE, even though both received the same number of examples to study. Moreover, students in the PEPE condition –despite receiving two examples– did not outperform students in the PPPP condition. This finding, which has since been replicated in two other studies (e.g., Kant et al., 2017; Leppink et al., 2014¹), suggests that the order in which example study and practice problem solving is alternated, matters: if novice learners start with a practice problem, example study loses its effectiveness (see also Reisslein et al., 2006). Van Gog et al. (2011) suggested –but did not test– that motivational aspects of learning might explain this finding: “students may not be motivated to study the example because of the negative experience of a failed problem-solving attempt” (p. 217). That is, when novices have to learn how to solve a complex task that requires domain-specific knowledge and is not particularly intrinsically rewarding (such as the physics task in the study by Van Gog et al., 2011), then starting the training phase with a practice problem (i.e., PEPE) might lead to a decrease in student motivation. When the practice problem is being experienced as so difficult that students lose confidence in their own abilities or lose interest to learn the task, they may not be motivated to study the subsequent example (and possibly also the tasks that follow). Starting with an example (EPEP) gives students a basis for how to approach the subsequent practice problem and may therefore prevent students from becoming demotivated.

2.1.4. The Role of Motivation

Three aspects of motivation that may be affected by example-problem sequences are self-efficacy, perceived competence, and topic interest.

¹ Note though, that EPEP > PEPE was not found when the examples and problems remained fully identical throughout the sequences (Van Gog, 2011).

Self-efficacy is a key construct in Bandura’s (1986) social learning theory and can be defined as a person’s belief in their own capacity to organize or accomplish a specific task or challenge (see also Bandura, 1997; Schunk, 1987). Self-efficacy has been shown to have a positive effect on factors such as academic motivation, study behavior, and learning outcomes (e.g., Bandura, 1997; Bong & Skaalvik, 2003; Schunk, 2001). Perceived competence plays a central role in Deci and Ryan’s (2002) self-determination theory of motivation and has also shown to have significant influence on academic motivation and learning outcomes (e.g., Bong & Skaalvik, 2003). This construct is related to self-efficacy but covers more general knowledge and perceptions (Bong & Skaalvik, 2003; Hughes et al., 2011; Klassen & Usher, 2010). Topic interest is a motivational construct that can be described as the level of interest generated by a specific topic (Ainley et al., 2002; Renninger, 2000; Schiefele & Krapp, 1996) and seems to have a positive effect on (deeper) learning and engagement (e.g., Benton et al., 1995; Schiefele & Krapp, 1996; Tobias, 1996). Although research has shown that example study only can foster students’ self-efficacy and perceived competence (e.g., Bandura, 1997; Crippen et al., 2009; Hoogerheide et al., 2014; Hoogerheide et al., 2018), it is an open question how different sequences would affect motivational aspects of learning.

In sum, it is both theoretically and practically relevant to address whether different sequences of examples and practice problems would differentially affect not only cognitive (i.e., effectiveness and efficiency), but also motivational aspects of learning (i.e., self-efficacy, perceived competence, and topic interest). The present study set out to do so, by performing a conceptual replication of Van Gog et al. (2011), extended with motivational measures.

2.1.5. The Present Study

The purpose of Experiment 1 was to investigate the effects of different example and problem-solving sequences on motivational and cognitive aspects of learning. We conducted a conceptual replication of the Van Gog et al. (2011) study, with the same task sequences (i.e., EEEE, EPEP, PEPE, and PPPP), but a different population (i.e., higher education students rather than secondary education students), different training tasks (i.e., mathematics tasks rather than physics tasks), and a different example format (i.e., video modeling examples consisting of screen recordings with voice-over, rather than worked examples; cf. McLaren et al., 2008). In addition to performance on posttest tasks and reported effort investment in the training phase, we measured time-on-task in the training phase, as well as mental effort and time-on-task in the posttest phase as (explorative) indicators of efficiency of the learning process and learning outcomes. We added a procedural and conceptual transfer task and a delayed posttest one week later to investigate any effects on transfer and whether effects would remain stable over time (cf. Van Gog et al., 2015). The most important novelty of Experiment 1 was that we measured the following motivational aspects of learning before and after the training phase: self-efficacy, perceived competence, and topic interest.

The main aim was to investigate how the different example and problem solving sequences (i.e., EEEE, EPEP, PEPE, and PPPP) would affect a) motivational aspects of learning (i.e., self-efficacy, perceived competence, and topic interest), and b) cognitive aspects of learning (i.e., effectiveness and efficiency). We expected that an EPEP sequence would result in higher levels of self-efficacy, perceived competence, and topic interest than an EEEE sequence (cf. the suggestion by Sweller & Cooper, 1985 and & Trafton & Reiser, 1993; Hypothesis 1a). Based on the motivational explanation for the effectiveness and efficiency of EPEP over PEPE proposed by Van Gog et al. (2011), we also expected EPEP to be more beneficial for self-efficacy, perceived competence, and topic interest than PEPE (Hypothesis 1b). We had no hypotheses for the other condition comparisons, so we examined them in an exploratory manner (**Question 1c**).

Regarding cognitive aspects of learning, we expected to replicate the findings by Van Gog et al. (2011) regarding both isomorphic problem-solving performance (i.e., EEEE = EPEP > PEPE = PPPP; Hypothesis 2) and mental effort invested in the training phase (i.e., EEEE = EPEP > PEPE = PPPP; Hypothesis 3). That is, we expected the EEEE and EPEP condition to attain greater posttest performance with less effort investment in the training phase than the PEPE and PPPP condition, and no differences to arise on these variables between the EEEE and EPEP condition and between the PEPE and PPPP condition. Because example-based learning has been found to be effective not only for learning to solve similar problems, but also for solving transfer problems (e.g., Cooper & Sweller 1987; Paas 1992; Paas & Van Merriënboer, 1994), we expected the same pattern of results for performance on the procedural transfer task (i.e., EEEE = EPEP > PEPE = PPPP; Hypothesis 4) and conceptual transfer task (i.e., EEEE = EPEP > PEPE = PPPP; Hypothesis 5).

2.2. Experiment 1

2.2.1. Method

2.2.1.1. Participants and design

An a-priori-power analysis was conducted to determine how many participants we would need to be able to reliably detect the effect sizes reported by Van Gog et al. (2011). Inserting $\eta^2_p = 0.23$ (i.e., effect size for test performance found in the study by Van Gog et al., 2011) into G*Power and performing an a-priori-power analysis for a one-way ANOVA with four groups, with an alpha of 0.05, and a power of 0.95, yielded a total sample of 64. Participants were 124 first-year students from a Dutch University of Applied Sciences, enrolled in an electrical and electronic or mechanical engineering program ($M^{age} = 19.25$, $SD = 1.90$; 117 male, 7 female). At the time of the experiment, students were novices to the task being taught in this study (i.e., approximating the definite integral using the trapezoidal rule) as this topic had not yet been taught in their curriculum. They received study credits for their participation. The experiment consisted of 3 phases, namely: the pretest, training, and immediate posttest phase. Participants

were randomly assigned to one of four conditions: 1) examples only ($n = 34$; EEEE), 2) example-problem pairs ($n = 25$; EPEP), 3) problem-example pairs ($n = 30$; PEPE), or 4) practice problems only ($n = 35$; PPPP).

2.2.1.2. Materials

All materials were presented in a web-based online learning environment.

Training tasks. The training phase consisted of four math tasks that were developed in collaboration with three mathematics teachers of a Dutch University of Applied Sciences. The tasks required the use of the trapezoidal rule (i.e., a numerical integration method which divides a specific region under the graph of a function into trapezoids and calculates its area) to approximate the region under the graph of a function.

Each task had a different cover story (i.e., task 1: fitness, task 2: energy measurement, task 3: soapsuds, and task 4: running). These cover stories were randomly distributed over the four tasks that were used in the training phase. The four tasks were divided in two pairs (i.e., pair 1: fitness and energy measurement, pair 2: soapsuds and running), based on their complexity level. In the first pair of tasks (complexity level 1), only positive numbers were used in constructing the graph of a function, whereas in the second pair (complexity level 2), negative numbers were used in constructing the graph of a function. Requiring students to calculate with negative numbers made the second pair of tasks slightly more complex than the first pair of tasks. Within each pair, the two tasks were isomorphic (i.e., a similar problem-solving procedure was required, but surface features such as the cover stories and numbers used in functions were slightly different).

Two versions of each task were created, a video modeling example and a practice problem. The video modeling example, a video screen capture, showed a digital recording of a female model's computer screen demonstrating step-by-step how to solve a problem using the trapezoidal rule. The visual demonstration was supported by verbal explanations and handwritten notes. The screen capture started with a brief introduction of the purpose of the trapezoidal rule, followed by an explanation of the problem state. For example, the problem state in the example format of 'Energy measurement' read as follows: "Jalil has bought a solar cell and wants to know how much energy the solar cell supplies during a certain amount of time. Jalil has used an energy meter to examine how much energy the solar cell produces during a specific amount of time. Jalil has measured the energy at different time points and plotted the results in a graph. The time (in minutes) is plotted on the horizontal axis and the power the solar cell supplies (Joule per minute) is plotted on the vertical axis of the graph. By calculating the area under the graph, Jalil can determine how much energy the solar cell has produced during a certain amount of time." Subsequently, the remainder of the example showed and explained how to interpret the corresponding graph of a function with information that was given (i.e., the left border and right border of the area, the number of intervals, the trapezoidal rule), and showed and explained how to

solve the problem by using four steps: 1) 'compute the step size of each subinterval', 2) 'calculate the x-values', 3) 'calculate the function values for all x-values', and 4) 'enter the function values into the formula and calculate the area' (for an example of a video modeling example, see Supplementary Materials B).

In the problem format, participants first received a short introduction describing the problem state, along with the graph of a function, the left border and right border of the area to be calculated, the number of intervals, and the trapezoidal rule formula. This information was the same as in the video modeling example. It was, however, not explained how to use this information to solve the problem. Participants had to solve the problem themselves by completing four steps: 1) 'compute the step size of each subinterval', 2) 'calculate the x-values', 3) 'calculate the function values for all x-values', and 4) 'enter the function values into the formula and calculate the area'. In addition, they were asked to write down their solution steps (for an example of a practice problem, see Supplementary Materials A).

The order of the four tasks was kept constant across conditions (i.e., the training phase always started with 'fitness' and ended with 'running'), only the format of each task varied among conditions (i.e., EEEE and EPEP started with 'fitness' as an example, whereas PEPE and PPPP started with 'fitness' as a practice problem).

Test tasks. The pretest consisted of two tasks that were isomorphic (i.e., same difficulty, different cover stories) to the training tasks (Cronbach's $\alpha = .63$). The complexity level of the first and second pretest tasks was identical to the first and second pair of training tasks, respectively. Students were asked to approach the region under the graph using the information that was given and to write down their solution steps. In these problems, the intermediate (four) steps were not explicitly displayed such as in the training tasks.

The immediate posttest consisted of four tasks. The first two tasks were isomorphic to the pretest and training tasks and were used to measure 'learning' (Cronbach's $\alpha = .66$). Students needed to apply the exact problem-solving procedure that they learned during the training phase, but tasks were used that differed in terms of surface features such as cover story and numbers used in the function. The third task was a procedural transfer problem in which participants were asked to use the Simpson rule instead of the trapezoidal rule to approximate the definite integral under a graph. The problem-solving procedure of the Simpson rule is similar to that of the problem-solving procedure of the trapezoidal rule. However, Simpson's rule uses a different formula to calculate the area under a graph and approximates the curve with a sequence of quadratic parabolic segments instead of straight lines (such as the trapezoidal rule). The final conceptual transfer task consisted of five questions that aimed to measure students' conceptual understanding of the underlying principles of the trapezoidal rule as a technique to approximate the area under a graph (Cronbach's $\alpha = .54$).

Each question consisted of a multiple-choice part (from which students had to choose the right answer) and an 'explanation' part (where students had to substantiate their chosen answer). An example of an isomorphic posttest task, procedural transfer task and a question concerning conceptual transfer can be found in the Supplementary Materials F, G, and H.

Mental effort. Participants were asked to rate how much mental effort they had invested after each task on the pretest, the training phase, and the immediate posttest, using the 9-point mental effort rating scale developed by Paas (1992), with answer options ranging from (1) "very, very low mental effort" to (9) "very, very high mental effort". Research has shown that this measure is an indicator of experienced cognitive load that is sensitive to variations in task complexity (Paas et al., 2003).

Self-efficacy, perceived competence, and topic interest. Self-efficacy was measured just before and directly after the training phase by asking participants to rate to what extent they were confident that they could approximate the definite integral of a graph using the trapezoidal rule on a 9-point rating scale, ranging from (1) "very, very unconfident" to (9) "very, very confident". This was an adapted version of the item used by Hoogerheide and colleagues (2016).

Perceived competence was measured using an adapted version of the *Perceived Competence Scale for Learning* (Williams & Deci, 1996; Williams, et al., 1988). The original scale consists of 4 items, namely: "I feel confident in my ability to learn this material", "I am capable of learning the material in this course", "I am able to achieve my goals in this course", and "I feel able to meet the challenge of performing well in this course". We adapted the scale by removing the third question on the topic of personal goals because this question was not relevant for the present study. For the remaining questions, we rephrased the word 'course' to focus on approximating the definite integral of a graph using the trapezoidal rule. Participants could rate from (1) "not at all true" to (7) "very true" to what degree the items applied to them. The adapted scale had a good reliability in our sample (Cronbach's $\alpha = .96$). It should be noted, however, that Cronbach's α has a limited degree of precision due to the large sampling error. Nevertheless, a high α measure (i.e., above .80) has been demonstrated in previous studies (e.g., Williams & Deci, 1996; Williams et al., 1998).

To measure students' topic interest, we developed a topic interest scale. Our scale consisted of 7 items (Cronbach's $\alpha = .80$) adapted from the topic interest scale by Mason and colleagues (2008) and from the Perceived Interest Scale developed by Schraw and colleagues (1995). We selected the items from both scales that focused on feelings and emotions towards a specific topic and adjusted the items to the context of using or practicing the trapezoidal rule. Each item asked participants to rate on a 7-point scale, ranging from 1 (not at all) to 7 (very true), to what degree each of the items applied to them. The items are shown in the Supplementary Materials I.

2.2.1.3. Procedure

The experiment was run in sixteen sessions (i.e., eight first sessions and eight second sessions) in a computer classroom with 5 to 25 participants present per session. In the first session, which lasted 100 minutes on average, the experimenter first provided participants with a general introduction in which she explained the aim and procedure of the experiment. Participants were told that they would be able to work on the tasks at their own pace with a max. of 130 minutes, and that they would be provided with a headset, a pen, and scrap paper on which they could write down their calculations. Participants were instructed to do their best but could write down an 'X' if they really did not know the answer. Then, the experimenter provided students a form with a link to the learning environment so students could enter the environment as soon as the instruction was finished.

The learning environment presented each task and questionnaire on a separate page, which ensured that participants could not go back to previous tasks or questionnaires, nor look ahead until the current task was completed. Time-on-task was logged. The learning environment first randomly assigned participants to one of the four conditions, and then presented participants with a short demographic questionnaire (e.g., age and gender) and the pretest. Both pretest tasks were followed by the mental effort rating scale, and the pretest was followed by the self-efficacy, perceived competence, and topic interest questionnaires. During the training phase, participants were provided with a combination of examples and/or practice problems, depending on their assigned condition. Each task was followed by the mental effort rating scale, and the training phase was followed by the self-efficacy, perceived competence, and topic interest questionnaires. Finally, participants were provided with the immediate posttest, which consisted of two tasks that were isomorphic to the tasks in first and second pair of the training phase, a procedural transfer task (i.e., Simpson's rule task) and five open-ended questions to measure conceptual knowledge. Again, after each immediate posttest task participants were provided with the mental effort rating scale. Before starting with the immediate posttest, students were asked to put away the scrap paper they used in the training phase and received a new scrap paper to make notes. After completing the immediate posttest, participants gave the scrap paper containing their calculations to the experimenter.

2.2.1.4. Data analysis.

For each training task, a maximum of 8 points could be earned: two points for correctly computing the step size of each subinterval (step 1), two points for correctly calculating all the x-values (step 2), two points for correctly calculating the function values for all x-values (step 3), and 2 points for correctly calculating the area by using the correct formula and giving the right answer (step 4). Students received 1 point when half or more of the solution steps were correct in step two, three, and four. If fewer than half of the solution steps were correct, 0 points were granted. These scoring standards were also used to score the pretest (max. 16 points). The same procedure was used for the

procedural transfer problem, so a maximum of 8 points could be earned for this task. A maximum of 9 points could be earned on the 5 open-ended questions in the conceptual transfer problem: one point for the first open-ended question (1 point for the correct answer, 0 points for an incorrect answer) and 2 points for the other open-ended questions. The maximum score of 2 points was only granted when participants got the answer right and provided correct reasoning. Only 1 point was granted if the answer was correct but not substantiated by reasoning and 0 points were granted when the answer was completely incorrect.

The data was scored by the experimenter (i.e., first author) based on a standard developed by the authors in collaboration with the mathematic teachers. To measure the reliability of the ratings, two raters independently scored 15% of the tests. The intra-class correlation coefficient was high, with respectively scores of .91 on the pretest tasks, .94 on the training phase tasks, and .98 on the posttest tasks.

Average mental effort was computed separately for the pretest tasks, training tasks, isomorph tasks, and transfer tasks on the immediate posttest. Average scores on self-efficacy, perceived competence, and topic interest were computed separately for the measurement that took place before the training phase and the measurement directly after the training phase.

2.2.2. Results

Because several variables were not normally distributed, we analyzed the data with nonparametric tests (cf. Field, 2009). We tested the main effects of Test Moment with the Wilcoxon signed-rank test and the main effects of Instruction Condition with the Kruskal-Wallis test. For post-hoc tests, we used Mann-Whitney U tests, with a Bonferroni corrected significance level of $p < .013$ (i.e., $0.05/4$) for the Wilcoxon signed-rank tests and a Bonferroni corrected alpha level of $p < .008$ (i.e., $0.05/6$) for the Kruskal-Wallis test. For the post-hoc tests, the effect size of Pearson r correlation is reported (i.e., Z/\sqrt{N}) with values of 0.10, 0.30, and 0.50 representing a small, medium, and large effect size, respectively (Cohen, 1988). Relevant descriptive statistics of self-efficacy, perceived competence, and topic interest scores are presented in Table 2.1, and performance scores, mental effort scores, and time-on-task scores are presented in Table 2.2.

Unfortunately, the delayed posttest data had to be excluded from the analyses. We had made a mistake in designing the delayed posttest, as the delayed posttest was not entirely isomorphic to the immediate posttest. The complexity level of the tasks used in the delayed posttest did not correspond to the complexity level of the tasks used in the immediate posttest (i.e., the tasks used in the delayed posttest were less complex than the tasks used in the immediate posttest because students did not have to calculate with fractions or negative numbers). Please see the Supplementary Materials J for the raw data of the delayed posttest (i.e., means, standard deviations, and medians per condition).

We first checked for prior knowledge differences among conditions. A Kruskal-Wallis test showed no significant differences among conditions on pretest performance, $H(3) = 3.54$, $p = .315$, and on pretest scores of self-efficacy, $H(3) = 2.36$, $p = .501$, perceived competence, $H(3) = 2.47$, $p = .480$, and topic interest, $H(3) = 6.68$, $p = .083$.

2.2.2.1. Does the sequencing of examples and problems affect self-efficacy, perceived competence, and topic interest?

Self-efficacy. Firstly, we analyzed whether students' self-efficacy increased from before to after the training phase. We found a main effect of Test Moment, $Z = 5.79$, $p < .001$, $r = .520$. Follow-up tests showed that the self-efficacy medians of the EEEE ($Z = 4.57$, $p < .001$, $r = .834$), EPEP ($Z = 3.00$, $p = .003$, $r = .514$), and PEPE Condition ($Z = 2.91$, $p = .004$, $r = .582$) significantly increased over time, whereas the medians of the PPPP Condition did not significantly increase over time ($p = .821$, $r = .038$). Regarding the main question of whether there would be differences among instructional conditions on reported self-efficacy measured after the training phase, we found a main effect of Instruction Condition, $H(3) = 43.46$, $p < .001$. Our findings were not in line with Hypothesis 1a (EPEP > EEEE) and Hypothesis 1b (EPEP > PEPE). Post-hoc tests showed that self-efficacy ratings did not differ between the EPEP and PEPE Condition ($p = .094$, $r = .218$) and that self-efficacy was even significantly higher in the EEEE Condition than in the EPEP Condition ($U = 293.50$, $p = .003$, $r = .375$). Further explorations showed that self-efficacy was significantly higher in the EEEE ($U = 66$, $p < .001$, $r = .759$), EPEP ($U = 287$, $p < .001$, $r = .450$), and PEPE Condition ($U = 144$, $p < .001$, $r = .573$) than in the PPPP Condition.

Perceived Competence. Perceived competence showed the same pattern of results. We found a main effect of Test Moment, $Z = 6.03$, $p < .001$, $r = .542$. Perceived competence increased in the EEEE ($Z = 4.48$, $p < .001$, $r = .818$), EPEP ($Z = 3.23$, $p = .001$, $r = .554$), and PEPE Condition ($Z = 3.23$, $p = .001$, $r = .646$), but not in the PPPP Condition ($p = .455$, $r = .133$). We also found a main effect of Instruction Condition on perceived competence measured after the training phase, $H(3) = 38.76$, $p < .001$. In contrast to our expectations (i.e., Hypothesis 1a: EPEP > EEEE; Hypothesis 1b: EPEP > PEPE), perceived competence was significantly higher in the EEEE Condition compared to the EPEP Condition ($U = 315.50$, $p = .008$, $r = .331$) and there was no significant difference between the EPEP and PEPE Condition ($p = .042$, $r = .264$). Further explorations showed that perceived competence scores were significantly higher in the EEEE ($U = 100.50$, $p < .001$, $r = .697$), EPEP ($U = 299.50$, $p = .001$, $r = .428$), and PEPE Condition ($U = 142.50$, $p = .001$, $r = .573$) compared to the PPPP Condition.

Topic Interest As for topic interest measured after the training phase, we found a main effect of Test Moment, $Z = -3.62$, $p < .001$, $r = .325$. Students' topic interest significantly decreased over time in the EPEP ($Z = -3.23$, $p = .001$, $r = .554$) and PPPP Condition ($Z = -2.69$, $p = .007$, $r = .455$). There was no main effect of Instruction Condition on topic interest measured after the training phase ($p = .143$), indicating that there

were no differences among conditions on topic interest. Hence, the topic interest results contrasted Hypothesis 1a (EPEP > EEEE) and Hypothesis 1b (EPEP > PEPE).

2.2.2.2. Does the sequencing of examples and problems affect learning and transfer?

Isomorphic Tasks. When analyzing whether performance on the test tasks isomorphic to the training phase improved significantly from pretest to posttest, we found a main effect of Test Moment, $Z = 3.86$, $p < .001$, $r = .311$. Numerically, performance increased over time in all example conditions (see Table 2.1), but follow-up tests showed that only the EEEE Condition performed significantly better on the posttest than on the pretest ($Z = 3.02$, $p = .003$, $r = .551$). The other conditions did not show a significant increase (EPEP: $p = .061$, $r = .321$; PEPE: $p = .047$, $r = .397$; PPPP: $p = .029$).

To answer our second main question of whether there would be differences among instructional conditions on learning, we analyzed whether there were any differences among conditions regarding isomorphic posttest performance. We found a main effect of Instruction Condition, $H(3) = 20.63$, $p < .001$. In line with our expectations (i.e., Hypothesis 2: EEEE = EPEP > PEPE = PPPP), we found that participants in the EEEE Condition scored significantly higher than those in the PPPP Condition ($U = 189$, $p < .001$, $r = .551$). The results of other post-hoc comparisons were not in line with our expectations, however, because we found no significant differences between the EEEE and PEPE Condition ($p = .469$, $r = .097$), between the EPEP and PPPP Condition ($p = .218$, $r = .148$), and between the EPEP and PEPE Condition ($p = .131$, $r = .197$). Performance was even higher in the EEEE Condition compared to EPEP Condition ($U = 293.50$, $p = .003$, $r = .366$), and the PEPE Condition performed better than the PPPP Condition ($U = 254.50$, $p = .006$, $r = .356$).

Transfer Tasks. To test our hypotheses regarding the effects on transfer, we analyzed the differences among conditions on the procedural transfer task (i.e., Hypothesis 4: EEEE = EPEP > PEPE = PPPP) and the conceptual transfer task (i.e., Hypothesis 5: EEEE = EPEP > PEPE = PPPP). Contrary to our hypotheses, we found no significant performance differences among conditions on the procedural transfer task ($p = .276$) and the conceptual transfer task ($p = .104$).

2.2.2.3. Does the sequencing of examples and problems affect invested mental effort and time-on-task in the training phase?

Mental Effort. With regard to learning efficiency, we analyzed our results on self-reported effort invested in the training tasks and found a main effect of Instruction Condition, $H(3) = 51.48$, $p < .001$. In line with Hypothesis 3 (EEEE = EPEP < PEPE = PPPP), we found that students in the EEEE ($U = 971.50$, $p < .001$, $r = .730$) and EPEP Condition ($U = 964.50$, $p < .001$, $r = .535$) reported significantly lower effort investment than students in the PPPP Condition.

The average reported effort investment was also lower in the EEEE Condition than the PEPE Condition ($U = 571.50, p = .001, r = .449$), but – in contrast to our expectations – no significant difference was found between the EPEP and PEPE Condition ($p = .419, r = .105$). In addition, we found that students in the EEEE Condition reported significantly lower effort investment than those the EPEP Condition ($U = 807, p < .001, r = .501$), and students in the PEPE Condition invested significantly less effort than students in the PPPP Condition ($U = 741, p < .001, r = .588$).²

Time-on-task. When exploring time-on-task in the training phase, we found a main effect of Instruction Condition, $H(3) = 59.70, p < .001$. The average time invested in the training phase was significantly shorter in the EEEE Condition than in the PEPE ($U = 691.50, p < .001, r = .722$), and PPPP Condition ($U = 976, p < .001, r = .737$). Surprisingly, students in the EEEE Condition ($U = 990, p < .001, r = .808$) and PEPE Condition ($U = 214, p = .001, r = .422$) invested significantly less time in the training tasks than students in the EPEP Condition. Other post-hoc comparisons were not significant ($ps > .012, rs < .322$).

2.2.2.4. Does the sequencing of examples and problems affect mental effort and time-on-task in the posttest phase?

Mental Effort. As for the exploration of self-reported effort invested in solving the isomorphic posttest tasks, we found a significant main effect of Instruction Condition, $H(3) = 21.88, p < .001$. In line with our findings regarding effort invested in the training phase, reported effort investment while solving the isomorphic posttest tasks was significantly lower in the EEEE ($U = 817, p < .001, r = .478$), EPEP ($U = 845, p = .003, r = .363$), and PEPE Condition ($U = 690.50, p < .001, r = .492$) compared to the PPPP Condition. No other post-hoc comparisons were significant ($ps > .141, rs < .184$). We found the same pattern of results on students' effort invested in solving the procedural transfer task: A significant main effect of Instruction Condition, $H(3) = 15.86, p = .001$, and perceived effort investment was lower in the EEEE ($U = 759, p = .002, r = .388$), EPEP ($U = 812.50, p = .008, r = .319$), and PEPE Condition ($U = 659.50, p = .001, r = .436$) than in the PPPP Condition. None of the other comparisons were significant ($ps > .193, rs < .170$). Lastly, results showed a main effect of Instruction Condition regarding students' effort invested in the conceptual posttest task, $H(3) = 10.02, p = .018$. Invested effort was significantly lower in the PEPE Condition than in the PPPP Condition ($U = 621, p = .005, r = .365$). Again, no other post-hoc comparisons were significant ($ps > .020, rs < .281$).

2 Upon a reviewer's request, we explored whether the mental effort invested in and performance on the two practice problems in the training phase differed between the EPEP and PEPE Condition. On the first practice problem (i.e., EPEP vs. PEPE), we found no performance difference ($p = .257, r = .148$), but the EPEP Condition reported significantly lower effort investment ($p = .024, r = .294$). On the second practice problem (i.e., EPEP vs. PEPE), the PEPE Condition attained greater performance ($p = .001, r = .428$) with less effort investment than the EPEP Condition ($p = .001, r = .436$). These results are not in line with the motivational hypothesis, as the initial disadvantage of starting with a practice problem disappeared (and even reversed) on the second practice problem.

Time-on-task. As for the invested time-on-task during the posttest phase, a main effect of Instruction Condition was found for the isomorphic posttest tasks, $H(3) = 39.34, p < .001$. The average time-on-task was significantly longer in the EEEE Condition than in the EPEP ($U = 218.50, p < .001, r = .491$), PEPE ($U = 199, p = .002, r = .427$), and PPPP Condition ($U = 86, p < .001, r = .717$). In addition, students in the PEPE Condition invested significantly less time than students in the PPPP Condition ($U = 218.50, p = .001, r = .424$). The other post-hoc comparisons were not significant ($ps > .011, rs < .305$). Concerning the transfer tasks, we found a main effect of Instruction Condition for the procedural transfer task, $H(3) = 15.47, p < .001$. The average time-on-task was significantly longer in the EEEE Condition compared to the PPPP Condition ($U = 250.50, p < .001, r = .450$). No other post-hoc comparisons were significant ($ps > .015, rs < .292$). We found no main effect of Instruction Condition for the conceptual transfer task ($p = .057$).

Table 2.1.
Mean (M), Standard Deviation (SD) and Median (Med) of Self-Efficacy (range 1 to 9) Perceived Competence (range 1 to 7), and Topic Interest (range 1 to 7) per Condition in Experiment 1.

	EEEE Condition			EPEP Condition			PEPE Condition			PPPP Condition		
	M	SD	Med	M	SD	Med	M	SD	Med	M	SD	Med
Pretest												
Self-efficacy	4.27	1.91	5.00	4.26	1.83	5.00	4.68	2.17	5.00	3.89	2.05	4.00
Perceived Competence	3.84	1.40	4.00	3.62	1.51	3.67	3.88	1.79	4.00	3.42	1.63	3.33
Topic Interest	4.15	0.89	4.29	3.92	1.12	4.29	3.52	1.01	3.29	4.21	1.04	4.00
Posttest												
Self-efficacy	7.07	0.87	7.00	5.71	2.01	6.00	6.48	2.14	7.00	3.80	1.98	4.00
Perceived Competence	5.70	0.66	5.83	4.76	1.52	5.00	5.45	1.50	5.67	3.32	1.61	3.00
Topic Interest	4.05	0.92	4.14	3.59	1.03	4.71	3.49	1.12	3.57	3.92	0.93	3.86

Table 2.2.
Mean (M), Standard Deviation (SD), and Median (Med) of Pretest (range 0 to 16),
Training Performance (range 0 to 24), Isomorphic Tasks Performance (range 0 to 16),
Procedural Transfer (range 0 to 8), Conceptual Transfer (range 0 to 9), Mental Effort
(range 1 to 9), and Time-on-Task per Condition in Experiment 1.

	EEEE Condition			EPEP Condition			PEPE Condition			PPPP Condition		
	M	SD	Med	M	SD	Med	M	SD	Med	M	SD	Med
Pretest												
Performance	7.17	5.27	8.00	5.26	4.57	4.00	6.92	4.93	7.00	5.57	5.00	5.00
Training												
Performance				4.06	2.68	3.75	4.82	2.01	5.00	3.64	2.14	3.50
Mental Effort	2.51	1.32	2.13	4.02	1.33	4.25	3.75	1.21	4.00	6.01	1.84	6.25
Time-on-Task	3.40	1.17	3.25	9.56	2.95	9.88	7.07	2.65	6.75	9.32	3.60	7.50
Posttest												
Isomorphic Tasks	10.43	2.40	10.00	7.24	4.88	7.00	9.08	4.69	10.00	5.69	4.13	6.00
Procedural Transfer	3.77	2.78	2.00	3.12	3.24	2.00	3.00	2.97	2.00	2.49	2.89	2.00
Conceptual Transfer	4.60	2.22	5.00	4.74	2.81	5.00	5.28	2.48	5.00	3.74	2.31	3.00
Mental Effort												
Isomorphic Tasks	4.53	1.71	4.50	5.13	1.70	5.50	4.60	1.51	5.00	6.39	1.96	6.50
Procedural Transfer	4.00	1.97	3.00	4.44	2.08	5.00	3.72	1.97	3.00	6.03	2.65	6.00
Conceptual Transfer	4.40	1.83	5.00	3.88	1.87	3.00	3.64	1.29	3.00	5.03	2.16	5.00
Time-on-Task												
Isomorphic Tasks	16.25	5.45	15.00	10.94	5.17	10.25	12.16	5.24	11.50	7.74	4.05	7.50
Procedural Transfer	8.63	3.60	8.00	6.97	2.90	7.00	6.72	3.94	6.00	5.14	4.19	5.00
Conceptual Transfer	8.90	4.40	8.00	7.06	3.66	6.00	8.04	3.36	7.00	6.69	2.55	7.00

2.2.3. Discussion

Our main aim was to investigate how different example and practice problem sequences (i.e., EEEE, EPEP, PEPE, and PPPP) would affect motivational (self-efficacy, perceived competence, and topic interest) and cognitive (effectiveness and efficiency) aspects of learning. The results were largely inconsistent with our hypotheses (see also Figure 2.1 for a graphical overview of the median scores on self-efficacy, perceived competence, and performance on isomorphic test tasks). We predicted that the EPEP condition would show higher levels of self-efficacy, perceived competence, and topic interest than the EEEE Condition (Hypothesis 1a) and the PEPE condition (Hypothesis 1b), but that was not the case. Instead, we found no differences among conditions on topic interest. Moreover, all three example conditions showed significantly higher self-efficacy and perceived competence than the PPPP Condition, which is interesting and extends prior research showing that example study only can foster self-efficacy and perceived competence (e.g., Bandura, 1997; Crippen, et al., 2009; Hoogerheide et al., 2014, 2018). Unexpectedly, given that problem solving after example study was implemented in early research on the worked example effect because it was considered to be more motivating (Sweller & Cooper, 1985; Trafton & Reiser, 1993), students in the EEEE Condition showed significantly higher self-efficacy than those in the EPEP Condition.

As for cognitive aspects of learning, we did not find the pattern of results that we expected based on the findings of Van Gog et al. (2011) either. In contrast to Hypothesis 2 (i.e., isomorphic test performance; EEEE/EPEP > PEPE/PPPP) and Hypothesis 3 (mental effort invested in the training phase; EEEE/EPEP < PEPE/PPPP), we found no significant differences between the EPEP and PEPE conditions on both variables. We did find, however, that studying EEEE was more effective (i.e., higher isomorphic posttest performance) and efficient (i.e., with lower effort investment in the training phase) for learning than studying EPEP and PPPP. Furthermore, studying PEPE was more effective and efficient than studying PPPP. Also, while the differences in posttest performance (i.e., mean performance of EPEP seemed higher than PPPP) were not significant, studying EPEP was more efficient than PPPP.

With regards to our exploration of mental effort invested in the isomorphic posttest tasks, our results suggest that all example conditions were more efficient than the PPPP condition. Our exploration of time-on-task in the training phase showed that the EEEE condition spent significantly less time on the learning phase compared to all the other conditions. Although this seems to be an efficiency benefit, we must note that time-on-task was not experimenter-paced and watching video modeling examples probably took less time than solving the practice problems. In contrast, we found that the EEEE condition spent most time in the posttest phase compared to all other conditions. On the one hand, this might indicate that students in the EEEE condition may have needed more time to solve the procedure on the posttest because they did not have the

possibility to practice during the training phase. On the other hand, it might be possible that students in the EEEE condition mastered the procedure so well and therefore spent more time on solving the posttest tasks. If you cannot figure out how to solve such problems during the training phase, you might drop out and spend less time on the remaining practice and posttest tasks. In sum, our findings show that all example conditions were more efficient than practice problem solving only, as equal or higher performance on isomorphic posttest problems was attained with less investment of effort in the training phase. Example study only was most efficient, requiring less effort (and time) investment in the training phase than all other conditions but attaining the highest scores on the isomorphic posttest tasks.

Our expectations regarding procedural transfer (i.e., Hypothesis 4; EEEE/EPEP > PEPE/PPPP) and conceptual transfer (i.e., Hypothesis 5; EEEE/EPEP > PEPE/PPPP) were not confirmed either. Our results showed no significant differences among conditions on procedural transfer and conceptual transfer.

A possible explanation for finding that EPEP was not more effective and efficient nor more motivating than PEPE, might lie in our participant sample. Although students were novices to this mathematical task, the fact that they were enrolled in a higher technical education program makes it likely that they had experience with learning similar types of mathematical problems that require complex mathematical calculations. This might have shielded those who had started with a practice problem from motivational issues, alleviating the negative effects of starting with a practice problem on motivation (i.e., self-efficacy and perceived confidence). The fact that self-efficacy and perceived competence increased in all example conditions and did not differ between the EPEP condition and the PEPE condition indeed suggests that the confidence of those who were provided with problem-example pairs either was unaffected by starting with a practice problem or recovered quickly once provided with the opportunity of studying examples.

If this explanation is correct, the results might be different (i.e., EPEP > PEPE) with a sample of students who are less experienced with these types of mathematical tasks and who would generally be less confident about their mathematical abilities. Therefore, we reran the experiment with a sample of primary education teacher training students, who are normally much less experienced with mathematician problems such as learning how to approximate the definite integral by using the trapezoidal rule. Our hypotheses were identical to those in Experiment 1 (see paragraph 1.1).

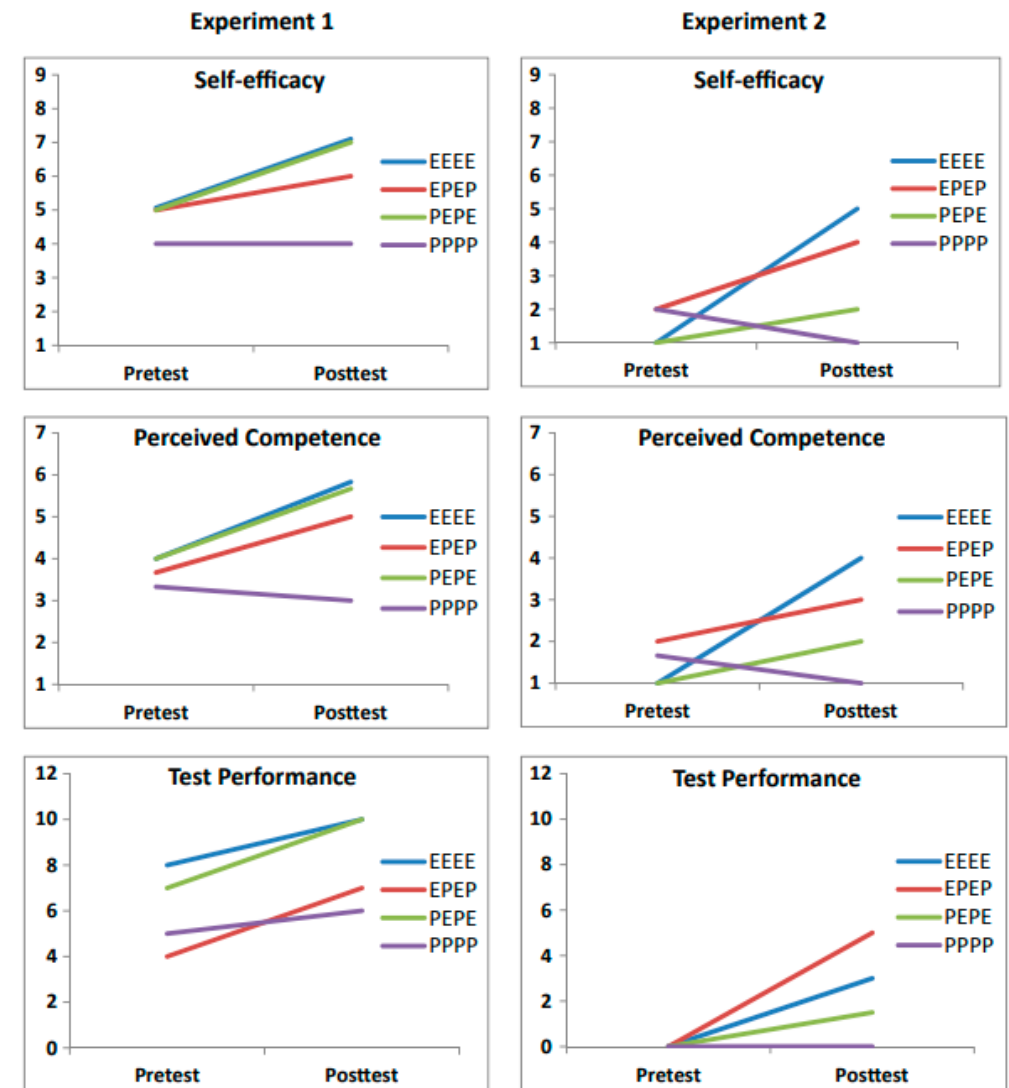


Figure 2.1. Median scores on self-efficacy (top row; range 1 to 9), perceived competence (middle row; range 1 to 7), and performance on the isomorphic test tasks (range 0 to 16) on the pretest and immediate posttest in Experiment 1 (left) and 2 (right).

2.3. Experiment 2

2.3.1. Method

2.3.1.1. Participants and design

Participants were 81 first year students from two Primary Education Teacher Training programs from two Dutch Universities of Applied Sciences ($M^{age} = 18.98$, $SD = 1.64$; 17 male, 65 female). Students could earn a monetary reward (eleven 20-euro bills were raffled among participants). Following the design of Experiment 1, students were randomly assigned to one of four conditions: 1) examples only ($n = 24$; EEEE), 2) example-problem pairs ($n = 19$; EPEP), 3) problem-example pairs ($n = 22$; PEPE), or 4) practice problems only ($n = 16$; PPPP). In this experiment, only a pretest, training phase, and (immediate) posttest were used.

2.3.1.2. Materials, procedure, and data analysis

The materials, procedure, and data analysis were the same as in Experiment 1. The reliability of the test tasks was measured again and showed the following Cronbach's alpha values: .81 for pretest tasks, .33 for isomorphic posttest tasks, and .39 for the conceptual transfer task. The only difference with Experiment 1 was that Experiment 2 was run in seven group sessions in a computer classroom with 5 to 16 participants per session instead of eight group sessions in a computer classroom with 5 to 25 participants present per session in Experiment 1.

2.3.2. Results

Again, most of the variables were not normally distributed, so we analyzed the data with nonparametric tests (cf. Experiment 1). Relevant descriptive statistics of self-efficacy, perceived competence, and topic interest scores are presented in Table 2.3, and performance scores, mental effort scores and time-on-task scores are presented in Table 2.4.

We checked for prior knowledge differences among conditions. Kruskal-Wallis test showed that there were no significant differences among conditions in terms of (pretest) performance, $H(3) = 3.35$, $p = .341$, perceived competence, $H(3) = 5.44$, $p = .142$, and topic interest, $H(3) = 5.40$, $p = .145$. We did, however, find significant differences among conditions on the pretest scores of self-efficacy, $H(3) = 9.98$, $p = .019$, and post-hoc test revealed that pretest scores of self-efficacy were significantly lower in the EEEE Condition than in the EPEP Condition ($U = 347$, $p = .002$, $r = .473$).

2.3.2.1 Does the sequencing of examples and problems affect self-efficacy, perceived competence, and topic interest?

Self-efficacy. Again, we started with the analysis of whether students' self-efficacy increased from before to after the training phase. We found a main effect of Test Moment, $Z = 4.61$, $p < .001$, $r = .512$. Follow-up tests showed that the self-efficacy scores

of the EEEE Condition ($Z = 3.96$, $p < .001$, $r = .807$) increased significantly over time, whereas the EPEP ($p = .044$, $r = .463$), PEPE ($p = .121$, $r = .331$), and the PPPP Condition ($p = .729$, $r = .087$) did not show a significant increase over time. Regarding the main question of whether there would be differences among instructional conditions on reported self-efficacy measured after the training phase, we found a main effect of Instruction Condition, $H(3) = 16.48$, $p = .001$. In contrast to Hypothesis 1a (EPEP > EEEE) and Hypothesis 1b (EPEP > PEPE), we found no differences between the EPEP and EEEE Condition ($p = .200$, $r = .195$) and the EPEP and PEPE Condition ($p = .152$, $r = .224$). Further explorations showed that self-efficacy was significantly higher in the EEEE Condition than in the PEPE ($U = 121$, $p = .001$, $r = .473$) and PPPP Condition ($U = 65$, $p < .001$, $r = .570$). No other post-hoc comparisons were significant ($ps > .056$, $rs < .007$). Note that these results have to be interpreted with caution, because there were pre-existing differences among the conditions on self-efficacy before the training phase (i.e., EEEE > EPEP).

Perceived Competence. Perceived competence showed the same pattern of results. We found a main effect of Test Moment, $Z = 4.64$, $p < .001$, $r = .516$, indicating that perceived competence increased significantly over time in the EEEE Condition ($Z = 4.02$, $p < .001$, $r = .821$) but not in the EPEP ($p = .028$, $r = .505$), PEPE, ($p = .151$, $r = .306$), and PPPP Condition ($p = .593$, $r = .134$). We also found a main effect of Instruction Condition on perceived competence measured after the training phase, $H(3) = 15.08$, $p = .002$. These results were not in line with our expectations (Hypothesis 1a: EPEP > EEEE; Hypothesis 1b: EPEP > PEPE), because we did not find any differences between the EPEP and EEEE Condition ($p = .641$, $r = .071$) and EPEP and PEPE Condition ($p = .063$, $r = .290$). Further explorations showed that scores were significantly higher in the EEEE Condition than in the PEPE ($U = 131$, $p = .003$, $r = .436$) and PPPP Condition ($U = 77.5$, $p = .001$, $r = .508$). No other post-hoc comparisons were significant ($ps > .015$, $rs < .002$).

Topic Interest. Concerning topic interest, we found that scores did not increase over time, since there was no significant main effect of Test Moment ($p = .196$). Unlike our expectations (i.e., Hypothesis 1a: EPEP > EEEE and Hypothesis 1b: EPEP > PEPE), we found no main effect of Instruction Condition ($p = .562$), meaning that there were no differences among conditions on topic interest measured after the training phase.

2.3.2.2. Does the sequencing of examples and problems affect learning and transfer?

Isomorphic Tasks. Subsequently, we checked whether performance on the test tasks isomorphic to the training phase improved significantly from pretest to posttest. We found a main effect of Test Moment, $Z = 3.76$, $p < .001$, $r = .418$. Follow-up tests showed that scores increased over time in the EEEE ($Z = 3.51$, $p < .001$, $r = .717$) and EPEP Condition ($Z = 2.56$, $p = .010$, $r = .586$), but not in the PEPE ($p = .052$, $r = .414$), and PPPP Condition ($p = .173$, $r = .341$). Regarding our second main aim, namely to examine if there are any differences among conditions on the isomorphic posttest tasks, our

results showed a main effect of Instruction Condition, $H(3) = 17.82, p < .001$. In line with our expectations (i.e., Hypothesis 2: $EEEE = EPEP > PEPE = PPPP$), participants scored significantly higher in the EEEE ($U = 71, p < .001, r = .414$) and EPEP Condition ($U = 55.50, p = .001, r = .576$) than in the PPPP Condition. However, we did not find any differences between the EEEE and PEPE Condition ($p = .144, r = .215$) and EPEP and PEPE Condition ($p = .019, r = .366$). As expected, no differences were found between the EEEE and EPEP Condition ($p = .166, r = .211$) and PEPE and PPPP Condition ($p = .033, r = .377$).

Transfer Tasks. Surprisingly, no significant performance differences were found among conditions on the procedural transfer task ($p = .257$) and the conceptual transfer task ($p = .841$). Hence, the results on our transfer measures also contrasted Hypothesis 4 ($EEEE = EPEP > PEPE = PPPP$) and Hypothesis 5 ($EEEE = EPEP > PEPE = PPPP$).

2.3.2.3. Does the sequencing of examples and problems affect mental effort and time-on-task in the training phase?

Mental Effort. When analyzing self-reported effort investment during the training phase as measure of efficiency, we found a main effect of Instruction Condition, $H(3) = 28.28, p < .001$. In line with Hypothesis 3 ($EEEE = EPEP < PEPE = PPPP$), reported effort investment was significantly lower in the EEEE ($U = 349, p < .001, r = .689$) and EPEP Condition ($U = 277.50, p < .001, r = .707$) compared to the PPPP Condition. Moreover, students in the EEEE Condition reported significantly lower effort investment than the PEPE Condition ($U = 394, p = .004, r = .422$), but no significant differences were found between the EPEP and PEPE Condition ($p = .059, r = .295$). As expected, we found no significant differences in effort investment during the training phase between the EEEE and EPEP Condition ($p = .470, r = .110$), but we did find the PEPE Condition to report significant less effort than the PPPP Condition while solving the tasks in the training phase ($U = 271, p = .004, r = .458$).³

Time-on-Task. Subsequently, we analyzed the average time spent on the tasks during the training phase and found a main effect of Instruction Condition, $H(3) = 26.17, p < .001$. We found that average time-on-task while solving the training tasks was significantly longer in the EPEP Condition than in the EEEE ($U = 430.50, p < .001, r = .757$), PEPE ($U = 69, p = .001, r = .573$), and PPPP Condition ($U = 64, p = .003, r = .493$). We found no differences between the EEEE and PPPP Condition ($p = .774, r = .122$) and PEPE and PPPP Condition ($p = .455, r = .122$).

3 Like for Experiment 1, we explored whether the mental effort invested in and performance on the two practice problems in the training phase differed between the EPEP and PEPE Condition. As one would expect, the EPEP Condition performed significantly better on the first practice problem ($p = .001, r = .513$) and invested less effort ($p < .001, r = .574$) than the PEPE Condition. However, on the second practice problem, we found no difference between the two conditions in terms of performance ($p = .178, r = .210$) or effort investment ($p = .813, r = .037$). Again, there was an advantage in favor of the EPEP Condition at the start, but no sign of a lasting disadvantage of starting with a (failed) practice problem solving attempt for those in the PEPE Condition.

2.3.2.4. Does the sequencing of examples and problems affect mental effort and time-on-task in the posttest phase?

Mental Effort. While exploring the differences among conditions on reported effort investment when solving the posttest tasks, we found no differences among conditions on the isomorphic posttest tasks ($p = .165$), procedural transfer task ($p = .238$), and conceptual transfer task ($p = .201$).

Time-on-task. We did find a main effect of Instruction Condition for average time invested in the isomorphic posttest tasks, $H(3) = 27.64, p < .001$. The average time-on-task was significantly longer in the EEEE Condition than in the PEPE ($U = 117, p = .001, r = .478$) and PPPP Condition ($U = 33, p < .001, r = .699$), and significantly longer in the EPEP Condition compared to the PEPE ($U = 108, p = .008, r = .410$) and PPPP Condition ($U = 41.50, p < .001, r = .629$). No other post-hoc comparisons were significant ($ps > .026, rs < .370$). Concerning the transfer tasks, we found a main effect of Instruction Condition for the procedural transfer task, $H(3) = 9.53, p = .023$. Average time-on-task was, however, only significantly longer in the EEEE Condition than in the EPEP Condition ($U = 337.50, p = .004, r = .442$). No other post-hoc comparisons were significant ($ps > .048, rs < .353$). We found no main effect of Instruction Condition for the conceptual transfer task ($p = .086$).

Table 2.3.
Mean (M), Standard Deviation (SD) and Median (Med) of Self-Efficacy (range 1 to 9) Perceived Competence (range 1 to 7), and Topic Interest (range 1 to 7) per Condition in Experiment 2.

	EEEE Condition			EPEP Condition			PEPE Condition			PPPP Condition		
	M	SD	Med	M	SD	Med	M	SD	Med	M	SD	Med
Pretest												
Self-Efficacy	1.67	1.27	1.00	2.84	1.86	2.00	2.18	2.08	1.00	2.31	1.66	2.00
Perceived Competence	1.54	1.03	1.00	2.19	1.10	2.00	1.98	1.62	1.00	2.06	1.39	1.66
Topic Interest	3.53	0.88	3.43	4.20	1.26	4.29	3.49	1.35	3.43	3.46	1.13	3.29
Posttest												
Self-Efficacy	4.88	1.83	5.00	3.84	2.61	4.00	2.68	2.01	2.00	2.19	2.04	1.00
Perceived Competence	3.72	1.40	4.00	3.47	2.11	3.00	2.35	1.55	2.00	1.94	1.53	1.00
Topic Interest	3.69	0.98	3.93	3.72	1.42	3.57	3.25	1.14	3.36	3.39	1.29	3.57

Table 2.4.
Mean (M), Standard Deviation (SD), and Median (Med) of Pretest (range 0 to 16),
Training Performance (range 0 to 24), Isomorphic Tasks Performance (range 0 to 16),
Procedural Transfer (range 0 to 8), Conceptual Transfer (range 0 to 9), Mental Effort
(range 1 to 9), and Time-on-Task per Condition in Experiment 2.

	EEEE Condition			EPEP Condition			PEPE Condition			PPPP Condition		
	M	SD	Med	M	SD	Med	M	SD	Med	M	SD	Med
Pretest												
Performance	0.63	1.71	0.00	1.47	3.82	0.00	1.27	3.06	0.00	1.75	2.74	0.00
Training												
Performance				2.82	1.95	3.00	1.98	1.98	1.50	1.28	1.15	1.25
Mental Effort	4.79	1.83	4.50	5.09	1.93	5.00	6.40	1.94	6.25	7.99	1.35	8.50
Time-on-Task	3.58	1.56	3.63	8.16	3.09	7.25	5.03	2.00	4.88	4.75	2.86	3.63
Posttest												
Isomorphic Tasks	3.00	2.64	3.00	4.26	3.28	5.00	1.91	1.97	1.50	0.56	1.03	0.00
Procedural Transfer	0.83	0.28	0.00	0.37	0.68	0.00	0.41	0.80	0.00	0.19	0.54	0.00
Conceptual Transfer	1.46	1.64	1.00	1.63	1.54	1.00	1.41	1.30	1.00	1.13	1.09	1.00
Mental Effort												
Isomorphic Tasks	6.81	1.90	7.00	6.87	2.31	8.00	7.07	2.35	7.50	7.81	2.22	9.00
Procedural Transfer	8.04	1.57	9.00	7.11	2.51	8.00	7.86	2.44	9.00	8.00	2.22	9.00
Conceptual Transfer	7.00	1.87	7.00	6.58	2.63	8.00	7.59	2.46	9.00	7.50	2.13	8.00
Time-on-Task												
Isomorphic Tasks	7.23	5.18	6.75	6.74	5.11	5.50	2.80	2.87	2.50	0.91	1.20	0.25
Procedural Transfer	0.50	0.93	0.00	3.68	4.84	1.00	1.59	2.38	0.50	0.81	1.38	0.00
Conceptual Transfer	3.08	2.28	3.00	5.26	4.07	4.00	2.68	2.06	2.50	3.06	2.17	3.00

2.3.3. Discussion

The main aim of this experiment was to investigate whether the results of Experiment 1 (i.e., EPEP = PEPE) on performance, mental effort invested in the learning tasks, and motivation would be different (i.e., EPEP > PEPE) with a sample of students who are less experienced with these types of mathematical tasks and who would generally be less confident about their mathematical abilities (i.e., teacher training students). Despite the different sample (i.e., primary education teacher training students), the results of Experiment 2 also did not show evidence in favor of our hypothesis (see also Figure 2.1); we expected that the EPEP condition would show higher levels of self-efficacy, perceived competence, and topic interest than the EEEE Condition (Hypothesis 1a) and the PEPE condition (Hypothesis 1b), but we found no significant differences between these conditions. As in Experiment 1, no differences were found among conditions regarding topic interest. Results did show that self-efficacy and perceived competence were significantly higher in the EEEE condition than in the PEPE and PPPP condition (as in Experiment 1).

With regard to cognitive aspects of learning, we partially replicated the results from Van Gog et al., 2011. In line with our expectations on isomorphic posttest performance (i.e., Hypothesis 2; EEEE/EPEP > PEPE/PPPP) and invested mental effort in the training phase (i.e., Hypothesis 3; EEEE/EPEP < PEPE/PPPP), we found that starting with an example (EEEE and EPEP) was more effective and efficient for learning than problem solving only (PPPP). Also, while the differences in isomorphic posttest performance (i.e., mean performance of PEPE seemed higher than PPPP) were not significant, studying PEPE was more efficient than PPPP. In contrast to our expectations, we did not find any significant differences on both variables between the EEEE and EPEP and between the EPEP and PEPE conditions.

When exploring mental effort on the isomorphic posttest tasks, we found no differences among conditions. Our exploration of time-on-task revealed that the EPEP Condition spent significantly more time in the training phase than all the other conditions. In addition, both conditions starting with an example (i.e., EEEE, EPEP) spent significantly more time on the isomorphic posttest tasks than the conditions starting with a problem (i.e., PEPE, PPPP). This might indicate that, considering the performance on the isomorphic posttest tasks, students understood the procedure and therefore spent more time on solving the posttest tasks. In sum, our findings show that all example conditions were more efficient than practice problem solving only, as equal or higher performance on isomorphic posttest problems was attained with less investment of effort in the training phase. Finally, our expectations regarding procedural transfer (i.e., Hypothesis 4; EEEE/EPEP > PEPE/PPPP) and conceptual transfer (i.e., Hypothesis 5; EEEE/EPEP > PEPE/PPPP) were not confirmed. Our results showed no significant differences among conditions on procedural transfer and conceptual transfer.

2.4. General Discussion

Two experiments were conducted to conceptually replicate and extend the study by Van Gog and colleagues (2011) in order to investigate how different example study and practice problem solving sequences would affect learning and motivation. Our main aim was to investigate how example study only (EEEE), example-problem pairs (EPEP), problem-example pairs (PEPE), and problem-solving only (PPPP) sequences would affect motivational (i.e., self-efficacy, perceived competence, and topic interest) and cognitive (i.e., effectiveness and efficiency) aspects of learning.

2.4.1. Example Study Only versus Example-Problem Pairs

First, we were interested in looking from a motivational perspective at the finding by Van Gog et al. (2011) that EEEE did not differ from EPEP in terms of learning outcomes. We expected EPEP to be more motivating for students than passively studying EEEE, as suggested -but not tested- by Sweller and Cooper (1985; see also Trafton & Reiser, 1993). Interestingly, our findings showed that EEEE was not less motivating than EPEP. In Experiment 1, students in the EEEE condition even showed higher self-efficacy (and better performance) than students in the EPEP condition. This finding might indicate that (at least when short training phases are used), the benefits of engaging in practice problem solving instead of further example study, seem limited for both learning and motivation. In general, this calls for further research into the role of practice problem solving in example-based learning, especially as findings from Baars, Van Gog, De Bruin, and Paas (2014) and Van der Meij et al. (2018) showed that even *additional* problem solving practice did not have a positive effect on learning. However, all those studies used relatively short training phases. It is possible that motivational differences will start to arise and affect learning when training phases are longer and consist of more training tasks, as students might get bored with studying examples only. The effects of longer sequences should therefore be addressed in future research.

2.4.2. Example-Problem Pairs versus Problem-Example Pairs

Second, we aimed to investigate whether motivational aspects of learning could account for the finding by Van Gog and colleagues (2011) that EPEP led to better test performance with less effort investment in the training phase than PEPE. However, in contrast to our expectations, we did not replicate these findings across two experiments with different populations. We also did not find any significant differences between EPEP and PEPE concerning students' self-efficacy and perceived competence. We thought we had a potential explanation for this null-finding in Experiment 1, because those higher technical education students, despite being novices, presumably had experience with similar types of mathematical problems and these problems were relevant for them (so they would not get frustrated that easily). However, the results of Experiment 2 indicated that this explanation does not hold. In Experiment 2, we again failed to find significant differences in learning or motivation between the EPEP and

PEPE condition, even though the sample consisted of student teachers for whom the tasks were less relevant, who had less experience with these types of mathematical tasks, and who felt less confident about their mathematical abilities, as evidenced by the pretest scores on performance, self-efficacy, and perceived competence. Note that, numerically, the EPEP and PEPE conditions did differ on isomorphic posttest performance (i.e., EPEP > PEPE) and average invested mental effort in the training phase as a whole (i.e., EPEP < PEPE). Importantly, exploratory analyses of students' performance on and effort invested in the two practice problems also suggest that starting with a practice problem did not have a demotivating effect in either experiment. Whereas EPEP was more effective and/or efficient than PEPE on the first practice problem (i.e., equal or higher performance attained with less effort), we found no performance or effort advantage of EPEP over PEPE on the second practice problem.

Given that our findings regarding the EPEP vs. PEPE comparison were not in line with other studies, and that the direction of the difference between conditions seemed to vary in our experiments (i.e., we found a non-significant, medium-sized [according to the Cohen's *d* criterion] negative effect of EP on learning in Experiment 1, and a large but non-significant, positive effect of EP on learning in Experiment 2), we entered all EP-PE comparisons from the published studies we are aware of in a small-scale random effect meta-analysis, to get a better estimate of the EP-PE effect size and its heterogeneity (see Figure 2.2). We used Cumming's (2012) ESCI software (www.thenewstatistics.com). This small-scale meta-analysis showed a significant, small to medium-sized advantage of EP over PE (Cohen's *d* of the meta-analytic effect was 0.350). This advantage has to be interpreted with caution because there was substantial heterogeneity among the comparisons (i.e., heavy variation in the results among studies). That is, of the 10 comparisons, 8 showed an EP advantage and 2 showed a PE advantage, and effect sizes varied from -0.397 (the first experiment in this study) to 0.862 (the study by Van Gog et al., 2011).

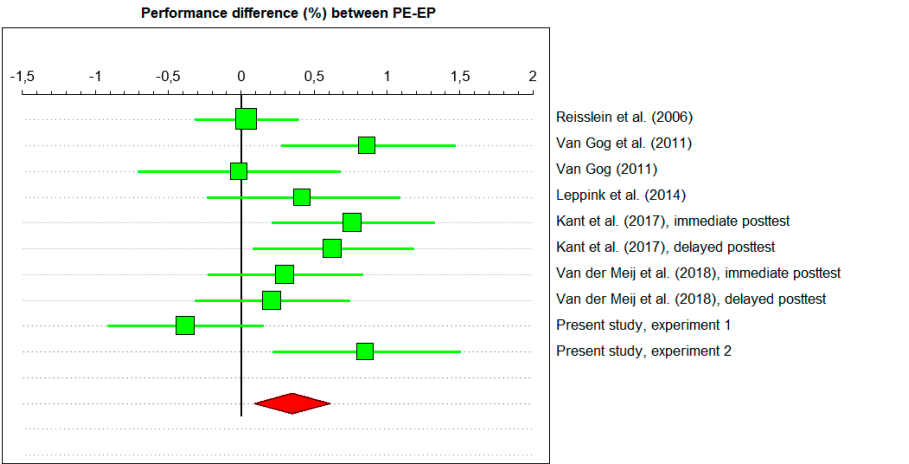


Figure 2.2. Results of the meta-analysis

A possible reason why there is substantial variation in results among the example-problem vs. problem-example comparisons (i.e., EP = PE vs. EP > PE) might lie in study characteristics that vary across studies, such as the learning material, target group, pair type (identical vs. isomorphic pairs), sequence length (two vs. four training tasks), and example format (worked vs. video modeling examples). For instance, when viewing the results of the studies in the small-scale meta-analysis that used video modeling examples, it seems that almost all of these studies did not find any learning differences between EP and PE⁴, whereas the studies that used worked examples did find EP to be more effective than PE. One could assume that after starting with a problem, demotivated learners would not pay attention to worked examples, but would study video modeling examples. Worked examples can be overwhelming because all the information is presented simultaneously, and it might be easy to ignore written text. Video modeling examples, however, present information step-by-step and the combination of dynamic visual information and the model's narration takes the learner by the hand. Thus, it is possible that students might find it more motivating to study a video modeling example after starting with a practice problem than studying a worked example, which might partially explain the differences in findings.

Another factor that might explain the differences in findings is that problem-example pairs may become more effective when the number of training tasks increases. When two tasks are presented in the training phase, a failed practice problem solving attempt means that students in the PE condition only effectively have one task to learn from, whereas those in the EP condition have the opportunity to first build a schema with the example and then learn again from problem solving (and repeat this again). Using four (or more) training tasks means that learners in the PE condition have more opportunities for learning. Note that we checked all studies used in the small-scale meta-analysis on whether one of these factors could explain the variation in results, and found that none of these factors could solely account for the mixed findings. Future research is recommended to investigate which (combination of) factors might moderate the EP-PE effect.

4 Except for the study of Kant, Scheiter, and Oschatz (2017) that did find EP to be more effective than PE.

2.4.3. Limitations

A limitation of the present study is the sample size of Experiment 2. While a power analysis indicated that our sample size was more than sufficient to reliably detect the effect sizes found by Van Gog and colleagues (2011), our small-scale meta-analysis suggests that the EP-PE effect might be significantly smaller than previously believed. The sample size of Experiment 2 was not sufficient to reliably detect small to medium-sized effects, and therefore, the results of Experiment 2 should be interpreted with caution. A second limitation of this study is that the reliability of our test tasks (i.e., isomorphic posttest tasks and conceptual transfer task) was rather low, particularly in Experiment 2. A possible explanation for the low reliability of the test tasks might be the low scores on the isomorphic posttest tasks and transfer tasks. Together with the high mental effort scores in the training phase, this might indicate that these test tasks were more difficult for the students in Experiment 2 than the students in Experiment 1. Another explanation may lie in the small number of tasks that was used to measure isomorphic posttest performance (i.e., 2 tasks) and conceptual transfer (i.e., 5 open-ended questions).

2.4.4. Practical Implications

Nevertheless, our findings are very interesting and relevant for educational practice, where example study and practice problem solving are frequently used to acquire new knowledge and skills (e.g., Atkinson & Renkl, 2007; Van Gog, et al., 2014). The results of this study suggest that, when short training phases are used, studying examples (only) is more preferable than problem solving only. These results complement previous findings on the 'worked example effect', that have (also) shown example study to result in higher learning outcomes with less reported effort investment than problem solving only (for reviews, see Atkinson et al., 2000; Renkl, 2014; Sweller et al., 2011; Van Gog & Rummel, 2010). A novel finding, however, is that example study also enhances motivational aspects of learning, such as believing in one's own competence when mastering a task, whereas problem solving only does not positively affect students' motivation at all. These results could be used by teachers during their classroom practice when instructing novices on new knowledge or skills, or as guidelines for instructional designers when designing new learning materials (such as books or online learning environments). In addition, students could be given the advice to study examples (only) instead of practice problem solving only when learning new knowledge or skills during self-study (for example when selecting own training tasks in online self-paced learning environments).

Chapter 3

Examples, practice problems, or both?
Effects on motivation and learning
in shorter and longer sequences



This chapter was published as:

Van Harsel, M., Hoogerheide, V., Verkoeijen, P. P. J. L., & Van Gog, T. (2020). Examples, practice problems, or both? Effects on motivation and learning in shorter and longer sequences. *Applied Cognitive Psychology*, 34(4), 793-812.
<https://doi.org/10.1002/acp.3649>

Acknowledgement of author contributions:

MvH, VH, PV, and TvG designed the study, MvH recruited participants and collected the data, MvH analyzed the data, VH checked the data package, MvH drafted the manuscript, all authors contributed to critical revision of the manuscript, VH, PV, and TvG supervised the study.

Abstract

Research suggests some sequences of examples and problems (i.e., EE, EP) are more effective (higher test performance) and efficient (attained with equal/less mental effort) than others (PP, sometimes also PE). Recent findings suggest this is due to motivational variables (i.e., self-efficacy), but did not test this during the training phase. Moreover, prior research used only short task sequences. Therefore, we investigated effects on motivational variables, effectiveness, and efficiency in a short (Experiment 1; 4 learning tasks; $N = 157$) and longer task sequence (Experiment 2; 8 tasks; $N = 105$). With short sequences, all example conditions were more effective, efficient, and motivating than PP. With longer sequences, all example conditions were more motivating and efficient than PP, but only EE was more effective than PP. Moreover, EE was most efficient during training, regardless of sequence length. These results suggest that example study (only) is more effective, efficient and more motivating than PP.

Keywords: example-based learning, video modeling examples, problem solving, self-efficacy, mental effort

3.1. Introduction

It is well-established that for novices who have little or no prior knowledge of a task, studying worked-out examples of problem solutions – or studying examples alternated with practice problem solving – is a more effective and efficient instructional strategy than practice problem solving only (for a review, see Van Gog et al., 2019). Effective means it often results in higher posttest performance, and efficient that this higher performance is often attained with equal or less effort investment in the learning and test phases. Example study is more effective and efficient for novices than practice problem solving because it gives novices the opportunity to devote all available cognitive capacity to study the step-by-step explanation of the solution procedure, which helps them develop a schema on how to solve this type of problem in the future (e.g., Sweller & Cooper, 1985). When solving practice problems, in contrast, novices (lacking prior knowledge) have to resort to weak problem-solving strategies (e.g., via trial-and-error, means-ends analysis), which is very effortful and time consuming, yet hardly contribute to learning (e.g., Sweller, 1988). For learners with higher prior knowledge, however, instructional strategies with a high level of support may be less efficient, because they have already developed proper cognitive schemata to guide their problem solving (cf. expertise-reversal effect; Kalyuga et al. 2001; Kalyuga and Sweller 2004; Kalyuga & Renkl, 2010; Roelle & Berthold, 2013). These learners might gain more from practice problem solving than example study.

Despite the multitude of studies on *example-based learning*, an important open question that remains is how example study and practice problem solving should be sequenced to be most effective (i.e., for students' posttest performance), most efficient (i.e., posttest performance considered in light of mental effort investment in the training and test tasks), and most motivating for learning.

3.1.1. Short Task Sequences of Example Study and Practice Problem Solving

Van Gog, Kester, and Paas (2011) were the first to compare the four most commonly used sequences of examples and practice problems to uncover which sequence would be most effective and efficient for learning. Secondary education students (novices) learned how to diagnose a fault in electrical circuits with the help of four training tasks presented as examples only (EEEE), example-problem pairs (EPEP), problem-example pairs (PEPE), or practice problems only (PPPP). Results showed that EEEE and EPEP were more effective and efficient than PEPE and PPPP. No differences were found, however, between the conditions starting with an example (i.e., EEEE and EPEP) and between the conditions starting with a practice problem (i.e., PEPE and PPPP).

Since then, follow-up research has investigated whether these findings would replicate and how they could best be explained. However, studies attempting to replicate the differences between the example-problem pairs (EP-pairs) and problem-example pairs (PE-pairs) conditions showed mixed results (see Table 3.1 for the characteristics of these studies). Whereas some studies also found that EP-pairs were more effective and efficient for learning than PE-pairs (e.g., Kant et al., 2017; Leppink et al., 2014), others did not find any test performance and/or effort investment differences (e.g., Van Harsel et al., 2019; Coppens et al., 2019; Van Gog, 2011; Van der Meij et al., 2018). A small-scale meta-analysis by Van Harsel et al. (2019) on all (published) studies available at that time showed a significant, small-to-medium meta-analytic advantage of EP over PE on final test performance (Cohen's d of 0.350), albeit with a large heterogeneity between effects.

Table 3.1.
Characteristics of Studies Investigating the Effectiveness and Efficiency of EP-Pairs and PE-Pairs.

	Van Harsel et al. 2019; Exp. 1	Van Harsel et al. 2019; Exp. 2	Coppens et al. 2019	Kant et al. 2017	Leppink et al. 2014	Van der Meij et al. 2018	Van Gog, 2011	Van Gog et al., 2011
Learner Characteristics								
Average Age	19.3	19	10.6	12.5	-	11.2	20.2	16.2
Educational Level	First-year students from a university of applied sciences, enrolled in an electrical and electronic or mechanical engineering program	First-year students from a university of applied sciences, enrolled in a teacher training program	Elementary school students	Seventh grade students	First-year university students, enrolled in a social and health sciences program	Fifth-grade and sixth-grade classrooms from elementary school	Students enrolled in programs at the Faculty of Social Sciences	Students in their fourth or fifth year of pre-university education
Type of Knowledge in Learning and Test Materials	Procedural knowledge	Procedural knowledge	Procedural knowledge	Conceptual and procedural knowledge	Proce- dural know- ledge	Proce- dural know- ledge	Proce- dural know- ledge	Procedural knowledge
Topic of Learning Materials	Mathe- matics, trapezoidal rule	Mathe- matics, trapezoidal rule	Mathe- matics, water jug problems	Science, Scientific reasoning and inquiry tasks	Statistics, appli- cation of Bayes' theorem	Software training on Word	Mathe- matics, Frog leap	Science, applying Ohm's law to reason about faults in electrical circuits
Learning Setting	Classroom experiment at school, not part of the curriculum	Classroom experiment at school, not part of the curriculum	Classroom experiment at school, not part of the curriculum	Computer room experiment at school, not part of the curriculum	Classroom experiment part of statistics course	Computer room experiment at school, not part of the curriculum	Individual experi- ment in the lab of the University	Classroom experiment at school, not part of the curriculum

3.1.2. The Role of Motivation during Example Study and Practice Problem Solving

An explanation for these mixed findings might lie in motivational aspects of learning. That is, when novices have to learn how to solve a complex task that requires domain-specific knowledge and that is not particularly intrinsically rewarding or enjoyable, then starting the training phase with a practice problem (PE-pairs) might decrease their motivation. Solving such a practice problem could be experienced as so difficult that learners lose interest in the topic of the learning materials (i.e., topic interest) or confidence in their ability to learn the task (e.g., self-efficacy and perceived competence). As a consequence, learners may not be motivated to study the subsequent example (and possibly also the tasks that follow). In this case, PE-pairs are probably less effective for learning than EP-pairs. However, when the complex task is experienced as intrinsically rewarding or enjoyable, starting the training phase with a practice problem (PE) might not have a detrimental effect on students' interest or confidence in their ability to learn the task. In this case, studying EP is probably equally effective for learning as studying PE.

This motivational explanation was tested in two recent studies in which novices learned to solve mathematical problems (i.e., Van Harsel et al., 2019; Coppens et al., 2019). In these studies, aspects of motivation such as topic interest, self-efficacy, and perceived competence were measured before and after the training phase to investigate whether students lose interest in the task (i.e., topic interest) or confidence in their ability to learn the task (i.e., self-efficacy and perceived competence) as a result of starting the training phase with a practice problem. Self-efficacy is defined as a personal judgment of one's own capacities to organize or accomplish a specific task or challenge and has shown to have a positive effect on factors such as academic motivation, study behavior, and learning outcomes (e.g., Bandura, 1997; Schunk, 2001). Perceived competence is related to the construct of self-efficacy but comprises more general knowledge and perceptions of people's self-concept towards one's own competence (e.g., Deci & Ryan, 2002; Hughes et al., 2011). Like self-efficacy, perceived competence is also positively linked to factors such as academic motivation and learning outcomes (e.g., Bong & Skaalvik, 2003). Finally, topic interest can be described as personal interest in a domain or activity based on previously acquired knowledge, personal experiences, and emotions (e.g., Ainley et al., 2002; Renninger, 2000). Topic interest has positive effects on cognitive functioning, (deep) learning, and engagement (e.g., Hidi, 1990; Schiefele & Krapp, 1996; Tobias, 1996).

In contrast to the motivational explanation, Van Harsel et al. (2019) and Coppens et al. (2019) found no differences between EP-pairs and PE-pairs on test performance, or on self-efficacy, perceived competence, and topic interest. However, in these studies, these motivational constructs were only measured before and after the training phase. Measuring self-efficacy after each task in the training phase would be more insightful, because it could reveal whether self-efficacy was not negatively affected at

all when starting the training phase with a practice problem or whether it recovered quickly once provided with an example. Another improvement that would allow for a more sensitive test is to use a conceptual pretest rather than a procedural one, as was the case in the study by Van Harsel et al. (2019; i.e., two practice problems isomorphic to the training phase). With such a procedural pretest, one could argue that all participants started with practice problem solving (also the example conditions: PPEEEE and PPEPEP). Therefore, the first aim of the present study was to investigate students' self-efficacy during the training phase in four task sequences (EEEE, EPEP, PEPE, PPPP). The second aim was to address the open question of how motivational and cognitive aspects of learning would be affected by those task sequences in longer training phases.

3.1.3. Longer Tasks Sequences of Example Study and Practice Problem Solving

Previous sequencing research often used a small number of training tasks (i.e., two tasks: Kant et al., 2017; Leppink et al., 2014; four tasks: Van Gog, 2011; Van Gog et al., 2011; Van Harsel et al., 2019). In such short sequences, EE was found to be equally or more effective (and efficient) for learning as EP on an immediate posttest (e.g., Kant et al., 2017; Leppink et al., 2014; Van der Meij et al., 2018; Van Harsel et al., 2019) and a delayed posttest (e.g., Leahy et al., 2015; Van Gog & Kester, 2012; Van Gog et al., 2015). Moreover, no differences between EE and EP were found on motivational aspects of learning (i.e., self-efficacy, perceived competence, and topic interest; Van Harsel et al., 2019).

However, in educational practice students may encounter (much) longer study sequences. Because students will gain knowledge as training progresses, longer task sequences may affect motivational and cognitive aspects of learning differently than shorter sequences. That is, studying examples only might not only become boring but also redundant as students gain knowledge from the first few tasks. This in turn might have negative effects on motivational aspects of learning (and performance; see Kalyuga et al., 2001) as compared to sequences in which examples and problems are alternated. It might be more engaging for learners to actively attempt to solve practice problems than to continuously study examples, which is more passive learning (as suggested –but not tested– by Sweller & Cooper, 1985). Examples alternated with practice problems might be more engaging than example study only in longer sequences as the interspersed practice problems give learners the opportunity to actively apply what they have learned and allow them to identify gaps in their knowledge (cf. Baars et al., 2014, 2017), which they can repair when studying subsequent examples.

3.1.4. The Present Study

In sum, the present study aimed to examine how short (i.e., Experiment 1: EEEE, EPEP, PEPE, and PPPP) and longer (i.e., Experiment 2: EEEEEEEE, EPEPEPEP, PEPEPEPE, and PPPPPPPP) task sequences of examples and/or practice problems would affect motivational and cognitive aspects of learning on an immediate posttest. With regard to short sequences, we added a delayed posttest to see whether effects remained stable over time. Furthermore, we measured self-efficacy after each task in the training phase (instead of only before and after the training phase). In this way, we were able to explore whether and how motivation was affected by the order of examples and practice problems in the training phase. Finally, a conceptual pretest was used instead of a procedural pretest as in the study by Van Harsel et al. (2019).

3.2 Experiment 1

In Experiment 1, it was investigated how short task sequences of examples and/or practice problems (i.e., EEEE, EPEP, PEPE, and PPPP) would affect motivational (i.e., self-efficacy, perceived competence, and topic interest measured before and after the training phase) and cognitive aspects of learning (i.e., invested mental effort in the training phase and performance on isomorphic and transfer tasks). We explored effects on time-on-task (training phase and posttest phases) and mental effort (posttest phases), because when combined with test performance, these measures are indicators of the efficiency of the learning process and learning outcomes (Van Gog & Paas, 2008). We also administered a delayed posttest to explore whether the pattern of results would remain stable after a one-week delay. We expect to replicate the pattern of results found by Van Harsel et al. (2019), because the same materials and population are used (see Table 3.2 for results found by Van Harsel et al., 2019). Note that we used a conceptual pretest instead of a procedural pretest to rule out the alternative explanation that when a procedural pretest is used (e.g., two practice problems in Van Harsel et al., 2019), one could argue that all participants start with practice problem solving (also the example conditions: PPEEEE and PPEPEP). As a result, if the motivational explanation would be valid, even students in the example-first conditions would lose interest and confidence in their own abilities before the first example. Therefore, it is possible that EPEP becomes more motivating, effective, and efficient for learning compared to PEPE when using a conceptual pretest (instead of EPEP = PEPE as found by Van Harsel et al., 2019).

Table 3.2.

Main Results of Experiment 1 of Van Harsel et al. (2019) Regarding the Effects of Short Sequences of Examples and Problems (EEEE, EPEP, PEPE, and PPPP) on Isomorphic Tasks, Transfer Tasks, Mental Effort, Self-Efficacy, Perceived Competence, and Topic Interest.

Main results	
Training phase	
Mental effort	EE, EP, PE < PP / EE < EP, PE / EP = PE
Immediate posttest phase	
Isomorphic tasks	EE, PE > PP / EE > EP / EE = PE / EP = PE, PP
Procedural transfer task	EE = EP = PE = PP
Conceptual transfer task	EE = EP = PE = PP
Self-efficacy	EE, EP, PE > PP / EE > EP / EE = PE / EP = PE
Perceived competence	EE, EP, PE > PP / EE > EP / EE = PE / EP = PE
Topic interest	EE = EP = PE = PP

Note. Acronyms for groups: EE = example study only; EP = example-problem pairs; PE = problem-example pairs; PP = problem solving only.

Regarding self-efficacy after each training task, it was expected that students in the EEEE and EPEP condition would show significantly higher levels of self-efficacy after the first training task than students in the PEPE and PPPP condition (H1a). We assumed that the PEPE condition would 'recover' after receiving an example as second training task (given that prior research with these tasks showed no differences in motivation and learning outcomes after training), and therefore we expected no significant differences on self-efficacy scores among the EEEE, EPEP, and PEPE conditions from the second training task onwards (H1b). Since students in the PPPP condition were not provided with an opportunity to study an example, it was predicted that self-efficacy scores would be significantly higher in the EEEE, EPEP, and PEPE condition than in the PPPP condition from the second training task onwards (H1c).

3.2.1. Method

3.2.1.1. Participants and design

Participants were 157 Dutch higher education students enrolled in the first year of an electrical and electronic mechanical engineering program ($M^{age} = 19.13$, $SD = 1.75$; 155 male, 2 female). Participants were randomly assigned to one of four conditions: examples only ($n = 33$; EEEE), example-problem pairs ($n = 45$; EPEP), problem-example pairs ($n = 40$; PEPE), or practice problems only ($n = 39$; PPPP). The experiment consisted of four phases: (1) pretest, (2) training phase, (3) immediate posttest phase, and (4) delayed posttest phase. At the delayed posttest, which was completed after one week, 25 participants were absent, so these data are based on 132 participants ($M^{age} = 19.04$, $SD = 1.71$; 130 male, 2 female). Participants were assumed to be novices to the modelled task (i.e., approximating the definite integral of a function using the trapezoidal rule) as this subject had not (yet) been a part of their study program. Participants gave their informed consent prior to their inclusion in the study and received study credits for their participation.

3.2.1.2. Materials

All materials were presented using a web-based learning environment. The materials were based on the materials developed by Van Harsel et al. (2019).

Pretest. The pretest was a conceptual prior knowledge test that consisted of seven multiple-choice questions ($\alpha = .49$)⁵ and was developed in collaboration with two math teachers from a higher education institute. This test was used to check whether participants' ability to recognize and name the basic principles of the trapezoidal rule was low and whether prior knowledge did not differ among conditions. An example of a conceptual prior knowledge question was given in the Supplementary Materials D.

Training phase. The training phase consisted of four tasks that required participants to use the trapezoidal rule. The trapezoidal rule is a numerical integration method that is used to give a quantitative approximation of the region under the graph of a specific function. Each task had its own cover story (i.e., task 1: fitness, task 2: energy measurement, task 3: washing machine, and task 4: soapsuds). To ensure that only the task format differed across conditions, the task order was identical for all participants (i.e., in order: fitness, energy measurement, washing machine, and soapsuds). Each task was part of a task pair (i.e., pair 1: fitness and energy measurement, pair 2: washing machine and soapsuds).

⁵ A possible explanation for the low reliability of the pretest could be the unfamiliarity of the participants with the subject matter (indeed, it was meant as a check that students were indeed novices regarding those tasks). Because the pretest consisted of multiple-choice questions and "I do not know" was not included as an answer option, students would have had to guess, which likely resulted in low reliability of the pretest.

Within a task pair, the tasks were isomorphic (i.e., a similar problem-solving procedure, but surface features such as the cover stories and numbers used in functions were slightly different). There was a minor complexity difference between the first and second task pair. The first pair of tasks required Participants to calculate with positive numbers. The second pair was slightly more complex because Participants had to calculate with both positive and negative numbers.

Regarding the design of the tasks, the practice problems started with a short description of the problem state. Then, some additional information was provided on how to solve the problem, such as the trapezoidal rule formula, the graph of a function, the left border and right border of the area to be calculated, and the number of intervals. It was, however, not explained how to use the information to solve the practice problem. At the end of the problem format, participants received the following assignment: "Approach the area under the graph using the information that is given. Write down all your intermediate steps and calculations". Participants could solve the problem by completing the four steps: 1) 'compute the step size of each subinterval', 2) 'calculate the x-values', 3) 'calculate the function values for all x-values', 4) 'enter the function values into the formula and calculate the area'. An example of a problem format is given in the Supplementary Materials A.

Each video modeling example displayed a screen capture of a female model's computer screen, in which she demonstrated in a stepwise manner how to solve a practice problem with the help of the trapezoidal rule. While solving the problem, the model provided verbal explanations and on-screen handwritten notes. At the start of the video, the model first explained the purpose of the trapezoidal rule and then provided an explanation of the problem state. The problem state was exactly the same as in the problem format. Subsequently, the model demonstrated and explained how one could interpret the corresponding graph of a function with information that was given (i.e., the left border and right border of the area, the number of intervals, and the trapezoidal rule) and eventually showed how to solve the problem by calculating the four steps listed in the description of the problem format. A screenshot of a video modeling example is given in the Supplementary Materials B.

Immediate and delayed posttest. The immediate and delayed posttest presented four tasks, two isomorphic and two transfer tasks. Of the two isomorphic tasks (immediate posttest: $\alpha = .71$; delayed posttest: $\alpha = .77$), one was isomorphic to the first pair of training tasks and the other to the second pair of training tasks. The third posttest task measured procedural transfer and asked participants to use the Simpson rule instead of the trapezoidal rule to approximate the definite integral under a graph. The Simpson rule is also a numerical method for approximating the integral of a function. The problem-solving procedure of Simpson's rule is comparable to that of the trapezoidal rule, however, Simpson's rule uses a different formula to approximate the definite integral of a function (i.e., with a sequence of quadratic parabolic segments

instead of straight lines such as the trapezoidal rule). The fourth posttest task measured conceptual transfer and consisted of five open-ended questions that aimed to measure Participants' understanding of the trapezoidal rule. All five questions comprised a multiple-choice part with four options an 'explanation' part (where participants had to justify their chosen answer). Hence, these questions were more complex than the conceptual pretest items, which only required participants to select the correct answer. Unfortunately, the data regarding the conceptual transfer questions had to be excluded from the analyses due to a programming error. An example of an isomorphic posttest task, procedural transfer task and conceptual transfer question can be found in the Supplementary Materials F, G, and H.

Mental effort. After each task on the pretest, the training phase, the immediate posttest, and the delayed posttest, participants rated their mental effort on a 9-point mental effort rating scale (Paas, 1992), with answer options ranging from (1) "very, very low mental effort" to (9) "very, very high mental effort".

Self-efficacy, perceived competence, and topic interest. Self-efficacy was measured before, during (i.e., after each training task), and after the training phase by asking participants to rate to what extent they were confident that they could approximate the definite integral of a graph using the trapezoidal rule on a 9-point rating scale, ranging from (1) "very, very unconfident" to (9) "very, very confident" (Van Harsel et al., 2019; adapted from Hoogerheide et al., 2016).

Perceived competence was measured using the *Perceived Competence Scale for Learning* (Van Harsel et al., 2019; based on Williams & Deci, 1996; Williams et al., 1988). This perceived competence scale (immediate posttest: $\alpha = .98$; delayed posttest: $\alpha = .97$) consisted of three items: "I feel confident in my ability to learn how to approximate the definite integral of a graph using the trapezoidal rule", "I am capable of approximating the definite integral of a graph using the trapezoidal rule", and "I feel able to meet the challenge of performing well when I have to apply the trapezoidal rule". Participants were asked to rate on a scale of (1) "not at all true" to (7) "very true" to what degree these three items applied to them.

The topic interest scale (Van Harsel et al., 2019; adapted from the topic interest scale by Mason et al., 2008, and the perceived interest scale by Schraw et al., 1995) were used to measure participants' interest in the topic (i.e., the trapezoidal rule). The topic interest scale (immediate posttest: $\alpha = .81$; delayed posttest: $\alpha = .82$) consisted of 7 items and participants had to rate on a 7-point scale, ranging from 1 (not at all) to 5 (very true), to what degree each of the items applied to them. All items are shown in the Supplementary Materials I.

3.2.1.3. Procedure

The experiment was run in sixteen sessions (i.e., eight first sessions and eight second sessions) and took place in a computer classroom at the participants' institute of higher education. The number of participants ranged from 2 to 23 per session. Prior to the first session, headsets, pens, and scrap paper (to write down calculations) were distributed. Once participants were seated in the computer classroom, the first session (ca. 106 minutes) started with a general introduction by the experimenter explaining the aim and procedure of the experiment. Participants were told they could work at their own pace (with a maximum of 135 minutes) on mathematical tasks in an online learning environment by means of different instructional formats (i.e., examples and/or practice problems). They were instructed to write down as much as possible when solving a training task or test task, and that if they really did not know what to answer, to write an X. After the instruction, participants received a paper with a link and a password that gave access to the online learning environment.

The learning environment was designed in such a way that each task and questionnaire were presented on a separate page. Participants were unable to go back to previous pages and had to complete each task or questionnaire before they could go to the next page. Time was logged for each task. When participants entered the learning environment, they were assigned to one of the four conditions (i.e., EEEE, EPEP, PEPE, or PPPP). Participants started with a short demographic questionnaire (e.g., age, gender, and preliminary education), followed by the conceptual pretest. After the pretest, participants completed the self-efficacy, perceived competence, and topic interest questionnaires before they started the training phase. During the training phase, participants received four tasks that were presented as examples and/or practice problems (depending on their assigned condition). After each task, participants were asked to indicate their perceived mental effort and self-efficacy. After the training phase, participants completed the self-efficacy, perceived competence, and topic interest questionnaires again. Lastly, participants took the immediate posttest. Participants had to rate their invested mental effort after each posttest task. Participants handed in their scrap paper before working on the posttest phase and received new ones to make notes.

The delayed posttest took place exactly 7 days later (ca. 40 minutes) and started with a general introduction in which the procedure was explained. Again, participants were told they could work at their own pace, write down everything they could, and note an X if they were not able to answer a question. Participants were provided with scrap paper and a password that gave them access to the online learning environment. They first completed the self-efficacy, perceived competence, and topic interest questionnaires. Subsequently, they took the delayed posttest, which consisted of four tasks that were isomorphic to the tasks used in the immediate posttest phase. After each task, participants were asked to indicate their invested mental effort.

3.2.1.4. Data analysis

The data was scored by the experimenter (i.e., first author) and a second encoder based on a scoring protocol that was developed by Van Harsel et al. (2019) in collaboration with higher education mathematics teachers. Participants could earn a maximum of 8 points per training problem. Two points could be earned for calculating the step size of each subinterval, two for correctly calculating all x-values, two for correctly calculating the function values for all x-values, and two for using the correct formula for the area under the graph and providing the correct answer. If half or more of the solution steps were correct in step two, three, and four, then one point was granted. If less than half of the solution steps were correct in step two, three and four, 0 points were granted. These scoring standards were also used to score the two isomorphic posttest tasks (i.e., max. score = 16 points) and the procedural transfer problem (i.e., max. score = 8 points). The intra-class correlation coefficient was .98 for the training tasks, .98 for the isomorphic posttest tasks, and .93 for the delayed posttest tasks.

The average mental effort invested in the training phase and on the isomorphic posttest tasks was calculated. In addition, the average self-efficacy, perceived competence, and topic interest ratings were calculated.

3.2.2. Results

Nonparametric tests were used to analyze our main research questions and explorative questions, because with the exception of topic interest on pretest and delayed posttest, and self-efficacy and perceived competence on the delayed posttest, none of our main variables were normally distributed (cf. Field, 2009), with either the kurtosis, skewness, or both coefficients being (substantially) below -1.96 or above +1.96. Therefore, effects of Instruction Condition (EEEE, EPEP, PEPE, and PPPP) were tested on motivational (i.e., self-efficacy, perceived competence, and topic interest) and cognitive aspects of learning (i.e., isomorphic test performance, procedural transfer, conceptual transfer, mental effort and time-on-task in learning and posttest phases) with Kruskal-Wallis tests. Significant main effects of Instruction Condition were followed by six Mann-Whitney U tests (EEEE vs. EPEP, EEEE vs. PEPE, EEEE vs. PPPP, EPEP vs. PEPE, EPEP vs. PPPP, and PEPE vs. PPPP) with a Bonferroni-corrected significance level of $p < .008$ (i.e., $0.05/6$). Results are presented in the main text and Table 3.6. Effects of Test Moment (Immediate Posttest and Delayed Posttest) for each condition (EEEE, EPEP, PEPE, and PPPP) were tested with Wilcoxon signed-rank tests and we used four Mann-Whitney U tests as post-hoc tests (see Table 3.6), with a Bonferroni corrected significance level of $p < .013$ (i.e., $0.05/4$). The effect size of *Pearson r correlation* is reported (i.e., Z/\sqrt{N}) with values of 0.10, 0.30, and 0.50 representing a small, medium, and large effect size, respectively (Cohen, 1988) for the post-hoc tests. The self-efficacy, perceived competence, and topic interest scores can be found in Table 3.4, and the test performance scores, mental effort scores, and time-on-task scores in Table 3.5.

Before the differences within and among conditions were analyzed, we checked for prior knowledge differences. Kruskal-Wallis tests showed no significant differences among conditions on pretest performance, $H(3) = 2.58$, $p = .460$, or on pretest scores of self-efficacy, $H(3) = 2.59$, $p = .460$, perceived competence, $H(3) = 2.18$, $p = .536$, and topic interest, $H(3) = 3.22$, $p = .360$.

3.2.2.1. How do short sequences of examples and problems affect self-efficacy, perceived competence, and topic interest?

Self-efficacy. Self-efficacy ratings measured after each training task are presented in Figure 3.1. It was analyzed whether participants' self-efficacy reported after each training task differed among conditions (see Table 3.3 for post-hoc comparisons). With regard to the *first training task*, there was a main effect of Instruction Condition, $H(3) = 83.13$, $p < .001$. As predicted (H1a), self-efficacy levels were higher in the EEEE and EPEP Condition than the PEPE and PPPP Condition. No significant differences were found between the EEEE and EPEP Condition or between the PEPE and PPPP Condition.

Regarding self-efficacy from the *second training task onwards*, there was also a main effect of Instruction Condition (task 2: $H(3) = 59.48$, $p < .001$; task 3: $H(3) = 68.37$, $p < .001$; task 4: $H(3) = 68.61$, $p < .001$). As expected (H1b, H1c), results showed that for all three tasks the self-efficacy ratings were higher in the EEEE, EPEP and PEPE Condition compared to the PPPP Condition. No differences were found, however, between the EPEP and PEPE Condition. Self-efficacy ratings were also higher after task 2 and task 3 in the EEEE Condition compared to the EPEP and PEPE Condition, but not after training task 4.

Analyses of participants' self-efficacy after the training phase revealed a main effect of Instruction Condition, $H(3) = 66.55$, $p < .001$, and self-efficacy ratings were higher in the EEEE, EPEP, and PEPE Condition compared to the PPPP Condition. No significant differences were found between the EEEE, EPEP and PEPE Condition. Measuring self-efficacy at the start of the delayed posttest phase revealed the same pattern of results. There was a main effect of Instruction Condition, $H(3) = 46.08$, $p < .001$, and follow-up tests showed that self-efficacy scores were higher in the EEEE, EPEP, and PEPE Condition compared to the PPPP Condition. Again, there was no significant difference between EEEE and EPEP or between EPEP and PEPE.

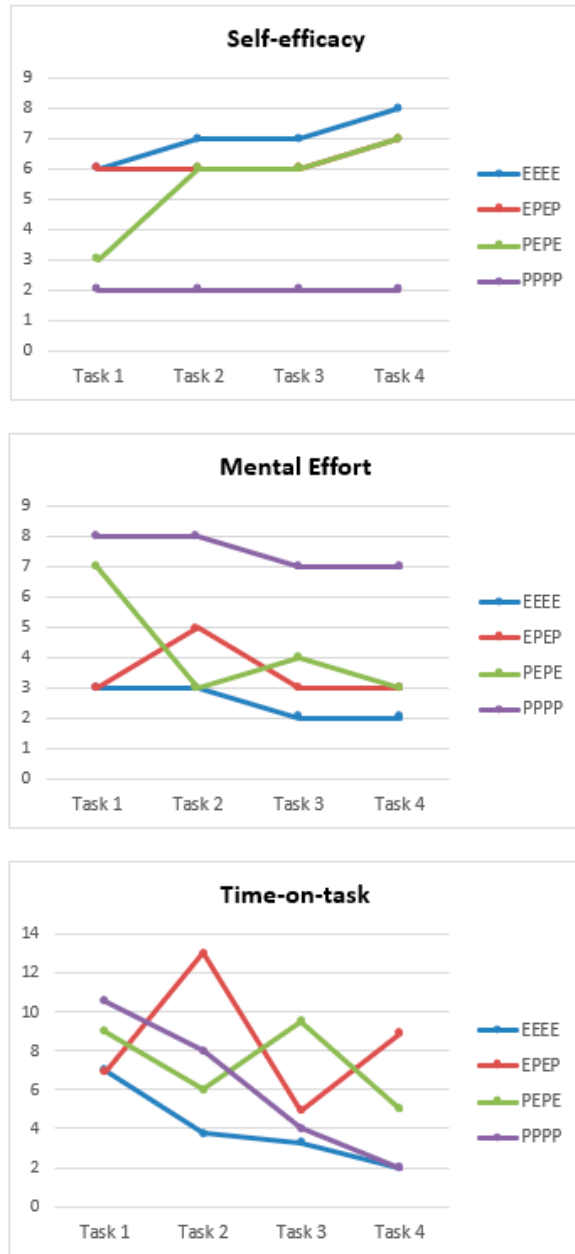


Figure 3.1. Median scores on self-efficacy (top row; range 1 to 9) and mental effort (top row; range 1 to 9) and time-on-task for each training task in Experiment 1.

Table 3.3.
Post-Hoc Comparisons of Self-Efficacy Reported after each Training Task
(see Figure 3.1) in Experiment 1.

T	EE vs. EP			EE vs. PE			EP vs. PE			EP vs. PP			PE vs. PP		
	U	p	r	U	p	r	U	p	r	U	p	r	U	p	r
Training task 1	524	.022	.258	80	<.001	.761	79	<.001	.760	198	.678	.191	.680	.716	.520
Training task 2	413	.001	.384	430	.008	.309	96.5	<.001	.740	1041.5	.140	307.5	.567	186.5	<.001
Training task 3	441.5	.002	.355	359	.001	.399	70.5	<.001	.772	810.5	.087	175	.698	193	<.001
Training task 4	506	.015	.276	479	.039	.242	68	<.001	.775	986	.840	173	.696	113.5	<.001
															.744

Note. Acronyms for groups: EE = example study only; EP = example-problem pairs; PE = problem-example pairs; PP = problem solving only.

Perceived competence. Analysis of perceived competence measured after the training phase showed a main effect of Instruction Condition, $H(3) = 67.41, p < .001$. Perceived competence was higher in the EEEE, EPEP, and PEPE Condition than in the PPPP Condition, and scores in the EPEP and PEPE Condition did not differ significantly. There was also no significant difference between the EEEE and EPEP Condition. The pattern of results was similar for the delayed posttest. There was a main effect of Instruction Condition, $H(3) = 41.19, p < .001$, as perceived competence was higher in the EEEE, EPEP, and PEPE Condition than in the PPPP Condition. There was no statistically significant difference between the EEEE and EPEP Condition or the EPEP and PEPE Condition.

Topic interest. There was a main effect of Instruction Condition, $H(3) = 8.93, p = .030$, and there were no differences between the EEEE and EPEP Condition or between the EPEP and PEPE Condition. However, results showed that topic interest scores were lower in the EEEE than in the PPPP Condition. As for topic interest measured before the delayed posttest, there was no main effect of Instruction Condition.

3.2.2.2. How do short sequences of examples and problems affect learning and transfer?

Isomorphic test tasks. Analyzing whether performance on the isomorphic tasks on the immediate posttest differed among conditions showed a main effect of Instruction Condition, $H(3) = 36.63, p < .001$. Results showed that the EEEE, EPEP, and PEPE Condition scored significantly higher than the PPPP Condition. No differences were found between the EEEE and EPEP, EPEP and PEPE, or EEEE and PEPE Condition.

The pattern of results was the same for the isomorphic tasks on the delayed posttest. There was a main effect of Instruction Condition, $H(3) = 24.76, p < .001$, and follow up tests showed that performance on the isomorphic tasks was significantly higher for the EEEE, EPEP, and PEPE Condition than the PPPP Condition. No differences were found between the EEEE and EPEP, EPEP and PEPE Condition, or EEEE and PEPE Condition.

Procedural transfer task. Analyzing whether performance differed among conditions on the procedural transfer task revealed a main effect of Instruction Condition, $H(3) = 27.41, p < .001$. Results showed that the EEEE, EPEP, and PEPE Condition significantly outperformed the PPPP Condition. No differences were found, however, in the other condition comparisons. On the delayed posttest, there was a main effect of Instruction Condition, $H(3) = 10.58, p = .014$, and follow-up tests showed that only the EEEE and PEPE Condition, but not the EPEP Condition scored significantly higher than the PPPP Condition on procedural transfer. Again, other comparisons were not significant.

3.2.2.3. How do short sequences of examples and problems affect mental effort and time-on-task in the training phase?

Mental effort. Mental effort ratings measured after each training task (see Figure 3.1) were used as a measure of learning efficiency. Results showed a main effect of Instruction Condition for self-reported effort ratings invested in the training tasks, $H(3) = 64.19, p < .001$, and the EEEE, EPEP, and PEPE Condition reported less effort during the training phase than the PPPP Condition. Moreover, the EEEE Condition reported less effort than the EPEP and PEPE Condition. Finally, the EPEP Condition also reported significantly less effort than the PEPE Condition.

Time-on-task. Time-on-task invested in each task in the training phase is presented in Figure 3.1 and exploratory analyses are presented in the Supplementary Materials K.

3.2.2.4. How do short sequences of examples and problems affect mental effort and time-on-task in the posttest phases?

Exploratory analyses of mental effort and time-on-task invested in the posttest phases are presented in the Supplementary Materials K.

3.2.3. Discussion

Regarding the main aim of uncovering how self-efficacy develops during the training phase, results showed, as expected, that self-efficacy was reported to be significantly higher after the first task for the example-first conditions compared to the problem-first conditions (i.e., EEEE and EPEP > PEPE and PPPP). Throughout the rest of the training phase (i.e., tasks 2 to 4), all example conditions reported significantly higher self-efficacy than the problem solving only condition, and the EEEE condition reported higher self-efficacy ratings than the EPEP and PEPE condition with regards to training task 2 and 3.

Furthermore, we (partly) replicated the results of Van Harsel et al. (2019) regarding motivational and cognitive aspects of learning measured after the training phase. All example conditions showed higher self-efficacy and perceived competence ratings and test performance (i.e., isomorphic and transfer tasks), while investing less mental effort in the training phase compared to the PPPP condition. All example conditions showed lower effort investment but longer time investment on the isomorphic posttest tasks during the immediate posttest than the PPPP condition. This pattern remained stable on the delayed posttest. Topic interest scores were lower in the EEEE than the PPPP condition on the immediate posttest, but this difference was no longer present on the delayed measurement. There were also no other differences among conditions on topic interest. Importantly, we found no differences on motivational variables (i.e., self-efficacy, perceived competence, or topic interest) or on posttest performance between the EEEE and EPEP, or between the EPEP and PEPE condition. We did find that

reported effort investment in the training phase was lower in the EEEE condition than in the EPEP (and PEPE) condition. Effort invested in the training phase was also significantly lower in the EPEP condition than in the PEPE condition.

The results of Experiment 1 provide some evidence for the motivational explanation of differences between EP and PE on learning. Starting the training phase with a practice problem (PE) affected self-efficacy negatively compared to starting with an example. However, this did not lead students in the PE condition to disengage in the present study; they studied the example and after that, their self-efficacy increased to the level of the EP (and EE) condition.

It is an important open question whether the findings on both cognitive and motivational aspects of learning would be different when the training phase is longer (i.e., consists of more training tasks). For example, one might expect that passively studying examples would become redundant and (therefore) boring when task sequences are longer, which in turn might lead to disengagement and lower learning outcomes. Hence, example-problem pairs might be more engaging and effective than example study only, because example-problem pairs provide the benefits of examples but also allow students to actively apply what they have learned. Therefore, a second experiment was conducted with the aim to investigate how motivational and cognitive aspects of learning would be affected by longer task sequences of examples and problems (i.e., 8 instead of 4 tasks: EEEEEEEE, EPEPEPEP, PEPEPEPE, and PPPPPPPP).

Table 3.4.
Mean (M), Standard Deviation (SD) and Median (Med) of Self-Efficacy (range 1 to 9) Perceived Competence (range 1 to 7), and Topic Interest (range 1 to 7) per Condition in Experiment 1.

	EEEE Condition			EPEP Condition			PEPE Condition			PPPP Condition		
	M	SD	Med	M	SD	Med	M	SD	Med	M	SD	Med
Pretest												
Self-efficacy	2.18	1.84	1.00	2.40	1.86	2.00	2.33	1.31	2.00	2.00	1.34	1.00
Perceived Competence	1.77	1.28	1.33	2.17	1.48	1.67	1.98	1.11	1.67	2.07	1.20	1.67
Topic Interest	4.57	0.78	4.86	4.43	0.73	4.29	4.45	0.84	4.36	4.23	0.89	4.43
Training												
Self-efficacy	7.09	1.39	7.26	6.06	1.36	6.00	5.53	1.11	5.38	2.72	1.90	2.00
Immediate Posttest												
Self-efficacy	7.39	1.27	7.00	6.73	1.64	7.00	7.10	1.28	7.00	2.79	2.19	2.00
Perceived Competence	5.83	0.88	6.00	5.35	1.30	5.67	5.66	0.87	6.00	2.29	1.61	2.00
Topic Interest	4.68	0.86	4.86	4.45	0.93	4.43	4.50	0.98	4.57	4.03	0.98	4.29
Delayed Posttest												
Self-efficacy	5.12	1.59	6.00	5.18	1.66	5.00	5.69	1.17	6.00	2.39	1.69	2.00
Perceived Competence	4.37	1.32	4.67	4.42	1.23	4.50	4.69	0.98	4.83	2.24	1.52	1.67
Topic Interest	4.26	0.97	4.14	4.13	0.88	4.00	4.11	0.86	4.00	3.95	0.86	4.14

Note. Acronyms for groups: EE = example study only; EP = example-problem pairs; PE = problem-example pairs; PP = problem solving only.

Table 3.5
Mean (M), Standard Deviation (SD), and Median (Med) of Pretest (range 0 to 16), Isomorphic Tasks Performance (range 0 to 16), Procedural Transfer (range 0 to 8), Mental Effort (range 1 to 9), and Time-on-Task per Condition in Experiment 1.

	EEEE Condition			EPEP Condition			PEPE Condition			PPPP Condition		
	M	SD	Med	M	SD	Med	M	SD	Med	M	SD	Med
Pretest												
Performance	2.94	2.03	4.00	2.31	1.41	2.00	2.60	1.63	3.00	2.46	1.59	2.00
Training												
Mental Effort	2.57	1.05	2.50	3.42	1.18	3.25	4.21	0.96	4.13	6.44	2.41	6.75
Time-on-Task	4.35	1.63	4.50	8.68	5.07	11.00	7.67	2.07	7.00	6.27	5.02	5.50
Immediate Posttest												
Isomorphic Tasks	9.67	4.06	10.00	9.89	5.07	11.00	10.20	3.34	10.50	3.77	4.64	2.00
Procedural Transfer	1.91	2.34	1.00	1.73	1.68	1.00	1.63	1.53	1.00	0.33	0.74	0.00
Mental Effort												
Isomorphic Tasks	4.89	1.52	5.00	4.73	1.69	4.50	4.94	1.38	5.00	6.51	2.56	7.00
Procedural Transfer	5.36	2.41	5.00	5.98	2.15	6.00	5.10	2.37	5.00	6.62	2.56	8.00
Time-on-Task												
Isomorphic Tasks	16.87	6.39	14.50	10.61	4.99	10.50	11.90	3.34	11.25	4.99	4.79	4.00
Procedural Transfer	9.27	4.87	9.00	8.38	5.29	8.00	7.88	4.29	7.00	3.87	4.13	2.00
Delayed Posttest												
Isomorphic Tasks	9.28	5.30	11.00	9.60	4.42	10.00	10.00	4.16	10.50	4.16	4.75	2.00
Procedural Transfer	1.32	1.70	1.00	1.15	1.53	1.00	1.08	1.23	1.00	0.52	1.48	0.00
Mental Effort												
Isomorphic Tasks	4.80	1.90	4.00	4.55	1.52	4.50	4.81	1.65	5.00	6.76	2.00	7.50
Procedural Transfer	5.36	2.33	5.00	5.23	2.07	5.00	5.03	2.18	5.00	6.71	2.52	8.00
Time-on-Task												
Isomorphic Tasks	12.56	4.48	12.00	11.69	4.82	11.50	10.85	4.37	10.50	7.31	5.29	7.50
Procedural Transfer	7.52	4.72	6.00	7.45	4.91	7.00	7.53	3.56	8.00	4.71	4.17	5.00

Note. Acronyms for groups: EE = example study only; EP = example-problem pairs; PE = problem-example pairs; PP = problem solving only.

Table 3.6.
Post-hoc comparisons of Mental Effort, Self-Efficacy, Perceived Competence, Topic Interest, Isomorphic Tasks Performance, and Procedural Transfer on Immediate and Delayed Posttests in Experiment 1.

	EE vs. EP		EE vs. PE		EE vs. PP		EP vs. PE		EP vs. PP		PE vs. PP	
	U	p	U	p	U	p	U	p	U	p	U	p
Training												
Mental Effort	1035.5	.003	.337	1309	<.001	.392	.039	232	1179.5	<.001	.651	1186
Immediate posttest												
Isomorphic Tasks ¹	800.5	.556	.067	690	.738	.039	.026	363	850.5	.662	.047	233.5
Procedural Transfer ²	774	.742	.037	679	.826	.026	.132	76	874	.810	.026	332
Self-efficacy ³	574.5	.082	.765	562	.260	.093	.765	997	171	.381	.095	109
Perceived Competence ⁴	609.5	.175	.154	589	.425	.093	.769	1000	151.5	.373	.097	92
Topic Interest ⁵	618	.207	.143	562.5	.279	.127	.336	942	679.5	.711	.040	574.5
Delayed post-test												
Isomorphic Tasks ¹	503	.986	.005	470	.769	.038	.457	754.5	256.5	.719	.041	206
Procedural Transfer ²	471	.677	.052	442.5	.907	.002	.353	749.5	426	.743	.038	341
Self-efficacy ³	482.5	.808	.030	518.5	.291	.017	.642	869.5	146	.106	.185	85
Perceived Competence ⁴	475.5	.739	.014	503	.432	.013	.832	832	168	.241	.135	113
Topic Interest ⁵	464.5	.631	.060	418	.638	.060	.723	723	566.5	.975	.004	532

Note. Acronyms for groups: EE = example study only; EP = example-problem pairs; PE = problem-example pairs; PP = problem solving only.

1 Isomorphic task performance did not differ statistically between the immediate and delayed posttest ($Z = 2821.5$, $p = .766$, $r = .026$).
2 Procedural transfer task performance statistically differed between the immediate and delayed posttest ($Z = 739.5$, $p = .006$, $r = .239$), however, follow-up tests showed that changes within conditions were not significant ($p > .031$, $n < 359$).
3 Self-efficacy statistically differed between the pretest and immediate posttest ($Z = 9.85$, $p < .001$, $r = .786$) and increased in EE, EP, and PE Condition ($p < .001$), not in the PP Condition ($p = .015$).
4 Self-efficacy statistically differed between the immediate and delayed posttest ($Z = 714$, $p < .001$, $r = .627$) and decreased in EE, EP, and PE Condition ($p < .001$), not in PP Condition ($p = .954$).
5 Perceived competence statistically differed between the pretest and immediate posttest ($Z = 9.92$, $p < .001$, $r = .760$) and increased in EE, EP, and PE Condition ($p < .001$), not in PP Condition ($p = .015$).
6 Perceived competence statistically differed between the immediate and delayed posttest ($Z = -6.034$, $p < .001$, $r = .625$) and decreased in EE, EP, and PE Condition ($p < .001$), not in PP Condition ($p = .954$).
7 Topic interest did not differ statistically between the pretest and immediate posttest ($p = .736$, $r = .325$). Topic interest statistically differed between the immediate and delayed posttest ($Z = -5.32$, $p < .001$, $r = .463$) and decreased in EP and PE Condition ($p < .011$), but not in EE Condition ($p = .147$) and PP Condition ($p = .030$).

3.3. Experiment 2

In Experiment 2, we investigated how longer task sequences of examples and/or practice problems (i.e., EEEEEEEE, EPEPEPEP, PEPEPEPE, and PPPPPPPP) would affect motivational (i.e., self-efficacy, perceived competence, and topic interest measured before and after the training phase) and cognitive aspects of learning (i.e., invested mental effort in the training phase). Time-on-task in the training phase, as well as mental effort and time-on-task in the posttest phases were again measured as (explorative) indicators of efficiency of the learning process and learning outcomes (Van Gog & Paas, 2008). Because example study only might become redundant and boring when task sequences are longer and therefore might lead to disengagement and lower performance scores, we expected that the EPEPEPEP condition would show significantly higher levels of self-efficacy (H2), perceived competence (H3), and topic interest (H4) after the training phase than the EEEEEEEE condition, and that the EPEPEPEP condition would attain higher levels of isomorphic posttest performance (H5), procedural transfer performance (H6), and conceptual transfer performance (H7), while investing less effort in the training phase (H8) compared to the EEEEEEEE condition. All other comparisons were considered exploratory.

3.3.1. Method

3.3.1.1. Participants and design

Participants were 105 Dutch higher education students in their first year of an electrical and electronic, mechanical engineering, or mechatronics program ($M^{age} = 19.30$, $SD = 1.80$; 105 male). Participants were randomly assigned to one of four conditions and received eight training tasks: 1) examples only ($n = 32$; EEEEEEEE), 2) example-problem pairs ($n = 28$; EPEPEPEP), 3) problem-example pairs ($n = 23$; PEPEPEPE), or 4) practice problems only ($n = 22$; PPPPPPPP). The experiment consisted of three phases: (1) pretest, (2) training phase, and (3) immediate posttest phase. At the time of the experiment, participants were novices to the modelled task as this subject had not (yet) been a part of their study program. Participants gave their informed consent prior to their inclusion in the study and received study credits for their participation.

3.3.1.2. Materials and procedure

The materials were presented using a web-based learning environment. The materials, procedure, and data analysis were the same as in Experiment 1 with the following exceptions. First, the training phase consisted of eight tasks; in addition to the four tasks also used in Experiment 1 two additional pairs of tasks were added. All eight tasks were paired based on their complexity (i.e., pair 1: fitness and energy measurement, pair 2: washing machine and soapsuds, pair 3: drinking water and running, and pair 4: the carousel and coffee consumption). The first pair of tasks required participants to calculate with positive numbers. The second and third pair of tasks were slightly more complex because participants had to calculate with both positive and negative numbers. The fourth pair of tasks was most complex and asked participants to calculate with a cubic function (polynomial of degree 3) instead of the quadratic function (polynomial of degree 2) that was used in the first three task pairs. The design of the formats (i.e., video modeling examples and practice problems) was similar to the formats used in Experiment 1. Second, the immediate posttest consisted of five instead of four tasks as in Experiment 1. Three isomorphic posttest tasks were used ($\alpha = .73$): one isomorphic to the first pair of training tasks, one to the second and third pair of training tasks, and one to the fourth pair of training tasks. The fourth task was a procedural transfer task (i.e., Simpson rule), followed by the conceptual transfer questions ($\alpha = .59$).

The procedure was the same as in Experiment 1, with the exception that Experiment 2 did not have a delayed posttest (i.e., in Experiment 1, results were consistent across both test moments and therefore we did not include a delayed posttest). This resulted in 10 single sessions with 2 – 21 participants per session that lasted ca. 116 minutes. As for the data analysis, we used the same scoring standards as in Experiment 1 for the training tasks, the three isomorphic posttest tasks (max. score = 24 points), and the procedural transfer task. Regarding the five conceptual transfer questions, participants could earn a maximum of 9 points: one point for the first open-ended question (0 points for an incorrect answer; 1 point for the correct answer) and 2 points for the other open-ended questions (0 points for an incorrect answer; 1 point for the correct answer, 2 points for the correct answer and a correct explanation).

3.3.2. Results

Again, with the exception of pretest performance and topic interest on the immediate posttest, all of the main variables were not normally distributed, with either the kurtosis, skewness, or both coefficients being (substantially) below -1.96 or above +1.96. Relevant descriptive statistics of self-efficacy, perceived competence, and topic interest scores are presented in Table 3.8, and performance scores, mental effort scores, and time-on-task scores are presented in Table 3.9. Again, we used Mann-Whitney U tests as post-hoc tests (see Table 3.10). Kruskal-Wallis tests showed that there were no significant differences among conditions on pretest performance, $H(3) = 2.86$, $p = .414$, and pretest scores of self-efficacy, $H(3) = 3.94$, $p = .268$, perceived competence, $H(3) = 3.42$, $p = .331$, and topic interest, $H(3) = 1.29$, $p = .731$.

3.3.2.1. How do longer sequences of examples and problems affect self-efficacy, perceived competence, and topic interest?

Self-efficacy. Self-efficacy ratings measured after each training task are presented in Figure 3.2. First it was explored whether self-efficacy ratings reported after each training task differed among conditions (see Table 3.7 for post-hoc comparisons). With regard to the *first training task*, there was a main effect of Instruction Condition, $H(3) = 33.45$, $p < .001$, and self-efficacy levels were higher in the EEEEEEEE and EPEPEPEP Condition than the PEPEPEPE and PPPPPPPP Condition. There were no significant differences between the EEEEEEEE and EPEPEPEP Condition or between the PEPEPEPE and PPPPPPPP Condition.

There was also a main effect of Instruction Condition for the *second training task onwards* (task 2: $H(3) = 18.58$, $p < .001$; task 3: $H(3) = 29.12$, $p < .001$; task 4: $H(3) = 32.35$, $p < .001$; task 5: $H(3) = 28.00$, $p < .001$; task 6: $H(3) = 29.52$, $p < .001$; task 7: $H(3) = 30.42$, $p < .001$; task 8: $H(3) = 30.69$, $p < .001$). Results showed that the self-efficacy scores were higher in the EEEEEEEE, EPEPEPEP, and PEPEPEPE Condition compared to the PPPPPPPP Condition. No differences were found, however, between the EPEPEPEP and PEPEPEPE Condition. Also, no differences were found between the EEEEEEEE and EPEPEPEP Condition, except for training task 8, where self-efficacy ratings were higher in the EEEEEEEE than EPEPEPEP Condition.

Concerning the main question of whether there would be differences among conditions on self-efficacy ratings measured after the training phase, there was a main effect of Instruction Condition, $H(3) = 29.49$, $p < .001$. Self-efficacy ratings were significantly higher in the EEEEEEEE, EPEPEPEP, and PEPEPEPE Condition compared to the PPPPPPPP Condition. Contrary to our expectations (H2), there were no differences between the EPEPEPEP and EEEEEEEE Condition. Further explorations showed that no other condition comparisons were significant.

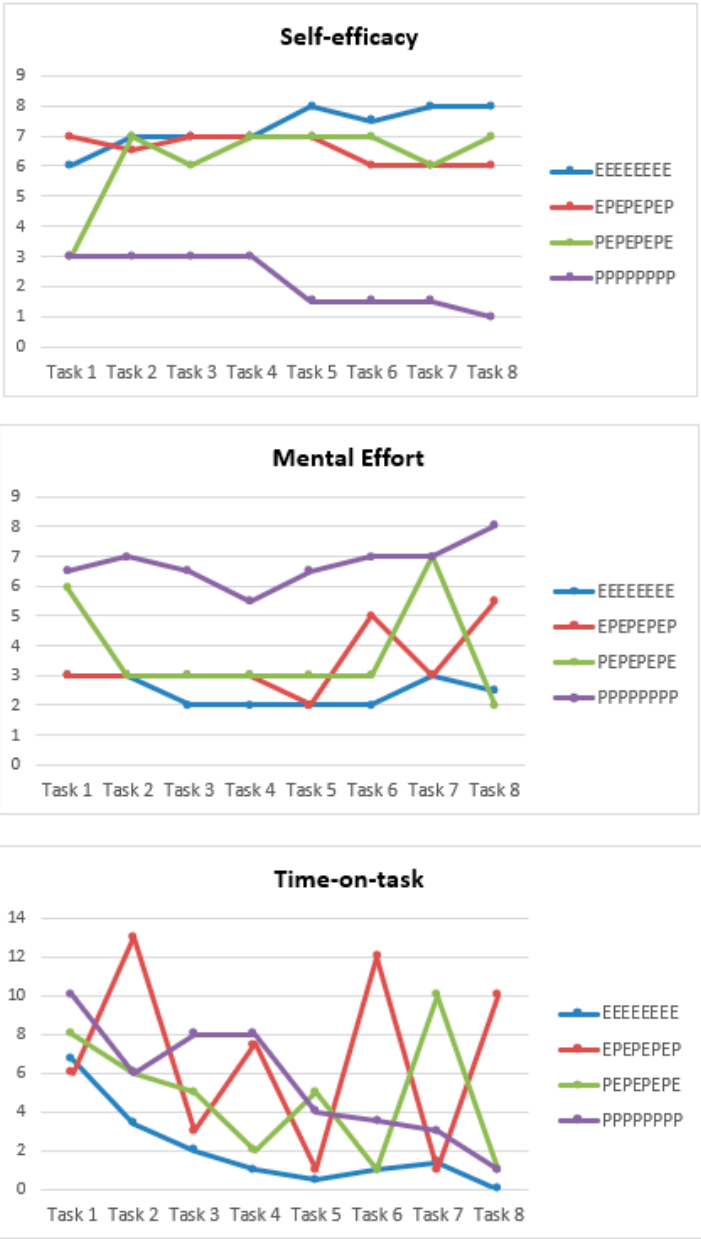


Figure 3.2. Median scores on self-efficacy (top row; range 1 to 9) and mental effort (top row; range 1 to 9) and time-on-task for each training task in Experiment 2.

Table 3.7.

Post-Hoc Comparisons of Self-Efficacy Reported after each Training Task (see Figure 3.2) in Experiment 2.

	EE vs. EP			EE vs. PE			EE vs. PP			EP vs. PE			EP vs. PP			PE vs. PP		
	U	p	r	U	p	r	U	p	r	U	p	r	U	p	r	U	p	r
Training task 1	566	.075	.230	186	.002	.423	138.5	<.001	.516	105.5	<.001	.581	71.5	<.001	.661	220.5	.455	.112
Training task 2	419.5	.668	.055	349.5	.749	.043	138.5	<.001	.516	323.5	.997	.004	130.5	<.001	.496	105.5	.001	.505
Training task 3	443.5	.946	.009	274.5	.105	.219	97.5	<.001	.615	223.5	.058	.266	74.5	<.001	.651	100	<.001	.523
Training task 4	426	.738	.043	319	.393	.115	74	<.001	.672	287	.494	.096	66	<.001	.677	69	<.001	.629
Training task 5	405.5	.519	.083	339	.613	.068	83.5	<.001	.650	323.5	.977	.004	100	<.001	.582	69	<.001	.618
Training task 6	279.5	.011	.328	322	.423	.108	85	<.001	.630	397.5	.144	.205	110	<.001	.557	83.5	<.001	.583
Training task 7	274.5	.009	.338	233	.019	.316	83.5	<.001	.647	317	.923	.013	112	<.001	.548	99.5	<.001	.527
Training task 8	265.5	.006	.356	299	.230	.162	93.5	<.001	.629	403.5	.116	.220	115.5	<.001	.553	78.5	<.001	.599

Note. Acronyms for groups: EE = example study only; EP = example-problem pairs; PE = problem-example pairs; PP = problem solving only.

Perceived competence. The pattern of results was similar for perceived competence. There was a main effect of Instruction Condition regarding perceived competence measured after the training phase, $H(3) = 23.83$, $p < .001$, and the EEEEEEEE, EPEPEPEP, and PEPEPEPE Condition showed higher perceived competence ratings than the PPPPPPPP Condition. In contrast to our expectations (H3), there was no difference between the EEEEEEEE and EPEPEPEP Condition ($p = .799$, $r = .033$). Further explorations revealed that no other comparisons were significant.

Topic interest. Analyzing whether conditions differed in topic interest scores measured after the training phase revealed a main effect of Instruction Condition, $H(3) = 8.30$, $p = .040$, however, follow-up tests showed no significant differences among any of the condition comparisons (H4).

3.3.2.2. How do longer sequences of examples and problems affect learning and transfer?

Isomorphic test tasks. Analysis revealed a main effect of Instruction Condition for performance on the isomorphic posttest tasks, $H(3) = 12.86$, $p = .005$. Results showed that the EEEEEEEE Condition showed significantly higher performance on the isomorphic test tasks than the PPPPPPPP Condition. However, the EPEPEPEP and PEPEPEPE Condition did not significantly differ from the PPPPPPPP Condition. Although we expected EPEPEPEP > EEEEEEEE (H5), there were no performance differences on the isomorphic posttest tasks between the EEEEEEEE and EPEPEPEP Condition. Our explorative analyses showed no other condition comparisons were significant.

Procedural transfer task and conceptual transfer questions. Subsequently, we analyzed whether conditions differed in scores on the procedural transfer task and conceptual transfer questions (H6, H7). Analysis showed there was no main effect of Instruction Condition for the procedural transfer task, $H(3) = 6.04$, $p = .110$, and for the conceptual transfer questions, $H(3) = 2.85$, $p = .415$.

3.3.2.3. How do longer sequences of examples and problems affect mental effort and time-on-task in the training phase?

Mental effort. The average of self-reported effort investment after each task in the training phase (see Figure 3.2) was analyzed as a measure of efficiency. There was a main effect of Instruction Condition, $H(3) = 34.85$, $p < .001$, and the EEEEEEEE, EPEPEPEP, and PEPEPEPE Condition invested less effort in the training tasks than the PPPPPPPP Condition. As expected (H8), the EEEEEEEE Condition invested significantly less effort in the training tasks compared to the EPEPEPEP Condition, and less effort than the PEPEPEPE Condition. No differences were found between the EPEPEPEP and PEPEPEPE Condition.

Time-on-task. Time-on-task invested in each task in the training phase is presented in Figure 3.2 and exploratory analyses are presented in the Supplementary Materials L.

3.3.2.3. How do short sequences of examples and problems affect mental effort and time-on-task in the posttest phase?

Exploratory analyses of mental effort and time-on-task invested in the posttest phase are presented in the Supplementary Materials L.

3.3.3. Discussion

The main aim of Experiment 2 was to investigate how longer training task sequences of examples and problems (i.e., EEEEEEEE, EPEPEPEP, PEPEPEPE, and PPPPPPPP) would affect motivational and cognitive variables. It was expected that example study only would result in lower scores on performance and motivational variables than example-problem pairs. In contrast to our hypotheses, however, there were no motivational or test performance differences between the EEEEEEEE and EPEPEPEP condition. As hypothesized, the effort that students reported to invest in the training phase was lower in the EEEEEEEE than the EPEPEPEP condition. However, exploring effort on the posttest phase revealed that levels of perceived effort when solving the isomorphic posttest tasks were higher in EEEEEEEE than EPEPEPEP. This might be explained by the fact that students in the EEEEEEEE condition did not have the opportunity to practice problem solving in the training phase, whereas the EPEPEPEP condition did have the opportunity to practice problem solving in the training phase and therefore could apply and automate the procedure several times.

With regard to our exploratory question of how the other conditions would compare to each other, the pattern of results regarding motivational aspects of learning was similar as in Experiment 1. Our exploration of self-efficacy during the training phase showed that there were differences in self-efficacy ratings between the conditions starting with an example and the conditions starting with a practice problem (i.e., EEEEEEEE, EPEPEPEP > PEPEPEPE, PPPPPPPP) regarding the first training task. From the second training task onwards, however, self-efficacy ratings in the PEPEPEPE condition increased to the same level as in the conditions starting with an example, whereas self-efficacy in the PPPPPPPP condition remained low. This pattern of results remained stable during and after the training phase and was also similar for perceived competence. There were no differences among conditions on topic interest.

Regarding performance, only the EEEEEEEE condition significantly outperformed the PPPPPPPP condition on isomorphic test performance, and there was no effect of condition on procedural transfer and conceptual transfer. All example conditions were more efficient in the sense that they reported to invest less effort in the training phase than the PPPPPPPP condition. Again, the EEEEEEEE condition was most efficient considering that they reported to invest the lowest effort levels (and time-on-task) in the training phase. Lastly, no differences in motivational aspects of learning, test performance, or effort investment were found between the EPEPEPEP and PEPEPEPE condition.

Table 3.8.
Mean (M), Standard Deviation (SD) and Median (Med) of Self-Efficacy (range 1 to 9) Perceived Competence (range 1 to 7), and Topic Interest (range 1 to 7) per Condition in Experiment 2.

	EEEEEEEE Condition			EPEPEPEP Condition			PEPEPEPE Condition			PPPPPPPP Condition		
	M	SD	Med	M	SD	Med	M	SD	Med	M	SD	Med
Pretest												
Self-efficacy	2.50	1.85	2.00	1.93	1.09	2.00	2.91	1.88	2.00	2.59	1.56	2.50
Perceived Competence	2.23	1.41	2.00	1.73	1.00	1.00	2.36	1.54	2.00	1.98	0.91	2.00
Topic Interest	4.30	0.87	4.43	4.35	0.70	4.43	4.47	0.91	4.57	4.43	0.81	4.43
Training												
Self-efficacy	6.94	1.45	7.13	6.57	1.19	6.50	6.18	1.57	5.88	3.32	2.24	2.36
Posttest												
Self-efficacy	7.03	1.38	7.00	6.29	1.63	6.00	6.52	1.86	7.00	3.05	2.54	2.00
Perceived Competence	5.47	1.94	5.67	5.50	1.40	5.67	5.41	1.25	6.00	2.86	2.04	2.00
Topic Interest	4.51	0.69	4.57	4.39	0.68	4.50	3.87	1.00	4.14	4.04	0.95	4.21

Note. Acronyms for groups: EE = example study only; EP = example-problem pairs; PE = problem-example pairs; PP = problem solving only.

Table 3.9.
Mean (M), Standard Deviation (SD), and Median (Med) of Pretest (range 0 to 16), Isomorphic Tasks Performance (range 0 to 24), Procedural Transfer (range 0 to 8), Conceptual Transfer (range 0 to 9), Mental Effort (range 1 to 9), and Time-on-Task per Condition in Experiment 2.

	EEEEEEEE Condition			EPEPEPEP Condition			PEPEPEPE Condition			PPPPPPPP Condition		
	M	SD	Med	M	SD	Med	M	SD	Med	M	SD	Med
Pretest												
Performance	2.03	1.33	2.00	2.00	1.12	2.00	2.74	1.81	3.00	2.36	1.94	2.50
Training												
Mental Effort	2.70	1.22	2.56	3.65	1.23	3.81	3.80	1.36	3.75	6.06	2.07	6.31
Time-on-Task	2.50	1.29	2.25	7.86	3.16	7.63	5.51	2.51	5.00	6.51	4.26	5.38
Posttest												
Isomorphic Tasks	11.94	6.40	12.00	10.43	7.25	11.00	8.22	5.50	8.00	5.63	6.41	5.00
Procedural Transfer	2.03	2.56	1.00	1.21	1.97	0.00	2.17	3.23	0.00	0.77	1.97	0.00
Conceptual Transfer	3.97	2.48	4.00	3.14	2.66	2.50	4.09	2.02	4.00	3.50	2.72	3.50
Posttest Mental Effort												
Isomorphic Tasks	5.29	1.70	5.67	4.13	1.84	4.17	3.80	1.73	4.00	6.05	2.51	6.33
Procedural Transfer	4.78	2.51	5.00	4.82	2.33	5.00	4.00	2.26	5.00	6.59	2.68	7.00
Conceptual Transfer	4.22	1.75	5.00	4.00	2.07	3.00	4.13	1.49	5.00	5.18	2.82	5.00
Posttest Time-on-Task												
Isomorphic Tasks	16.13	7.15	16.33	6.69	4.70	6.83	6.07	3.63	4.33	4.00	3.11	3.12
Procedural Transfer	5.94	5.12	6.00	2.82	3.17	1.00	3.48	3.36	3.00	2.36	2.98	2.00
Conceptual Transfer	7.97	5.43	6.50	4.54	3.42	4.50	5.78	2.75	6.00	5.77	4.02	5.00

Note. Acronyms for groups: EE = example study only; EP = example-problem pairs; PE = problem-example pairs; PP = problem solving only.

Table 3.10.
Post-hoc comparisons of Mental Effort, Self-Efficacy, Perceived Competence, Topic Interest, Isomorphic Tasks Performance, Procedural Transfer, and Conceptual Transfer on the Immediate Posttest in Experiment 2.

	EE vs. EP			EE vs. PE			EE vs. PP			EP vs. PE			EP vs. PP			PE vs. PP		
	U	p	r	U	p	r	U	p	r	U	p	r	U	p	r	U	p	r
Training																		
Mental Effort	652	.002	.391	531.5	.005	.377	644.5	<.001	.701	338.5	.755	.044	502	.001	.537	407	.001	.522
Immediate posttest																		
Isomorphic tasks	396.5	.444	.099	238	.026	.300	167.5	.001	.445	267	.295	.147	188	.018	.335	177.5	.083	.258
Procedural Transfer	369	.203	.164	345.5	.677	.056	233	.017	.325	348.5	.568	.080	249	.154	.201	194	.094	.250
Conceptual Transfer	361.5	.196	.167	378	.863	.023	314	.500	.092	405.5	.111	.223	331	.650	.064	216	.397	.126
Self-efficacy ¹	328.5	.071	.233	319.5	.397	.114	81.5	<.001	.655	360	.464	.103	103	<.001	.574	75.5	<.001	.606
Perceived ² Competence	465	.799	.033	368.5	.993	.001	112.5	<.001	.577	311	.833	.030	103.5	<.001	.569	82	<.001	.583
Topic Interest ³	429	.777	.037	221.5	.012	.338	256	.090	.231	204	.025	.314	241.5	.192	.184	281	.524	.095

Note. Acronyms for groups: EE = example study only; EP = example-problem pairs; PE = problem-example pairs; PP = problem solving only.

1 Self-efficacy statistically differed between the pretest and immediate posttest ($Z = 8.16, p < .001, r = .796$) and increased in EE, EP, and PE Condition ($ps < .001$), not in PP Condition ($p = .303$).
2 Perceived competence statistically differed between the pretest and immediate posttest ($Z = 8.30, p < .001, r = .810$) and increased in EE, EP, and PE Condition ($ps < .001$), not in PP Condition ($p = .020$).
3 Topic interest did not differ statistically between the pretest and immediate posttest ($p = .297, r = .102$).

3.4. General Discussion

Two experiments were conducted to investigate how different sequences of example study and practice problem solving (i.e., example study only [EE], example-problem pairs [EP], problem-example pairs [PE], problem solving only [PP]) would affect motivational (i.e., self-efficacy, perceived competence, and topic interest) and cognitive aspects of learning (i.e., performance on isomorphic and transfer tasks, and mental effort). A short sequence of four training tasks was used in Experiment 1 and a longer sequence of eight training tasks in Experiment 2. We were particularly interested in how participants' self-efficacy would develop during the training phase and whether the pattern of results would remain stable on a delayed posttest (Experiment 1), as well as whether findings would change when the training phase comprised more training tasks (Experiment 2).

In a training phase with four training tasks, example study (alternated with practice problem solving) was a more effective (in terms of performance on isomorphic and procedural transfer tasks) and efficient (in terms of mental effort invested in the training and posttest phases) strategy for learning than problem solving only. We also replicated the findings of Van Harsel et al. (2019): self-efficacy and perceived competence scores were significantly higher after the training phase in all three example conditions compared to problem solving only. We did find, however, that studying example-problem pairs resulted in lower mental effort investment during the training phase than studying problem-example pairs in Experiment 1. A novel finding is that these effects persisted on a delayed test one week later. Experiment 2 showed that with longer sequences, example study (alternated with practice problem solving) resulted in lower mental effort ratings during the training phase and higher ratings on self-efficacy and perceived competence than problem solving only. Whereas mental effort was lower during the training phase in the example-problem pairs condition compared to the problem-example pairs condition in Experiment 1, no differences were found between these conditions when sequences were longer as in Experiment 2.

3.4.1. Effects of Different Short Task Sequences on Motivation

The findings of Experiment 1 provide evidence for the first part of the motivational explanation regarding the differential effects of EP vs. PE comparisons reported in the literature (cf. Van Harsel et al., 2019; Coppens et al., 2019). That is, starting the training phase with a practice problem (PE, PP) affected self-efficacy negatively compared to starting with an example (EE, EP). However, we found no evidence for the second part of the motivational explanation (i.e., as a consequence of lower self-efficacy levels, students might not be motivated to study subsequent example and probably also the tasks that follow). It seems that in our study, learners did not disengage after starting with a practice problem and studied the example that was provided as a second

training task. As a consequence, their levels of self-efficacy increased to the level of the EP (and EE) condition and remained stable during the entire training phase. We must note, though, that using a complex math task might not have resulted in lasting detrimental effects on students' self-efficacy (and perceived competence), because our sample of technical higher education students had experience with similar types of tasks and did not find these tasks unpleasant or uninteresting (topic interest scores were relatively high). Further research is recommended to investigate whether these findings replicate with different learning materials and student populations.

These findings indicate that the benefit of an EP-sequence over a PE-sequence is likely not as large as previously believed (e.g., Van Gog et al., 2011) and may only occur under specific conditions. It is, however, an open question what factor or combination of factors moderate(s) the (small) differential effects of EP versus PE on learning (see small-scale meta-analysis by Van Harsel et al., 2019). In other words, what factors determine whether students will or will not disengage after starting with a practice problem (as they presumably did in prior studies, in which their learning outcomes did not benefit from the examples presented to them; e.g., Kant et al., 2017; Leppink et al., 2014; Van Gog et al., 2011)? It is still possible that other (motivational) variables play a role in determining whether students would disengage. For instance, students in PE conditions might disengage when interest in the learning material is very low, or when the second task consists of a text-based worked example (cf. Van Gog et al., 2011) rather than a video example as used in the present study (which might more easily grab and hold their attention). Hence, future research should further explore what (combination of) factors might moderate the EP-PE effect. We recommend the use of large sample sizes, because a recent meta-analysis indicated that the effectiveness of example-problem pairs as compared to problem-example pairs is rather small (Van Harsel et al., 2019).

3.4.2. Effects of Longer Task Sequences on Performance and Motivation

Another noteworthy finding is that longer task sequences did not necessarily result in better learning outcomes when we visually compare the results of Experiments 1 and 2, except in the examples only condition. Studying examples only remained very effective, efficient, and motivating even with longer sequences. This is at first glance surprising in light of the expertise-reversal effect, which proposes that examples become less conducive to learning than practice problems for learners with more prior knowledge (e.g., Kalyuga et al. 2001). Moreover, one might expect that studying examples only, which is more passive, could be less motivating (i.e., more boring) than alternating examples and problems (cf. Sweller & Cooper, 1985), especially with longer sequences. This could, in turn, lead to disengagement and lower learning outcomes, but we found no evidence that this was the case. It should be noted, though, that the training tasks increased in complexity during the training phase (i.e., after the second

task in Experiment 1 and 2 and after the sixth task in Experiment 2). Although the problem-solving procedure remained the same, this may have prevented students from experiencing the examples as too repetitive. Moreover, we provided participants the opportunity to study examples and/or solve practice problems in a self-paced instead of a system-paced learning environment. Although participants were instructed to watch the entire example, it was possible to skip (parts of) the video modeling example. As evidenced by the time-on-task data that was obtained during training phase, time spent on the examples decreased as the learning phase progressed, and this control over the video examples may also explain why participants did not disengage during example study only. Further research should investigate whether the overall findings replicate, and under what circumstances studying longer sequences of examples only remains effective, efficient, and motivating for learning.

3.4.3. Limitations

There are also some limitations to this study. The first limitation is that we did not directly manipulate sequence length (i.e., four vs. eight training tasks) as a between-subject factor in one single experiment, which would have allowed us to test for interaction effects between the length of the task sequence and the outcome variables. That being said, the pattern of results in Experiment 1 and 2 is highly similar and thus seems to reinforce each other. Secondly, a strength of our study was the use of a conceptual pretest. A procedural pretest (as used in the prior study by Van Harsel et al., 2019) might have led students in the example-first conditions to feel that they started the learning phase with practice problem solving. Yet, we did not experimentally vary the type of pretest within the present experiments, and therefore cannot draw definite conclusions about the potential effects of a procedural vs. conceptual pretest. That being said, when we compare the findings from the present study (with a conceptual pretest) to the prior study (with a procedural pretest; Van Harsel et al., 2019) the results are highly similar: There is no evidence for an advantage of example-problem pairs. Thirdly, we 'only' used two different task length sequences. The findings might be different with short training phases comprised of two tasks, where students provided with a PE-sequence would only have one example to study after starting with failed practice problem. Lastly, it is as of yet an open question whether example study would become less effective and motivating with even longer sequences. Hence, future research is recommended to experimentally manipulate how many tasks students receive during the training phase and to cover a broader range of possible sequence length manipulations which take into account the (increase of) complexity level of the training tasks.

Another limitation concerns the self-efficacy and perceived competence measures. The use of a 9-point scale for the self-efficacy measurement raises the question of whether students are really able to report their task-specific confidence on

such a fine-grained level – the same question arises when asking students to report their effort investment on a 9-point scale. A factor that might also have influenced the self-efficacy measurements during the learning phase is whether students could estimate their task-specific confidence based on an actual attempt to solve the problem or just the imagination of doing so after studying the example. Moreover, it has been questioned whether or not (task-specific) self-efficacy and perceived competence are really separate constructs. Literature shows that perceived competence may be a common core component of both self-efficacy and measures of self-concept (e.g., Bong & Skaalvik, 2003; Marsh et al., 2019; Schunk & Pajares, 2005). In line with this notion, the pattern of results on self-efficacy and perceived competence was nearly identical in both experiments and the correlations between these two constructs on the measurement after the training phase were extremely high in Experiment 1 (.96) and Experiment 2 (.92). As such, the use of one of the measures might suffice in future research in this area.

3.4.4. Conclusions

In sum, our results have shown that studying examples only – possibly alternated with practice problem solving – is more effective and efficient for novices' learning than practice problem solving only. These results were established with higher technical education students and a mathematical problem-solving task. However, based on the large body of research on the worked example effect (see for a review Van Gog et al., 2019), it seems safe to assume that these effects would generalize to other problem-solving tasks and populations as well. A new finding of our study was that examples had clear effects on motivational aspects of learning (i.e., self-efficacy and perceived competence); so far, little is known about the effects of different example and problem sequences on motivation (Renkl, 2014; Sweller et al., 2011; Van Gog & Rummel, 2010). Moreover, a new and interesting finding both from a theoretical and practical perspective, is that example study only can remain more effective, efficient, and motivating for learning than solving practice problems only when longer sequences are studied. However, because our study is among the first to examine the effects of different short and longer sequences of examples and problems on student motivation, an open question that needs to be addressed in future research is whether these results generalize to other populations, domains, and materials.

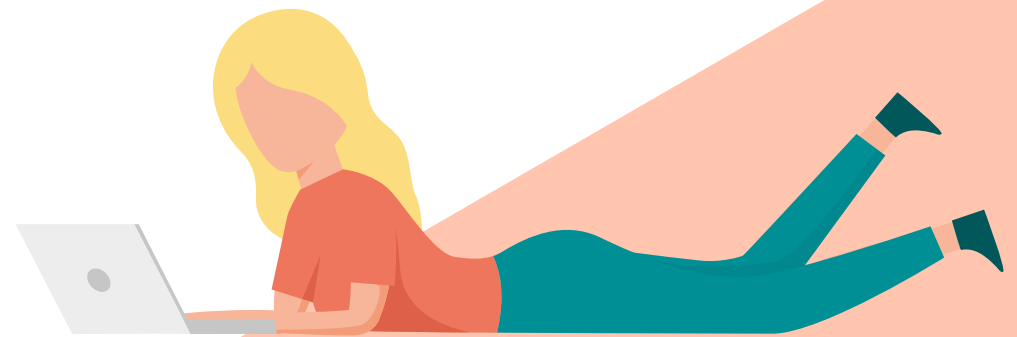
3.4.5. Implications for Practice

Our results could be interesting and relevant for educators who are instructing new knowledge or skills to novices, for students who have to learn new knowledge or skills through self-study, and also for instructional designers who are designing learning materials. Our results suggest that, when studying short sequences of examples and problems, it is more preferable to study or provide examples (probably alternated with

problem solving) instead of practicing problem solving only, from both a cognitive and a motivational perspective. Moreover, even with longer sequences, example study remains very effective, efficient and motivating, however, future research should further investigate under what specific conditions example study remains effective in longer learning phases. Secondly, it is advisable to start training phases with an example instead of a problem. Although we did not find any differences in test performance and student motivation between example-problem pairs and problem-example pairs, our results showed that starting the training phase with an example is more efficient for learning than starting with a practice problem.

Chapter 4

How do higher education students regulate their learning in an online environment with video modeling examples, worked examples, and practice problems?



This chapter was submitted as:

Van Harsel, M., Hoogerheide, V., Janssen, E. M., Verkoeijen, P. P. J. L., & Van Gog, T. (2020). *How do higher education students regulate their learning in an online environment with video modeling examples, worked examples, and practice problems?* Manuscript submitted for publication.

Acknowledgement of author contributions:

MvH, VH, PV, and TvG designed the study, MvH recruited participants and collected the data, MvH and EJ analyzed the data, VH checked the data package, MvH drafted the manuscript, all authors contributed to critical revision of the manuscript, VH, PV, and TvG supervised the study.

Abstract

Presenting novices with examples and problems is an effective and efficient way to acquire new problem-solving skills. Examples and problems are increasingly presented in online learning environments, in which learners often have to self-regulate their learning (i.e., choose what type of task to work on and when). Yet, it is questionable how novices self-regulate their learning from examples and problems, and to what extent their choices match with effective principles from instructional design research. In this study, 147 higher education students had to learn how to solve problems on the trapezoidal rule. During the self-regulated learning phase, they were free to select six tasks from a database of 45 tasks that varied in task format (video examples, worked examples, practice problems), complexity level (level 1, 2, 3), and cover story. Almost all students started with (video) example study at the lowest complexity level. The number of examples selected gradually decreased and task complexity gradually increased during the learning phase. However, examples and lowest level tasks remained relatively popular throughout the entire learning phase. There was no relation between students total score on how well their behavior matched with the instructional design principles and learning outcomes, mental effort, and motivational variables.

Keywords: example-based learning, self-regulated learning, self-efficacy, mental effort, problem solving

4.1 Introduction

Problem solving is important in many curricula, especially in the domains of science, technology, engineering, and mathematics (STEM; Van Gog et al., 2020). Most problems students encounter in (the initial years of) STEM curricula are algorithmic problems, in which students have to learn to perform the procedure to get from an initial state to a described goal state (Newell & Simon, 1972). Different types of tasks are commonly provided to help students learn to solve new problems, and nowadays, this is often done in online or blended learning environments. Such tasks include video modeling examples (i.e., a model demonstrating and possibly explaining the solution procedure step by step on video), worked examples (i.e., a written step-by-step explanation of a full and correct solution procedure of how to solve a problem), and practice problems that students have to try to solve themselves. A popular example of such an environment is Khan Academy (www.khanacademy.org), where students can decide for themselves which type of tasks to work on (i.e., examples or problems), for how long, and in which order.

When acquiring problem-solving skills in online learning environments, it is important that students can adequately self-regulate their learning from examples and problems, especially when guidance or support is not (directly) available. Although there are different theoretical models of self-regulated learning (see Panadero, 2017), these models all agree that self-regulated learning requires students to plan, execute, monitor (i.e., track), evaluate, and control their learning (i.e., adapt their study behavior in response to their evaluation; e.g., Nelson & Narens, 1990; Winne & Hadwin, 1998; Zimmerman, 1990). Self-regulated learning of problem-solving tasks requires students to decide which task they want to perform, monitor their progress while performing the task (and possibly adjust their strategies while working on the task), to judge their performance after the task is completed, and to use this as input for deciding what subsequent task to work on (i.e., which task suits their learning needs best, e.g., De Bruin & Van Gog, 2012; Van Gog et al., 2020).

However, little is known about how students regulate their learning of problem-solving tasks, that is, about when they choose examples or practice problems, and why (i.e., what reasons underlie their choices; e.g., Van Gog et al., 2020; Van Gog et al., 2019). Moreover, it is an open question how well students' choices align with what we know to be effective, efficient, and motivating task sequences for acquiring new problem-solving skills from many years of instructional design research. Therefore, the present study addresses those questions.

4.1.1. Learning from Examples and Problems at Different Complexity Levels

Instructional design research has uncovered several principles on how to optimize the acquisition of new problem-solving skills for novices (i.e., students with little if any prior knowledge). These principles are concerned with how to ensure that novices work on tasks that provide an optimal level of instructional support and complexity given their current level of knowledge (4C/ID Model; Van Merriënboer, 1997; Van Merriënboer & Kirschner, 2013). With regard to instructional support, the *worked examples principle* (Renkl, 2014; Sweller et al., 2011; Van Gog et al., 2019) states that for novices studying several examples (possibly alternated with solving practice problems) leads to better test performance (i.e., is more effective) attained with less time and/or effort investment (i.e., is more efficient) than practice problem solving only. Since this applies not only to worked examples (i.e., a written step-by-step explanation of how to solve a problem) but also to video modeling examples (i.e., a person demonstrating and/or explaining a problem-solving procedure on video), we call this the *example-based-learning-principle*.

A second robust finding is the *example-first-principle*, which says that novice learners should not only study several examples while learning, but also start the learning phase with an example (or several examples) instead of practice problem solving, because research has consistently shown that sequences of example study only or example-problem pairs –but not problem-example pairs– are more effective and efficient for learning than solving practice problems only (e.g., Van Harsel et al., 2020; Kant, et al., 2017; Van Gog et al., 2011). Recent studies have also shown that example study only, or example study alternated with problem solving, is also more beneficial for motivational aspects of learning than problem solving only (e.g., Van Harsel et al., 2019, 2020; Coppens et al., 2019): Example-based learning led to increased self-efficacy (i.e., a personal judgment of one's own capacities to organize or accomplish a specific task or challenge; e.g., Bandura, 1977) and perceived competence (i.e., related to the construct of self-efficacy, but comprises more general knowledge and perceptions of people's self-concept towards their own competence; Deci & Ryan, 2002), but it did not affect topic interest (i.e., the level of interest triggered in an individual by a specific topic, which is relatively stable across time; e.g., Ainley et al., 2002). Such motivational effects become especially important in environments where students can self-regulate their learning (e.g., Pajares, 1996), because when novices start the learning phase with a task that demotivates them (such as a failed problem-solving attempt), they might lose confidence in their ability to learn the task and that could cause students to quit studying. We must note, though, that learners do not need the instructional support provided by examples anymore when their knowledge increases. From that point onwards, they learn more from solving problems than from example study (i.e., the expertise reversal effect; Kalyuga et al., 2003).

Finally, problem-solving tasks that are presented in school curricula (either via online environments or in textbooks/workbooks) often span multiple complexity levels. Task complexity is determined by the number of elements in a learning task and the interaction between those elements (e.g., Sweller & Chandler, 1994). Simple learning tasks consist of a few information elements and a small number of interactions between elements that need to be processed simultaneously in working memory. With increasing numbers of information elements and interactions between elements, task complexity (and working memory load) increases (e.g., Pollock et al., 2002; Van Zundert et al., 2012).

When tasks span multiple complexity levels, learners should not only be working on tasks that provide optimal support given their current level of knowledge, but also on tasks that are at an optimal level of complexity (see the 4C/ID model; Van Merriënboer, 1997; Van Merriënboer & Kirschner, 2013). When learners work on tasks that are too complex given their prior knowledge, their learning outcomes and motivation might suffer (Van Merriënboer et al., 2003). Therefore, novices should start with a task at the lowest complexity level (i.e., *lowest-level-first-principle*) and build up tasks in such way that the level of complexity gradually increases (i.e., *simple-to-complex-principle*). When the choice is made to move up a complexity level, learners often need instructional support again (cf. 4C/ID model). Therefore, it becomes important to start each new complexity level with example study (i.e., *start-each-level-with-example-principle*).

4.1.2. Self-Regulated Learning with Examples and Problems

These instructional design principles provide clear guidelines on what works best when learning from examples and problems at different complexity levels. However, the question is whether students would spontaneously apply these principles when selecting tasks (i.e., examples and problems at different complexity levels) during self-regulated learning in an online environment. As mentioned earlier, for effective self-regulated learning of problem-solving tasks, students need to be able to self-assess their performance on a task just completed and then select a next task with the right level of support and complexity (e.g., De Bruin & Van Gog, 2012). There are, however, both empirical and theoretical reasons to believe that learners will engage in suboptimal task selection when they are left to their own devices (e.g., Azevedo et al., 2008; Niemiec et al., 1996).

Firstly, self-regulated learning research has shown that learners' estimation of their own task performance (or knowledge) is often not in line with their actual performance (e.g., Bjork, 1999, 1994; Kostons et al., 2010, 2012; Rawson & Dunlosky, 2007), particularly for novices (e.g., Dunning et al., 2004; Koriat & Bjork, 2005). Inaccurate self-assessments are a major problem when learners are in control of task selection, because for

learning to be effective and efficient, learners need to select a task at an optimal level of instructional support and complexity given their current level of performance. Novices who overestimate their performance might select a task that is too complex and/or does not provide the necessary instructional support, while those who underestimate their performance might select a task that is too easy (e.g., Dunlosky & Rawson, 2012). As a result, learners will end up working on tasks that are not aligned with their learning needs, which might negatively affect their performance on domain specific knowledge or skills and motivation.

Secondly, research has shown that novices often experience difficulties discerning which task aspects are relevant for learning when selecting their own learning tasks (e.g., Quilici & Mayer, 2002), probably because they lack domain knowledge and/or task-selection skills (i.e., knowing about relevant task-selection aspects and combining this with characteristics of available learning tasks; e.g., Van Merriënboer et al., 2006). As a consequence, novices might select tasks based on surface features used to exemplify the problem-solving procedure (e.g., cover story) rather than structural features that are (more) relevant for learning (e.g., the level of complexity and instructional support; Corbalan et al., 2008).

A recent study conducted by Foster and colleagues (2018) provided some evidence for the idea that learners also show suboptimal behavior when they can select their own task format in the form of examples and practice problems. In their study, university students (novices) learned how to solve probability calculation problems in an online learning environment. In Experiment 1, students received 12 probability problems (with different cover stories). The self-regulated learning group (i.e., SRL-group) was given a choice with each problem on whether they wanted to study it in the form of a worked example or a practice problem. In Experiment 2 and 3, students received 24 probability problems (with different cover stories). Again, one group (i.e., SRL-group) was given a choice with each problem of whether they wanted a worked example or practice problem, and another group (i.e., SRL-completion group) could additionally opt for completion problems, which are partially worked-out examples that provide a medium level of support but require learners to complete some steps themselves (e.g., Paas, 1992; Renkl & Atkinson, 2003; Van Merriënboer et al., 2002).

Based on the effectiveness and efficiency of the example-based-learning-principle and the example-first-principle, one would expect that novices learn most when they select more examples than problems and start the self-regulated learning phase with a worked example rather than a (completion) problem. However, Experiment 1 of Foster et al. (2018) showed that problems were selected more frequently on average than examples (at odds with the example-based-learning-principle), and that only one third of the participants in the SRL-group started the learning phase with a worked example (at odds with the example-first-principle). Also, in Foster et al.'s Experiment 2

and 3, example study remained the least picked strategy and problem solving the most popular one in both the SRL-group and the SRL-completion group. Moreover, students rarely chose a worked example as a first task.

In sum, little research has been conducted on how learners regulate their learning from examples and problems (at different complexity levels) and how well their selection behavior matches with evidence-based principles from instructional design research. The few studies available, suggest that learners' task-selection behavior does not align with these principles (Foster et al., 2018). It is important to get more insight in what learners do (and why) when they can determine themselves how to learn new problem-solving skills, as this will provide information that can be used by teachers and instructional designers to determine whether and what instructional support or advice learners might need to optimally self-regulate their learning from examples and practice problems (at different complexity levels).

4.1.3. The Present Study

This study investigated how higher education students (novices on the to-be-learned topic) regulate their learning in an online environment in which they could select their own learning tasks from a task database comprising video modeling examples, worked examples, and practice problems of varying levels of complexity and with different cover stories. We decided to provide the option of worked examples *and* video modeling examples because both are widely used in online learning environments, yet it is largely unclear which example types students prefer (at which phase in the learning process). For example, Hoogerheide and colleagues (2014) compared the effects of worked examples to video modeling examples with two samples of secondary education students. Although they found no differences between the two example formats on cognitive (i.e., test performance, mental effort) and motivational aspects of learning (i.e., self-efficacy, perceived competence), there was an effect on the degree to which students preferred to receive instruction in a similar manner in the future. When only one example was studied (Experiment 2), the video modeling example condition gave a higher preference rating (at least numerically; $p = .07$), but when two examples were studied (Experiment 1), the worked example condition gave a higher preference rating ($p = .03$).

These findings might suggest that students would prefer video modeling examples at the beginning of a learning phase and worked examples later in the training phase. A possible explanation for why students would prefer to start with a video modeling example instead of a worked example could be that in video modeling examples, information is demonstrated in a step-by-step manner and that the combination of dynamic visual information and the model's narration take the learner by the hand. Worked examples can be overwhelming because all the information is presented

simultaneously, and it might be easy to ignore written text. However, because all the information is presented simultaneously, worked examples do allow for efficiently looking up difficult problem-solving steps more easily than video modeling examples (in which the information is presented in succession).

This study had three research questions. First, what tasks do technical higher education students select and why, when learning from examples and problems at different complexity levels? Second, to what extent do students' task selections match with principles of effective, efficient, and motivating task sequences derived from instructional design research? Given the paucity of research on what learners do when they are in charge of learning a new problem-solving skill with the help of (different) examples and problems at different complexity levels, we refrain from formulating explicit hypotheses and consider this study as exploratory in nature. Third, we investigated whether there is a positive relation between the extent to which students' choices match with the instructional design principles and their scores on learning outcomes, mental effort, and motivational variables. Given how much evidence there is for the instructional design principles, one would expect a positive relationship between the extent to which learners' choices match with these principles and scores on learning outcomes (i.e., isomorphic tasks, procedural transfer task, and conceptual questions) and motivational aspects of learning (i.e., self-efficacy, perceived competence, and topic interest), and a negative relationship with mental effort during the learning phase.

4.2. Method

4.2.1. Participants and design

Participants were 180 Dutch higher education students enrolled in the first year of an electrical and electronic mechanical engineering program ($M^{age} = 19.00$, $SD = 1.64$; 169 male, 11 female). All participants were assigned to an online self-regulated learning environment in which they had to learn a mathematical problem that required them to approximate the region under a graph using the trapezoidal rule. The environment consisted of three phases: (1) pretest, (2) self-regulated learning phase, (3) and posttest. We excluded 10 participants who did not finish the isomorphic (and transfer) items on the posttest on time, and 7 participants of whom (part of the) learning phase data was missing due to a programming error. Because we were interested in the task-selection behavior of novice learners, we also excluded 16 participants who had too much prior knowledge, indicated by a score of 5 or more (out of 10) on the prior knowledge test. Therefore, the final sample consisted of 147 participants ($M^{age} = 18.90$, $SD = 1.64$; 139 male, 8 female). Participants gave their informed consent in the online learning environment before the study began and received study credits for their participation.

4.2.2. Materials

The materials were based on the materials developed by Van Harsel et al. (2019, 2020) and presented in a web-based learning environment.

4.2.2.1. Learning tasks.

The task database contained 45 learning tasks. These tasks varied in complexity level, task format, and cover story (for an overview, see Figure 4.1).

Complexity level. Tasks could be selected at three levels of complexity. *Level 1* tasks required participants to approximate the region under a graph using the trapezoidal rule in problems that always contained a polynomial degree of 2. These problems also required participants to calculate more than two intervals and calculate with fractions and positive numbers only. *Level 2* tasks were more complex than *Level 1* tasks, because participants were asked to calculate with negative numbers. The negative number changes the relation between information elements. That is, calculating with negative numbers requires students to take into account an additional rule (i.e., relation between elements) than calculating with positive numbers (i.e., subtracting a negative number from a positive number, turns the two signs into a plus sign; $5 - - 7 = 12$). Moreover, in more complex calculations with negative numbers (i.e., large functions using brackets, exponents, different arithmetic operations), the order of the arithmetic operations is important. *Level 3* tasks were, in turn, more complex than *Level 2* tasks, because students had to calculate with a cubic polynomial instead of a quadratic polynomial. A cubic polynomial has a term more than a quadratic polynomial, which increases the number of information elements and relations students have to calculate with.

DE PROEF OP DE SOM								
Level 1			Level 2			Level 3		
Video modeling examples	Worked examples	Practice problems	Video modeling examples	Worked examples	Practice problems	Video modeling examples	Worked examples	Practice problems
Drinking beer	Drinking beer	Drinking beer	Running	Running	Running	Carousel	Carousel	Carousel
Energy measurement	Energy measurement	Energy measurement	Drinking water	Drinking water	Drinking water	Rowing	Rowing	Rowing
Fitness	Fitness	Fitness	Washing machine	Washing machine	Washing machine	Perfume	Perfume	Perfume
Fuel consumption	Fuel consumption	Fuel consumption	Soapsuds	Soapsuds	Soapsuds	Coffee consumption	Coffee consumption	Coffee consumption
Traffic	Traffic	Traffic	Plasterer	Plasterer	Plasterer	Chocolate party	Chocolate party	Chocolate party

Figure 4.1. Screenshot of the task database.

Task format. Within each complexity level, participants could choose from three task formats, namely, video modeling examples, worked examples, and conventional practice problems (see Figure 4.1). Each *video modeling example* displayed a computer screen recording of a female model who demonstrated (with handwritten notes) and verbally explained how to solve a mathematical problem step-by-step, using the trapezoidal rule. The screen recording started with a brief introduction on the trapezoidal rule and an explanation of a specific problem state. Subsequently, the model explained how to interpret the information that was given to solve the problem (i.e., the graph of a function, the left border and right border of the area, the number of intervals, and the formula of the trapezoidal rule). Finally, she demonstrated and explained how to solve the problem by undergoing four steps: 1) 'compute the step size of each subinterval', 2) 'calculate the x-values', 3) 'calculate the function values for all x-values', and 4) 'enter the function values into the formula and calculate the area'. The written information on previously completed steps remained visible on the screen while the model worked on and explained the next step.

Each *worked example* was presented on one page. The worked examples also started with a short description of the problem state and participants received some additional information that was needed to solve the problem (i.e., the graph of a function, the left border and right border of the area, the number of intervals, and the formula of the trapezoidal rule). This was followed by the written-out solution procedure that showed students how to solve each step of the problem (the problem state, additional information, and written explanations and correct answers on all steps were simultaneously visible on the screen).

The *practice problems* were presented on one page and consisted of the problem state and some additional information on how to solve the problem (i.e., the graph of a function, the left border and right border of the area, the number of intervals, and the formula of the trapezoidal rule), with by the following assignment: "Approach the area under the graph using the information that is given. Write down all your intermediate steps and calculations". Participants did not receive any feedback on their answers. A screenshot of a practice problem, video modeling example, and worked example are given in the Supplementary Materials A, B, and C.

Cover story. In addition to selecting a complexity level and instructional format, students could also choose their own cover story. At each complexity level, participants could choose between five different cover stories (see Figure 4.1). For example, they could approximate how many liters of beer is tapped within a certain amount of time (i.e., drinking beer) or approximate how often the circular platform of a carousel rotates in a given period of time (i.e., carousel). The cover stories were similar for each task format that was provided within a complexity level (e.g., drinking beer could be selected as video modeling example, worked example, and practice problem), yet the numbers used differed per task format.

4.2.2.2. Test tasks

The *pretest* was a conceptual prior knowledge test that consisted of five questions (i.e., multiple choice questions with explanation part) that aimed to measure participants' understanding of the trapezoidal rule. Cronbach's alpha in the current sample was .33. Each multiple-choice question had four answer options (i.e., a, b, c, and d) and an 'explanation' part where participants had to explain their answer. The *posttest* consisted of five tasks. The first three tasks were isomorphic to the tasks in the self-regulated learning phase (i.e., a level 1, 2, and 3 task). Cronbach's alpha in the current sample was .74. The fourth task was a procedural transfer task that required participants to use the Simpson rule (instead of the trapezoidal rule) to approximate the definite integral under a graph. Simpson's rule is also a numerical integration method to approximate the integral of a function. Although both procedures look almost similar, Simpson's rule uses quadratic polynomials (instead of the straight-line segments). The fifth task consisted of five open-ended conceptual questions that aimed to measure participants' understanding of the trapezoidal rule, and these were isomorphic to the questions in the pretest. Cronbach's alpha in the current sample was .44. An example of a conceptual pretest item, an isomorphic posttest task, a procedural transfer task, and a conceptual posttest item is shown in the Supplementary Materials E, F, G, and H.

4.2.2.3. Mental effort

Participants rated their mental effort on a 9-point mental effort rating scale (Paas, 1992), with answer options ranging from (1) "very, very low mental effort" to (9) "very, very high mental effort". Mental effort was rated after each task in the self-regulated learning phase and the posttest phase, except for the five conceptual posttest questions (where it was rated only once after the last question).

4.2.2.4. Self-efficacy

After the pretest, during the self-regulated learning phase (i.e., after each learning task), and before the posttest, participants were asked to rate to what extent they were confident that they could approximate the definite integral of a graph using the trapezoidal rule. A 9-point rating scale was used, ranging from (1) "very, very unconfident" to (9) "very, very confident" (Van Harsel et al., 2019, 2020; adapted from Hoogerheide et al., 2016).

4.2.2.5. Perceived competence

Perceived competence was measured using the *Perceived Competence Scale for Learning* (Van Harsel et al., 2019, 2020, 2020b; based on Williams & Deci, 1996; Williams et al., 1988). This perceived competence scale consisted of three items: "I feel confident in my ability to learn how to approximate the definite integral of a graph using the trapezoidal rule", "I am capable of approximating the definite integral of a graph using the trapezoidal rule", and "I feel able to meet the challenge of performing well when I have to apply the trapezoidal rule". Participants were asked to rate on a scale of (1) "not at all true" to (7) "very true" to what degree these three items applied to them. Cronbach's alpha in the current sample was .93.

4.2.2.6. Topic interest.

Finally, participants' interest in the topic was measured with a topic interest scale, comprised of 7 items (Van Harsel et al., 2019, 2020; adapted from the topic interest scale by Mason et al., 2008, and the perceived interest scale by Schraw et al., 1995). Participants were asked to rate to what degree each of the items applied to them on a 7-point scale (1: totally disagree, to 7: totally agree). Cronbach's alpha in the current sample was .82. All items are shown in the Supplementary Materials I.

4.2.2.7. Task-selection questionnaire.

To shed light on why participants selected the learning tasks that they did, we developed a questionnaire. This questionnaire consisted of five questions, each with a multiple-choice (mc) and open-answer part, namely: 1) What was the format of the first task you chose (mc: video modeling example, worked example, practice problem) and why (open answer)?, 2) What was the level of complexity of the first task (mc: level 1, level 2, level 3) and why (open answer)?, 3) What was the format of the second task you chose (mc: video modeling example, worked example, practice problem) and why (open answer)?, 4) What was the level of complexity of the second task and why (open answer)?, and 5) Which task format did you choose most often and why (open answer)?

4.2.3. Procedure

The study was run in sixteen sessions with 7 to 28 participants per session. The sessions lasted 116 minutes on average and took place in a computer classroom at participants' higher education institute. Each participant received a headset, pen, and scrap paper to write down their calculations. The session started with the experimenter explaining the aim and procedure of the study. Participants were told that they were going to learn a mathematical task in an online learning environment by selecting their own learning tasks. Participants were also instructed that they could work at their own pace (with a maximum of 135 minutes). Moreover, they received the instructions to write down as much as possible and to write an "X" if they really did not know what to answer.

After the instructions, participants entered the online learning environment. In the environment, tasks and questionnaires were presented on a separate page, and participants were unable to go back to the previous pages or to look forward to the next page before completing the current task or questionnaire. Time was logged for each task. Participants were first presented with, in order, a short demographic questionnaire (e.g., age, gender, and prior education), the pretest, and the self-efficacy, perceived competence, and topic interest questionnaire. Then, participants entered the self-regulated learning phase. To ensure that participants had some

knowledge of the task database and how to select their own tasks, they first received an explanation of the task database. A picture of the task database was presented on the screen. Participants were instructed to select 6 learning tasks of their own choice from a task database containing 45 tasks that differed in format (video modeling examples, worked examples, and practice problems), complexity level (level 1, 2, and 3) and cover story. They were also told that each task could only be selected once, and that there was a maximum of 10 minutes to watch, study, or solve each task.

After the self-regulated learning phase, participants completed the self-efficacy, perceived competence, and topic interest questionnaires again. Participants were instructed to turn their scrap paper upside down and given a new scrap paper to use during the posttest. After each task on the posttest, participants rated their mental effort. Lastly, participants completed the task-selection questionnaire.

4.2.4. Data Analysis

4.2.4.1. What tasks do technical higher education's students select and why?

To shed light on participants' task-selection behavior, we counted the task format (video modeling example, worked example, practice problem) and complexity level (1, 2, or 3) of the six learning tasks each participant had selected and converted these scores into percentages. Then, we counted the task formats and complexity levels participants *said* they selected on the first and second learning task, and the task format participants *said* they selected most often during the entire learning phase. We used Chi-Square Tests to analyze whether there was a significant relation between task format or complexity level and the order of the learning tasks.

To evaluate participants' answers on the task-selection questionnaire, we coded their explanations (open coding) and grouped these codes into categories (axial coding). Two coders scored about 20% of the data and the interrater reliability of their scores was assessed by calculating Cohen's Kappa (Cohen, 1960). A Kappa value of 0 would mean no agreement, values between 0.01–0.20 slight agreement, values between 0.21–0.40 fair agreement, values between 0.41–0.60 moderate agreement, values between 0.61–0.80 substantial agreement, and values between 0.81–1.00 almost perfect agreement (Landis & Koch, 1977). The agreement between the coders was moderate to almost perfect: Cohen's Kappa was .65 for question 1, .82 for question 2, .59 for question 3, .84 for question 4, and .95 for question 5.

4.2.4.2. How do novices' task selections match with instructional design principles?

We scored for each participant whether their task-selection behavior matched with the instructional design principles (i.e., example-based-learning-principle, example-first-principle, simple-to-complex-principle, lowest-level-first-principle, and start-each-

complexity-with-example-principle). For each of these principles, participants could earn 1 point in total. More specifically, for the example-based-learning-principle, simple-to-complex-principle, and start-each-complexity-with-example-principle, 1 point was awarded when students' choices matched the principle entirely, 0.5 points when their choices matched the principle only partially, and 0 points when their choices did not match the principle at all. For the example-first-principle and lowest-level-first-principle, 1 point was awarded when students' choices matched the principle entirely and 0 points when their choices did not meet the principle at all. For each participant, a total score was computed (maximum: 5 points). For an extended version of the scoring protocol and an example, see the Supplementary Materials M. Two coders scored about 20% of the data and the interrater reliability of their scores was assessed by calculating a two-way mixed, consistency, single-measures intra-class correlation (ICC; McGraw & Wong, 1996). According to Cicchetti (1994), ICC values that are below .40 are classified as poor, values between .40 and .59 are classified as fair, values between .60 and .74 are classified as good, and values between .75 and 1.0 are classified as excellent. With a score of .96, the ICC was in the good and excellent range for the principles.

4.2.4.3. Is there a positive relation between the extent to which students' choices match with the instructional design principles and their scores on learning outcomes, mental effort, and motivational variables?

Lastly, we explored the extent to which students' choices match with the instructional design principles correlated with cognitive (i.e., performance on the isomorphic posttest tasks, procedural transfer task, and conceptual questions, and mental effort) or motivational aspects of learning (i.e., self-efficacy, perceived competence, and topic interest).

We computed averages for the perceived competence and topic interest measurements before and after the learning phase, as well as for the reported effort invested in the learning tasks and the isomorphic posttest tasks. Test performance was scored by the first author and the third author based on a scoring protocol that was developed in collaboration with higher education mathematics teachers by Van Harsel et al. (2019). On the conceptual pretest and conceptual posttest items, participants could earn a maximum of 9 points. One point could be earned for the first open-ended question (1 point for the correct answer, 0 points for an incorrect answer) and 2 points for the other open-ended questions. Participants were rewarded with the maximum of 2 points when they got the answer right and provided correct explanations. Only 1 point was awarded if the answer was correct, but the explanation was incorrect or missing, and 0 points were given when both the answer and explanation were incorrect. On the isomorphic posttest items, a maximum of 8 points could be earned for each task (i.e., three tasks, max. score = 24 points), with 2 points for calculating each step correctly: 1) the step size of each subinterval, 2) all x-values, 3) the function

values for all x-values, and 4) using the correct formula for the area under the graph and providing the correct answer. In step two, three, and four, one point was granted if half or more of the solution steps were correct and zero points were granted if less than half of the solution steps were correct. The same scoring standard was used to score the procedural transfer task (i.e., max. score = 8 points). Again, two coders scored about 20% of the data and the interrater reliability of their scores was assessed by calculating a two-way mixed, consistency, single-measures intra-class correlation (ICC; McGraw & Wong, 1996). According to Cicchetti (1994), all our ICCs were all in the excellent range, with a score of .77 for the conceptual pretest tasks, .95 for the conceptual posttest tasks, .99 for the isomorphic posttest tasks, and .91 for the procedural transfer task.

4.3. Results

To answer our research questions on what tasks students selected and why (i.e., question 1) and how well their behavior matched with evidence-based principles from instructional design research (i.e., question 2), we report descriptive statistics. Regarding the correlational analyses (i.e., question 3), the effect size of Pearson r correlation is reported with values of 0.10, 0.30, and 0.50 representing a small, medium, and large effect size, respectively (Cohen, 1988). We must note, though, that we used an uncorrected significance level ($p < .05$) for the correlational analyses reported in this paper and that significant findings should be regarded with caution as we could not control the false-positive rate in the present study.

4.3.1. What tasks do technical higher education's students select and why?

Participants' task-selection behavior during the learning phase was explored. The percentages of selected formats are presented in Table 4.1 and Table 4.2, and the percentages of selected complexity levels in Table 4.3. The results of the Chi-Squared Tests are presented in the text.

4.3.1.1. Examples and problems.

On average, participants selected more examples to study (64.3%) than practice problems to solve (35.7%). The large majority of participants started the learning phase with an example instead of a practice problem. However, there was a strong decrease in the number examples selected from task 1 to task 2, and a strong increase in the selection of practice problems. Surprisingly, example study remained the preferred task format on task 2, 3, 4, and 5 (55% or higher). Only on the last learning task (i.e., task 6) did more participants select a practice problem than an example. Chi-Squared Tests revealed that the proportion selected examples depends on the order of the tasks, $\chi^2(5) = 92.48$, $p < .001$. This suggests that students selected fewer examples (and more problems) as the learning phase progressed.

4.3.1.2. Example format.

Results showed that participants, on average, selected more video modeling examples (36.9%) than worked examples (25.8%) during the learning phase. The majority of the participants selected a video example as the first learning task. However, the percentage of selected video modeling examples dropped considerably on the second learning task. This percentage remained relatively stable up to and including the fifth learning task but decreased further on the last learning task. The percentage of selected worked examples increased from the first to the second learning task and stayed relatively constant during the rest of the learning phase. Chi-Squared Tests revealed that the proportion selected video modeling examples depends on the order of the tasks $\chi^2(5) = 52.48, p < .001$, meaning that students selected fewer video modeling examples (and more worked examples) as the learning phase progressed.

4.3.1.3. Complexity level.

The results showed that the majority of participants started the learning phase with a task at the lowest complexity level instead of selecting a level 2 or level 3 task. The complexity of the selected tasks seemed to increase as the learning phase progressed. That is, the percentage of level 1 tasks was highest on the first and second learning task but declined from the third learning task onwards. The percentage of selected level 2 learning tasks, on the other hand, was relatively low on the first and second learning task, was highest on the third and fourth learning task and declined again on the fifth and sixth learning task. Level 3 tasks were selected seldomly during the first half of the learning phase and were selected most often on the last two learning tasks. Surprisingly, results showed that during the second half of the learning phase (i.e., learning task 4, 5, and 6), almost one third of the total sample still selected tasks at the lowest complexity level. Chi-Squared Tests revealed that the proportion selected lowest level tasks (i.e., level 1) depends on the order of the tasks $\chi^2(10) = 285.92, p < .001$, suggesting that students selected fewer level 1 tasks (and more level 2 or level 3 tasks) as the learning phase progressed.

Table 4.1.
Percentages of Selected Examples and Practice Problems During the Self-regulated Learning Phase for the Total Sample (N = 147).

Total sample		
	E	P
Learning task 1	95.9%	4.1%
Learning task 2	55.8%	44.2%
Learning task 3	59.9%	40.1%
Learning task 4	63.9%	36.1%
Learning task 5	65.1%	34.9%
Learning task 6	45.6%	54.4%

Note. E = example, P = problem.

Table 4.2.
Percentages of Selected Video Modeling Examples, Worked Examples, and Practice Problems During the Self-regulated Learning Phase for the Total Sample (N = 147).

Total sample			
	VME	WE	P
Learning task 1	76.9%	19.0%	4.1%
Learning task 2	24.5%	31.3%	44.2%
Learning task 3	37.4%	22.5%	40.1%
Learning task 4	35.4%	28.6%	36.0%
Learning task 5	30.2%	34.9%	34.9%
Learning task 6	17.0%	28.6%	54.4%

Note. VME = video modeling example, WE = worked example, P = problem.

Table 4.3.
Percentages of Selected Complexity Levels (Level 1, 2, and 3) in the Self-regulated Learning Phase for the Total Sample (N = 147).

Total sample			
	Complexity Level 1	Complexity Level 2	Complexity Level 3
Learning task 1	88.4%	8.2%	3.4%
Learning task 2	75.5%	19.7%	4.8%
Learning task 3	47.6%	41.5%	10.9%
Learning task 4	35.4%	44.2%	20.4%
Learning task 5	28.1%	26.0%	45.9%
Learning task 6	26.5%	23.8%	49.7%

4.3.1.4. Reasons for task selections.

We also analyzed participants' answers to the questions that asked them which tasks they selected and why. As shown in Table 4.4, participants reported that they predominantly started the learning phase with a video modeling example, because this format was most comfortable for them or provided the most support. The reason why it was common to start with tasks of the lowest complexity level is that participants believed that this could help them build up their level of expertise or that this suited their current level of expertise. Note that these were also the most common reasons why participants chose the lowest complexity level as a second learning task. Regarding the format of the second learning task, participants selected practice

problems most often, followed by worked examples. They mentioned that these formats helped them to assess their level of expertise or provided the most support (this reason was especially mentioned by those who selected worked examples). Finally, video modeling examples (followed by practice problems) were preferred most on average during the learning phase. Student often said it was the most comfortable way of learning (especially for video modeling examples because this format was most familiar, suited their learning preference, was most clear, etc.) or said they learned most from these format (especially for practice problems because participants felt practice helped them master the procedure). Note that participants' memory regarding what format and level they selected for the first task matched their actual choice (see Table 4.2 and 4.3), but for the second task participants only correctly remembered the task format and not the complexity level.

Table 4.4.
Students' Answers to the Open Questions Regarding the Selection of Different Formats and Complexity Levels in the Self-regulated Learning Phase.

	What was the format of the first task you chose and why?	What was the level of the first task you chose and why?	What was the format of the second task you chose and why?	What was the level of the second task you chose and why?	Which task format did you choose most often and why?
Format					
Video modeling example	75.5%		24.5%		40.0%
Worked example	17.7%		35.4%		15.7%
Practice problem	2.0%		38.1%		29.3%
All formats equally often	-		-		19.0%
No answer	4.8%		2.0%		2.0%
Complexity Level					
Level 1		87.1%		57.1%	
Level 2		9.5%		30.7%	
Level 3		2.7%		10.2%	
No answer		0.7%		2.0%	
Reason					
This way of learning is (most) comfortable	34.7%	-	12.9%	-	20.4%
This way of learning is (most) effective	2.7%	-	2.0%	-	25.9%
This way of learning is (most) supportive	36.7%	0.7%	17.0%	1.4%	10.2%
This way, I can assess my level of expertise	1.4%	9.5%	36.1%	18.4%	4.8%
This way, I can build up the level of complexity	-	33.3%	1.4%	21.7%	-
It fits my current level of expertise	-	35.4%	2.0%	21.8%	1.4%
It fits the order in the task database	2.7%	1.4%	4.1%	2.7%	-
Unclear reason	9.6%	4.7%	10.2%	12.2%	16.3%
No reason	12.2%	15.0%	14.3%	21.8%	21.1%

4.3.2 Do novices’ task selections match with the instructional design principles?

Thirdly, we analyzed how well students’ behavior matched the principles known to be effective and efficient based on instructional design research. Results showed that students’ choices matched with many of the principles when selecting their own learning tasks. As shown in Table 4.5, the majority of the students had a total score of 4 or higher (out of 5), which means that their choices matched with (almost all of) the principles. When exploring how well participants’ task selections matched with the individual principles, results showed that most of participants’ choices matched with the example-based-learning-principle, example-first-principle, and lowest-level-first-principle. Moreover, the majority of the students started each complexity level with an example and another 21.1% did this only partially. Finally, only half of participants’ choices aligned with the simple-to-complex-principle entirely, and more than a quarter of participants’ choices aligned with this principle only partially.

Table 4.5.
Percentages of Principles from Example-based Learning Research Applied in the Self-regulated Learning Phase for the Total Sample (N = 147).

	Total sample		
	Fully	Partially	Not at all
Example-based learning principle	89.8%	9.5%	0.7%
Example first principle	95.9%	X	4.1%
Lowest level first principle	88.4%	X	11.6%
Simple-to-complex principle	49.0%	26.5%	24.5%
Start each level with example principle	76.9%	21.1%	2.0%
Total score Principles	39.4%	53.8%	6.8%

Note. X = not a scoring option for this principle.

4.3.3 Is there a positive relation between the extent to which students’ choices match with the instructional design principles and their scores on learning outcomes, mental effort, and motivational variables?

Finally, we explored how whether the degree of spontaneously applying the instructional design principles correlated with cognitive or motivational aspects of learning. As shown in Table 4.6, total scores of how well students task selections matched with the principles did not correlate with any of the cognitive or motivational variables. However, some of the individual principles did correlate with some of the cognitive or motivational variables. Firstly, there was a positive relation between spontaneously applying the example-first-principle and average scores on self-efficacy in the learning phase ($r = .183$). Secondly, spontaneously applying the lowest-level-first-principle negatively correlated with average scores of self-efficacy in the learning phase ($r = -.257$) and self-efficacy ($r = -.268$) and perceived competence ($r = -.219$) after the learning phase. The lowest-level-first-principle also negatively correlated with the scores on the procedural transfer task ($r = -.235$). A positive correlation was shown, however, between the lowest-level-first-principle and average ratings of mental effort invested in the conceptual questions in the posttest ($r = .232$). Finally, spontaneously applying the simple-to-complex-principle positively correlated with average scores of mental effort invested in the learning ($r = -.180$) and isomorphic posttest tasks ($r = -.183$). We must note, though, that the strength of these correlations can be referred to as small (because the absolute values of r are below .30; Cohen, 1988).

Table 4.6.

Correlation Coefficients (*r*) from Pearson Correlation Analysis Between Principles from Example-based Learning and Different Cognitive and Motivational Variables.

	Example-based learning principle		Example first principle		Lowest level first principle		Simple-to-complex principle		Start each level with example principle		Total score principles	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Learning phase												
Mental Effort	-.002	.983	-.069	.408	.147	.075	-.180	.029	-.114	.168	-.084	.312
Self-efficacy	-.045	.590	.183	.026	-.257	.002	.136	.100	.072	.383	.025	.767
Posttest												
Conceptual Questions	.094	.258	.148	.074	.032	.699	.022	.795	.037	.658	.090	.279
Isomorphic Tasks	.004	.966	.052	.535	-.064	.442	.067	.422	.019	.817	.028	.740
Procedural Transfer Task	.037	.752	.123	.292	-.235	.041	.117	.316	.165	.153	.050	.666
Mental Effort												
Conceptual Questions	.023	.787	-.134	.106	.232	.005	.031	.708	-.020	.811	.073	.378
Isomorphic Tasks	.085	.304	-.046	.583	.134	.106	-.183	.027	.000	.996	-.033	.690
Procedural Transfer Task	.101	.225	.017	.841	.161	.051	.004	.966	-.010	.906	.087	.294
Self-efficacy												
Perceived Competence	-.085	.305	.038	.647	-.268	.001	.081	.327	.000	.997	-.073	.381
Topic Interest	-.056	.504	.076	.360	-.219	.008	.115	.165	.036	.667	-.010	.900
	-.086	.300	-.004	.964	.089	.285	.149	.072	.084	.312	.117	.159

Note. Significant *p*-values are bolded.

4.4. Discussion

The aim of this study was to explore the task-selection choices of first year higher education students (i.e., novices to the learning materials) when engaging in self-regulated learning in an online learning environment. Students had to learn how to solve problems using the trapezoid rule and could select learning tasks from a database comprising different task formats (i.e., video modeling examples, worked examples, and practice problems), levels of complexity (i.e., three levels), and cover stories. We were particularly interested in which tasks students would choose (and why), and how students' task-selection decisions would adhere to the robust principles from instructional design research.

4.4.1. Students' Task-Selection Patterns

Results showed that the selection of video modeling examples significantly decreased and the selection of worked examples and problems significantly increased during the learning phase. In addition, the selection of lowest level tasks significantly decreased, whereas the selection of level 2 and level 3 tasks increased. Also, findings showed that students' choices matched quite well with the principles derived from instructional design research on the effectiveness and efficiency of different fixed sequences of examples and problems. The vast majority of students selected many examples during the learning phase (i.e., example-based-learning-principle) and started the learning phase with an example instead of a problem (i.e., example-first-principle). Although the choices of approximately half our sample aligned with the simple-to-complex principle, almost all participants started the learning phase with a task at the lowest complexity level (i.e., lowest-level-principle). Moreover, most participants started each complexity level with example study (i.e., start-each-level-with-example-principle).

That students spontaneously applied almost all of the instructional design principles (with the exception of the simple-to-complex-principle) is surprising. Although there is relatively little research on this issue, the available evidence suggested that novices underutilize example study with respect to the amount (i.e., about 40 percent worked examples versus 60 percent practice problems) and timing (i.e., students rarely started the learning phase with example study) of their use (e.g., Foster et al., 2018).

There are several possible reasons for why students' choices were so well aligned with the instructional design principles in our study compared to the study of Foster et al. (2018). Firstly, our sample consisted of technical higher education students instead of a mixed group of students obtained from the university's participant pool (as in the study of Foster et al., 2018). In the study programs of our sample, mathematics is an important subject and as a result, students might have already had experience with

learning new mathematical problem-solving skills with the help of examples (since examples are frequently used to learn new mathematical procedures). It is possible that the students used in the study of Foster et al., (2018) had less experience with example study when learning new (mathematical) problem-solving procedures, for example because mathematics might not have been part of their courses.

Secondly, it is possible that being able to select video modeling examples might have motivated our students to start the learning phase with example study. Studying video modeling examples could be a more familiar way of learning new problem-solving skills for students than studying worked examples (Hoogerheide & Roelle, 2020). With the rise of popular video-sharing platforms such as YouTube (where people can, for example, watch videos to learn new knowledge and skills on many different subjects), it is likely that many students have at least gained some experience with (the effectiveness of) learning new skills by studying video examples. This explanation is partly supported by the answers to the open questions, where students said they selected video modeling examples most often during the learning phase, because this format was most comfortable (i.e., most recognizable, most preferred, most clear, etc.) for learning.

Another explanation might be that video modeling examples are more preferred at the beginning of a learning phase (compared to worked examples), because information is demonstrated in a step-by-step manner and the combination of dynamic visual information and the model's narration take the learner by the hand. In contrast, worked examples might be preferred later in the learning phase, because they allow for efficiently looking up difficult problem-solving steps. Indeed, our findings revealed that almost all students started with a video modeling example. However, this number rapidly decreased while the selection of worked examples gradually increased. These results correspond with the results of Hoogerheide et al. (2014) that worked examples were also more preferred than video modeling examples when more tasks had to be studied. Tentative evidence was provided by the coding of the answers on the open questions, as students said they selected worked examples as a second task mostly because this format provided them the opportunity to assess their level of expertise and easily check to what extent they had understood the procedure.

Thirdly, a more likely and more practical explanation for why our sample relied so heavily on example study is that we provided the opportunity to choose between two example formats (i.e., video modeling examples and worked examples) next to practice problems. As a result, two thirds of the learning tasks were examples (i.e., 67%) and only one third were practice problems (i.e., 33%). This might have increased the likelihood of selecting an example rather than a practice problem. In the study of Foster et al. (2018), students could only choose between worked examples next to (completion) problems.

These possible explanations provide several interesting avenues for future research in self-regulated learning settings. For instance, it would be interesting to investigate in further detail whether or not familiarity with example study in one domain would affect the degree to which novices opt for example study relative to practice problem solving in the same or in a different domain. Moreover, as comparisons between (different sequences of) worked and video modeling examples are scarce (e.g., Hefter et al. 2019; Hoogerheide et al., 2014), future research could investigate whether starting the learning phase with a video modeling example and switching to worked examples is not only a more preferred way of learning but also more effective, efficient, and motivating than the other way around. Lastly, another interesting avenue for future research would be to determine whether providing the option of both video modeling and worked examples indeed helps to optimize the frequency and timing of example study.

4.4.2. Limitations

This study also has some limitations. First, because performance on the practice problems was not logged, it was not possible to examine the degree to which students' task-selections were adaptive to their needs. The optimal sequence (length) of examples and problems differs for each individual learner because there is variance in the speed to which students learn (e.g., due to differences in cognitive abilities). If we had access to practice problem solving performance, we could score whether students made accurate decisions following a practice problem. For instance, students who just failed to solve a problem should ideally select an example or another practice problem of the same complexity level, while students should select a more complex task after successfully solving a practice problem. That task-selections should ideally be tailored to individual progress could also explain the lack of correlations between the extent to which students' task selections matched with the instructional design principles and learning outcomes. Moreover, had we successfully logged performance on the practice problems, we could determine whether students' knowledge during the learning phase was so high that they would benefit more from problem solving than example study (cf. the expertise reversal effect; Kalyuga et al., 2003). However, it is unlikely that this expertise-reversal effect occurred, because performance on the isomorphic posttest was not that high and students were only allowed to select six learning tasks while having to learn three different complexity levels.

A second limitation of the present study concerns the measurement of self-efficacy and perceived competence. There is research showing overlap between these two constructs, and more specifically that perceived competence may be a common core component of both self-efficacy and self-concept (e.g., Marsh et al., 2019; Schunk & Pajares, 2005). The results of our study confirm this idea, as correlational analyses of these two constructs measured after the self-regulated learning phase revealed a

score of $r = .86$. One could wonder to what extent both measures differ or measure the same general feeling of competence regarding to what has been learned and how well someone considers him/herself capable of solving a similar task. Therefore, it might be sufficient for future research in this area to use of one of the questionnaires.

4.4.3. Conclusions and Practical Implications

In sum, our explorative study showed that students' task-selection patterns corresponded fairly well with principles derived from instructional design research. This seems promising, because it would mean that students know quite well how to use examples and problems (at different complexity levels) when learning new problem-solving skills and therefore might need little support. However, given the paucity of research on self-regulated learning of examples and problems (at different levels of complexity), the mixed findings regarding the use of examples and problems (i.e., Foster et al., 2018 vs. our study), and the open question of whether students' task selections are adapted to their levels of expertise, we cannot say this with absolute certainty. Moreover, regarding the task selections (of some of the students) and test performance scores in our study and the study of Foster et al. (2018), there seems to be (some) room for improvement in how students regulate their learning from examples and problems. Therefore, more research is needed to gain insight in how and how well (novice) learners regulate their learning from examples and practice problems, and whether and how they can benefit from support. Moreover, future research should investigate to what extent the findings of this study regarding students' task selections are problem-specific or generalizable, for example by using similar procedures but different problem-solving tasks.

Chapter 5

Instructing students on effective sequences of examples and problems: Does self-regulated learning improve from knowing what works and why?



This chapter was submitted as:

Van Harsel, M., Hoogerheide, V., Verkoeijen, P. P. J. L., & Van Gog, T. (2020). *Instructing students on effective sequences of examples and problems: Does self-regulated learning improve from knowing what works and why?* Manuscript submitted for publication.

Acknowledgement of author contributions:

MvH, VH, PV, and TvG designed the study, MvH recruited participants and collected the data, MvH analyzed the data, VH checked the data package, MvH drafted the manuscript, all authors contributed to critical revision of the manuscript, VH, PV, and TvG supervised the study.

Abstract

Nowadays, students often practice problem-solving skills in online learning environments. This requires them to self-regulate their learning. The present study investigated the open questions (1) to what extent students' task selections align with instructional design principles and (2) whether informing them about these principles would improve their task selections, learning outcomes, and motivation. Higher education students ($N = 150$) learned a problem-solving procedure by fixed sequences of examples and problems (FS-condition), or by self-regulated learning (SRL). The SRL participants selected tasks from a database, varying in format, complexity, and cover story, either with (ISRL-condition) or without (SRL-condition) watching a video detailing the instructional design principles. Students' task-selection patterns in both SRL conditions largely corresponded to the principles, although tasks were built up in complexity more often in the ISRL-condition than in the SRL-condition. There were no test performance or motivational differences among conditions. Implications for practice and theory are discussed.

Keywords: example-based learning, self-regulated learning, problem-solving, task selection, motivation, higher education.

5.1. Introduction

Decades of research on instructional design have resulted in several principles for optimizing the acquisition of new problem-solving skills for novices (i.e., students with little if any prior knowledge). These principles are concerned with how to ensure that novices work on tasks that provide an optimal level of instructional support and complexity given their current level of knowledge (4C/ID Model; Van Merriënboer, 1997; Van Merriënboer & Kirschner, 2013). Following these principles should make novices' learning process more effective and efficient, and make them feel more self-efficacious. Nowadays, students often have to self-regulate their learning of problem-solving skills in online learning environments, and it is an open question to what extent their task selections during self-regulated learning align with those principles. Therefore, the present study investigates to what extent students' task selections align with instructional design principles and whether informing students about such principles would improve their task selections, motivation, and learning outcomes.

5.1.1 Instructional Design Principles to Support Acquisition of Problem-Solving Skills

The *example-based-learning-principle* postulates that replacing all or a substantial number of practice problems with worked examples (i.e., a written step-by-step explanation of how to solve a problem; e.g., Van Gog et al., 2011) or video modeling examples (i.e., a person demonstrating and/or explaining a problem-solving procedure on video; e.g., Kant et al., 2017) helps novices to learn more (i.e., is more effective) with less time and effort investment (i.e., is more efficient) than solving practice problems without any instructional support (e.g., Sweller et al., 2011; Van Gog et al., 2019). Moreover, recent findings show that studying examples also increases students' self-efficacy during learning compared to only solving practice problems (e.g., Van Harsel et al., 2019, 2020; Coppens et al., 2019). When alternating examples and problems, research has shown that novices should start with an example (instead of practice problem-solving), as this was found to be more efficient for learning than starting with problem-solving only (e.g., Van Harsel et al., 2019, 2020; Van Gog et al., 2011). We refer to this as the *example-first-principle*.

These principles should be considered in relation to task complexity. Students should ideally be working on tasks that are at an optimal level of complexity given their current level of knowledge (4C/ID Model; Van Merriënboer, 1997; Van Merriënboer & Kirschner, 2013). The *lowest-level-first-principle* postulates that novices should start with a task at the lowest level of complexity. From there, the level of complexity should gradually increase as their knowledge increases: the *simple-to-complex-principle* (cf. 4C/ID model). According to the 4C/ID model, students should receive a high level of instructional support (like an example) at the start of each new complexity level: The *start-each-level-with-example-principle*.

These principles have been mainly derived from research with fixed (or adaptive) sequences of examples and problems (at different complexity levels), where the learning environment or the experimenter determined whether, when, and for how long a learner should study examples or solve practice problems. However, learners, particularly in higher education, spend much of their time on self-study activities where they need to self-regulate their learning as effectively and (given that they also need to devote time to other coursework, part-time jobs, et cetera) as efficiently as possible. It is still an open question, however, how (well) learners regulate their learning from examples and problems when they can make their own choices (e.g., Van Gog et al., 2020; Van Gog et al., 2019), and whether and how they could be guided in making choices that fit their learning needs.

5.1.2 Self-Regulated Learning of Examples and Problems

Self-regulated learning of problem-solving tasks is notoriously difficult for novices, because they need to be able to accurately assess their understanding or performance on the just completed task and subsequently select a new task with the right level of complexity and support (De Bruin & Van Gog, 2012; Van Gog et al., 2020). Research has shown that novices often experience difficulties in accurately assessing their performance (e.g., Dunning et al., 2004; Koriath & Bjork, 2005) and mostly overestimate (though sometimes underestimate) their own performance (e.g., Hacker & Bol, 2019; Kostons et al., 2010, 2012). Consequently, tasks might be selected that are too complex or too simple, or do not provide the necessary instructional support (e.g., Dunlosky & Rawson 2012). Moreover, novices do not always seem to be aware which task aspects influence how much they learn (e.g., Kostons et al., 2010; Nugteren et al., 2018), and therefore tend to select tasks based on irrelevant (e.g., cover stories) instead of relevant task aspects (i.e., complexity and support; e.g., Corbalan et al., 2008).

Based on these findings, one might expect that novices experience difficulties when self-regulating their learning from examples and problems. Indeed, a recent study conducted by Foster and colleagues (2018) found that novices make suboptimal choices when they are in control of selecting tasks to work on. In their study, university students had to learn how to solve probability problems and were repeatedly given the choice of whether to study a worked example or to practice solving a (completion) problem. Results showed that on average, students opted more often for (completion) problems than examples and rarely started the learning phase with example study. These choices are at odds with the example-based-learning-principle and the example-first-principle, as studying examples, especially at the start of the learning phase, is more efficient (and effective), and motivating for learning than (starting with) problem-solving only (e.g., Van Gog et al., 2011; Van Harsel et al., 2020).

In contrast, Van Harsel et al. (submitted) found other results. Higher education students learned how to solve a math problem by selecting 6 learning tasks from a database that consisted of 45 learning tasks that differed in format (worked examples, video modeling examples, and practice problems), complexity (three levels), and cover story. Results showed that most of the learners' choices matched with the instructional design principles: the vast majority of students selected many examples during the learning phase, as they started the learning phase with an example at the lowest complexity level and often started a new complexity level with example study as well. However, the complexity of tasks was built up less well by only half of the sample: Particularly those who performed poorly on the posttest kept selecting examples or practice problems at the lowest complexity level. It is, however, an open question whether self-regulated learning would be as effective as fixed sequences of tasks based on those principles. Moreover, as there still was room for improvement in learners' task selections (and test performance scores), they might benefit from instructional support to help them self-regulate their learning, for instance by explicitly informing learners prior to self-regulated learning about the principles derived from instructional design research.

5.1.3 Strategy Instruction to Support Self-Regulated Learning of Examples and Problems

Explicitly informing students about learning strategies has been found to be successful for increasing learners' metacognitive beliefs and/or knowledge (e.g., Endres et al., 2021; Lineweaver et al., 2011; McCabe, 2011; Yan et al., 2016) and their use of these strategies (e.g., Biwer et al., 2019). Ariel and Karpicke (2017) even found that explicitly informing students about learning strategies also improved their learning outcomes. They asked university graduates to learn Lithuanian-English word-pairs. Students could decide for themselves whether to restudy word-pairs, whether to retrieve already learnt word-pairs from memory (i.e., retrieval practice, a proven effective study strategy for word-pair learning; Rowland, 2014), or to stop learning. The experimental condition received a short-written instruction with information about the effectiveness and mnemonic benefits of repeated retrieval practice and how to use it, while the control condition did not receive this information. Results showed that students in the experimental condition used the repeated retrieval practice strategy more often than those in the control condition, and subsequently outperformed the control group on an immediate cued-recall test. Students in the experimental condition even (spontaneously) used retrieval practice to learn novel materials a week later more often than the control condition.

A possible explanation for why informing students about effective learning strategies can improve the use of such strategies and learning (cf. Ariel & Karpicke, 2017) could be that this information helps learners become (more) aware of the value

of a strategy and increases their metacognitive knowledge (i.e., knowledge about why and which strategies are [not] beneficial for learning). In turn, this could increase the likelihood that an individual will search for, modify, and apply that strategy (e.g., Tullis et al., 2003; Yan et al., 2014).

These findings are promising, given that this approach of informing students about effective strategies would be relatively easy to use across a variety of learning materials and contexts. However, it is an open question whether this approach would also be effective for improving self-regulated learning of problem-solving skills with examples and problems. Therefore, the present study investigates whether informing students on effective strategies for learning new problem-solving skills with examples and problems (at different complexity levels) would increase the likelihood that they select the right tasks (according to their level of expertise), and would increase performance and motivation.

5.1.4 The Present Study

The first aim of this study was to investigate whether the finding that students' choices during self-regulated learning aligned quite well with the instructional design principles for learning from examples and problems (cf. Van Harsel et al., submitted) would replicate (Research Question 1), because this result is rather surprising in light of other related research (e.g., Foster et al., 2018). To shed further light on the quality of students' task selections, the present study also explored what tasks learners select after solving a practice problem, which was not possible in Van Harsel et al. (submitted) as practice problem performance data were unavailable.

The second aim was to examine whether self-regulated learning would be as effective, efficient, and motivating as a fixed task sequence based on the principles derived from instructional design research (Research Question 2). We consider this an open question. It is possible that self-regulated learning would have motivational benefits, given that related research suggests that allowing learners to select their own tasks can improve other motivational variables such as interest (e.g., Corbalan et al., 2008), however, it is questionable whether this also applies to motivational variables such as self-efficacy and perceived competence. At the same time, self-regulated learning might be less conducive to learning outcomes than a predetermined sequence, as self-regulated learning has been found to impair learning outcomes relative to teacher- or computer-controlled fixed or personalized instruction (see e.g., Azevedo et al., 2008; Lawless & Brown, 1997; Niemiec et al., 1996).

Thirdly, given that there still was room for improvement in learners' task selections and test performance scores in the study of Van Harsel et al. (submitted), we investigated whether explicitly informing learners about instructional design principles

would enhance their self-regulated learning of examples and problems (at different complexity levels) compared to self-regulated learning without such information (Research Question 3). Assuming that students in the 'informed self-regulated learning condition' actually adopt these principles (cf., studies on other learning strategies: Ariel & Karpicke, 2017; Biwer et al., 2019), one could expect their choices to be better aligned with the principles than students' choices in the self-regulated learning condition and therefore show higher test performance (i.e., on conceptual questions, isomorphic tasks, and procedural transfer tasks), attained with lower effort investment and time-on-task in the learning and posttest phase. As for the comparison between the informed self-regulated learning and fixed sequences condition, we consider this an open question. If informing students about effective principles for learning from examples and problems would help students select better tasks, they might show similar performance as the fixed sequences condition. Effects on self-efficacy and perceived competence are explored.

5.2 Method

5.2.1 Participants and design

Participants were 241 students from a Dutch university of applied sciences ($M^{age} = 18.84$, $SD = 1.76$; 232 male), enrolled in the first year of an electrical and electronic mechanical engineering program. Participants had to learn how to approximate the definite integral of a function using the trapezoidal rule. They were randomly allocated to one of three conditions, namely the 1) informed self-regulated learning condition (ISRL; $n = 109$), 2) self-regulated learning condition (SRL; $n = 60$), and fixed sequences condition (FS; $n = 72$). More participants were assigned to the ISRL-condition to increase the chances of having a sufficiently large subset of students who would follow the advice and to be able to explore differences between students who did and did not follow the advice. The experiment consisted of three phases: (1) pretest, (2) learning phase, and (3) posttest. Participants who did not finish the isomorphic (and transfer) items on the posttest on time were excluded from further analysis ($n = 42$). Moreover, we also excluded 49 participants with too much prior knowledge (indicated by a score of 5 or more out of 9 on the prior knowledge test), because we were specifically interested in the selection behavior of novice learners. Therefore, the final sample consisted of 150 participants ($M^{age} = 18.68$, $SD = 1.57$; 143 male) divided over the ISRL-condition ($n = 66$), SRL-condition ($n = 32$), and FS-condition ($n = 52$). Students could earn study credits for participation. All participants gave informed consent in the learning environment.

5.2.2. Materials

All materials were based on the materials developed by Van Harsel et al. (2019, 2020) and Van Harsel et al. (submitted), and presented in a web-based learning environment.

5.2.2.1 Pretest

The pretest consisted of five conceptual knowledge questions that measured participants' understanding of the trapezoidal rule ($\alpha = -.73$). These questions consisted of a multiple-choice part with four answer options and an explanation part where participants had to explain their answer (see Supplementary Materials E for an example of a pretest question). Note that a possible reason for the poor reliability of the pretest is that students had (very) low prior knowledge but a 25% chance to guess the right answer.

5.2.2.2 Instructional video

The instructional video that was used to inform students in the ISRL-condition on effective instructional design principles started with a brief explanation of the procedure of the experiment. Students were informed that they were going to select learning tasks themselves and would receive help on how to select the most effective and efficient learning task, based on well-established findings from scientific research. Then, a total of four 'rules' were presented: a) "At the start of the learning phase, choose a task at the lowest complexity level", b) "When you mastered a complexity level, choose a task one complexity level higher", c) "Start each new complexity level with example study and alternate with practice problems when you want to check whether you understand how to solve the problem", d) "Start the learning phase with a video modeling example and continue with written examples when more example study is necessary". We added the fourth rule as there are some indications that a video modeling example is preferred at the start and worked examples later in the training phase (e.g., Van Harsel et al., submitted; Hoogerheide et al., 2014). This might be explained by the fact that in video modeling examples, information is demonstrated step-by-step and the combination of dynamic visual information and the model's verbal explanations take the learner by the hand. In worked examples, information is also demonstrated step-by-step, however, shown all at once. This allows for efficiently looking up difficult problem-solving steps and therefore might be preferred later in the learning phase. Each rule was accompanied with the necessary background information about why this rule would help students learn more and when/how to apply it (see Supplementary Materials N). The instructional video lasted 223 seconds.

5.2.2.3. Task database

Together with three mathematics teachers from the university of applied sciences where the study was conducted, a task database consisting of 45 learning tasks was developed (see Figure 5.1). The tasks required participants to approximate a specific region under the graph of a function using the trapezoidal rule. The tasks varied in complexity level, task format, and cover story.

Complexity level. The learning tasks were developed at three levels of complexity. Tasks at *complexity level 1* required participants to use the trapezoidal rule to approximate the region under the graph of a polynomial function of degree 2 (i.e., quadratic function). Moreover, functions were constructed in such a way that participants had to calculate more than two intervals and calculate with fractions and positive numbers only. Tasks at *complexity level 2* were more difficult, since they asked participants to calculate with negative numbers as well. Tasks at the highest complexity level (i.e., *complexity level 3*) additionally asked participants to use the trapezoidal rule to approximate the region under the graph of a polynomial function of degree 3 (i.e., cubic function).

Format. The learning tasks were developed in three different formats: video modeling examples, worked examples, and (conventional) practice problems. Video modeling examples consisted of a screen recording of a female model's computer screen, where she demonstrated (with PowerPoint slides and handwritten notes) and explained step-by-step how to solve a problem using the trapezoidal rule. The model started with an introduction on the trapezoidal rule, followed by an explanation of the problem state and an explanation of how to use the information that was presented on the screen to solve the problem (i.e., the graph of a function, the left border and right border of the area, the number of intervals, and the formula of the trapezoidal rule). Subsequently, she showed and explained how to solve the problem by calculating four steps: 1) 'compute the step size of each subinterval', 2) 'calculate the x-values', 3) 'calculate the function values for all x-values', and 4) 'enter the function values into the formula and calculate the area', and ended the video by providing the final answer.

Worked examples were presented on one page and consisted of a written step-by-step explanation of the solution procedure. Worked examples also started with a short description of the problem state and some additional information that was needed to solve the problem (i.e., the graph of a function, the left border and right border of the area, the number of intervals, and the formula of the trapezoidal rule). Subsequently, written explanations (and correct answers) were given for each of four steps on how to solve the problem.

Practice problems also started with a short description of the problem state and the additional information that was needed to solve the problem. However, it was not explained how to use the information that was given to solve the problem. Participants received the following assignment: "Approach the area under the graph using the information that is given. Write down all your intermediate steps and calculations". Screenshots of the three task formats are presented in the Supplementary Materials A, B, and C.

Cover story. Finally, tasks varied in cover story. For example, participants could solve a problem that asked them to approximate how many liters of beer were tapped within a certain amount of time (i.e., drinking beer) or approximate how often the circular platform of a carousel rotates in a given period of time (i.e., carousel). The cover stories were similar for each task format that was provided within a complexity level (e.g., drinking beer could be selected as video modeling example, worked example, and practice problem), yet the numbers used differed per task format.

DE PROEF OP DE SOM								
Level 1			Level 2			Level 3		
Video modeling examples	Worked examples	Practice problems	Video modeling examples	Worked examples	Practice problems	Video modeling examples	Worked examples	Practice problems
Drinking beer	Drinking beer	Drinking beer	Running	Running	Running	Carousel	Carousel	Carousel
Energy measurement	Energy measurement	Energy measurement	Drinking water	Drinking water	Drinking water	Rowing	Rowing	Rowing
Fitness	Fitness	Fitness	Washing machine	Washing machine	Washing machine	Perfume	Perfume	Perfume
Fuel consumption	Fuel consumption	Fuel consumption	Soapsuds	Soapsuds	Soapsuds	Coffee consumption	Coffee consumption	Coffee consumption
Traffic	Traffic	Traffic	Plasterer	Plasterer	Plasterer	Chocolate party	Chocolate party	Chocolate party

Figure 5.1. Screenshot of the task database.

Task sequences. During the learning phase, participants in the two SRL conditions could select six tasks from the task database (see Figure 5.1; each task could be selected only once). They were instructed that the posttest would include tasks at all three complexity levels. Participants in the FS-condition received 6 tasks from the task database in the following order: (1) video modeling example at complexity level 1, (2) worked example at complexity level 1, (3) practice problem at complexity level 1, (4) worked example at complexity level 2, (5) practice problem at complexity level 2, and (6) worked example at complexity level 3. The cover stories of these tasks were randomly chosen.

5.2.2.4. Posttest

The *posttest* consisted of five tasks. The first three tasks concerned a level 1, 2, and 3 task; these were isomorphic to the learning phase tasks ($\alpha = .81$). The fourth task was a procedural transfer task that required participants to use the Simpson rule to approximate the definite integral under a graph. Simpson’s rule is also a numerical integration method, however, uses quadratic polynomials (instead of the straight-line segments) to approximate the region under a graph. The final task consisted of five questions that aimed to measure participants’ understanding of the trapezoidal rule ($\alpha = .48$), and these were isomorphic to the pretest questions. Examples of test tasks are shown in the Supplementary Materials F, G, and H.

5.2.2.5. Mental effort

Mental effort was measured using a 9-point rating scale (Paas, 1992), asking participants to rate how much mental effort they invested in studying an example or solving a practice problem. Answer options ranged from (1) “very, very low mental effort” to (9) “very, very high mental effort”. Mental effort was rated after each learning and posttest task, with the exception of the five conceptual posttest questions (where it was rated only once after the final item).

5.2.2.6. Self-efficacy

Self-efficacy was measured by asking participants for their confidence in that they could approximating the definite integral of a graph using the trapezoidal rule. Answer options ranged from (1) “very, very unconfident” to (9) “very, very confident” (Van Harsel et al., 2019, 2020; adapted from Hoogerheide et al., 2016).

5.2.2.7. Perceived competence

Perceived competence was measured using an adapted version of the *Perceived Competence Scale for Learning* (Van Harsel et al., 2019, 2020, based on Williams & Deci, 1996), consisting of three items (instead of the 4 items), such as “I feel confident in my ability to learn how to approximate the definite integral of a graph using the trapezoidal rule”. Participants had to rate on a scale of (1) “not at all true” to (7) “very true” to what degree the items applied to them ($\alpha = .95$).

5.2.3. Procedure

Fourteen single sessions (with 9 to 24 participants per session) that lasted 102 minutes on average were run in a computer classroom at participants’ university of applied sciences. Before each session, a headset, pen, and scrap paper were placed on the tables. After participants arrived, the experimenter first explained the aim and procedure of the experiment. Then, participants were told that they could work at their own pace (with a maximum of 135 minutes), and that they had to write down as much as possible and to write an “X” if they really did not know what to answer. Students could use a calculator (different from Van Harsel et al., submitted).

After the instructions, participants entered the online learning environment. Each task and questionnaire were presented on a separate page. Participants were unable to go to the next page before completing the current task/questionnaire and were unable to go back to any previously completed pages. Time-on-task was logged.

Participants were first provided with a short demographic questionnaire (e.g., age and gender), the pretest, and self-efficacy and perceived competence questionnaires. Next, the learning environment provided written instructions about the learning phase. For the SRL conditions, these instructions explained that 6 tasks had to be selected from the task database and how to select a task to work on. Participants in the ISRL-

condition additionally were told that they had to watch an instructional video that explained how to select tasks to learn most effectively and efficiently. In the FS-condition, participants were told that they had to study or solve the tasks that were provided to them. In all conditions, participants had to rate their mental effort and self-efficacy after each task in the learning phase. After the learning phase, participants had to turn their scrap paper upside down and were provided with a new scrap paper. Then, participants completed the self-efficacy and perceived competence questionnaires and the posttest. After the posttest, participants handed in their materials and left the classroom.

5.2.4. Data Analysis

To answer our first research question, we used the same approach as Van Harsel et al. (submitted). We first analyzed what tasks participants selected in the SRL conditions and coded the task format (video modeling example, worked example, practice problem) and complexity level (1, 2, or 3) of the selected learning tasks, and converted the scores into percentages. We then coded to what degree participants' task-selection behavior matched with the evidence-based instructional design principles (i.e., example-based-learning, example-first, simple-to-complex, lowest-level-first, and start-each-level-with-example-principle). For each principle, participants could earn 1 point in total. Following the example-first-principle and lowest-level-first-principle was awarded with 1 point and 0 points were assigned when the principle was not followed. For the example-based-learning-principle, simple-to-complex-principle, and start-each-level-with-example-principle, 1 point was granted when the principle was followed entirely, 0.5 points when the principle was followed partially, and 0 points when the principle was not followed at all (same scoring protocol as used in Van Harsel et al., submitted). For each participant, a total score was computed that represented how well all principles were followed (maximum: 5 points).

Then, we scored participants' performance on the practice problems. A maximum of 8 points could be earned for each practice problem, with 2 points for calculating each step correctly: 1) the step size of each subinterval, 2) all x-values, 3) the function values for all x-values, and 4) using the correct formula for the area under the graph and providing the correct answer. In step two, three, and four, 2 points were given when all solution steps were correct, 1 point was given if half or more of the steps were correct, and 0 points when less than half of the steps were correct. To explore how well students' task-selection behavior matched with their performance on the practice problems, we scored whether students selected a new task (i.e., video modeling example, worked example, or practice problem) on a higher complexity level when a practice problem was graded with 6 or more out of 8 points (75% or more correct). We also scored whether students selected a new task (i.e., video modeling example, worked example, or practice problem) on a similar (or lower) complexity level when a practice problem was graded with less than 6 out of 8 points (less than 75% correct).

To answer our second and third research question, we also used the same approach as Van Harsel et al. (submitted). We first scored performance on the conceptual pretest and posttest items. On the conceptual pretest questions and conceptual posttest questions, participants could earn a maximum of 9 points. One point could be earned for the first open-ended question (correct answer: 1 point; incorrect answer: 0 points) and 2 points for the other open-ended questions. Participants were only rewarded with the maximum of 2 points when they got the answer right and provided correct explanations. Only 1 point was awarded when the answer was correct but the explanation was incorrect or missing, and 0 points were given when the answer and explanation were incorrect. With regard to performance on the posttest, the isomorphic posttest items (i.e., three tasks, max. score = 24 points) and procedural transfer item (i.e., 1 task, max. score = 8 points) were scored similarly to the learning tasks. Averages of mental effort invested in the learning tasks and posttest tasks were calculated, as well as the averages of participants' self-efficacy and perceived competence ratings before, during (only self-efficacy), and after the learning phase.

5.3. Results

Descriptive statistics were used to evaluate the first research questions on how students behaved in the SRL conditions, how well their behavior matched with evidence-based principles from instructional design research, and whether students made the right choices in the learning phase according to their performance on the practice problems. Percentages are only mentioned in the text if they cannot be found in the Tables.

Non-parametric tests were used to answer the second and third research question because the main variables were not normally distributed (i.e., the kurtosis and/or skewness values, divided by their standard error, were below -1.96 or above +1.96; cf. Field, 2009). The effects of Test Moment (Pretest and Posttest) were tested with Wilcoxon signed-rank tests. Differences between the SRL conditions in following the instructional design principles were tested with Mann-Whitney U tests. Differences between the ISRL-Condition, SRL-Condition, and FS-Condition regarding cognitive (i.e., performance on the conceptual tests, isomorphic test, procedural transfer test, as well as mental effort and time-on-task in learning and posttest phases) and motivational aspects of learning (i.e., self-efficacy and perceived competence) were tested with Kruskal-Wallis tests. For post-hoc tests, we used Mann-Whitney U tests, with a Bonferroni corrected significance level of $p < .017$ (i.e., $0.05/3$) for the Wilcoxon signed-rank tests and a Bonferroni-corrected alpha level of $p < .017$ (i.e., $0.05/3$) for the Kruskal-Wallis tests. For the post-hoc tests, the effect size of Pearson's correlation (r) is reported (i.e., Z/\sqrt{N}), with values of 0.10, 0.30, and 0.50 representing a small, medium, and large effect size, respectively (Cohen, 1988).

Before non-parametric analyses were conducted, we checked for pre-existing differences among the three conditions. Kruskal-Wallis tests showed no significant differences among conditions on pretest performance $H(2) = 2.60, p = .273$, nor on self-efficacy $H(2) = 0.73, p = .696$, or perceived competence $H(2) = 1.70, p = .919$. We also checked whether participants in the ISRL-Condition actually watched the instructional video detailing the instructional design principles. Results showed that 57.6% of the participants watched the entire video (i.e., $n = 38$), 21.2% watched between half and three quarter of the video ($n = 14$), and 21.2% watched less than half of the video instruction ($n = 14$). Additionally, we explored whether there were differences between these three subgroups in terms of following the instructional design principles and cognitive and motivational aspects of learning (see Supplementary Materials O).

5.3.1. To what extent do novices’ task-selection patterns match with the findings from example-based learning research?

We first checked the percentages of selected examples and problems (see Table 5.1 and Table 5.2) and complexity levels (Table 5.3) and analyzed how well students’ choices matched with the instructional design principles (Table 5.4) in the ISRL and SRL-Condition. Almost all participants in both conditions started the learning phase with an example instead of a practice problem. On the second learning task, the percentage of selected examples rapidly decreased whereas the percentage of practice problems increased in both conditions. In de ISRL-Condition, problem-solving was preferred over example study on the second and third learning task, however, example study became most popular again on the fourth and fifth learning task. In the SRL-Condition, example study remained most popular up to and including the fifth learning task. Only on the last learning task, practice problems were preferred over example study in both conditions. The findings of the SRL-condition seem to be in line with the findings of Van Harsel et al. (submitted).

Moreover, participants in the SRL-Condition preferred video modeling examples (37.4%) over worked examples (27.0%), however, formats were almost equally preferred in the ISRL-Condition (video modeling examples: 29.6%; worked examples: 28.8%). In both conditions, participants clearly preferred a video modeling example as the first learning task compared to a worked example (or a practice problem). These percentages dropped considerably on the second learning task, as worked examples became more popular. Nevertheless, the popularity of the video modeling examples rose again on the third and fourth learning task (especially in the SRL-Condition); however, these percentages dropped again on the last two learning tasks. The selection of worked examples remained fairly stable in the ISRL-Condition from the second learning task onwards, with an outlier on the fifth learning task. In the SRL-Condition, the selection of worked examples dropped after the second learning task yet increased again on the last two learning tasks. The findings of the SRL-condition again replicate the results of Van Harsel et al. (submitted).

Findings also showed that the level of complexity was gradually built up in both conditions. The lowest complexity level was selected most on the first two learning tasks, the second complexity level was selected most on the third and fourth learning task, and the most difficult complexity level was selected most on the last two learning tasks. The results of the SRL-condition are again in line with the findings of Van Harsel et al. (submitted).

Table 5.1.
Percentages of Selected Examples and Practice Problems in the Informed Self-Regulated Learning Condition (n = 66) and Self-Regulated Learning Condition (n = 32).

	Informed Self-Regulated Learning Condition		Self-Regulated Learning Condition	
	<i>Example</i>	<i>Practice problem</i>	<i>Example</i>	<i>Practice problem</i>
Learning task 1	95.5%	4.5%	100%	0%
Learning task 2	43.9%	56.1%	56.3%	43.7%
Learning task 3	46.2%	53.8%	59.4%	40.6%
Learning task 4	57.6%	42.4%	58.1%	41.9%
Learning task 5	66.7%	33.3%	71.0%	29.0%
Learning task 6	40.6%	59.4%	41.9%	58.1%

Table 5.2.
Percentages of Selected Video Modeling Examples, Worked Examples, and Practice Problems in the Informed Self-Regulated Learning Condition (n = 66) and Self-Regulated Learning Condition (n = 32).

	Informed Self-Regulated Learning Condition			Self-Regulated Learning Condition		
	Video modeling example	Worked example	Practice problem	Video modeling example	Worked example	Practice problem
Learning task 1	84.8%	10.6%	4.6%	81.3%	18.7%	0.0%
Learning task 2	9.1%	34.8%	56.1%	18.8%	37.5%	43.7%
Learning task 3	20.0%	26.2%	53.8%	43.8%	15.6%	40.6%
Learning task 4	34.8%	22.7%	42.4%	41.9%	16.1%	41.9%
Learning task 5	18.2%	48.5%	33.3%	29.0%	42.0%	29.0%
Learning task 6	10.9%	29.7%	59.4%	9.7%	32.3%	58.1%

Table 5.3.
Percentages of Selected Complexity Levels (Level 1, 2, and 3) in the Informed Self-Regulated Learning Condition (n = 66) and Self-Regulated Learning Condition (n = 32).

	Informed Self-Regulated Learning Condition			Self-Regulated Learning Condition		
	Complexity Level 1	Complexity Level 2	Complexity Level 3	Complexity Level 1	Complexity Level 2	Complexity Level 3
Learning task 1	97.0%	0.0%	3.0%	87.5%	6.3%	6.2%
Learning task 2	89.4%	4.5%	6.1%	75.0%	18.8%	6.2%
Learning task 3	47.7%	46.2%	6.1%	40.6%	40.6%	18.8%
Learning task 4	18.2%	60.6%	21.2%	22.6%	45.2%	32.2%
Learning task 5	15.2%	30.3%	54.5%	16.2%	29.0%	54.8%
Learning task 6	14.1%	15.6%	70.3%	16.1%	19.4%	64.5%

Analyzing how well students’ choices matched with the instructional design principles revealed that participants’ choices matched very well with these principles (see Table 5.4). Moreover, results revealed that in both SRL conditions, many participants followed (almost all of) the principles entirely (as their total score was between 4.5 and 5 out of a maximum of 5 points; ISRL: 57.6%, SRL: 46.0%) or partially (as their total score was between 3 and 4.5 out of a maximum of 5 points; ISRL: 37.9%, SRL: 50.1%). As a result, there were no significant differences between the ISRL-Condition and the SRL-Condition on the ‘total score’ ($U = 930$, $p = .294$, $r = .106$), nor in the degree to which both conditions followed the example-based-learning-principle ($U = 1022.5$, $p = .570$, $r = .057$), example-study-first-principle ($U = 1104$, $p = .223$, $r = .123$), lowest-level-first-principle ($U = 956$, $p = .068$, $r = .184$), or start-each-level-with-example-principle ($U = 1170$, $p = .209$, $r = .127$). There was, however, a significant difference between conditions in following the simple-to-complex-principle ($U = 838.5$, $p = .042$, $r = .205$), which was followed entirely by 77.3% of the participants in the ISRL-Condition and only by 53.1% of the participants in the SRL-Condition.

Table 5.4.
Percentages of Example-Based Learning Principles Applied in the Informed Self-Regulated Learning Condition (n = 66) and Self-Regulated Learning Condition (n = 32).

	Informed Self-Regulated Learning Condition			Self-Regulated Learning Condition		
	Principle followed entirely	Principle followed partially	Principle followed not at all	Principle followed entirely	Principle followed partially	Principle followed not at all
Example-based-learning-principle	93.9%	4.6%	1.5%	90.6%	9.4%	0.0%
Example-first-principle	95.5%	X	4.5%	100.0%	X	0.0%
Lowest level-first-principle	97.0%	X	3.0%	87.5%	X	12.5%
Simple-to-complex-principle	77.3%	6.0%	16.7%	53.1%	28.1%	18.8%
Start-each-level-with-example-principle	77.3%	18.2%	4.5%	87.5%	12.5%	0.0%

Note. X = not a scoring option for this principle.

5.3.2. To what extent do novices make effective task selections after a practice problem-solving attempt?

To shed further light on the quality of students’ task selections, we categorized the type of decisions students made after problem-solving practice, taking into account whether they performed well on the practice problem (75%-100% correct; ‘standard achieved’) or not (less than 75% correct, ‘standard not achieved’). The results are presented in Table 5.5. The ISRL-Condition made many more task selections after a practice problem than the SRL-Condition (i.e., 125 vs. 48, respectively). Both conditions made more effective task-decisions (i.e., moving up a complexity level after achieving the standard, or not moving up a complexity level after failing to achieve the standard) than ineffective task-decisions (all other choices, classified as “other task-selection decisions”). However, there was definitely room for improvement in both conditions, as approximately 40% of the task selections were likely classified as ineffective for learning

Table 5.5.
Percentages of Task-Selection Decisions After Practice Problem Solving in the Informed Self-Regulated Learning Condition (n = 66) and Self-Regulated Learning Condition (n = 32).

	Informed Self-Regulated Learning Condition	Self-Regulated Learning Condition
	125 task-selection decisions after practice problem solving	48 task-selection decisions after practice problem solving
Standard achieved, video modeling example on higher complexity level	16.8%	18.7%
Standard achieved, worked example on higher complexity level	18.4%	12.5%
Standard achieved, practice problem on higher complexity level	7.2%	2.1%
Standard not achieved, video modeling example on similar or lower complexity level	4.0%	14.6%
Standard not achieved, worked example on similar or lower complexity level	8.0%	6.3%
Standard not achieved, practice problem on similar or lower complexity level	5.6%	8.3%
Other task-selection decisions*	40.0%	37.5%

Note. Standard achieved = performance 75% or higher. Standard not achieved = performance lower than 75%.

*Other task-selection decisions concern ineffective decisions, such as selecting a task at a higher complexity level when the standard was not achieved or selecting a task at a similar or lower complexity level when the standard was achieved.

5.3.3. Comparison of conditions on cognitive and motivational aspects of learning?

Then, we analyzed whether conditions differed on cognitive and motivational aspects of learning (see Table 5.6). Note that we explored whether the results would change if we exclude those participants in the ISRL-Condition who did not watch the entire video detailing the instructional design guidelines. It was decided to keep these students in the sample, because removing them would not change the findings.

5.3.3.1 Cognitive aspects of learning

Performance on test tasks. Analyses revealed that conceptual knowledge increased from pretest to posttest ($Z = 5.86, p < .001, r = .478$). Post-hoc analyses showed significant increases in the FS-Condition ($Z = 3.94, p < .001, r = .547$) and ISRL-Condition ($Z = 1.97, p = .001, r = .482$), but not in the SRL-Condition ($Z = 5.86, p = .049, r = .348$). There was, however, no effect of Instruction Condition on students’ performance on the conceptual knowledge posttest, ($H(2) = 0.21, p = .900$), the isomorphic posttest tasks ($H(2) = 1.34, p = .511$), or the procedural transfer task ($H(2) = 0.97, p = .616$).

Mental effort. There was no significant effect of Instruction Condition on (average) self-reported mental effort invested in the learning tasks, ($H(2) = 3.16, p = .206$), nor on the average invested mental effort in the conceptual knowledge posttest questions ($H(2) = 3.23, p = .199$), isomorphic posttest tasks ($H(2) = 5.87, p = .053$), or procedural transfer task ($H(2) = 0.50, p = .780$).

Time-on-task. There was also no significant effect of Instruction Condition on average time-on-task invested in the learning tasks, ($H(2) = 4.22, p = .121$), conceptual knowledge questions ($H(2) = 0.08, p = .961$), or procedural transfer task ($H(2) = 5.92, p = .052$). Conditions differed in the average time-on-task invested in the isomorphic posttest tasks, $H(2) = 7.86, p = .020$. Follow-up analyses revealed that the ISRL-Condition invested less time in the isomorphic posttest tasks compared to the FS-Condition ($U = 1247.5, p = .011, r = .234$). No differences were found between the ISRL-Condition and SRL-Condition ($U = 1326, p = .041, r = .207$), nor between the SRL-Condition and FS-Condition ($U = 812.5, p = .857, r = .020$).

5.3.3.2 Motivational aspects of learning

Self-efficacy. We found a main effect of Test Moment on students' self-efficacy from before to after the learning phase (at a sample level), ($Z = 10.45, p < .001, r = .853$), indicating that the self-efficacy medians significantly increased over time in the ISRL-Condition ($Z = 7.09, p < .001, r = .873$), SRL-Condition ($Z = 4.73, p < .001, r = .836$), and FS-Condition ($Z = 6.09, p < .001, r = .845$). Self-efficacy after the learning phase did not differ among conditions ($H(2) = 4.67, p = .097$). We did, however, find a main effect of Instruction Condition on average self-efficacy ratings *during* the learning phase ($H(2) = 7.86, p = .020$). Post-hoc analyses revealed that average self-efficacy ratings were higher in the ISRL-Condition than the SRL-Condition ($U = 702.5, p = .007, r = .271$). No significant differences were found between the ISRL-Condition and FS-Condition ($U = 1881.5, p = .369, r = .083$), nor between the SRL-Condition and FS-Condition ($U = 680.5, p = .162, r = .129$).

Perceived competence. Analyzing whether perceived competence increased from pretest to posttest revealed a main effect of Test Moment ($Z = 10.54, p < .001, r = .861$). Perceived competence significantly increased in the ISRL ($Z = 7.07, p < .001, r = .870$), SRL ($Z = 4.94, p < .001, r = .874$), and FS-Condition ($Z = 6.11, p < .001, r = .847$). No differences were found among conditions, however, with regard to perceived competence rated after the learning phase, $H(2) = 5.34, p = .069$.

Table 5.6.
Mean (M), Standard Deviation (SD), and Median (Med) of Conceptual Questions (range 0 to 9), Isomorphic Tasks (range 0 to 24), Procedural Transfer Task (range 0 to 8), Mental Effort (range 1 to 9), Self-Efficacy (range 1 to 9), and Perceived Competence (range 1 to 7) for the Informed Self-Regulated Learning Condition (n = 66), Self-regulated Learning Condition (n = 32), and Fixed Sequences Condition (n = 52).

	Informed Self-Regulated Learning Condition			Self-Regulated Learning Condition			Fixed Sequences Condition		
	M	SD	Med	M	SD	Med	M	SD	Med
Pretest									
Conceptual Questions	2.50	1.13	3.00	2.66	1.23	3.00	2.25	1.25	2.00
Self-Efficacy	3.09	1.65	3.00	2.91	1.97	2.00	3.04	1.86	3.00
Perceived Competence	2.62	1.50	2.33	2.49	1.40	2.17	2.65	1.47	2.17
Learning Phase									
Self-Efficacy	6.71	1.00	6.83	6.14	1.16	6.08	6.42	1.45	6.67
Mental Effort	3.20	0.99	3.33	3.71	1.04	3.50	3.45	1.50	3.33
Time-on-Task	8.28	2.43	8.50	7.38	2.32	8.00	7.50	2.17	7.25
Posttest									
Conceptual Questions	3.73	2.23	3.00	3.63	2.49	3.50	3.77	2.21	4.00
Isomorphic Tasks	14.62	7.84	16.50	13.41	7.05	14.50	13.38	7.67	15.50
Procedural Transfer Task	3.29	3.48	2.00	2.41	3.12	0.00	2.77	2.95	2.00
Mental Effort	3.70	1.62	3.00	4.09	1.73	4.00	3.46	1.78	3.00
Conceptual Questions	3.66	1.87	3.33	4.60	2.01	4.33	3.66	2.03	3.00
Mental effort	4.59	2.78	3.00	4.81	2.82	4.00	4.35	2.52	3.00
Isomorphic Tasks	4.92	2.15	5.00	4.88	2.88	5.00	5.15	2.99	5.00
Procedural Transfer Task	4.92	2.15	5.00	4.88	2.88	5.00	5.15	2.99	5.00
Time-on-Task	8.82	3.29	8.33	10.17	3.15	10.00	10.51	3.61	9.67
Conceptual Questions	4.64	2.99	5.00	5.59	4.11	6.00	6.56	4.23	6.00
Time-on-Task	4.64	2.99	5.00	5.59	4.11	6.00	6.56	4.23	6.00
Isomorphic Tasks	7.27	1.22	7.00	6.72	1.11	6.50	7.06	1.65	7.00
Procedural Transfer Task	5.86	0.89	6.00	5.52	0.77	5.33	5.62	1.23	6.00
Self-Efficacy									
Perceived Competence									

5.4. Discussion

This study investigated higher education students' self-regulated learning of problem-solving tasks in an online learning environment. We investigated whether: 1) the findings of Van Harsel et al. (submitted) that students regulate their learning from examples and practice problems quite well (i.e., in alignment with what we know to be effective task sequences from instructional design research) would replicate (Research Question 1), 2) self-regulated learning of examples and problems would be as effective as fixed sequences of examples and problems (Research Question 2), and 3) informing learners prior to self-regulated learning about the principles for effective task sequences derived from instructional design research would enhance their task selections, and thereby learning and motivation compared to self-regulated learning without such information and studying fixed sequences of tasks (Research Question 3).

Regarding the first research question, our results replicated the findings of Van Harsel et al. (submitted), as task selections of students in the self-regulated learning condition (and informed self-regulated learning condition) largely aligned with the instructional design principles. Almost all students followed the example-first-principle and lowest-level-first-principle by starting the learning phase with an example (predominantly a video modeling example) at the lowest complexity level. Also, the majority of students started each new complexity level with example study (and therefore selected more examples than problems) and built up the level of complexity of the learning tasks reasonably well, adhering to the start-each-level-with-example-study, example-based-learning, and simple-to-complex-principle. When exploring in more detail what task selections learners make after having solved a practice problem, we found that students made more effective (e.g., selecting a task at a higher complexity level when sufficiently high performance was achieved) than ineffective task selections (e.g., selecting a task at a higher complexity level when sufficiently high performance was not yet achieved), however, there seemed to be room for improvement.

As for the second research question, we found no performance or motivation differences between the self-regulated learning and fixed sequences condition. This is somewhat surprising in light of previous findings that fixed learning paths are often more effective for novices' learning than self-regulated learning (see e.g., Azevedo et al., 2008; Lawless & Brown, 1997; Niemiec et al., 1996). That self-regulated learning did not have additional motivational benefits might also seem surprising, as previous research suggests that providing learners with control over task selection can increase their motivation in terms of interest and involvement (e.g., Corbalan et al., 2008). However, self-regulated learning might not foster students' motivation in terms of perceptions of their own abilities (e.g., self-efficacy and perceived competence), possibly because this is much more related to learning outcomes (e.g., Collins, 1982), where we also found no effect.

With regard to the third research question, no significant differences were found between the two self-regulated learning conditions in how their task selections matched with the instructional design principles. The only exception was that students in the informed self-regulated learning condition followed the simple-to-complex-principle more often than students in the self-regulated learning condition did spontaneously. However, this facilitative effect did not enhance the informed self-regulated learning condition's learning or motivation. There were two exceptions: relative to the self-regulated learning condition, the informed self-regulated learning condition showed more confidence in their own abilities during the learning phase (but this effect was not found after the learning phase) and invested less time in the posttest tasks isomorphic to the tasks in the learning phase. Finally, we did not find any differences on cognitive and motivational aspects of learning between the informed self-regulated learning condition and fixed sequences condition.

5.4.1. Students' Task-Selection Patterns

These findings raise an important question: Why did we (and Van Harsel et al., submitted) find that students were already quite good at regulating their learning of examples and problems, while other studies found that having control over what information to study or what tasks to work on is not (entirely) effective for novices' self-regulated learning (e.g., Foster et al., 2018), and often less effective than learning from computer pre-structured or personalized sequences of tasks (e.g., Azevedo et al., 2008; Lawless & Brown, 1997; Niemiec et al., 1996)? A possible explanation is that our sample may have had substantial prior experience with learning from examples (cf. Van Harsel et al., submitted). Although we cannot corroborate this idea with data, our students were likely quite experienced with example-based learning, because their electrical and electronic mechanical engineering programs rely heavily on mathematics. Example-based learning indeed is a very common strategy for learning mathematical problem-solving skills (Hoogerheide & Roelle, 2020). If students were accustomed to studying examples when learning new math problem-solving skills, this would explain why examples were selected earlier and more often. By contrast, Foster et al. (2018) tested a mixed student population (from the university's participant pool) that possibly had less experience with mathematics in their curricula and therefore with example-based learning. If true, this could explain why students in that study chose example study less early and often.

5.4.2. Explicitly Informing Students about Effective Sequences

Another important question is: Why did our intervention only show a minimal effect on students' task selections and why did it not enhance their learning and motivation? One possible explanation could be that approximately 40% of the students in the informed self-regulation condition did not watch the entire instructional video. However,

additional exploratory analyses revealed no significant performance or motivational differences among conditions when those who did not watch the entire video were excluded (see Supplementary Materials O). A more likely explanation is that there was not that much room for students' task-selection skills to improve (with the exception of the simple-to-complex-principle, which those who studied the video did follow more often). Indeed, the self-regulated learning condition already did quite a good job at regulating their learning.

A second potential explanation, given that there was still room for some improvement, is that informing students about the principles only once may not have been sufficient to improve task-selection behavior to such an extent that it enhances students' learning and motivation. For example, students received a lot of information they had to both understand and memorize in order to apply it during the learning phase later. Although we provided a short review/reminder at the end of the instructional video, it is possible they forgot some of the principles and/or how/when to use them during the learning phase. A solution might be to allow students to go back to the description and explanation of the principles during the learning phase.

It is also questionable whether 'merely' informing students about instructional design principles and how to apply them would improve task-selection behavior to such an extent that it enhances students' learning and motivation. Although providing knowledge is considered an important component in theories on behavioral change (e.g., Theory of Planned Behavior; Fishbein & Ajzen, 2011), and although it had a small effect on behavior in our study (i.e., the simple-to-complex-principle was followed more often by those who studied the video), it is questionable whether it is sufficient for large scale behavioral change. To achieve those changes, it is for example also important to allow people to experience what the planned behavior actually brings them (i.e., to enhance their beliefs and commitment; McDaniel & Einstein, 2020), which might be achieved by additionally having students practice with and/or reflect on the information that is provided to them (e.g., Biwer et al., 2019; Endres et al., 2021). Moreover, as still 40% of the task selections made after solving a problem were likely not effective for learning, it might be necessary to link the principles more strongly to students' (self-assessments of) their understanding and performance (e.g., emphasize what to do when they do not yet master a task at a certain complexity level) or to target their self-assessment ability (cf. Kostons et al., 2012; Raaijmakers et al., 2018).

5.4.3. Limitations

This study does have several important limitations. First, it is an open question to which extent our findings are generalizable, because our sample –despite being novices– might have had prior experience with example-based learning and/or similar types of (math) tasks. Less experienced samples would likely show different (i.e., worse)

task-selection behavior. Therefore, it is unclear whether our instructional video intervention would have a more pronounced effect (i.e., improve task-selection, and thereby motivation and learning) under different circumstances, such as with less experienced samples. A particularly interesting avenue for future research would be to test this intervention with a sample that has previously been shown to show suboptimal task-selection behavior when learning from examples and problems (e.g., Foster et al., 2018).

A second limitation is that, although our students had more responsibility and control over their learning relative to most example-based learning research, they still did not have full control over their learning. For instance, students had to select six learning tasks, yet in real learning settings there would likely be much more variation in the number of tasks selected, due to differences in motivation (e.g., motivated students would be willing to work on more tasks than unmotivated students) and abilities (e.g., faster learners would not need as many tasks relative to slower learners). This is an interesting avenue for future research.

5.4.4. Conclusion and Implications for Practice

To conclude, the findings suggest that the higher education students who participated in this study were relatively good at regulating their own learning with examples and problems in online learning environments (cf. Van Harsel et al., submitted). This is an important finding because providing students with control over their own learning is becoming more and more common, especially in higher education. Given that earlier studies painted a less rosy picture of students' self-regulated learning of problem-solving skills using examples and problems (e.g., Foster et al., 2018), and that the sample used in this study might have had some prior experience with similar mathematics problem-solving tasks, future research is needed to uncover under which circumstances students can and cannot regulate their learning of new problem-solving skills. Our findings also suggest that there is still room for some improvement in students' task selections. Informing students about evidence-based instructional design principles via an instructional video can help them to apply the simple-to-complex-principle more often, however, not to such an extent that it results in performance or motivational benefits. Therefore, future research should examine how the design and/or implementation of this intervention can be improved to (further) improve self-regulated learning of our population and other populations of higher education students.

Chapter 6

Summary and discussion



A vast body of instructional design research has shown that example study is a very effective and efficient instructional strategy for acquiring new problem-solving skills (for a review, see Van Gog et al., 2019). During the past years, example-based learning has become increasingly popular in formal and informal educational settings, as it has become much easier to create and share examples thanks to modern technological advances (e.g., Hoogerheide & Roelle, 2020). However, the technological possibilities are far ahead of our understanding of what the optimal sequences of (different types of) examples and practice problems would be to foster students' learning and motivation. Therefore, the first aim of this dissertation was to examine *what sequences of examples and practice problems are most effective, efficient, and motivating for first year higher education students' learning of new mathematical problem-solving skills* (**Chapter 2** and **3**).

Furthermore, with the emerging popularity of educational concepts such as flipping the classroom, blended learning, and massive online open courses, students nowadays acquire new knowledge and skills increasingly via online learning environments (in blended or fully online courses), in which worked examples, video modeling examples, and practice problems are often embedded (e.g., Roll et al., 2011). These online environments usually require students to self-regulate their learning (e.g., when doing homework or studying for a test) and determine themselves where (i.e., at school or at home), when, and how they want to study. However, relatively little is known about how (well) learners regulate their own learning from examples and practice problems and whether and how they need support for that. Therefore, the second aim of this dissertation was to *examine how and how well first year higher education students regulate their learning from examples and practice problems in an online learning environment and whether informing them about effective, efficient, and motivating instructional design principles helps to improve their task-selections, and thereby their motivation and learning outcomes* (**Chapter 4** and **5**).

This chapter summarizes and discusses the findings of the studies in each part of the dissertation, along with implications and suggestions for future research.

6.1. Part 1: Sequencing Example Study and Practice Problem Solving

6.1.1. Summary of findings

Chapter 2 reported two experiments that investigated whether different short sequences of examples and practice problems (i.e., 4 learning tasks) differ in effectiveness, efficiency, and how they affect motivational aspects of learning. In Experiment 1 ($N = 124$), technical higher education students learned how to approximate the region under a graph using the trapezoidal rule (a math task) by

means of example study only, example-problem pairs, problem-example pairs, or problem solving only. Experiment 2 ($N = 81$) used the same materials and design as Experiment 1, but with a sample of primary teacher training students in order to examine whether results would replicate with a different sample (i.e., students with a non-technical background). Effectiveness was measured by assessing performance on the isomorphic test tasks, procedural transfer task, and conceptual transfer task. Efficiency was measured by logging time-on-task and rating invested mental effort after each task in the learning and posttest phase. Motivation was measured by means of short self-efficacy, perceived competence, and topic interest questionnaires, provided to learners before and after the learning phase.

Results of Experiment 1 showed that students in all three example conditions attained equal or higher performance on the isomorphic posttest tasks while investing less effort in the learning phase. Students in all three example conditions also reported higher levels of self-efficacy and perceived competence after the learning phase relative to students who engaged in practice problem solving only. Also, example study only resulted in higher performance on the isomorphic posttest tasks with less time and effort investment in the learning phase compared to the other example conditions, and higher levels of self-efficacy than the condition with example-problem pairs. This pattern was somewhat different in Experiment 2, however. In Experiment 2, students in all three example conditions attained equal or higher performance on the isomorphic posttest tasks while investing less effort in the learning phase. However, only students in the examples only condition reported higher self-efficacy and perceived competence than students in both the problem-example pairs and problem-solving only condition. Finally, there were no differences between example-problem and problem-example pairs on any of the outcome variables, except for a time-on-task advantage for the problem-example pairs condition in Experiment 1. No differences among conditions were found on the transfer measures or on topic interest in either experiment.

Chapter 3 reported two experiments that examined whether different short (i.e., 4 tasks) and longer (i.e., 8 tasks) sequences of example study only, example-problem pairs, problem-example pairs, and practice problem-solving practice show differences in effectiveness, efficiency, and motivation. Experiment 1 aimed to investigate whether the results of Experiment 1 described in **Chapter 2** would replicate with a conceptual pretest (instead of a procedural pretest) and would remain stable on a delayed test one week later. Experiment 2 examined whether the effects found with short sequences would generalize to longer sequences. In both experiments, it was investigated how self-efficacy develops during the learning phase. Therefore, technical higher education students learned a mathematical procedure with the help of four (Experiment 1; $N = 157$) or eight learning tasks (Experiment 2; $N = 105$) sequenced as (video modeling) examples only, example-problem pairs, problem-example pairs, or problems only. The outcome measures were identical to the ones described in **Chapter 2**, except that self-efficacy was also measured after each task in the learning phase.

In both experiments, self-efficacy was higher after the first learning task when participants started with an example than when they started with a problem. However, after the second learning task, students in all example conditions (also problem-example pairs) reported higher self-efficacy than students in the problem-solving only condition, and this pattern remained stable during and after the learning phase (cf. perceived competence). In Experiment 1, students in all three example sequences also attained higher performance on the isomorphic and procedural transfer tasks than students in the problem-solving only condition, while investing less effort in the learning phase. In Experiment 2, students in the example study only condition (but not in the other two example conditions) attained higher isomorphic posttest performance with lower effort investment in the learning phase than students who engaged in problem-solving only. No differences were found among conditions on conceptual transfer or on topic interest in either experiment. Finally, the amount of mental effort and time students reported to have invested in the learning tasks was lower in the example only than the other example conditions in Experiment 1 and 2, and mental effort was lower in the example-problem pairs than in the problem-example pairs condition in Experiment 1.

6.1.2. Implications of results

The experiments reported in **Chapter 2** and **3** made several important contributions to the literature on example-based learning. Firstly, all four experiments replicated prior findings that example study – possibly alternated with problem-solving practice – is very effective and efficient for novices' acquisition of new problem-solving skills (cf. Van Gog et al., 2019). Secondly, the findings extend existing example-based learning research by showing that short sequences of example study (interspersed with problem-solving practice) are also highly beneficial for students' expectancies of their own abilities (i.e., self-efficacy and perceived competence), but not for topic interest. Moreover, it was found that sequences containing example study remained effective, efficient, and motivating on a delayed test administered one week later.

6.1.2.1. Motivational effects of example-based learning

That all three example sequences were more conducive to self-efficacy and perceived competence than problem solving only is in line with Bandura's (1977, 1986) social learning theory, and adds to the sequencing literature. Not much was known about the motivational effects of different example and problem sequences, because sequencing research has mostly been conducted against the backdrop of cognitive theories (e.g., cognitive load theory) that tend to ignore student motivation (Sweller et al., 2011; Van Gog & Rummel, 2010). Self-efficacy and perceived competence are important to consider in sequencing research: Self-efficacy has been shown to enhance factors such as academic motivation, study behavior, and learning outcomes (e.g., Bandura, 1997; Bong & Skaalvik, 2003; Schunk, 2001), and perceived competence

has also been shown to have significant influence on academic motivation and learning outcomes (e.g., Bong & Skaalvik, 2003). Moreover, both self-efficacy and perceived competence positively affect the willingness to invest effort and task persistence (e.g., Pintrich, 2003; Schunk, 1995).

That sequences with examples were not more beneficial for enhancing students' interest in the topic than problem solving only is at first glance a surprising finding, because the example conditions learned more, and research has shown that an increase in knowledge can lead to an increase in interest (e.g., Schmidt & Rotgans, 2017). That being said, topic interest was relatively high in our samples and might not have been affected by the short (i.e., single session) interventions in the studies presented in this dissertation.

In sum, sequences with example study increased students' beliefs in their own abilities and did not affect topic interest relative to only practice problem-solving. These findings indicate that, from a motivational perspective, all three sequences containing example study are "safe" to use in educational practice when novices learn new problem-solving skills. A caveat to this interpretation is that motivation is an incredibly broad concept, and therefore future sequencing research is recommended to examine effects on other aspects of student motivation. A particularly interesting avenue for future research might be to measure effects of different sequences on perceptions of autonomy, which refers to the basic need to have choices and be free from control. For example, it is possible that example study (only) would be too prescriptive and thereby impair feelings of autonomy relative to sequences that include practice problem solving.

6.1.2.2. Starting with example study versus problem-solving practice?

Another important finding reported in **Chapter 2** and **3** is that starting with a problem worked better for learning and motivation than expected based on prior research, provided that problem-solving was followed by an example to study (i.e., problem-example pairs). That is, based on prior research (e.g., Kant et al., 2017; Leppink et al., 2014; Van Gog et al., 2011), it was expected that the problem-example condition would perform as poorly as the problem-solving only condition and therefore be less effective and efficient than example-problem pairs. However, none of the experiments using short sequences revealed performance or motivation differences between problem-example and example-problem pairs, and a problem-example pair sequence was generally more effective, efficient, and motivating than problem-solving only. We must note, though, that example-problem pairs were more efficient than problem-example pairs, as the example-problem pairs condition showed lower average scores on effort investment in the learning phase compared to the problem-example pairs condition. While conducting the four experiments, several other studies also did not find any test performance differences between example-problem and problem-example pairs (e.g., Coppens et al., 2019; Van der Meij et al., 2018).

These mixed findings raise the question of whether an example-problem sequence does or does not promote learning relative to a problem-example sequence. This question was explored in **Chapter 2** via a small-scale meta-analysis, which showed a significant, small-to-medium meta-analytic advantage of example-problem pairs over problem-example pairs on test performance. However, there was substantial variation in the results, suggesting that the benefit of an example-problem sequence over a problem-example sequence is probably not as large as previously believed, and may only occur under specific conditions. This raises the question of what factors determine when example-problem pairs are (not) advantageous over problem-example pairs.

One factor that might explain these mixed findings is student motivation. Van Gog et al. (2011) proposed that when problem-solving tasks require domain-specific knowledge and are not inherently rewarding or enjoyable (e.g., troubleshooting electrical circuits), starting the learning phase with a failed problem-solving attempt could cause students to lose confidence in their abilities or interest in the task. As a consequence, this would cause them to disengage and not be motivated to study tasks that follow. By contrast, rewarding and enjoyable tasks (e.g., a puzzle problem; Van Gog, 2011) might 'shield' students from such motivational issues, because failing at puzzle problems might challenge rather than demotivate learners to study an example that follows. As the studies reported in **Chapter 2** and **3** used problem-solving tasks that require domain-specific knowledge and were not intrinsically rewarding or enjoyable, one would expect to find performance and motivational benefits of example-problem pairs. Indeed, **Chapter 3** did reveal that starting with an example was more beneficial for self-efficacy than starting with a problem. However, this difference quickly disappeared after studying an example as a second learning task and therefore, none of the experiments revealed differential effects on overall motivation or performance. Although one could argue that the math task was relevant and potentially rewarding for technical higher education students because learning such integration methods is important in order to carry out their future profession, this argument does not apply to teacher training students for whom this is less relevant. It is important to note, though, that student motivation in terms of topic interest was fairly high in both samples. As learners cope with challenging tasks more effectively when they maintain their interest (e.g., Ainly et al., 2002) and enjoy what they are doing (e.g., Linnenbrink & Pintrich, 2004), it is possible that students who are interested in a topic would more easily persist after failing to solve a practice problem, because a failed problem-solving attempt elicits feelings of enjoyment or challenge. By contrast, starting with a failed practice problem might have a demotivating effect when students are not interested in a topic and cause students to disengage.

Another factor that might explain these mixed findings (and could interact with motivation) is the type of example format used, since studies with worked examples generally reported an advantage of example-problem pairs (e.g., Leppink et al., 2014; Van Gog et al., 2011) while most studies with video modeling examples did not find any

performance differences between example-problem and problem-example pairs (Coppens et al., 2019; Van Gog, 2011; Van Harsel et al., 2019, 2020; Van der Meij et al., 2018)⁶. It is possible that after starting with a failed practice problem, (demotivated) learners would be less willing to pay attention to worked examples than video modeling examples. Worked examples might feel more overwhelming because all the information is presented simultaneously. Video modeling examples, however, present information step-by-step and the combination of dynamic visual information and the model's narration takes the learner by the hand.

In short, the advantage of example-problem pairs over problem-example pairs is likely smaller than previously believed. Future research will have to uncover under which conditions example-problem pairs are (not) more beneficial for learning than problem-example pairs, for instance by comparing the effects of example-problem and problem-example pairs with worked vs. video modeling examples and with low vs. highly interesting material.

6.1.2.3. Effects of example study only

A final important finding is that studying examples only was particularly effective, efficient, and motivating for learning (**Chapters 2** and **3**). Example study only resulted in higher motivation and test performance than problem solving only. Moreover, the example study only condition attained equal or higher test performance scores with less effort and/or time investment in the learning phase relative to the other two example conditions. Even with longer task sequences, studying different examples remained very effective, efficient, and motivating. These findings are surprising in light of the expertise-reversal effect (e.g., Kalyuga et al., 2001), which states that learners who have (acquired) knowledge benefit more from activities with minimal instructional guidance such as problem solving than from activities with a lot of instructional guidance such as example study, and in light of Sweller & Cooper's (1985) suggestion that additional practice could be experienced as more motivating than continuing with the more passive activity of studying examples. So why did example study not lose its benefits?

A possible reason could be that the learning tasks increased in complexity, while the number of tasks per complexity level was rather limited (i.e., 2 or 4 tasks per complexity level). Consequently, students likely often did not qualify as advanced learners and may have not experienced an example only sequence as too repetitive or redundant. Although task complexity commonly also increases in expertise-reversal research, in that line of research students typically work on many more tasks that enable them to become advanced learners (e.g., Kalyuga et al., 2001).

⁶ Except for the study of Kant et al. (2017) that did find EP to be more effective than PE with video modeling examples.

With even longer task sequences and more learning tasks per complexity level, it indeed seems probable that example study would become less effective and motivating for learning, as studying examples without solving any practice problem might eventually become redundant or boring. Another contributing factor might be that a self-paced learning environment was used, in which learners could decide for themselves how long they wanted to work on tasks. Although participants were always instructed to watch the entire video modeling example, the online learning environment did allow students to skip (parts of) the example. The time-on-task data that was obtained during the learning phase showed that time spent on the examples decreased as the learning phase progressed, which in turn might have prevented learners from perceiving examples as redundant or demotivating.

In sum, results have shown that example study only is very effective, efficient, and motivating, even with longer sequences. However, given the paucity of sequencing research with longer task sequences, follow-up research should further investigate whether the findings replicate and under what circumstances example study holds or loses its effectiveness, efficiency, and motivational benefits (i.e., by increasing sequence length or comparing self-paced versus system- or experimenter-paced environments).

6.1.3 Interesting avenues for future sequencing research

In addition to the suggestions for future follow-up research raised above, there are several more general directions for future research that could help move the research on example-based learning forward.

6.1.3.1. Extending sequencing research with different formats and sequences.

Most sequencing research has focused on the same sequences (i.e., example study only, example-problem pairs, problem-example pairs, and problem solving) and the same task formats (i.e., video modeling examples, worked examples, and practice problems). Yet, these sequences and task formats might be suboptimal if the aim is to gradually reduce instructional support during learning. Therefore, it might be interesting to include other formats, such as erroneous examples or a mix of correct and erroneous examples (e.g., Große & Renkl, 2007; Kopp et al., 2008) and completion problems (e.g., Renkl & Atkinson, 2003; Van Merriënboer et al., 2002). For instance, there is already some research showing that sequences containing worked examples, completion problems, and practice problems are more effective than example-problem pairs (Atkinson et al., 2003; Renkl et al., 2002). Less is known, however, about how including these formats affect motivational aspects of learning and how sequences including erroneous examples, or a mix of correct and erroneous examples affect motivation and learning compared to the sequences that have been studied so far. Moreover, it might be more motivating and effective for novices to study a sequence that starts with several examples and then continues with example-problem pairs or completion

problems, so that students initially experience little if any extraneous load (i.e., demands on working memory that are due to poorly designed instructional procedures and do not contribute to learning; Paas et al., 2010) and can focus on learning the procedure before they start applying (parts of) it.

6.1.3.2. Extending sequencing research with more classroom studies.

Another important next step in sequencing research would be to examine the effects of different example and problem sequences in real classroom settings, such as during a course or an entire curriculum, and over a longer period of time (Hoogerheide & Roelle, 2020; Renkl, 2014; Van Gog et al., 2019). The experiments reported in **Chapter 2** and **3** of this dissertation already answered the call for more classroom studies by investigating the efficacy of different sequences in a classroom setting. However, these settings were still fairly controlled, and the experiments were conducted within a relatively short period of time (e.g., one or two sessions across a time period of approximately 2 hours). As a result, it is possible that motivation did not play a major role in determining whether students stayed engaged or disengaged while working on certain sequences. It would therefore be interesting to study the development of motivation during a multi-week course (although the use of a practice problem only control condition would raise ethical concerns in this context, given the findings presented in this dissertation).

6.2. Part 2: Self-regulating example study and practice problem solving

6.2.1. Summary of findings

Chapter 4 describes an explorative study that investigated what choices technical higher education students ($N = 147$) make and why when they learn a mathematical problem-solving skill by selecting six learning tasks themselves from a task-database with 45 tasks that varied in format (video modeling examples, worked examples, and practice problems), complexity level (level 1, 2, and 3), and cover story. Students were assumed to be novices to the modelled task. Subsequently, it was explored to what extent their task-selection decisions match with effective, efficient, and motivating principles derived from instructional design research (see Table 6). Finally, it was explored whether there is a positive relation between the extent to which students follow these instructional design principles and their scores on cognitive (isomorphic test tasks, procedural transfer task, and conceptual questions, mental effort, and time-on-task) and motivational variables (self-efficacy, perceived competence, and topic interest).

Results showed that students' choices aligned quite well with the instructional design principles, as the vast majority of students selected many examples during the learning phase and started the learning phase with an example instead of a problem.

Almost all students also started the learning phase with a task at the lowest complexity level. However, only half of the sample built up the learning tasks from simple to complex. Also, most students started each complexity level with example study. During the entire learning phase, examples were preferred over practice problems and students mostly selected items from the lowest complexity level. Finally, total scores of how well students task selections matched with all of the principles did not correlate with any of the cognitive or motivational variables.

Table 6.
Effective, efficient, and motivating principles derived from instructional design research on learning new problem-solving skills

Principle	Explanation	References
Example-based-learning-principle	Replacing all or a substantial number of practice problems with examples helps novices to learn more (i.e., is more effective) with less time and effort investment (i.e., is more efficient) than solving practice problems without any instructional support, and also more motivating.	e.g., Sweller et al. (2011), Van Gog et al. (2019) Van Harsel et al. (2019, 2020)
Example-study-first-principle	Novices should start the learning phase with an example instead of a practice problem, as this was found to be more efficient, than starting with problem-solving only, and also more motivating.	e.g., Van Gog et al. (2011) Van Harsel et al. (2019, 2020)
Lowest-level-first-principle	Novices should start with a task at the lowest level of complexity	Van Merriënboer (1997), Van Merriënboer & Kirschner (2013)
Simple-to-complex-principle	Novices should gradually increase the level of task complexity as their knowledge increases	Van Merriënboer (1997), Van Merriënboer & Kirschner (2013)
Start-each-level-with-example-principle	Novices should receive a high level of instructional support (like an example) at the start of each new complexity level	Van Merriënboer (1997), Van Merriënboer & Kirschner (2013)

The final study presented in **Chapter 5** investigated whether the results of **Chapter 4** would replicate, and additionally explored which task-selections were made after working on a practice problem. It was also investigated whether self-regulated learning would be as effective, efficient, and motivating for learning as a fixed task sequence, and whether explicitly informing learners about instructional design principles would enhance their task-selections during self-regulated learning, and thereby performance and motivation. The outcome measures were identical to **Chapter 4**, except that topic

interest was not measured. Technical higher education students ($N = 150$), who were assumed to be novices to the modelled task, learned a mathematical problem-solving procedure and were allocated to a fixed task sequences condition, a self-regulated learning condition, or an 'informed' self-regulated learning condition. In the fixed sequence condition, students received six learning tasks based on effective, efficient, and motivating instructional design principles. In both self-regulated learning conditions, students selected six learning tasks themselves from the task database (cf. **Chapter 4**). Before selecting their own learning tasks, students in the 'informed' self-regulated learning condition watched a video instruction concerning the principles.

Results were similar to those reported in **Chapter 4**, that is, students' task-selection patterns in the self-regulated learning condition aligned relatively well with the instructional design principles, except for building up task complexity from simple to complex. Students in the informed self-regulated learning condition followed the principles slightly better, as tasks were built up more often from simple to complex compared to the self-regulated learning condition. However, this did not enhance their motivation and learning outcomes. Analyses of students' task-selections made after working on a practice problem revealed that both conditions made more effective decisions (i.e., moving up a complexity level after achieving the standard, or not moving up a complexity level after failing to achieve the standard) than ineffective decisions (i.e., selecting a task at a higher complexity level when the standard was not achieved or selecting a task at a similar or lower complexity level when the standard was achieved) after working on a practice problem. Yet, there was still quite some room for improvement as more than one third of the task-selections made after solving a practice problem were qualified as ineffective for learning. Finally, no significant differences in learning outcomes or motivation were found between the self-regulated learning conditions and the fixed sequence condition.

6.2.2. Implications and future research

The studies reported in **Chapter 4** and **5** showed that students regulated their learning reasonably well, as their task selections were overall in line with what we know to be effective, efficient, and motivating sequences from instructional design research. Surprisingly, the study reported in **Chapter 5** showed that self-regulated learning of examples and problems was as effective, efficient, and motivating for learning as a fixed sequence of tasks. Although informing students about instructional design principles via a video instruction before self-regulated learning helped students to do a better job at building up the complexity of tasks (i.e., sequenced from simple to complex), this intervention had no effect on how well the other principles were followed and did not enhance learning or motivational outcomes. Moreover, no differences were found between the informed self-regulated learning condition and the fixed sequence condition on any of the outcome measures.

6.2.2.1. Students' self-regulated learning of examples and problems

That most students' task selections spontaneously matched with most of the evidence-based instructional design principles was surprising, given that Foster and colleagues (2018) found that novices underutilize example study with respect to the amount (i.e., about 40 percent worked examples versus 60 percent practice problems) and timing (i.e., students rarely started the learning phase with example study) of their use. Yet this might explain why self-regulated learning was found to be as effective and motivating as a fixed task sequence, which was a very interesting finding given that prior research has shown that providing novices with control over what information to study or what tasks to work on hampers learning relative to computer pre-structured or personalized task sequences (e.g., Azevedo et al., 2008; Lawless & Brown, 1997; Niemiec et al., 1996). Two possible (not mutually exclusive) explanations are discussed below for why students regulated their learning from examples and practice problems quite well in my studies (and therefore performed as well as students in the fixed sequence condition).

The first possible explanation is that the task database in the learning environment contained both video modeling examples and worked examples at each complexity level and therefore presented more examples than problems (67% vs. 33%, respectively), increasing the chance of an example being selected instead of a problem. This could also explain why the students of Foster et al. (2018) relied more heavily on problem solving, as students could either select worked examples and practice problems (i.e., 50% examples) or worked examples, completion problems, and practice problems (i.e., 33% examples).

A second potential explanation is that the sample used in this dissertation consisted only of technical higher education students instead of a mixed group of students obtained from a research university's participant pool (as in the study of Foster et al., 2018). In the study programs of the student samples in this dissertation, mathematics is an important subject and as a result, students might have already had experience with similar mathematical problem-solving procedures (i.e., using formulas or related integration methods). This prior experience might have helped students decide how much support they needed or what complexity level they should work on. As research has suggested, prior knowledge guides information selection, in a sense that individuals with extensive prior knowledge are better able to identify their knowledge needs and make their task-selections accordingly (e.g., Corbalan et al., 2006; Gall & Hannafin, 1994). Moreover, example-based learning is a very common strategy for learning mathematical problem-solving skills (e.g., Hoogerheide & Roelle, 2020) and therefore students might have been familiar with learning new mathematical problem-solving skills with the help of examples.

In sum, these findings suggest that higher education students can be quite capable of self-regulating their learning from examples and problems in an online

learning environment, and that when they do, their learning outcomes do not differ significantly compared to those students who work on a fixed task sequence. However, future research is needed to uncover to what extent novices' ability to regulate their learning from examples and problems is moderated by task database aspects (such as the ratio and type of examples) and sample characteristics (such as their experience with example-based learning and similar types of tasks).

6.2.2.2. Supporting students' self-regulated learning of examples and problems.

Another key finding reported in **Chapter 5** is that explicitly informing students about effective, efficient, and motivating instructional design principles helped students to do a better job at gradually building up the complexity of the tasks. However, this did not enhance students' performance or motivation. It is an open question why informing students about instructional design principles through an instructional video was not that effective.

One explanation could be that there was not that much room for students' task-selection skills to improve, as evidenced by the fact that students in the self-regulated learning condition who did not watch the video already did quite a good job regulating their learning from examples and practice problems. It must be noted, though, that the results reported in Chapter 5 showed that more than one third of the task-selections made after working on a practice problem were likely not effective for learning. These results could suggest that students experienced some difficulties with self-assessing their performance after working on a practice problem or selecting a suitable follow-up task, and that there is still some room for improvement.

Another possible explanation is that students may have had trouble remembering and therefore applying the content of the video during self-regulated learning, because the instructional video was only studied once. Allowing students to study the video several times before self-regulated learning or to revisit the video or its principles during self-regulated learning might therefore increase the effectiveness of the intervention. Moreover, to really help people change their behavior, it is considered important to ensure that they experience what the "planned behavior" actually brings them (i.e., to enhance their beliefs and commitment; McDaniel & Einstein, 2020). This could for instance be achieved by additionally having students practice with and/or reflect on the information that is provided to them (e.g., Biwer et al., 2019; Endres et al., 2021). Furthermore, as the results reported in **Chapter 5** showed that more than one third of the task-selections made after working on a practice problem were likely not effective for learning, it could also be suggested to investigate whether linking the principles more strongly to students' (self-assessments of) their understanding and performance (e.g., emphasize what to do when they do not yet master a task at a certain complexity level) or to target their self-assessment ability (cf. Kostons et al., 2012; Raaijmakers et al., 2018) enhances their task-selections, and thereby performance and motivation.

In sum, these findings revealed that informing students on how to sequences examples and problems with an instructional video helped students to make better task-selection decisions, but not to such an extent that it enhanced their learning and motivation. Therefore, research should further investigate whether adapting the (implementation of the) intervention would further enhance task-selections, performance, and motivation. Moreover, as the sample used in the studies of this dissertation probably already had some prior experience with example-based learning and with similar tasks, it would be interesting to investigate the effects of both interventions with a sample that lacks this experience (e.g., the sample of Foster et al., 2018).

6.3 Limitations

The research presented in this dissertation has some limitations. Firstly, given that all experiments relied on samples from the same student population (i.e., technical higher education students; except for Experiment 2, **Chapter 3**) and the same math problem-solving materials (i.e., learning how to use the trapezoidal rule), it is unclear to what extent the findings would generalize to other contexts. This is particularly true for novel findings, such as the effects of different sequences of examples and problems on student motivation (Part I of this dissertation) and how (well) students self-regulate their learning from examples and problems (Part II of this dissertation). All participants were likely quite motivated (as evidenced by the relatively high averages of topic interest). It is possible that students with lower interest would have been more heavily affected by (starting with) a failed problem-solving attempt and therefore would have learned less from problem-example pairs (Part I of this dissertation). Moreover, because math covers a large part of technical higher education students' curriculum, students probably had prior experience with similar types of tasks or learning from examples, which could have helped them in making relatively adequate task-selection decisions (Part II of this dissertation).

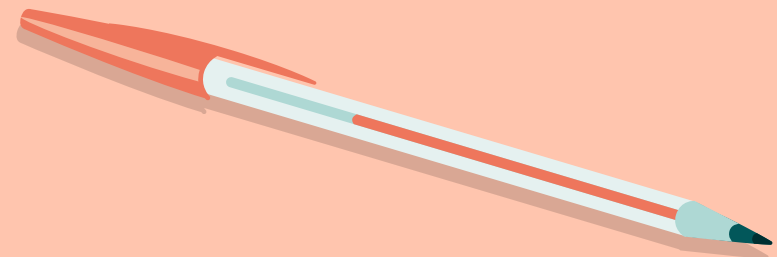
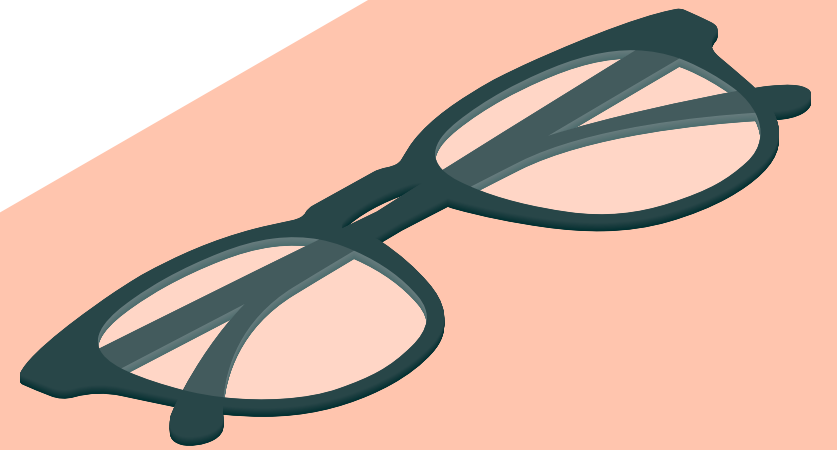
Another limitation is that students were limited in what types of tasks they could work on (**Chapters 2 and 3**, fixed sequences condition **Chapter 5**) and how many tasks they could work on (all conditions in **Chapters 2, 3, 4, and 5**). This choice was made to ensure that conditions would be comparable in all other respects. Yet in real learning settings there would likely be much more variation in the type and number of tasks selected, because an ideal task sequence hinges on students' prior knowledge, speed of learning, motivation, and effort investment and therefore varies from learner to learner. It is likely that some students provided with a task sequence worked on tasks that were not optimal given their level of expertise and motivation. Similarly, even students in the self-regulation conditions who received some control over their learning were likely not always able to make choices that truly fit their level of expertise, because they were only allowed to select six tasks while having to master three complexity levels. Therefore, future research could examine what choices students make when it is up to them how many learning tasks they select.

6.4 Conclusions and recommendations for practice

Based on the results of the first part of this dissertation and findings from previous research, it seems important to provide student with little or no prior knowledge (multiple) examples - possible alternated with practice problems - when learning new problem-solving skills. Although providing them with examples only is also effective, motivating and especially efficient for learning (according to these studies), in practice it is probably desirable to let students also practice problems themselves (as previous research has shown that this helps them to assess their own learning progress; Baars et al., 2014; 2017). The best order to provide examples and practice problems is to have students start their learning with an example prior to problem-solving practice instead of the other way around. However, this seems to apply when study time and learning tasks are limited. When sequences get longer, the motivation and efficiency benefit (in terms of effort) of starting with an example compared to starting with a practice problem seems to disappear. When task sequences are longer, it is still advisable to provide examples to students because this is more effective, efficient and motivating for learning than solving practice problems alone, as long as the complexity of the tasks builds up from simple to complex (if there is only one level of task complexity, we know from previous research that the advantage of studying examples disappears and that students benefit more from solving practice problems; Kalyuga 2001).

In view of the results of the second part of this dissertation, it seems 'safe' to allow students, perhaps after some instruction on effective instructional design principles (see Table 6), to independently start learning new problems-solving tasks using examples and problems, in the sense that this does not seem to harm their learning compared to a fixed set of learning tasks structured according to those principles. However, caution is warranted because it is unclear whether these findings generalize to other student populations, as the students participating in my studies may already have had some experience with similar tasks or learning with examples. Moreover, even after instruction on effective instructional design principles, there was still room for improvement in students' task selections, particularly when one considers what tasks students selected after working on a practice problem. Follow-up research should therefore investigate how we can further improve self-regulated learning from examples and practice problems, for example by showing learners the video instruction more often, or by training students in making better self-assessments based on which they can make appropriate follow-up tasks.

References



Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning and the psychological processes that mediate their relationship. *Journal of Educational Psychology, 94*(3), 545-561. <https://doi.org/10.1037/0022-0663.94.3.545>

Ariel, R., & Karpicke, J. D. (2017). Improving self-regulated learning with a retrieval practice intervention. *Journal of Experimental Psychology: Applied, 24*(1), 43-56. <https://doi.org/10.1037/xap0000133>

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*(2), 181-214. <https://doi.org/10.3102/00346543070002181>

Atkinson, R. K., & Renkl, A. (2007). Interactive example-based learning environments: using interactive elements to encourage effective processing of worked examples. *Educational Psychology Review, 19*(3), 375-386. <https://doi.org/10.1007/s10648-007-9055-2>

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Educational Psychology, 95*, 774-783. <https://doi.org/10.1037/0022-0663.95.4.774>

Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. G. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development, 56*, 45-72. <https://doi.org/10.1007/s11423-007-9067-0>

Baars, M., Van Gog, T., De Bruin, A., & Paas, F. (2017). Effects of problem solving after worked example study on secondary school children's monitoring accuracy. *Educational Psychology, 37*(7), 810-834. <https://doi.org/10.1080/01443410.2016.1150419>

Baars, M., Van Gog, T., De Bruin, A., & Paas, F. (2014). Effects of problem solving after worked example study on primary school children's monitoring accuracy. *Applied Cognitive Psychology, 28*(3), 382-391. <https://doi.org/10.1002/acp.3008>

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs: Prentice Hall.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215. <https://doi.org/10.1037/0033-295X.84.2.191>

Benton, S. L., Corkill, A. J., Sharp, J. M., Downey, R. G., & Khramtsova, I. (1995). Knowledge, interest, and narrative writing. *Journal of Educational Psychology, 87*(1), 66-79. <https://doi.org/10.1037/0022-0663.87.1.66>

Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. Virginia: International Society for Technology in Education.

Biwer, F., Oude Egbrink, M. G., Aalten, P., & De Bruin, A. B. (2020). Fostering effective learning strategies in higher education—A mixed-methods study. *Journal of Applied Research in Memory and Cognition, 9*(2), 186-203. <https://doi.org/10.1016/j.jarmac.2020.03.004>

Bjerrum, A. S., Hilberg, O., Van Gog, T., Charles, P., & Eika, B. (2013). Effects of modelling examples in complex procedural skills training: A randomised study. *Medical Education, 47*, 888-898. <https://doi.org/10.1111/medu.12199>

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriati (Eds.), *Attention and performance XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 435-459). Cambridge, MA: MIT Press.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (p. 185-205). Cambridge, MA: MIT Press.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417-444. <https://doi.org/10.1146/annurev-psych-113011-143823>

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: how different are they really? *Educational Psychology Review, 15*(1), 1-40. <https://doi.org/10.1023/A:1021302408382>

Carroll, W. M. (1994). Using worked out examples as an instructional support in the algebra classroom. *Journal of Educational Psychology, 86*(3), 360-367. <https://doi.org/10.1037/0022-0663.86.3.360>

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284. <https://doi.org/10.1037/1040-3590.6.4.284>

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46. <https://doi.org/10.1177/001316446002000104>

Collins, J. (1982). Self-efficacy and ability in achievement behavior. *Paper presented at the annual meeting of the American Educational Research Association*, New York.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79*(4), 347-362. <https://doi.org/10.1037/0022-0663.79.4.347>

Coppens, L. C., Hoogerheide, V., Snippe, E. M., Flunger, B., & Van Gog, T. (2019). Effects of problem-example and example-problem pairs on gifted and nongifted primary school students' learning. *Instructional Science, 47*(3), 279-297. <https://doi.org/10.1007/s11251-019-09484-3>

Corbalan, G., Kester, L., & Van Merriënboer, J. J. (2009). Dynamic task selection: Effects of feedback and learner control on efficiency and motivation. *Learning and Instruction, 19*(6), 455-465. <https://doi.org/10.1016/j.learninstruc.2008.07.002>

Corbalan, G., Kester, L., & Van Merriënboer, J. J. G. (2008). Selecting learning tasks: Effects of adaptation and shared control on learning efficiency and task involvement. *Contemporary Educational Psychology, 33*(4), 733-756. <https://doi.org/10.1016/j.cedpsych.2008.02.003>

Corbalan, G., Kester, L., & Van Merriënboer, J. J. (2006). Towards a personalized task selection model with shared instructional control. *Instructional Science, 34*(5), 399-422. <https://doi.org/10.1007/s11251-005-5774-2>

Crippen, K. J., Biesinger, K. D., Muis, K. R., & Orgill, M. (2009). The role of goal orientation and self-efficacy in learning from web-based worked examples. *Journal of Interactive Learning Research, 20*(4), 385-403.

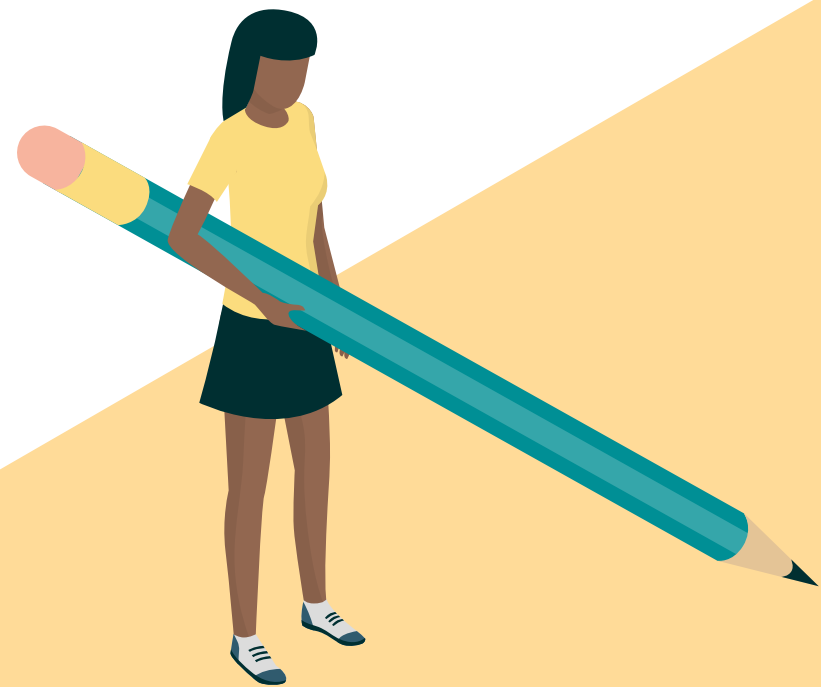
- Crippen, K. J., & Earl, B. L. (2007). The impact of web-based worked examples and self-explanation on performance, problem solving, and self-efficacy. *Computers & Education*, 49(3), 809-821. <https://doi.org/10.1016/j.compedu.2005.11.018>
- De Bruin, A. B. H., & Van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction*, 22, 245-252. <https://doi.org/10.1016/j.learninstruc.2012.01.003>
- Deci, E. L., & Ryan, R. M. (2002). *Handbook of Self-Determination Research*. New York: University of Rochester Press.
- De Koning, B. B., Hoogerheide, V., & Boucheix, J. M. (2018). Developments and trends in learning with instructional video. *Computers in Human Behavior*, 89, 395- 398. <https://doi.org/10.1016/j.chb.2018.08.055>
- Deng, R., Benckendorff, P., & Gannaway, D. (2019). Progress and new directions for teaching and learning in MOOCs. *Computers & Education*, 129, 48-60. <http://dx.doi.org/10.1016/j.compedu.2018.10.019>
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction*, 22(4), 271-280. <https://doi.org/10.1016/j.learninstruc.2011.08.003>
- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5(3), 69-106. <https://doi.org/10.1111/j.1529-1006.2004.00018.x>
- Endres, T., Leber, J., Böttger, C., Rovers, S., & Renkl, A. (2021). Improving lifelong learning by fostering students' learning strategies at university. *Psychology Learning & Teaching*, 20(1) 144-161. <https://doi.org/10.1177/1475725720952025>
- Field, A. (2009). *Discovering Statistics Using SPSS* (3rd Edition), Thousand Oaks, CA: Sage Publication.
- Fiorella, L., & Mayer, R. E. (2018). What works and doesn't work with instructional video. *Computers in Human Behavior*, 89, 465- 470. <https://doi.org/10.1016/j.chb.2018.07.015>
- Fishbein, M., & Ajzen, I. (2011). *Predicting and changing behavior: The reasoned action approach*. Psychology Press.
- Foster, N. L., Rawson, K. A., & Dunlosky, J. (2018). Self-regulated learning of principle based concepts: Do students prefer worked examples, faded examples, or problem solving? *Learning and Instruction*, 55, 124-138. <https://doi.org/10.1016/j.learninstruc.2017.10.002>
- Gall, J. E., & Hannafin, M. J. (1994). A framework for the study of hypertext. *Instructional Science*, 22(3), 207-232. <https://doi.org/10.1007/BF00892243>
- Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction*, 17, 612-634. <https://doi.org/10.1016/j.learninstruc.2007.09.008>
- Hacker, D. J., & Bol, L. (2019). Calibration and self-regulated learning. In J. Dunlosky and K. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 647-677). Cambridge University Press. <https://doi.org/10.1017/9781108235631.026>
- Hausmann, R. G. M., Van de Sande, B., & VanLehn, K. (2008). Are self-explaining and coached problem solving more effective when done by pairs of students than alone? In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2369-2374). Austin, TX: Cognitive Science Society.
- Heffer, M. H., ten Hagen, I., Krense, C., Berthold, K., & Renkl, A. (2019). Effective and efficient acquisition of argumentation knowledge by self-explaining examples: Videos, texts, or graphic novels? *Journal of Educational Psychology*, 111(8), 1396-1405. <https://doi.org/10.1037/edu0000350>
- Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research*, 60(4), 549-571. doi:10.3102%2F00346543060004549
- Hoffman, B., & Schraw, G. (2010). Conceptions of efficiency: Applications in learning and problem solving. *Educational Psychologist*, 45(1), 1-14. <https://doi.org/10.1080/00461520903213618>
- Hoogerheide, V., Loyens, S. M., & Van Gog, T. (2014). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior*, 41, 80-91. <https://doi.org/10.1016/j.chb.2014.09.013>
- Hoogerheide, V., & Roelle, J. (2020). Example-based learning: New theoretical perspectives and use-inspired advances to a contemporary instructional approach. *Applied Cognitive Psychology*. <https://doi.org/10.1002/acp.3706>
- Hoogerheide, V., Van Wermeskerken, M., Loyens, S. M. M. & Van Gog, T. (2016). Learning from video modeling examples: Content kept equal, adults are more effective than peers. *Learning and Instruction*, 44, p. 22-30. <https://doi.org/10.1016/j.learninstruc.2016.02.004>
- Hoogerheide, V., Van Wermeskerken, M., van Nassau, H., & Van Gog, T. (2018). Model-observer similarity and task-appropriateness in learning from video modeling examples: Do model and student gender affect test performance, self-efficacy, and perceived competence? *Computers in Human Behavior*, 89, 457-464. <https://doi.org/10.1016/j.chb.2017.11.012>
- Hughes, A., Galbraith, D., & White, D. (2011). Perceived competence: a common core for self-efficacy and self-concept? *Journal of Personality Assessment*, 93(3), 278-289. <https://doi.org/10.1080/00223891.2011.559390>
- Kalyuga, S., Ayres, P. L., Chandler, P. A., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23-31. doi:10.1207/S15326985EP3801_4
- Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, 93(3), 579-588. <https://doi.org/10.1037/0022-0663.93.3.579>
- Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science*, 38(3), 209-215. <https://doi.org/10.1007/s11251-009-9102-0>
- Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology*, 96(3), 558-568. <https://doi.org/10.1037/0022-0663.96.3.558>
- Kant, J., Scheiter, K., & Oschatz, K. (2017). How to sequence video modeling examples and inquiry tasks to foster scientific reasoning. *Learning and Instruction*, 52, 46-58. <https://doi.org/10.1016/j.learninstruc.2017.04.005>
- Klassen, R. M., & Usher, E. L. (2010). Self-efficacy in educational settings: Recent research and emerging directions. In T. C. Urdan, & S. A. Karabenick (Eds.), *The decade ahead: Theoretical perspectives on motivation and achievement* (pp. 1-33). Bingley, UK: Emerald Group Publishing Limited.
- Kopp, V., Stark, R., & Fischer, M. R. (2008). Fostering diagnostic knowledge through computer-supported, casebased worked examples: Effects of erroneous examples and feedback. *Medical Education*, 42, 823-829. <https://doi.org/10.1111/j.1365-2923.2008.03122.x>

- Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology*, 31, 187-194. <https://doi.org/10.1037/0278-7393.31.2.187>
- Kostons, D., Van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction*, 22, 121-132. <https://doi.org/10.1016/j.learninstruc.2011.08.004>
- Kostons, D., Van Gog, T., & Paas, F. (2010). Self-assessment and task selection in learner controlled instruction: Differences between effective and ineffective learners. *Computers & Education*, 54(4), 932-940. <https://doi.org/10.1016/j.compedu.2009.09.025>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. <https://doi.org/10.2307/2529310>
- Lawless, K. A., & Brown, S. W. (1997). Multimedia learning environments: Issues of learner control and navigation. *Instructional Science*, 25(2), 117-131. <https://doi.org/10.1023/A:1002919531780>
- Leahy, W., Hanham, J., & Sweller, J. (2015). High element interactivity information during problem solving may lead to failure to obtain the testing effect. *Educational Psychology Review*, 27(2), 291-304. <https://doi.org/10.1007/s10648-015-9296-4>
- Leppink, J., Paas, F., Van Gog, T., Van der Vleuten, C.P.M., & Van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction*, 30, 32-42. <https://doi.org/10.1016/j.learninstruc.2013.12.001>
- Lineweaver, T. T., Hall, A. C., Hilycord, D., & Vitelli, S. E. (2019). Introducing and evaluating a "study smarter, not harder" study tips presentation offered to incoming students at a four-year university. *Journal of the Scholarship of Teaching and Learning*, 19(2), 16-46. <https://doi.org/10.14434/josotl.v19i1.23734>
- Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology*, 111(2), 331-353. <https://doi.org/10.1037/edu0000281>
- Mason, L., Gava, M., & Boldrin, A. (2008). On warm conceptual change: The interplay of text, epistemological beliefs, and topic interest. *Journal of Educational Psychology*, 100(2), 291-309. <https://doi.org/10.1037/0022-0663.100.2.291>
- McAuley, A., Stewart, B., Siemens, G., & Cormier, D. (2010). *The MOOC model for digital practice*. Charlottetown: University of Prince Edward Island.
- McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition*, 39(3), 462-476. <https://doi.org/10.3758/s13421-010-0035-2>
- McDaniel, M. A., & Einstein, G. O. (2020). Training learning strategies to promote self-regulation and transfer: The knowledge, belief, commitment, and planning framework. *Perspectives on Psychological Science*, 15(6), 1363-1381. <https://doi.org/10.1177/1745691620920723>
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30-46. <https://doi.org/10.1037/1082-989X.1.1.30>
- 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: Cognitive Science Society.
- Mwangi, W., & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and Instruction*, 16(2), 173-199. https://doi.org/10.1207/s1532690xci1602_2
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G.H. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 125-173). New York, NY: Academic Press.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Niemiec, R. P., Sikorski, C., & Walberg, H. J. (1996). Learner-control effects: A review of reviews and a meta-analysis. *Journal of Educational Computing Research*, 15(2), 157-174. <https://doi.org/10.2190/JV1U-EQ5P-X2PB-PDBA>
- Nugteren, M. L., Jarodzka, H., Kester, L., & Van Merriënboer, J. J. (2018). Self-regulation of secondary school students: Self-assessments are inaccurate and insufficiently used for learning-task selection. *Instructional Science*, 46(3), 357-381. <https://doi.org/10.1007/s11251-018-9448-2>
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology*, 84(4), 429-434. <https://doi.org/10.1037/0022-0663.84.4.429>
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance Cognitive load theory. *Educational Psychologist*, 38(1), 63-71. https://doi.org/10.1207/S15326985EP3801_8
- Paas, F., Van Gog, T., & Sweller, J. (2010). Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review*, 22(2), 115-121. <https://doi.org/10.1007/s10648-010-9133-8>
- 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(1), 122-133.
- Paas, F., Tuovinen, J. E., Van Merriënboer, J. J., & Darabi, A. A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development*, 53(3), 25-34. <https://doi.org/10.1007/BF02504795>
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543-578. <https://doi.org/10.3102/00346543066004543>
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, 422. <https://doi.org/10.3389/fpsyg.2017.00422>
- Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, 95(4), 667.
- Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction*, 12(1), 61-86. [https://doi.org/10.1016/S0959-4752\(01\)00016-0](https://doi.org/10.1016/S0959-4752(01)00016-0) <https://doi.org/10.1037/0022-0663.95.4.667>
- Quilici, J. L., & Mayer, R. E. (2002). Teaching students to recognize structural similarities between statistics word problems. *Applied Cognitive Psychology*, 16(3), 325-342. <https://doi.org/10.1002/acp.796>

- Raaijmakers, S. F., Baars, M., Paas, F., Van Merriënboer, J. J. G., & Van Gog, T. (2018). Training self-assessment and task-selection skills to foster self-regulated learning: Do trained skills transfer across domains? *Applied Cognitive Psychology*, 32, 270-277. <https://doi.org/10.1002/acp.3392124>
- Rawson, K. A., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology*, 19(4-5), 559-579. <https://doi.org/10.1080/09541440701326022>
- 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(2), 92-103. <https://doi.org/10.1016/j.learninstruc.2006.02.008>
- Renkl, A. (2014). Towards an instructionally-oriented theory of example-based learning. *Cognitive Science*, 38(1), 1-37. <https://doi.org/10.1111/cogs.12086>
- Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist*, 38(1), 15-22. https://doi.org/10.1207/S15326985EP3801_3
- Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to problem solving: Smooth transitions help learning. *Journal of Experimental Education*, 70, 293-315. <https://doi.org/10.1080/00220970209599510>
- Renninger, K. A. (2000). Individual interest and its implications for understanding intrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 375-407). San Diego, CA: Academic Press.
- Roelle, J., & Berthold, K. (2013). The expertise reversal effect in prompting focused processing of instructional explanations. *Instructional Science*, 41(4), 635-656. <https://doi.org/10.1007/s11251-012-9247-0>
- Roll, I., Alevén, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction*, 21(2), 267-280. <https://doi.org/10.1016/j.learninstruc.2010.07.004>
- Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognize designers' styles. *Learning and Instruction*, 19(2), 185-199. <https://doi.org/10.1016/j.learninstruc.2008.03.006>
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432. <https://doi.org/10.1037/a0037559>
- Schiefele, U., & Krapp, A. (1996). Topic interest and free recall of expository text. *Learning and Individual Differences*, 8(2), 141-160. [https://doi.org/10.1016/S1041-6080\(96\)90030-8](https://doi.org/10.1016/S1041-6080(96)90030-8)
- Schmidt, H. G., & Rotgans, J. I. (2017). Like it or not: Individual interest is not a cause but a consequence of learning. Rejoinder to Hidi and Renninger (2017). *British Educational Research Journal*, 43(6), 1266-1268. <https://doi.org/10.1002/berj.3307>
- Schnackenberg, H. L., & Sullivan, H. J. (2000). Learner control over full and lean computer based instruction under differing ability levels. *Educational Technology Research and Development*, 48(2), 19-35. <https://doi.org/10.1007/BF02313399>
- Schnotz, W., Fries, S., & Horz, H. (2009). Motivational aspects of cognitive load theory. In M. Wosnitza, S. Karabenick, A. Efklides & P. Nenniger (Eds.), *Contemporary motivation research: From global to local perspectives* (pp. 69-96). Hogrefe & Huber.
- Schraw, G., Bruning, R., & Svoboda, C. (1995). Sources of situational interest. *Journal of Reading Behavior*, 27(1), 1-17. <https://doi.org/10.1080/10862969509547866>
- Schunk, D. H. (2001). Social cognitive theory and self-regulated learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 125-151). Mahwah, NJ: Erlbaum
- Schunk, D. H. (1995). Self-efficacy, motivation, and performance. *Journal of Applied Sport Psychology*, 7(2), 112-137. <https://doi.org/10.1080/10413209508406961>
- Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational Research*, 57(2), 149-174. <https://doi.org/10.2307/1170234>
- Schunk, D. H., & Pajares, F. (2005). Competence Perceptions and Academic Functioning. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (p. 85-104). New York: Guilford Press.
- Stark, R., Gruber, H., Renkl, A., & Mandl, H. (2000). Instruktionale Effekte einer kombinierten Lernmethode: Zählt sich die Kombination von Lösungsbeispielen und Problemlöseaufgaben aus? [Instructional effects of a combined learning method: Does the combination of worked-out examples and problem-solving tasks pay off?] *Zeitschrift für Pädagogische Psychologie*, 14(4), 205-217. <https://doi.org/10.1024//1010-0652.14.4.206>
- Stein, J., & Graham, C. R. (2020). *Essentials for blended learning: A standards-based guide*. Routledge. <https://doi.org/10.4324/9781351043991>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285. [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7)
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12(3), 185-233. https://doi.org/10.1207/s1532690xc1203_1
- Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, 2(1), 59-89. https://doi.org/10.1207/s1532690xc10201_3
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251-295. <https://doi.org/10.1023/A:1022193728205>
- Tempelaar, D. T., Rienties, B., & Nguyen, Q. (2020). Individual differences in the preference for worked examples: Lessons from an application of dispositional learning analytics. *Applied Cognitive Psychology*, 34(4), 890-905. <https://doi.org/10.1002/acp.3652>
- Tobias, S. (1996). Interest and metacognitive word knowledge. *Journal of Educational Psychology*, 87(3), 399-405. <https://doi.org/10.1037/0022-0663.87.3.399>
- Trafton, J. G., & Reiser, B. J. (1993). The contribution of studying examples and solving problems to skill acquisition. *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 1017-1022). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition*, 41, 429-442. <https://doi.org/10.3758/s13421-012-0274-5>

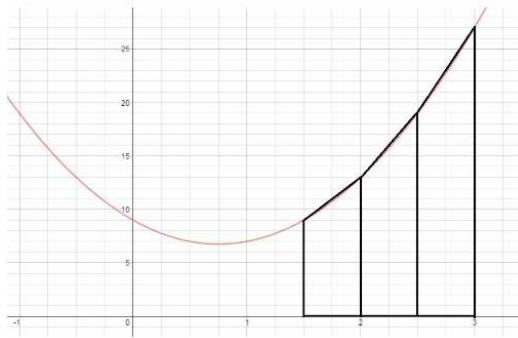
- Van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of Flipping the Classroom on Learning Outcomes and Satisfaction: a Meta-Analysis. *Educational Research Review*, 28, 1–18. <https://doi.org/10.1016/j.edurev.2019.05.003>
- Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., & Schmidt, H. G. (2002). Cognitive load theory and aging: Effects of worked examples on training efficiency. *Learning and Instruction*, 12(1), 87–105. [https://doi.org/10.1016/S0959-4752\(01\)00017-2](https://doi.org/10.1016/S0959-4752(01)00017-2)
- Van Gog, T. (2011). Effects of identical example–problem and problem–example pairs on learning. *Computers & Education*, 57(2), 1775–1779. <https://doi.org/10.1016/j.compedu.2011.03.019>
- Van Gog, T., Hoogerheide, V., & Van Harsel, M. (2020). The role of mental effort in fostering self-regulated learning with problem-solving tasks. *Educational Psychology Review*, 32, 1055–1072. <https://doi.org/10.1007/s10648-020-09544-y>
- Van Gog, T., & Kester, L. (2012). A test of the testing effect: Acquiring problem-solving skills from worked examples. *Cognitive Science*, 36(8), 1532–1541. <https://doi.org/10.1111/cogs.12002>
- Van Gog, T., Kester, L., Dirx, K., Hoogerheide, V., Boerboom, J., & Verkoeijen, P. P. J. L. (2015). Testing after worked example study does not enhance delayed problem-solving performance compared to restudy. *Educational Psychology Review*, 27(2), 265–289. <https://doi.org/10.1007/s10648-015-9297-3>
- 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(3), 212–218. <https://doi.org/10.1016/j.cedpsych.2010.10.004>
- Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, 43(1), 16–26. <https://doi.org/10.1080/00461520701756248>
- 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(2), 154–164. <https://doi.org/10.1016/j.learninstruc.2006.02.003>
- Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, 22(2), 155–174. <https://doi.org/10.1007/s10648-010-9134-7>
- Van Gog, T., Rummel, N., & Renkl, A. (2019). Learning how to solve problems by studying examples. In J. Dunlosky & K. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 183–208). New York: Cambridge University Press.
- Van Gog, T., Verveer, I., & Verveer, L. (2014). Learning from video modeling examples: Effects of seeing the human model's face. *Computers & Education*, 72, 323–327. <https://doi.org/10.1016/j.compedu.2013.12.004>
- Van Harsel, M., Hoogerheide, V., Verkoeijen, P. P. J. L., & Van Gog, T. (2020). Examples, practice problems, or both? Effects on motivation and learning in shorter and longer sequences. *Applied Cognitive Psychology*, 34(4), 793–812. <https://doi.org/10.1002/acp.3649>
- Van Harsel, M., Hoogerheide, V., Verkoeijen, P. P. J. L., & Van Gog, T. (2019). Effects of different sequences of examples and problems on motivation and learning. *Contemporary Educational Psychology*, 58, 260–275. <https://doi.org/10.1016/j.cedpsych.2019.03.005>
- Van der Meij, H., Rensink, I., & Van der Meij, J. (2018). Effects of practice with videos for software training. *Computers in Human Behavior*, 89, 439–445. <https://doi.org/10.1016/j.chb.2017.11.029>
- Van der Meij, H., & Van der Meij, J. (2013). Eight guidelines for the design of instructional videos for software training. *Technical Communication*, 60(3), 205–228.
- 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. (2013). *Ten steps to complex learning: A systematic approach to four-component instructional design* (2nd ed.). New York, NY: Taylor & Francis.
- Van Merriënboer, J. J., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist*, 38(1), 5–13. https://doi.org/10.1207/S15326985EP3801_2
- Van Merriënboer, J. J. G., Schuurman, J. G., de Croock, M. B. M., & Paas, F. (2002). Redirecting learners' attention during training: Effects on cognitive load, transfer test performance, and training efficiency. *Learning and Instruction*, 12(1), 11–37. [https://doi.org/10.1016/S0959-4752\(01\)00020-2](https://doi.org/10.1016/S0959-4752(01)00020-2)
- Van Merriënboer, J. J. G., Sluijsmans, D. M. A., Corbalan, G., Kalyuga, S., Paas, F., & Tattersall, C. (2006). Performance assessment and learning task selection in environments for complex learning. In J. Elen & R. E. Clark (Eds.), *Handling complexity in learning environments: Theory and research* (Advances in learning and instruction series) (pp. 201–220). Oxford, UK: Elsevier.
- Van Zundert, M. J., Könings, K. D., Sluijsmans, D. M. A., & Van Merriënboer, J. J. G. (2012). Teaching domain-specific skills before peer assessment skills is superior to teaching them simultaneously. *Educational Studies*, 38(5), 541–557. <https://doi.org/10.1080/03055698.2012.654920>
- Williams, G. C., & Deci, E. L. (1996). Internalization of biopsychosocial values by medical students: A test of self-determination theory. *Journal of Personality and Social Psychology*, 70(4), 767–779. <https://doi.org/10.1037/0022-3514.70.4.767>
- Williams, G. C., Freedman, Z. R., & Deci, E. L. (1998). Supporting autonomy to motivate glucose control in patients with diabetes. *Diabetes Care*, 21(10), 1644–1651. <https://doi.org/10.2337/diacare.21.10>
- Williams, G. C., Freedman, Z. R., & Deci, E. L. (1998). Supporting autonomy to motivate glucose control in patients with diabetes. *Diabetes Care*, 21(10), 1644–1651. <https://doi.org/10.2337/diacare.21.10>
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 279–306). Hillsdale, NJ: Erlbaum.
- Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General*, 145(7), 918–933. <https://doi.org/10.1037/xge0000177>
- Yan, V. X., Thai, K.-P., & Bjork, R. A. (2014). Habits and beliefs that guide self-regulated learning: Do they vary with mindset? *Journal of Applied Research in Memory and Cognition*, 3(3), 140–152. <https://doi.org/10.1016/j.jarmac.2014.04.003>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25, 3–17. https://doi.org/10.1207/s15326985ep2501_2

Supplementary materials



A: Example of a practice problem

Jalil has bought a solar cell and wants to know how much energy the solar cell supplies during a certain amount of time. Jalil has used an energy meter to examine how much energy the solar cell produces during a specific amount of time. Jalil has measured the energy at different time points and plotted the results in a graph. The time (in minutes) is plotted on the horizontal axis and the power the solar cell supplies (Joule per minute) is plotted on the vertical axis of the graph. By calculating the area under the graph, Jalil can determine how much energy the solar cell has produced during a certain amount of time.



Approaching the area under the graph can be done by using the trapezoidal rule:

$$\frac{(b-a)}{n} \left[\frac{1}{2}f(x_0) + f(x_1) + f(x_2) + \dots + \frac{1}{2}f(x_n) \right]$$

The trapezoidal rule divides the area under a graph into "strips". By adding up the surface of the "strips", you can approach the total area under the graph. To approach the area under the graph, you need the following information:

- a: this is the left x value of the area that has to be approached;
- b: this is the right x value of the area that has to be approached;
- n: this is the number of "strips" in which the area is divided;
- xi: this is the x-value that belongs to the left- or right border of a "strip" and it is calculated using the following function: $f(x) = 3x^2 - 6x + 9$

Approach the area under the graph using the information that is given. Write down all your intermediate steps and calculations.

B: Example of a video modeling example

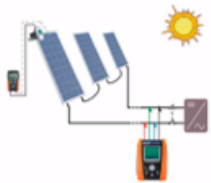
1

De trapeziumregel

Energimeting

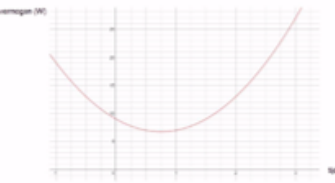
"In this video, we start working with the mathematics topic 'the trapezoidal rule'. The trapezoidal rule is a formula which you can use to approach an area under a graph. We will show you how this works with an example called 'energy measurement'."

2



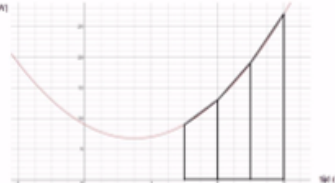
"Jalil has bought a solar cell and wants to know how much energy the solar cell supplies during a certain amount of time. Jalil uses an energy meter to examine how much energy the solar cell produced at a specific moment"

3



"Jalil has measured the energy at different time points and plotted the results in a graph. The time (in minutes) is plotted on the horizontal axis and the power the solar cell supplies (Joule per minute) is plotted on the vertical axis of the graph. By calculating the area under the graph, Jalil can determine how much energy the solar cell has produced during a certain amount of time"

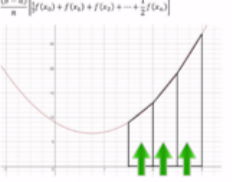
4



$$\frac{(b-a)}{n} \left[f(x_0) + f(x_1) + f(x_2) + \dots + f(x_n) \right]$$

"Approaching the surface can be done by using the trapezoidal rule. At this moment, the formula that belongs to the trapezoidal rule is shown on your screen."

5



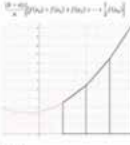
"The trapezoidal rule divides the area under a graph into 'strips'. By adding up the surface of the 'strips', you can approach the total surface under the graph. To approach the area under the graph, you need the following information:

- a: this is the left x value of the area that has to be approached; in this example 1 1/2
- b: this is the right x value of the area that has to be approached; in this example 3
- n: this is the number of 'strips' in which the area is divided; in this example 3
- xi: this is the x-value that belongs to the left- or right border of 'strip' and it is calculated using the function that is given under the image. We will now continue to approach the surface under the graph using these different 'components'.

We will now show you how to approach the area under the graph using the information that is given."

6

1



$$\frac{(b-a)}{n} \left[f(x_0) + f(x_1) + f(x_2) + \dots + f(x_n) \right]$$

$$\frac{(3 - 1.5)}{3} \left[f(1.5) + f(2) + f(2.5) \right]$$

$$\frac{1.5}{3} \left[f(1.5) + f(2) + f(2.5) \right]$$

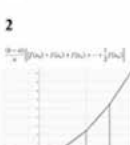
$$\frac{1}{2} \left[f(1.5) + f(2) + f(2.5) \right]$$

$$\frac{1}{2} \left[4 + 9 + 16 \right]$$

$$\frac{1}{2} \left[29 \right]$$

$$14.5$$

2



$$\frac{(b-a)}{n} \left[f(x_0) + f(x_1) + f(x_2) + \dots + f(x_n) \right]$$

$$\frac{(3 - 1.5)}{3} \left[f(1.5) + f(2) + f(2.5) \right]$$

$$\frac{1.5}{3} \left[f(1.5) + f(2) + f(2.5) \right]$$


$$\frac{1}{2} \left[f(1.5) + f(2) + f(2.5) \right]$$

$$\frac{1}{2} \left[4 + 9 + 16 \right]$$

$$\frac{1}{2} \left[29 \right]$$

$$14.5$$

3



$$\frac{(b-a)}{n} \left[f(x_0) + f(x_1) + f(x_2) + \dots + f(x_n) \right]$$

$$\frac{(3 - 1.5)}{3} \left[f(1.5) + f(2) + f(2.5) \right]$$

$$\frac{1.5}{3} \left[f(1.5) + f(2) + f(2.5) \right]$$

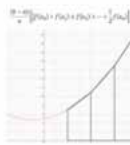
$$\frac{1}{2} \left[f(1.5) + f(2) + f(2.5) \right]$$

$$\frac{1}{2} \left[4 + 9 + 16 \right]$$

$$\frac{1}{2} \left[29 \right]$$

$$14.5$$

4



$$\frac{(b-a)}{n} \left[f(x_0) + f(x_1) + f(x_2) + \dots + f(x_n) \right]$$

$$\frac{(3 - 1.5)}{3} \left[f(1.5) + f(2) + f(2.5) \right]$$

$$\frac{1.5}{3} \left[f(1.5) + f(2) + f(2.5) \right]$$

$$\frac{1}{2} \left[f(1.5) + f(2) + f(2.5) \right]$$

$$\frac{1}{2} \left[4 + 9 + 16 \right]$$

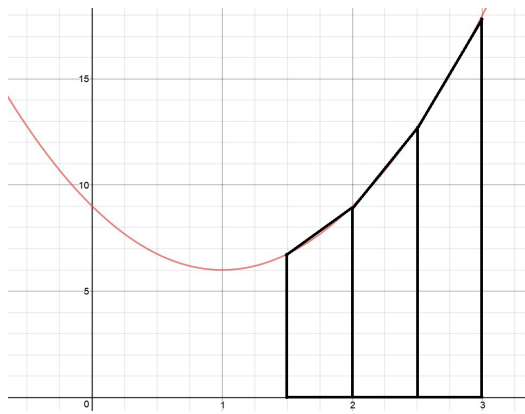
$$\frac{1}{2} \left[29 \right]$$

$$14.5$$

These screenshots show how the problem was solved step-by-step. Every step was explained verbally.

C: Example of a worked example

Jalil has bought a solar cell and wants to know how much energy the solar cell supplies during a certain amount of time. Jalil has used an energy meter to examine how much energy the solar cell produces during a specific amount of time. Jalil has measured the energy at different time points and plotted the results in a graph. The time (in minutes) is plotted on the horizontal axis and the power the solar cell supplies (Joule per minute) is plotted on the vertical axis of the graph. By calculating the area under the graph, Jalil can determine how much energy the solar cell has produced during a certain amount of time.



Approaching the area under the graph can be done by using the trapezoidal rule:

$$\frac{(b-a)}{n} \left[\frac{1}{2}f(x_0) + f(x_1) + f(x_2) + \dots + \frac{1}{2}f(x_n) \right]$$

The trapezoidal rule divides the area under a graph into "strips". By adding up the surface of the "strips", you can approach the total area under the graph. To approach the area under the graph, you need the following information:

- a: this is the left x value of the area that has to be approached, this is
- b: this is the right x value of the area that has to be approached, this is
- n: this is the number of "strips" in which the area is divided, this is
- xi: this is the x-value that belongs to the left- or right border of a "strip" and it is calculated using the following function: $f(x) = 3x^2 - 6x + 9$

Step 1: Compute the step of each subinterval $\frac{b-a}{n}$:

1. $b-a = 3 - 1\frac{1}{2} = 1\frac{1}{2}$
2. $\frac{1\frac{1}{2}}{3} = \frac{1}{2}$

Step 2: Calculate the x-values:

1. $x_0 = a$, so $x_0 = 1\frac{1}{2}$
2. $x_1 = x_0 + \text{subinterval}$, so $x_1 = 1\frac{1}{2} + \frac{1}{2} = 2$
3. $x_2 = x_1 + \text{subinterval}$, so $x_2 = 2 + \frac{1}{2} = 2\frac{1}{2}$
4. $x_3 = x_2 + \text{subinterval}$, so $x_3 = 2\frac{1}{2} + \frac{1}{2} = 3$

Step 3: Calculate the function values for all x-values

1. $f(x_0) = f(1\frac{1}{2}) = 3 \cdot 1\frac{1}{2}^2 - 6 \cdot 1\frac{1}{2} + 9 = 6\frac{3}{4} - 9 + 9 = 6\frac{3}{4}$
2. $f(x_1) = f(2) = 3 \cdot 2^2 - 6 \cdot 2 + 9 = 12 - 12 + 9 = 9$
3. $f(x_2) = f(2\frac{1}{2}) = 3 \cdot 2\frac{1}{2}^2 - 6 \cdot 2\frac{1}{2} + 9 = 18\frac{3}{4} - 15 + 9 = 12\frac{3}{4}$
4. $f(x_3) = f(3) = 3 \cdot 3^2 - 6 \cdot 3 + 9 = 27 - 18 + 9 = 18$

Step 4: Enter the function values into the formula and calculate the area

$$\text{Formula: } \frac{(b-a)}{n} \left[\frac{1}{2}f(x_0) + f(x_1) + f(x_2) + \frac{1}{2}f(x_3) \right]$$

Adjusted formula with step 1 and 2 included:

$$\frac{1}{2} \left[\frac{1}{2}f(x_0) + f(x_1) + f(x_2) + \frac{1}{2}f(x_3) \right]$$

Adjusted formula with step 3 included and calculated:

$$\frac{1}{2} \left[\frac{1}{2} \cdot 6\frac{3}{4} + 9 + 12\frac{3}{4} + \frac{1}{2} \cdot 18 \right] = \frac{1}{2} \left[3\frac{3}{8} + 9 + 12\frac{3}{4} + 9 \right] = \frac{1}{2} \cdot 34\frac{1}{8} = 17\frac{1}{16}$$

The approached area under the graph is $17\frac{1}{16}$.

D: Chapter 3 - Example of a conceptual prior knowledge test question

Question: What is the minimum required number of measurement points needed to be able to successfully apply the trapezoidal rule?

- a. 0
- b. 1
- c. 2
- d. 3

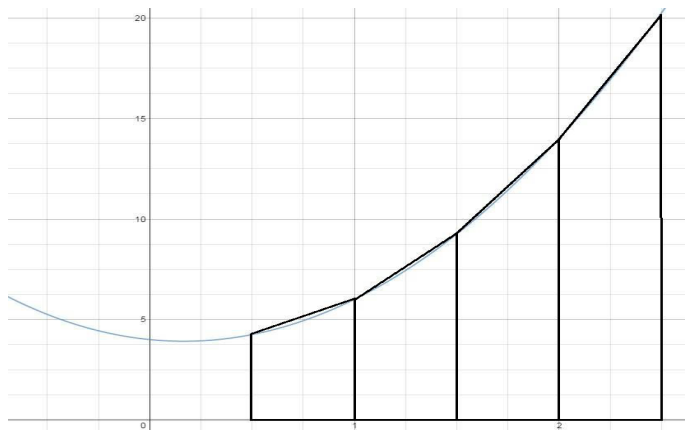
E: Chapter 4 and 5 - Example of a conceptual prior knowledge test question

Question 1: When the number of intervals increases, what can you say about the accuracy of the approximation of the area under a graph?

Answer the question and explain your answer

F: Example of an isomorphic test task

Rachel is an intern at a factory that produces different kinds of perfume. At one point, Rachel's supervisor asks her to examine how many liters of perfume is produced of the brand 'Scents' in two days. Rachel has measured this and plotted the results in a graph. The time (in days) is plotted on the horizontal axis and the liters (liter per day) are plotted on the vertical axis. By approaching the area under the graph, Rachel can determine how much liter has been produced during a certain amount of time.



Approaching the area under the graph can be done by using the trapezoidal rule:

$$\frac{(b-a)}{n} \left[\frac{1}{2}f(x_0) + f(x_1) + f(x_2) + \dots + \frac{1}{2}f(x_n) \right]$$

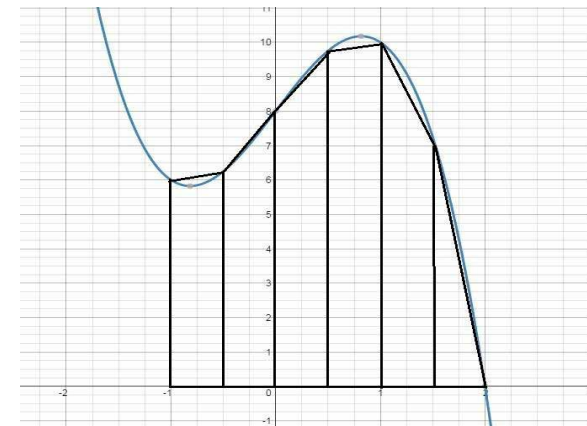
The trapezoidal rule divides the area under a graph into "strips". By adding up the surface of the "strips", you can approach the total area under the graph. To approach the area under the graph, you need the following information:

a: this is the left x value of the area that has to be approached;
 b: this is the right x value of the area that has to be approached;
 n: this is the number of "strips" in which the area is divided;
 xi: this is the x-value that belongs to the left- or right border of a "strip" and it is calculated using the following function: $f(x) = 3x^2 - x + 4$

Approach the area under the graph using the information that is given. Write down all your intermediate steps and calculations.

G: Example of a procedural transfer test task

It takes energy to stop an elevator at a certain level. This energy is proportional to the distance between the current and desired position. Jimmy wants to determine how much energy is used to stop the lift three levels higher by measuring the distance during a certain amount of time. Jimmy has plotted the results in a graph. The time (in seconds) is plotted on the horizontal axis and the distance (in meters) is plotted on the vertical axis. By approaching the area under the graph, Jimmy can determine the energy that is needed.



Approaching the area under the graph can be done by using the Simpson rule:

$$\int_a^b f(x) dx \approx \frac{b-a}{6} \left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b) \right]$$

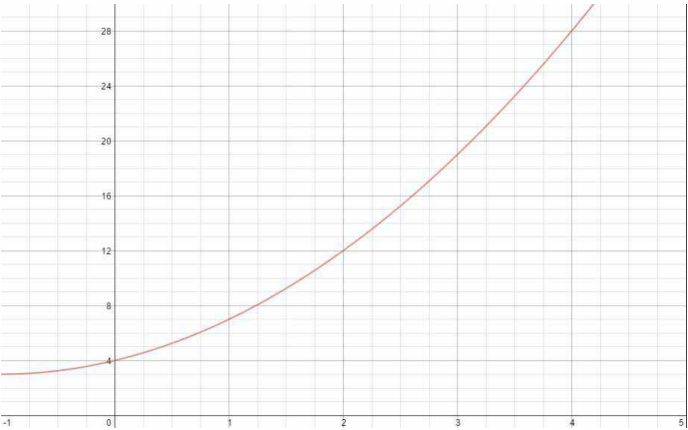
The Simpson rule divides the area under a graph into "strips". By adding up the surface of the "strips", you can approach the total area under the graph. To approach the area under the graph, you need the following information:

a: this is the left x value of the area that has to be approached;
 b: this is the right x value of the area that has to be approached;
 n: this is the number of "strips" in which the area is divided;
 xi: this is the x-value that belongs to the left- or right border of a "strip" and it is calculated using the following function: $f(x) = -2x^3 + 4x + 8$

Approach the area under the graph using the information that is given. Write down all your intermediate steps and calculations.

H: Example of a conceptual transfer test task

Study the graph below (this is a part of a parabola):



You can approach the area under this graph with help of the trapezoidal rule in two ways:

- A: Left border 2 and right border 4
- B: Left border 7 and right border 9

Which surface will approach the exact surface at best? Choose one of the options and explain your answer.

I: Topic interest scale

Items (in Dutch)	Translation (in English)
1. Ik vind de opdrachten over de trapeziumregel erg interessant	1. I think that the tasks about the trapezoidal rule are very interesting
2. Weten hoe de trapeziumregel werkt is niet belangrijk voor mij	2. Knowing how the trapezoidal rule works is not important to me
3. Het is gemakkelijk om mijn aandacht bij de opdrachten over de trapeziumregel te houden	3. It is easy to stay focused on tasks about the trapezoidal rule
4. Ik wil meer te weten komen over de trapeziumregel	4. I am keen to learn more about the trapezoidal rule
5. Ik vind de opdrachten over de trapeziumregel niet boeiend	5. I think that the tasks about the trapezoidal rule are uninteresting
6. Ik vind andere wiskunde onderwerpen relevanter dan de trapeziumregel	6. I think that other mathematics topics are more relevant than the trapezoidal rule
7. Ik vind dat tijdens de wiskundelessen aandacht besteed moet worden aan de trapeziumregel	7. I think that during math class, more attention should be paid to the trapezoidal rule

J: Chapter 2 - Means, standard deviation, and medians of delayed posttest
Mean (M), Standard Deviation (SD), and Median (Med) of Motivation, Performance, Mental Effort (range 0 to 9), and Time-on-Task per Condition of the Delayed Posttest in Experiment 1.

	EEEE Condition			EPEP Condition			PEPE Condition			PPPP Condition		
	M	SD	Med	M	SD	Med	M	SD	Med	M	SD	Med
Motivation												
Self-efficacy	5.78	1.31	6.00	5.32	2.14	5.00	5.94	1.95	6.00	4.26	2.13	5.00
Perceived Competence	4.57	1.11	4.67	4.12	1.53	4.00	4.77	1.20	4.83	3.55	1.43	3.33
Topic Interest	3.77	1.03	3.79	3.47	1.16	3.43	3.19	0.71	3.14	3.71	0.92	3.86
Performance												
Isomorphic Tasks	10.39	3.52	10.00	8.48	4.49	8.00	9.56	4.15	10.50	8.23	3.95	8.00
Procedural Transfer	2.50	2.71	1.00	2.16	3.12	1.00	2.50	3.31	1.00	2.32	1.98	1.00
Conceptual Transfer	4.72	2.19	4.50	4.32	2.53	5.00	5.69	1.92	5.50	3.47	2.15	3.00
Mental Effort												
Isomorphic Tasks	4.17	1.51	4.50	4.46	1.97	4.50	3.66	1.77	3.25	5.66	2.03	6.00
Procedural Transfer	4.22	2.10	3.00	4.96	2.13	5.00	4.50	2.37	4.00	4.84	2.41	5.00
Conceptual Transfer	3.56	1.46	3.00	4.04	2.13	3.00	3.38	1.54	3.00	4.32	1.81	5.00
Time-on-Task	12.03	3.66	11.25	9.20	3.77	9.50	10.25	3.79	9.75	8.37	3.58	8.00
Procedural Transfer	7.17	3.03	7.00	7.80	4.99	7.00	6.63	2.45	6.50	6.55	3.58	5.00
Conceptual Transfer	5.89	1.60	6.00	6.36	2.98	6.00	7.56	3.44	7.00	6.23	2.67	6.00

K: Chapter 3 - Exploratory analyses of mental effort in posttest phases and time-on-task in the training and posttest phases
Post-hoc comparisons of Mental Effort on Immediate and Delayed Posttests and Time-on-Task on Training Tasks and Immediate and Delayed Posttests in Experiment 1.

	EE vs. EP			EE vs. PE			EE vs. PP			EP vs. PE			EP vs. PP			PE vs. PP		
	U	p	r	U	p	r	U	p	r	U	p	r	U	p	r	U	p	r
Training																		
Time-on-Task	1409	<.001	.764	1209.5	<.001	.714	770	.152	.169	691.5	.066	.199	472	<.001	.403	506.5	.007	.302
Immediate posttest																		
Mental Effort	690.5	.597	.060	666.5	.942	.008	925.5	.001	.378	986	.446	.083	1289.5	<.001	.408	1116	.001	.373
Isomorphic Tasks	836	.337	.109	620	.651	.053	841	.023	.267	713.5	.095	.181	1046	.125	.167	1053	.006	.308
Procedural Transfer	345	<.001	.456	346.5	.001	.407	72.5	<.001	.761	1050	.186	.143	315.5	<.001	.550	173.5	<.001	.670
Isomorphic Tasks	650	.348	.106	548.5	.215	.145	182	<.001	.453	854	.685	.044	348	.002	.375	345	.007	.328
Procedural Transfer																		
Delayed posttest																		
Mental Effort	483	.817	.029	468	.791	.034	590	.001	.448	699.5	.406	.095	995.5	<.001	.519	805	<.001	.451
Isomorphic Tasks	496.5	.962	.006	420.5	.661	.056	415.5	.034	.284	676.5	.645	.053	863	.004	.339	793	.003	.365
Procedural Transfer	457	.561	.072	344	.119	.199	261.5	<.001	.511	594.5	.191	.150	435.5	<.001	.435	385.5	<.001	.439
Isomorphic Tasks	499.5	.995	.000	479	.669	.055	255	.028	.293	752.5	.734	.039	415.5	.017	.283	334	.005	.346
Procedural Transfer																		

Note. Acronyms for groups: EE = example study only; EP = example-problem pairs; PE = problem-example pairs; PP = problem solving only. P-values are Bonferroni corrected due to non-parametric testing. See text for details.

How do short sequences of examples and problems affect time-on-task in the training phase? There was a main effect of Instruction Condition, $H(3) = 52.45$, $p < .001$. The average time invested in the training tasks was shorter in the EEEE Condition than in the EPEP and PEPE Condition, and shorter in the PPPP Condition than in the EPEP and PEPE Condition. No other post-hoc comparisons were significant.

How do short sequences of examples and problems affect mental effort and time-on-task in the posttest phases?

Mental effort. Mental effort during the posttest phases was also explored as a measure of efficiency. There was a main effect of Instruction Condition, $H(3) = 18.11$, $p < .001$, and the average of perceived effort was lower in the EEEE, EPEP, and PEPE Condition than in the PPPP Condition. No other condition comparisons were significant. The pattern of results was similar for average mental effort invested in the isomorphic tasks on the delayed posttest. There was a main effort of Instruction Condition, $H(3) = 22.52$, $p < .001$, and mental effort ratings were again lower in the EEEE, EPEP, and PEPE Condition compared to the PPPP Condition. Again, no differences were found among other condition comparisons. Regarding mental effort invested while solving the procedural transfer task on the immediate posttest, there was a main effect of Instruction Condition, $H(3) = 9.38$, $p = .025$, and mental effort was significantly lower in the PEPE than the PPPP Condition. No other condition comparisons were significant. There was also a main effect of Instruction Condition for mental effort invested while solving the procedural transfer task on the delayed posttest, $H(3) = 11.09$, $p = .011$. Effort ratings were significantly higher in the EPEP and PEPE Condition compared to the PPPP Condition. Again, other comparisons were not significantly different.

Time-on-task. As for average time-on-task invested in the isomorphic tasks on the immediate posttest, there was a main effect of Instruction Condition, $H(3) = 64.06$, $p < .001$. Average time-on-task was longer in the EEEE, EPEP, and PEPE Condition than in the PPPP Condition. Moreover, the average time-on-task was longer in the EEEE Condition compared to the EPEP and PEPE Condition. No differences were found between the EPEP and PEPE Condition. Regarding average time-on-task invested in the isomorphic tasks on the delayed posttest, there was also a main effect of Instruction Condition, $H(3) = 26.01$, $p < .001$. Time-on-task was longer in the EEEE, EPEP, and PEPE Condition compared to the PPPP Condition. No other post-hoc comparisons were significant. There was also a main effect of Instruction Condition for average time-on-task spent on the procedural transfer task during the immediate posttest, $H(3) = 16.13$, $p = .001$. The average time-on-task was longer in the EEEE, EPEP, and PEPE Condition than in the PPPP condition. Other post-hoc comparisons were not significant. Concerning average time-on-task spent on the procedural transfer task during the immediate posttest, there was a main effect of Instruction Condition, $H(3) = 9.36$, $p = .025$. Time-on-task was only significantly longer in the PEPE Condition compared to the PPPP Condition and no other condition comparisons were significant.

L: Chapter 3 - Exploratory analyses of mental effort in posttest phase and time-on-task in the training and posttest phase
Post-hoc comparisons of Mental Effort on Immediate Posttest and Time-on-Task on Training Tasks and Immediate Posttest in Experiment 2.

	EE vs. EP			EE vs. PE			EE vs. PP			EP vs. PE			EP vs. PP			PE vs. PP		
	U	p	r	U	p	r	U	p	r	U	p	r	U	p	r	U	p	r
Training																		
Time-on-Task	856	<.001	.781	658	<.001	.668	570	<.001	.523	175	.386	.005	231.5	.135	.212	268	.733	.051
Immediate posttest																		
Mental Effort	264.5	.006	.294	192	.003	.406	431.5	.160	.191	291.5	.562	.081	449.5	.006	.392	384	.003	.444
Isomorphic Tasks	470	.739	.043	311	.316	.135	485	.017	.324	249	.154	.200	440.5	.009	.372	392	.001	.329
Mental Effort	414.5	.612	.065	359	.874	.021	422.5	.207	.172	342.5	.690	.056	383.5	.133	.213	303.5	.240	.175
Procedural Transfer	118.5	<.001	.631	77	<.001	.670	44.5	<.001	.737	304	.733	.048	202.5	.039	.292	160	.035	.315
Conceptual Transfer	283	.013	.322	265	.075	.240	205	.008	.358	362	.435	.109	292.5	.752	.045	207.5	.289	.158
Isomorphic Tasks	269	.008	.344	281	.136	.201	251.5	.075	.242	400	.138	.208	356.5	.340	.135	227	.552	.089
Time-on-Task																		
Procedural Transfer																		
Conceptual Transfer																		

Note. Acronyms for groups: EE = example study only; EP = example-problem pairs; PE = problem-example pairs; PP = problem solving only. P-values are Bonferroni corrected due to non-parametric testing. See text for details.

How do longer sequences of examples and problems affect time-on-task in the training phase? There was a main effect of Instruction Condition, $H(3) = 45.61$, $p < .001$, and time-on-task was significantly shorter in the EEEEEEEE Condition than in the EPEPEPEP, PEPEPEPE, and PPPPPPPP Condition. Moreover, post-hoc tests showed that time-on-task was longer in the EPEPEPEP Condition compared to the PEPEPEPE Condition. No other condition comparisons were significant.

How do longer sequences of examples and problems affect mental effort and time-on-task in the posttest phase?

Mental effort. Analyzing self-reported effort invested in solving the isomorphic posttest tasks revealed a significant main effect of Instruction Condition, $H(3) = 17.46$, $p = .001$. Average mental effort was significantly higher in the EEEEEEEE Condition compared to the EPEPEPEP and PEPEPEPE Condition. Moreover, effort ratings were significantly higher in the PPPPPPPP Condition compared to the EPEPEPEP and PEPEPEPE Condition. No other comparisons were significant. There was also a main effect of Instruction Condition regarding invested mental effort when solving the procedural transfer task, $H(3) = 12.23$, $p = .007$. Post-hoc tests showed that effort ratings were lower in the PEPEPEPE Condition compared to the PPPPPPPP Condition but other condition comparisons were not significant. Finally, there was no main effect of Instruction Condition for mental effort invested in the conceptual transfer questions.

Time-on-task. Exploring average time-on-task spent on the isomorphic posttest tasks revealed a main effect of Instruction Condition, $H(3) = 45.41$, $p < .001$. Participants in the EEEEEEEE Condition spent more time on the isomorphic posttest tasks than participants in the EPEPEPEP, PEPEPEPE, and PPPPPPPP Condition. No other comparisons showed significant results. As for average time spent on the procedural transfer task, we also found a main effect of Instruction Condition, $H(3) = 9.74$, $p = .021$. Average time-on-task was significantly longer in the EEEEEEEE Condition than in the PPPPPPPP Condition but no other post-hoc comparisons were significant. Finally, analysis revealed a main effect of Instruction Condition for average time spent on the conceptual transfer questions, $H(3) = 8.51$, $p = .037$. Post-hoc tests showed that time-on-task was significantly longer in the EEEEEEEE Condition compared to the EPEPEPEP Condition, however, no other post-hoc comparisons were significant.

M: Chapter 4 - Examples of scoring how well students followed the instructional design principles

Scoring protocol

Example-based-learning-principle

- 1 point for selecting 3 or more examples during the learning phase (50% or more)
- 0.5 point for selecting 2 examples during the learning phase (33%)
- 0 points for selecting less than 2 examples during the learning phase (less than 33%)

Example-study-first-principle

- 1 point for starting the learning phase with an example
- 0 points for starting the learning phase with a practice problem

Lowest-level-first-principle

- 1 point for starting the learning phase with a task at the lowest complexity level (level 1)
- 0 points for starting the learning phase with a task a higher complexity level (level 2 or 3)

Simple-to-complex-principle

- 1 point for selecting tasks at the same or a higher complexity level (i.e., never selecting a task of a lower level than already worked on)
- 0.5 points for building up the level of task complexity, but decreasing the level of complexity during the learning phase
- 0 points for students not building up the level of task complexity at all

Start-each-level-with-example-principle

- 1 point for always starting a new complexity level with an example
- 0.5 points for sometimes stating a new complexity level with an example and sometimes with a practice problem
- 0 points for always starting a new complexity level with a practice problem

Example 1

	Task selections of student X	
Learning tasks	Format	Complexity level
Task 1	Video modeling example	Level 1
Task 2	Worked example	Level 1
Task 3	Practice problem, level 1	Level 1
Task 4	Video modeling example	Level 2
Task 5	Practice problem	Level 2
Task 6	Worked example	Level 2

Principle	Score for student X
Example-based-learning-principle	1 point
Example-study-first-principle	1 point
Lowest-level-first-principle	1 point
Simple-to-complex-principle	1 point
Start-each-level-with-example-principle	1 point
Total score	5 points (out of 5)

Example 2

	Task selections of student Y	
Learning tasks	Format	Complexity level
Task 1	Practice problem	Level 2
Task 2	Practice problem	Level 2
Task 3	Worked example	Level 2
Task 4	Practice problem	Level 1
Task 5	Practice problem	Level 1
Task 6	Worked example	Level 1

Principle	Score for student Y
Example-based-learning-principle	0.5 points
Example-study-first-principle	0 points
Lowest-level-first-principle	0 points
Simple-to-complex-principle	0 points
Start-each-level-with-example-principle	0 points
Total score	0.5 points (out of 5)

N: Chapter 5 – Translated script for the video instruction (in English)

Soon, you will learn all about the mathematical subject “the trapezoidal rule” in the online learning environment. You are free to choose 6 tasks that will help you to learn the trapezoidal rule to the best of your abilities. As you know, there are tasks at 3 levels of complexity. You can choose tasks at each of these complexity levels in the form of video modeling examples, worked examples, or practice problems. Do you already know what tasks you want to select to be able to solve all the tasks on the posttest? Here are four tips, derived from scientific research, that can help you learn as much as possible.

Tip 1: First, choose a task at the lowest complexity level and build up the complexity of the tasks.

If you start learning and you don't know how to use the trapezoidal rule, it might be good start with a task that is not too difficult. Therefore, choose a task at the lowest complexity level. Do you feel you've mastered this level? Then, choose a task at a higher complexity level. This way, you build up the complexity of the tasks in such a way that it fits with what you already know.

Tip 2: Start with an example at each complexity level, especially when you feel you (still) know too little to solve the tasks.

If you don't know much about how to use the trapezoidal rule, it is not only useful to start with a task at the lowest complexity level, but also to learn more about how to solve such a task. By choosing an example, you will learn how to use the trapezoidal rule, because an example shows you how to solve a problem step-by-step. This prevents you from spending a lot of time figuring out the right solution procedure yourself. Starting with an example is therefore also very helpful when you want to choose a task at a higher complexity level.

Tip 3: Start at the very beginning with a video modeling example, then choose worked examples.

You can choose two different example formats. A video modeling example provides a lot of support during learning, because you can hear and see the solution procedure step-by-step. This is very useful if you are studying the trapezoidal rule for the first time. You can also opt for a worked example. In a worked example, you can only see the entire solution procedure. Whereas the information in a video modeling example quickly disappears, all steps are always visible in a worked example. This is very useful if you already understand part(s) of the solution procedure, but want to look up some more (difficult) steps.

Do you think you understand the solution procedure presented in the examples and want to check whether you do? Then, choose a practice problem so you can practice the task.

Tip 4: Alternate examples and practice problems

As said before, it is recommended to select an example first before solving a practice problem when you want to move to a higher complexity level for the first time. This way, you can study the steps that you might find difficult. Moreover, you get an impression of the complexity of the task. When you think you understand the problem-solving procedure, then select a practice problem so you can test whether you actually understand the problem-solving task.

Before you start, here is a short summary of the tips:

Tip 1: First, choose a task at the lowest complexity level and build up the complexity of the tasks.

Tip 2: Start with an example at each complexity level, especially when you feel you (still) know too little to solve the problem

Tip 3: Start at the very beginning with a video modeling example, then choose worked examples.

Tip 4: Alternate examples and practice problems

And now it is time to get started, good luck!

O: Chapter 5 – Explorative analyses of differences among subgroups in the ISRL-Condition

We explored to which degree participants in the ISRL-Condition followed the instructional design principles depended on whether they watched the entire, between half and three quarter, or less than half of the video. Results showed that these groups differed in following the lowest-level-principle ($H(2) = 7.55, p = .023$); however, post-hoc tests with a Bonferroni correction revealed no significant results ($ps = .019$; adjusted level of significance = .017). The groups also differed in following the simple-to-complex-principle ($H(2) = 13.45, p = .001$), and follow-up analyses showed that watching the entire ($U = 151, p = .002, r = .425$) or between half and three quarter of the video instruction ($U = 45.5, p = .014, r = .556$) resulted in higher scores on following the simple-to-complex-principle than watching less than half of the video. Finally, there was a difference among groups in the total score on following the principles ($H(2) = 7.77, p = .021$). Post-hoc tests showed that participants who watched the entire video instruction scored higher on following all of the principles than participants who watched less than half of the video instruction ($U = 152.5, p = .010, r = .357$).

We also explored whether there were performance and motivational differences among participants in the ISRL condition, depending on whether they watched the entire, between half and three quarter, or less than half of the video instruction. There were no significant differences among these groups on any of the outcome variables ($ps = 0.77$), except for participants' confidence in their own abilities before the learning phase, as indicated by a main effect on pretest self-efficacy ($H(2) = 10.46, p = .005$) and pretest perceived competence ($H(2) = 7.24, p = 0.27$). Post-hoc tests only showed one significant comparison on both motivational measures: Relative to the participants who only watched between half and three quarter of the video, those who watched the entire video indicated lower levels of pretest self-efficacy ($U = 407, p = .003, r = .413$) and perceived competence ($U = 380.5, p = .016, r = .334$).

Nederlandse samenvatting (Summary in Dutch)

Leren probleem-oplossen in het hoger onderwijs:
Het sequentiëren en zelfgestuurd leren van
voorbeelden en oefenproblemen.



Probleem-oplostaken vormen een belangrijk onderdeel van het curriculum op veel (hoge)scholen, bijvoorbeeld in vakken waarin natuurkunde, technologie, engineering en wiskunde de basis vormen. Veel van de probleem-oplostaken die studenten tegenkomen in deze vakken zijn algoritmisch, waarbij studenten moeten leren een procedure uit te voeren waarmee je van A (beschreven beginsituatie) naar B (beschreven eindsituatie) komt. Het oplossen van deze problemen vraagt om conceptuele en procedurele kennis over welke acties uit te voeren, hoe deze uit te voeren en waarom deze uit te voeren. Een effectieve manier voor novieten, dat wil zeggen lerenden met weinig tot geen voorkennis van een specifieke taak, om dergelijke kennis te verwerven is door middel van het leren van **voorbeelden** (Van Gog et al., 2019). Denk hierbij aan **tekst-gebaseerde voorbeelden** (Sweller et al., 2011), waarin stap voor stap is uitgewerkt hoe een probleem opgelost moet worden, bijvoorbeeld een uitgewerkte wiskunde opgave in een wiskundeboek. Maar denk ook aan **modelvoorbeelden** (Bandura, 1977), waarin een model stap voor stap de oplossingsprocedure demonstreert en daarbij eventueel mondelinge uitleg geeft, bijvoorbeeld een leraar die voordoeft en uitlegt hoe je een wiskunde opgave moet oplossen. Modelvoorbeelden kunnen live gegeven worden (de leraar die het oplossen van een wiskunde probleem in de les voordoeft en uitlegt), maar worden ook steeds vaker gegeven door middel van video (Van Gog et al., 2014).

Met de komst van moderne technologieën waarmee videovoorbeelden gemakkelijk gemaakt en gedeeld kunnen worden, is de populariteit van het leren van voorbeelden de afgelopen jaren alleen maar toegenomen (Hoogerheide & Roelle, 2020). Denk bijvoorbeeld aan YouTube, dat vol staat met *how-to* video's over een scala aan onderwerpen, die ook relevant zijn voor het onderwijs. Echter, de technologische mogelijkheden lopen vooruit op wat we weten over de wijze waarop voorbeelden het beste ingezet kunnen worden in het onderwijs om de prestaties en motivatie van studenten te bevorderen. Daarom was de eerste centrale vraag van dit proefschrift (**Deel 1, hoofdstuk 2 en 3**) om te onderzoeken **hoe effectief, efficiënt en motiverend verschillende kortere en langere sequenties van (video) voorbeelden en oefenproblemen zijn voor eerstejaars hbo-studenten tijdens het leren oplossen van nieuwe wiskunde problemen?**

Bovendien, met de toenemende populariteit van onderwijsconcepten zoals *flipping the classroom*, *blended learning* en het leren via *MOOC's* (*Massive Online Open Courses*) verwerven studenten tegenwoordig steeds meer kennis en vaardigheden via online leeromgevingen. In dit soort omgevingen zijn vaak ook videovoorbeelden, uitgewerkte voorbeelden en oefenproblemen ingebed (Roll et al., 2011). Deze online leeromgevingen vereisen doorgaans van studenten dat zij zelf kunnen bepalen waar (bijvoorbeeld op school of thuis), wanneer en hoe ze kennis en vaardigheden willen verwerven. Echter, er is relatief weinig bekend over hoe en hoe goed studenten hun leren reguleren met behulp van videovoorbeelden, uitgewerkte voorbeelden en oefenproblemen en of zij hierin ondersteuning nodig hebben. Daarom

was de tweede centrale vraag van dit proefschrift (**Deel 2, hoofdstuk 4 en 5**) om te onderzoeken **hoe (goed) eerstejaars hbo-studenten in het technisch onderwijs hun leren reguleren van voorbeelden en oefenproblemen in een online leeromgeving, en of hun taakselecties, leerresultaten en motivatie verbeteren wanneer we hen expliciet informeren over effectieve, efficiënte en motiverende instructieprincipes?**

Deel 1: Het sequentiëren van voorbeelden en oefenproblemen

In eerdere onderzoek is onderzocht of sommige sequenties (volgorde van - en verhouding tussen) van voorbeelden en problemen effectiever en efficiënter zijn dan andere. Effectiever betekent in dit geval dat studenten beter presteren op testtaken die vergelijkbaar zijn met wat zij hebben geoefend (isomorfe taken) en soms op testtaken die nieuw zijn maar waarvoor dezelfde procedure gevolgd moeten worden als voor de taken die geleerd zijn (transfer taken). Efficiënter betekent in dit geval dat gelijke of hogere prestaties worden behaald met minder mentale inspanning of tijdsinvestering tijdens de leerfase of tijdens het oplossen van de testtaken. Deze sequenties bestonden voornamelijk uit alleen voorbeelden, een voorbeeld gevolgd door een oefenprobleem (voorbeeld-probleem paar), een oefenprobleem gevolgd door een voorbeeld (probleem-voorbeeld paar) of alleen problemen (zie Figuur 1 voor een visuele weergave). Uit dit onderzoek is gebleken dat het bestuderen van alleen voorbeelden en voorbeeld-probleem paren even effectief en efficiënt is, maar dat beiden effectiever en efficiënter zijn dan het bestuderen van probleem-voorbeeld paren of alleen oefenproblemen oplossen (Van Gog et al., 2011).



Figuur 1: Verschillende sequenties van voorbeelden en problemen.

We weten echter nog weinig over wat de effecten van deze verschillende sequenties van voorbeelden en oefenproblemen zijn op de motivatie van studenten (Van Gog et al., 2011). Motivatie is belangrijk voor de onderwijspraktijk. Zeker in leeromgevingen of -situaties waarin studenten zelf keuzes kunnen maken is het belangrijk om rekening te houden met het effect van (sequenties van) leertaken op hun motivatie, omdat het van invloed kan zijn op de mate waarin een student begint, doorzet of stopt met leren (Pintrich, 2003). In dit deel van het proefschrift is de impact van verschillende sequenties op twee belangrijke aspecten van motivatie onderzocht, namelijk de mate van vertrouwen in eigen kunnen (Engels: *self-efficacy* en *perceived competence*) en interesse in de taak (Engels: *topic interest*).

Ook is onderzocht of deze aspecten van motivatie kunnen verklaren waarom het starten met een oefenprobleem doorgaans minder effectief en efficiënt is voor novieten dan het starten met een voorbeeld. Starten met een poging een probleem op te lossen, zonder te weten hoe dit aangepakt moet worden, zou er voor kunnen zorgen dat het vertrouwen in eigen kunnen en de taakinteresse van de student vermindert. Als gevolg daarvan bestudeert de student het opvolgende voorbeeld (en andere vervolgtaken) mogelijk minder goed, wat een negatief effect op het leerresultaat kan veroorzaken (Van Gog et al., 2011). Het zou in dat geval beter zijn te starten met een voorbeeld. Het voordeel van het starten met een voorbeeld geldt mogelijk alleen wanneer taken als oninteressant of onplezierig worden ervaren. Als taken interessant en plezierig zijn, wordt het starten met een probleem wellicht als uitdagend ervaren, waardoor studenten eerder gemotiveerd dan gedemotiveerd raken als zij (nog) niet weten hoe zij een probleem moeten aanpakken.

Ten slotte is het de vraag of resultaten op prestatie, motivatie, mentale inspanning en tijd anders zijn wanneer de verschillende sequenties langer worden en dus meer leertaken bevatten. De resultaten uit eerder onderzoek zijn namelijk vooral gevonden met korte taaksequenties van 2 of 4 leertaken. Het is mogelijk dat de resultaten anders zijn als taaksequenties langer worden. We weten bijvoorbeeld dat als voorkennis toeneemt, het bestuderen van voorbeelden voor studenten minder effectief is dan het oplossen van oefenproblemen (Kalyuga et al., 2001). Het bestuderen van alleen voorbeelden zou daarom minder effectief en efficiënt kunnen worden, en het bestuderen van probleem-voorbeeld paren effectiever en efficiënter. Daarnaast speelt motivatie mogelijk ook hier weer een rol. Dat wil zeggen, met langere taaksequenties wordt het enkel bestuderen van voorbeelden mogelijk als minder motiverend ervaren dan wanneer de geleerde kennis ook toegepast kan worden in oefenproblemen.

Onderzoeksresultaten

Het doel van de twee experimenten beschreven in **hoofdstuk 2** van dit proefschrift was om te onderzoeken hoe effectief, efficiënt en motiverend verschillende sequenties van (video) voorbeelden en oefenproblemen zijn. In het eerste experiment kregen eerstejaars hbo-studenten in het technisch onderwijs ($N = 124$) vier wiskunde leertaken aangeboden in één van de vier condities, namelijk 1) voorbeelden, 2) voorbeeld-probleem paren, 3) probleem-voorbeeld paren, of 4) problemen. Om te onderzoeken of de resultaten zouden repliceren met een andere studentpopulatie (studenten met een niet-technische achtergrond), werd een tweede experiment uitgevoerd met dezelfde opzet maar met pabo-studenten ($N = 81$). Effecten op isomorfe taken en een procedurele en conceptuele transfertaak werden onderzocht (effectiviteit), alsmede de geleverde inspanning en geïnvesteerde tijd na iedere taak in de leerfase en testfase (efficiëntie). Motivatie werd voor en na de leerfase gemeten met behulp van korte vragenlijsten gericht op vertrouwen in eigen kunnen en taakinteresse.

In het eerste experiment leidden de drie condities waarin voorbeelden (afgewisseld met oefenproblemen) werden aangeboden tot hogere prestaties op de isomorfe taken, met minder moeite en meer vertrouwen in eigen kunnen dan de conditie waarin alleen oefenproblemen werden aangeboden. Daarnaast leidde de conditie waarin alleen voorbeelden werden tot hogere prestaties op de isomorfe taken, met minder moeite, tijdsinvestering en meer vertrouwen in eigen kunnen dan de conditie met voorbeeld-probleem paren. In het tweede experiment leidden de drie condities waarin voorbeelden (afgewisseld met oefenproblemen) werden aangeboden ook tot hogere prestaties op de isomorfe taken, met minder moeite en meer vertrouwen in eigen kunnen dan de conditie waarin alleen oefenproblemen werden aangeboden. De resultaten op motivatie verschilden echter van het eerste experiment. In het tweede experiment leidde alleen de conditie met voorbeelden tot meer vertrouwen in eigen kunnen dan de condities waarin probleem-voorbeeld paren of alleen oefenproblemen werden aangeboden. Daarnaast behaalde de conditie met alleen voorbeelden dezelfde prestaties met minder moeite en tijdsinvestering dan de condities met voorbeeld-probleem paren en probleem-voorbeeld paren. In beide experimenten werden geen verschillen gevonden in prestatie, vertrouwen in eigen kunnen en mentale inspanning tussen voorbeeld-probleem paren en probleem-voorbeeld paren. Ook werd er in beide experimenten geen verschil gevonden tussen de vier condities op de transfertaken en taakinteresse.

Omdat in de experimenten beschreven in **hoofdstuk 2**, in tegenstelling tot eerder onderzoek (Van Gog et al., 2011), geen verschillen werden gevonden tussen voorbeeld-probleem paren en probleem-voorbeeld paren in prestatie en vertrouwen in eigen kunnen, en omdat studenten in de conditie met alleen voorbeelden een hogere prestatie en meer vertrouwen in eigen kunnen bereikte dan de conditie met voorbeeld-probleem paren, werd in het eerste experiment beschreven in **hoofdstuk 3** onderzocht of deze resultaten zouden repliceren en stabiel zouden blijven op een tweede test een week later. Daarnaast werd in het tweede experiment beschreven in **hoofdstuk 3** onderzocht of de resultaten met korte sequenties (4 leertaken) anders worden met langere sequenties (8 leertaken). In beide experimenten werd tevens in meer detail onderzocht hoe vertrouwen in eigen kunnen zich ontwikkelt tijdens het leren. Eerstejaarsstudenten uit het technisch hoger onderwijs leerden een wiskunde probleem oplossen met behulp van vier (Experiment 1; $N = 157$) of acht leertaken (Experiment 2; $N = 105$). Studenten werden toegewezen aan een conditie met 1) alleen voorbeelden, 2) voorbeeld-probleem paren, 3) probleem-voorbeeld paren, of 4) alleen oefenproblemen. De uitkomstmaten waren identiek aan de uitkomstmaten van het eerste en tweede experiment, behalve dat vertrouwen in eigen kunnen ook werd gemeten na elke taak in de leerfase.

Beide experimenten lieten zien dat het vertrouwen in eigen kunnen hoger was na de eerste leertaak in de condities waarin studenten startten met een voorbeeld in plaats van een probleem. Echter, na de tweede leertaak rapporteerden alle studenten

in de condities met voorbeelden (ook probleem-voorbeeld paren) meer vertrouwen in eigen kunnen dan de conditie met alleen oefenproblemen. Dit patroon bleef stabiel tijdens en na de leerfase. In het eerste experiment leidden alle condities met voorbeelden (afgewisseld met oefenproblemen) ook tot hogere prestaties op de isomorfe taken, met minder moeite dan de conditie waarin alleen oefenproblemen werden aangeboden. In het tweede experiment leidde alleen de conditie waarin enkel voorbeelden werden bestudeerd tot hogere prestaties op de isomorfe taken, met minder mentale inspanning tijdens de leerfase dan de conditie met alleen oefenproblemen. Er werd in beide experimenten geen verschil gevonden tussen condities op de conceptuele transfer vragen of op taakinteresse. Ten slotte werd in de conditie met alleen voorbeelden minder moeite en tijd geïnvesteerd in de leerfase dan in de condities waarin voorbeelden en problemen werden afgewisseld, en werd in de conditie met voorbeeld-probleem paren in het derde experiment minder moeite geïnvesteerd dan in de conditie met probleem-voorbeeld paren.

Conclusie

Samengevat lieten de resultaten van de experimenten beschreven in Deel 1 van dit proefschrift zien dat, net zoals in eerder onderzoek is aangetoond (Van Gog et al. 2011), het leren van voorbeelden – eventueel afgewisseld met oefenproblemen – een effectievere en efficiëntere strategie is om nieuwe probleemoplosvaardigheden aan te leren dan het alleen oplossen van oefenproblemen. Een nieuwe bevinding is dat het gebruik van voorbeelden (afgewisseld met oefenproblemen) ook tot meer vertrouwen in eigen kunnen leidt dan het alleen oplossen van oefenproblemen. Dat er geen effecten van verschillende sequenties zijn gevonden op taakinteresse komt mogelijk doordat taakinteresse al vrij hoog was en moeilijk veranderbaar is in een kort tijdsbestek.

Een tweede interessante bevinding was dat, in tegenstelling tot eerder onderzoek (Van Gog et al., 2011), het starten met een oefenprobleem beter werkte dan verwacht, mits gevolgd door een voorbeeld. Hierdoor rijst de vraag in welke situaties het starten met een voorbeeld voorafgaand aan een probleem effectiever is voor leren dan starten met een probleem voorafgaand aan een voorbeeld, en wanneer niet. In dit proefschrift is gekeken of motivatie, in termen van taakinteresse en vertrouwen in eigen kunnen, het verschil in resultaten zou kunnen verklaren. Echter, de experimenten beschreven in **Deel 1** van dit proefschrift gaven hiervoor geen duidelijk bewijs. Starten met een oefenprobleem resulteerde weliswaar in eerste instantie in minder vertrouwen in eigen kunnen, maar dit zorgde er niet voor dat studenten afhaakten bij het bestuderen van de voorbeelden (en oefenproblemen) die volgden. We moeten opmerken dat de taakinteresse van studenten redelijk hoog was: Mogelijk haken studenten bij een lagere taakinteresse wel af wanneer zij starten met een probleem. Ook is het mogelijk dat het gebruik van videovoorbeelden motiverender is dan het gebruik van schriftelijke voorbeelden (zoals meestal gebruikt in eerder onderzoek).

Ten slotte lieten de bevindingen zien dat het alleen bestuderen van voorbeelden nog steeds effectief is wanneer er meer taken in de leerfase (en dus langere sequenties) worden bestudeerd. Dit is een verrassende bevinding omdat we uit eerder onderzoek weten dat voorbeelden hun kracht verliezen naarmate lerenden meer voorkennis verkrijgen (Kalyuga et al., 2001). Echter, de taken namen toe in complexiteit en studenten bestudeerden slechts 1 of 2 voorbeelden per complexiteitsniveau, waardoor het leren van voorbeelden mogelijk nog steeds krachtig bleef. Daarnaast doorliepen studenten de leerfase in hun eigen tempo, waardoor zij zelf bepaalden hoe lang ze een voorbeeld wilden bestuderen of een taak wilden maken (ondanks dat studenten geïnstrueerd werden alles volledig te doorlopen). Hierdoor is het bestuderen van voorbeelden mogelijk niet als overbodig of demotiverend ervaren door studenten.

Deel 2: Zelfgestuurd leren met behulp van voorbeelden en oefenproblemen

Studenten leren tegenwoordig steeds vaker nieuwe probleemoplosvaardigheden in online leeromgevingen waarin voorbeelden en oefenproblemen worden aangeboden en waarin zij zelf mogen kiezen wanneer en hoe (vaak) zij van voorbeelden en oefenproblemen leren. Het is echter de vraag of studenten zelfstandig taken kunnen selecteren die passen bij hun leerbehoeften. Zelfgestuurd leren van probleemoplosvaardigheden is namelijk moeilijk, omdat studenten niet alleen hun prestaties op een specifieke leertaak moeten inschatten, maar deze informatie ook moeten gebruiken bij het selecteren van een geschikte vervolgtask (De Bruin & Van Gog, 2012). Het komt dan ook niet als een verrassing dat onderzoek heeft aangetoond dat studenten, met name novieten, vaak moeite hebben met het nauwkeurig inschatten van hun eigen kennislacunes en bepalen welke volgende taak hen helpt deze te overbruggen (Kostons et al., 2012). Deze bevindingen roepen de vraag op hoe en hoe goed studenten taken selecteren tijdens het verwerven van nieuwe probleemoplosvaardigheden met behulp van voorbeelden en problemen.

Eén recent onderzoek van Foster en collega's (2018) heeft aangetoond dat studenten ook suboptimale keuzes maken wanneer zij kunnen kiezen tussen het leren van voorbeelden en problemen. Zij kiezen vaker voor oefenproblemen dan voor voorbeelden, en starten zelden de leerfase met een voorbeeld. Er is echter meer onderzoek nodig naar het zelfgestuurd leren van voorbeelden en problemen, in het bijzonder naar welke keuzes studenten maken en hoe goed deze keuzes passen bij wat effectief is gebleken voor leren in onderzoek met vaste sequenties (onder andere gebleken uit de studies beschreven in **Deel 1** van dit proefschrift).

Daarnaast is het de vraag of en hoe het zelfgestuurd leren van voorbeelden en problemen ondersteund kan worden. Een manier die mogelijk goed werkt, en in de

praktijk vrij eenvoudig geïmplementeerd kan worden, is het expliciet informeren van studenten over effectieve instructieprincipes die zijn afgeleid uit onderzoek naar instructieontwerp (zie Tabel 1). Het expliciet informeren van studenten over deze principes zou kunnen helpen om de (metacognitieve) kennis van studenten te vergroten over welke principes gunstig zijn voor het leren, en welke niet. Als gevolg daarvan wordt de kans groter dat deze principes ook daadwerkelijk worden toegepast (Yan et al., 2014). Onderzoek naar leerstrategieën heeft laten zien dat het expliciet informeren van studenten inderdaad succesvol is gebleken in het vergroten van deze kennis (Endres et al., 2020) en het gebruik van deze strategieën (Biwer et al., 2019). Daarnaast heeft onderzoek laten zien dat het expliciet informeren van studenten over effectieve leerstrategieën zoals jezelf testen (Engels: *retrieval practice*) niet alleen de kennis over en het gebruik van deze strategie vergroot, maar ook de prestatie kan verbeteren (Ariel & Karpicke, 2017). Het is de vraag of deze benadering ook werkt om het zelfgestuurd leren van voorbeelden en oefenproblemen te verbeteren.

Tabel 1.
Effectieve, Efficiënte en Motiverende Instructieprincipes uit Onderzoek naar Instructieontwerp.

Principe	Uitleg	Referentie
Leren-van-voorbeelden-principe	Door het vervangen van alle of een deel van de oefenproblemen door voorbeelden kunnen beginners meer leren met minder tijd en moeite dan door alleen oefenproblemen op te lossen. Dit is ook motiverender voor leren.	Sweller et al. (2011), Van Gog et al. (2019) Van Harsel et al. (2019, 2020)
Voorbeeld-eerst-principe	Beginners moeten de leerfase starten met een voorbeeld in plaats van een probleem. Dit is efficiënter en motiverender voor leren	Van Gog et al. (2011) Van Harsel et al. (2019, 2020)
Laagste-complexiteits-niveau-eerst-principe	Beginners moeten starten met een taak op het laagste complexiteits-niveau	Van Merriënboer (1997), Van Merriënboer & Kirschner (2013)
Simpel-naar-complex-principe	Beginners moeten het complexiteitsniveau van een taak geleidelijk aan verhogen naarmate hun kennis toeneemt	Van Merriënboer (1997), Van Merriënboer & Kirschner (2013)
Ieder-nieuw-complexiteits-niveau-starten-met-voorbeeld-principe	Beginners moeten aan het begin van elk nieuw complexiteitsniveau een hoog niveau van instructie-ondersteuning krijgen (zoals via een voorbeeld)	Van Merriënboer (1997), Van Merriënboer & Kirschner (2013)

Onderzoeksresultaten

Hoofdstuk 4 beschrijft een exploratieve studie waarin is verkend welke keuzes eerstejaars technische hbo-studenten ($N = 147$) maakten wanneer zij een nieuw wiskunde probleem leerden oplossen door zelf zes leertaken te selecteren uit een taakdatabase met 45 leertaken. Deze taken varieerden in format (videovoorbeelden, tekst-gebaseerde voorbeelden en oefenproblemen), complexiteitsniveau (niveau 1, 2 en 3) en context (zie Figuur 2). Studenten hadden nog geen voorkennis over dit specifieke probleem. Ook is onderzocht in hoeverre hun taakselecties overeenkwamen met effectieve, efficiënte, en motiverende instructieprincipes die zijn afgeleid uit experimenteel instructieonderzoek. Ten slotte is onderzocht of er een relatie was tussen de mate waarin studenten deze principes volgden en hun prestaties op isomorfe en transfertaken, de geïnvesteerde mentale inspanning en tijd, en hun vertrouwen in eigen kunnen en taakinteresse.

De resultaten lieten zien dat de taakselecties van studenten redelijk goed overeenkwamen met de principes uit experimenteel onderzoek naar instructieontwerp. Dat wil zeggen, de overgrote meerderheid van de studenten selecteerde vooral voorbeelden tijdens de leerfase en startte de leerfase met een voorbeeld in plaats van een probleem. Vrijwel alle studenten startten ook de leerfase met een taak op het laagste complexiteitsniveau. Echter, slechts de helft van de studenten bouwde de complexiteit van de taken op van simpel naar complex (van level 1 naar level 2 naar level 3). Als studenten voor het eerst een taak op een hoger complexiteitsniveau bestudeerden, was dit vaak een voorbeeld. Al met al werden er tijdens de hele leerfase meer voorbeelden gekozen dan oefenproblemen en kozen studenten voornamelijk taken op het laagste complexiteitsniveau. Ten slotte lieten de resultaten zien dat er geen relatie was tussen het volgen van alle instructieprincipes en prestaties op isomorfe en transfertaken, de geïnvesteerde mentale inspanning en tijd, en hun vertrouwen in eigen kunnen en taakinteresse.

DE PROEF OP DE SOM								
Niveau 1			Niveau 2			Niveau 3		
Videovoorbeelden	Uitgewerkte voorbeelden	Oefenproblemen	Videovoorbeelden	Uitgewerkte voorbeelden	Oefenproblemen	Videovoorbeelden	Uitgewerkte voorbeelden	Oefenproblemen
Biet drinken	Biet drinken	Biet drinken	Hardlopen	Hardlopen	Hardlopen	Draaimolen	Draaimolen	Draaimolen
Energemeting	Brandstofverbruik	Brandstofverbruik	Water drinken	Water drinken	Water drinken	Roeien	Roeien	Roeien
Fitness	Energemeting	Energemeting	Wasmachine	Wasmachine	Wasmachine	Parfum	Parfum	Parfum
Brandstofverbruik	Fitness	Fitness	Zeeopplossing	Zeeopplossing	Zeeopplossing	Koffievetruik	Koffievetruik	Koffievetruik
Verkeersdrukte	Verkeersdrukte	Verkeersdrukte	Stukadoor	Stukadoor	Stukadoor	Chocoladefeest	Chocoladefeest	Chocoladefeest

Figuur 2: Taakdatabase

De laatste studie van deze dissertatie, beschreven in **hoofdstuk 5**, onderzocht of de resultaten van de studie beschreven in **hoofdstuk 4** zouden repliceren. Ook werd onderzocht welke taakselecties werden gemaakt nadat studenten geprobeerd hadden een oefenprobleem op te lossen (aangezien dit niet mogelijk was in de studie beschreven in **hoofdstuk 4**). Er werd ook onderzocht of het zelfgestuurd leren van voorbeelden en problemen (op verschillende niveaus) even effectief, efficiënt, en motiverend is als het leren van een door de onderzoeker vooraf vastgestelde sequentie van voorbeelden en problemen. Ten slotte werd onderzocht of het expliciet informeren van studenten over effectieve instructieprincipes hun taakselecties, prestaties en motivatie verbeterde. De uitkomstmaten waren identiek aan de uitkomstmaten in de studie beschreven in **hoofdstuk 4**, behalve dat taakinteresse niet werd gemeten. Eerstejaars hbo-studenten in het technische onderwijs ($N = 150$), leerden een voor hen nieuw wiskunde probleem oplossen, door middel van a) een vooraf vastgestelde sequentie van zes voorbeelden en oefenproblemen (vaste sequentie conditie), b) het zelf kiezen van voorbeelden en oefenproblemen na het bekijken van een instructievideo over effectieve instructieprincipes (geïnformeerde zelfregulatieconditie) of c) het zelf kiezen van voorbeelden en oefenproblemen zonder het bekijken van een instructievideo (zelfregulatieconditie). In beide zelfregulatiecondities moesten studenten zelf zes leertaken kiezen uit een taakdatabase (zie Figuur 2).

De resultaten t.a.v. taakselecties waren vergelijkbaar met de resultaten van de studie beschreven in **hoofdstuk 4**. Dat wil zeggen, de taakselecties van het gros van de studenten in de zelfregulatieconditie kwamen redelijk goed overeen met de principes uit onderzoek naar instructieontwerp. Uitzondering was opnieuw dat de taken slechts door ongeveer de helft van de studenten werden opgebouwd van simpel naar complex. Studenten in de geïnformeerde zelfregulatieconditie, die de instructievideo hadden bekeken, volgden de principes net wat beter, aangezien zij vaker de taken opbouwden van simpel naar complex. Echter, dit resulteerde niet in hogere prestaties of meer vertrouwen in eigen kunnen op de posttest dan bij studenten die deze video niet hebben bekeken. Ook bleek er nog ruimte voor verbetering in de taakselecties van studenten: Rekening houdend met de prestatie na een poging een oefenprobleem op te lossen, was ongeveer 40% van de taakselecties niet zo effectief voor het leren (bijvoorbeeld omdat studenten een nieuw oefenprobleem op een hoger complexiteitsniveau kozen terwijl zij het oefenprobleem op een lager complexiteitsniveau nog niet of niet helemaal konden oplossen). Ten slotte werden er geen verschillen in prestatie en motivatie gevonden tussen studenten in de twee zelfregulatie condities en studenten in de vaste sequentie conditie.

Conclusie

Samengevat suggereren de resultaten uit **Deel 2** van dit proefschrift dat studenten tijdens het zelfgestuurd leren van voorbeelden en oefenproblemen (op verschillende complexiteitsniveaus) redelijk goed in staat zijn om leertaken te selecteren. Veel van de taakselecties van studenten kwamen overeen met de instructieprincipes die in onderzoek effectief zijn gebleken voor het aanleren van nieuwe probleemoplosvaardigheden. Dat studenten de principes in grote lijnen volgden, betekende niet per se dat ze altijd de beste keuze op dat moment maakten. Gelet op hun prestaties op de oefenproblemen, was er nog behoorlijk wat ruimte voor verbetering in het selecteren van een passende vervolgtak. Een belangrijk resultaat van de studie beschreven in **hoofdstuk 5** was ook dat het leren van studenten die hun eigen leerproces reguleerden even effectief, efficiënt en motiverend bleek te zijn als dat van studenten die leerden via een vaste sequentie van voorbeelden en oefenproblemen op verschillende complexiteitsniveaus. Samengenomen zijn dit interessante bevindingen, omdat eerder onderzoek liet zien dat de taakselecties van novieten (met name in het begin van de leerfase) nauwelijks overeenkwamen met de effectieve instructieprincipes (Foster et al., 2018), en dat zelfgestuurd leren voor novieten vaak minder effectief is dan het bestuderen van vaste sequenties (Azevedo et al., 2008).

Een mogelijke verklaring voor het feit dat de deelnemers aan de studies uit dit proefschrift het leren van voorbeelden en problemen redelijk goed konden reguleren, en daar evenveel van leerden dan van een vaste sequentie aan leertaken, is dat de deelnemende studenten waarschijnlijk al enige ervaring hadden met soortgelijke taken (zoals het werken met formules) vanwege de grote hoeveelheid wiskunde in hun curriculum. Daarnaast is het leren van voorbeelden een veelvoorkomende strategie in het wiskundeonderwijs (Hoogerheide & Roelle, 2020), waardoor studenten hiermee mogelijk al enige ervaring hebben opgedaan. Deze 'voorkennis' kan hebben geholpen bij het maken van betere taakselectiebeslissingen, aangezien studenten met meer voorkennis betere taakselectiebeslissingen kunnen maken (Corbalan et al., 2006). Daardoor rijst de vraag of studenten zonder deze mogelijke 'voorkennis' dezelfde resultaten laten zien als zij zelf taken mogen selecteren.

Een andere bevinding is dat studenten die de instructievideo hadden bekeken betere taakselecties leken te maken (o.a. vaker de complexiteit van de taken op te bouwen van simpel naar complex) dan studenten die deze video niet hadden bekeken. Niettemin resulteerde dit niet in verschillen op prestatie en motivatie tussen beide zelfregulatiecondities, noch tussen de geïnformeerde zelfregulatieconditie en de vaste-sequentieconditie. Een mogelijke verklaring hiervoor is dat studenten ook zonder video-instructie al relatief goede taakselectiebeslissingen maakten gebaseerd op de instructieprincipes. Echter, zoals eerder genoemd, was er nog ruimte voor verbetering (de taakselecties na het proberen oplossen van een oefenprobleem

pasten niet altijd bij de prestatie op dat probleem). Daarom zou een andere verklaring kunnen zijn dat de interventie niet 'sterk' genoeg was om de taakselecties van studenten nog beter te maken, bijvoorbeeld doordat studenten de video-instructie maar één keer te zien kregen voorafgaand aan het maken van keuzes en niet konden oefenen met de principes. Hierdoor hadden ze mogelijk moeite met het herinneren hoe bepaalde principes toegepast moesten worden. Ook zouden studenten mogelijk baat hebben bij trainingen die helpen om betere inschattingen te maken van hun prestaties en het kiezen van passende vervolgtaken (Raaijmakers et al., 2018), aangezien een deel van de taakselecties die werden gemaakt na het proberen oplossen van een oefenprobleem gekwalificeerd werd als minder relevant voor het leren van de student. Vervolgonderzoek zou moeten uitwijzen of het aanpassen van de interventie de taakselecties, prestaties en motivatie van studenten (verder) zou kunnen versterken.

Praktische aanbevelingen en suggesties voor vervolgonderzoek

Op basis van de resultaten van het eerste deel van het proefschrift en bevindingen uit eerder onderzoek, lijkt het vooral van belang om studenten met weinig tot geen voorkennis (meerdere) voorbeelden - eventueel afgewisseld met oefenproblemen - aan te reiken als zij een nieuwe probleemoplosvaardigheid leren. Hoewel het enkel aanbieden van voorbeelden ook goed werkt (blijkens deze studies), is het in de praktijk waarschijnlijk wenselijk studenten wel te laten oefenen met zelf probleem-oplossen (ook omdat uit eerder onderzoek blijkt dat dit hen helpt met het beoordelen van hun eigen leerproces/vooruitgang; Baars et al. 2014, 2017). De beste volgorde om voorbeelden en oefenproblemen aan te bieden is om studenten te laten starten met een voorbeeld voorafgaand aan het oplossen van een oefenprobleem. Dit lijkt vooral tegelden wanneer de studietijden het aantal leertaken beperkt is. Het motivatievoordeel en het efficiëntievoordeel in termen moeite van het starten met een voorbeeld in plaats van een oefenprobleem verdwijnen wanneer sequenties langer worden. Bij langere sequenties is het nog steeds aan te raden voorbeelden aan te bieden tijdens het leren omdat dit effectiever, efficiënter en motiverender is voor het leren dan enkel het oplossen van oefenproblemen, mits de complexiteit van de taken opbouwt (als dit niet het geval is, weten we uit eerder onderzoek, verdwijnt het voordeel van het bestuderen van voorbeelden en hebben studenten meer baat bij het oplossen van oefenproblemen; Kalyuga 2001).

Gelet op de resultaten van het tweede deel van het proefschrift lijkt het 'veilig' om studenten, wellicht na enige instructie over effectieve, efficiënte en motiverende instructie-ontwerpprincipes (zie Tabel 1), zelfstandig aan de slag te laten gaan met het leren van nieuwe wiskunde vaardigheden met behulp van voorbeelden en problemen, in die zin dat dit hun leren niet lijkt te schaden vergeleken met een vaste set leertaken die vormgegeven is volgens die principes. Echter, er is enige voorzichtigheid geboden

met dit soort conclusies omdat we nog niet weten of deze bevindingen generaliseren naar andere studentpopulaties. De deelnemende studenten in mijn studies hadden mogelijk al enige ervaring met het leren van soortgelijke taken en het leren van voorbeelden. Bovendien was er, zelfs na instructie over effectieve instructie-ontwerpprincipes, nog ruimte voor verbetering in de taakselecties van studenten, en met name wanneer we keken naar hun feitelijke prestatie op de oefenproblemen. Vervolgonderzoek zou daarom moeten onderzoeken hoe we zelfgestuurd leren nog verder kunnen verbeteren, bijvoorbeeld door instructie over effectieve principes krachtiger te maken door deze vaker te laten zien of door studenten aanvullend te trainen in het maken van goede inschattingen van hun prestaties op basis waarvan zij een vervolgtask moeten kiezen.

Curriculum vitae



Milou van Harsel was born in Roosendaal, the Netherlands, on November 8th, 1988. After completing her (bilingual) secondary education at the Jan Tinbergen College in Roosendaal in 2006, she started studying Psychology at Tilburg University. After one year, she switched to Teacher Education for Primary Schools at Avans University of Applied Sciences in Breda, from which she obtained her bachelor's degree in 2010. Hereafter, she enrolled in the Pre-Master (2011) and Master Educational Sciences at Utrecht University (2012), from which she obtained her degree in 2012. During her Master, she completed an internship at the Learning and Innovation Centre of Avans University of Applied Sciences and subsequently remained working there as an education (policy) advisor. After three years (November 2015), she decided to combine her job with a part-time PhD trajectory at the Department of Education at Utrecht University (funded by Avans University of Applied Sciences), resulting in this dissertation. During her PhD-project, she presented her work at various international conferences and was the recipient of the best poster award at the biannual conference of Special Interest Groups 6 and 7 (Instructional Design & Technology Enhanced Learning and Instruction) of the European Association for Research on Learning and Instruction (EARLI SIG 6-7 2016). Moreover, she co-organized two meetings of Special Interest Groups 6 and 7 in Bonn (2018) and online (2020), as part of her junior coordinatorship for EARLI SIG 7. As a part of her job and her PhD-trajectory, she gave numerous workshops and presentations to various educational professionals in the Netherlands. Most of these workshops and presentations focused on translating findings from her own research project and other research in (technology enhanced) learning and instruction to educational practice. Milou is still employed at Avans University of Applied Sciences in Breda and currently working as a researcher and senior education and research policy advisor.

Publications

- Van Gog, T., Hoogerheide, V., & **Van Harsel, M.** (2020). The role of mental effort in fostering self-regulated learning with problem-solving tasks. *Educational Psychology Review*, 32, 1055–1072. <https://doi.org/10.1007/s10648-020-09544-y>
- Van Harsel, M.**, Hoogerheide, V., Verkoeijen, P. P. J. L., & Van Gog, T. (2020). Examples, practice problems, or both? Effects on motivation and learning in shorter and longer sequences. *Applied Cognitive Psychology*, 34(4), 793–812. <https://doi.org/10.1002/acp.3649>
- Van Harsel, M.**, Hoogerheide, V., Verkoeijen, P. P. J. L., & Van Gog, T. (2019). Effects of different sequences of examples and problems on motivation and learning. *Contemporary Educational Psychology*, 58, 260–275. <https://doi.org/10.1016/j.cedpsych.2019.03.005>

Manuscripts submitted for publication

- Van Harsel, M.**, Hoogerheide, V., Janssen, E. M., Verkoeijen, P. P. J. L., & Van Gog, T. (2020). *How do higher education students regulate their learning in an online environment with video modeling examples, worked examples, and practice problems?* Manuscript submitted for publication.
- Van Harsel, M.**, Hoogerheide, V., Verkoeijen, P. P. J. L., & Van Gog, T. (2020). *Instructioning students on effective sequences of examples and problems: Does self-regulated learning improve from knowing what works and why?* Manuscript submitted for publication.

Presentations

- Van Harsel, M.** (2019). Voorbeelding leren van voorbeelden [Exemplary learning from examples]. Presentation at Lectoraat Brein & Leren of Avans University of Applied Sciences, Breda, the Netherlands.
- Van Harsel, M.** (2018). Voorbeelding leren. [Exemplary learning]. Presentation at Trendsconferentie Effectief Leren of Open Universiteit, Heerlen, the Netherlands
- Van Harsel, M.** (2017). Voorbeelding leren: Leren en instrueren met behulp van voorbeelden [Exemplary learning: Learning and instructing while using examples]. Presentation at Math@Avans of Avans University of Applied Sciences, Breda, the Netherlands.
- Van Harsel, M.** (2017). Kijk en Leer: Onderzoek naar het afwisselen en zelf-reguleren van voorbeelden en oefenproblemen in het hoger onderwijs. [Watch and Learn: Research into alternating and self-regulating example study and problem-solving practice in higher education]. Presentation at Team Onderwijs of Leer- en Innovatiecentrum of Avans University of Applied Sciences, Breda, the Netherlands.
- Van Harsel, M.**, Hoogerheide, V., Verkoeijen, P.P.J.L., & van Gog, T. (2020, August). *Self-regulated learning of examples and problems in an online learning environment*. Presentation at the biannual conference of Special Interest Groups 6 and 7 (Instructional Design & Technology Enhanced Learning and Instruction) of the European Association for Research on Learning and Instruction (EARLI SIG 6-7 2020), online.
- Van Harsel, M.**, Hoogerheide, V., Verkoeijen, P.P.J.L., & van Gog, T. (2019, August). *How do higher education students use examples and practice problems in self-*

regulated learning? Poster presentation at the 23rd pre-conference of the Junior Researchers of the European Association for Research on Learning and Instruction (JURE 2019), Aachen, Germany.

Van Harsel, M., Hoogerheide, V., Verkoeijen, P.P.J.L., & van Gog, T. (2019, August). *Example study and practice problem solving: Effects of sequence length on motivation and learning*. Presentation at the biannual conference of the European Association for Research on Learning and Instruction (EARLI 2019), Aachen, Germany.

Van Harsel, M., Hoogerheide, V., Verkoeijen, P.P.J.L., & van Gog, T. (2018, August). *Example study and practice problem solving: Effects of sequence length on motivation and learning*. Presentation at the biannual conference of Special Interest Groups 6 and 7 (Instructional Design & Technology Enhanced Learning and Instruction) of the European Association for Research on Learning and Instruction (EARLI SIG 6-7 2018), Bonn, Germany.

Van Harsel, M., Hoogerheide, V., Verkoeijen, P.P.J.L., & van Gog, T. (2017, August). *A motivational perspective on the effects of alternating example study and problem solving*. Presentation at the biannual conference of the European Association for Research on Learning and Instruction (EARLI 2017), Tampere, Finland.

Van Harsel, M., Hoogerheide, V., Verkoeijen, P.P.J.L., & van Gog, T. (2017, August). *Example study and practice problem solving: Effects of sequence length on motivation and learning*. Presentation at the 22nd pre-conference of the Junior Researches of the European Association for Research on Learning and Instruction (JURE 2017), Tampere, Finland.

Van Harsel, M., Hoogerheide, V., Verkoeijen, P.P.J.L., & van Gog, T. (2016, August). *Sequencing example study and practice problem solving in higher technical education*. Poster presentation at the biannual conference of Special Interest Groups 6 and 7 (Instructional Design & Technology Enhanced Learning and Instruction) of the European Association for Research on Learning and Instruction (EARLI SIG 6-7 2016), Dijon, France. Junior Researcher Award for Best Poster, European Association for Research on Learning and Instruction - Special Interest Groups 6-7 Conference 2016.

van Harsel, M., Hoogerheide, V., Verkoeijen, P.P.J.L., & van Gog, T. (2016, June). *Sequencing example study and practice problem solving in higher technical education*. Presentation at the 9th International Cognitive Load Theory Conference, Bochum, Germany.

Workshops

Raaijmakers, L. H., Van Peppen, L. M., Tillema, M., & **Van Harsel, M.** (2019, February). *Motiverend lesgeven [Motivated teaching]*. Workshop at the Academie voor Industrie en Informatica (AI&I) of Avans University of Applied Sciences, Den Bosch, the Netherlands.

Van Harsel, M. (2018). *Voorbeelding leren: Leren en instrueren met behulp van voorbeelden [Exemplary learning: Learning and instructing while using examples]*. Workshop at Academie voor Engineering & ICT of Avans University of Applied Sciences, Breda, the Netherlands.

Van Harsel, M. (2018). *Voorbeelding leren: Leren en instrueren met behulp van voorbeelden [Exemplary learning: Learning and instructing while using examples]*. Workshop at Academie voor Industrie & Informatica of Avans University of Applied Sciences, Breda, the Netherlands.

Van Harsel, M. & Raaijmakers, L. (2017). *Voorbeelding leren: Leren en instrueren met behulp van voorbeelden [Exemplary learning: Learning and instructing while using examples]*. Workshop at Academie voor Algemeen en Financieel Management of Avans University of Applied Sciences, Breda, the Netherlands.

Van Harsel, M. & Tillema, M. (2017). *Leren van voorbeelden in het creatieve domein [Example-based learning in the creative domain]*. Workshop at Communication and Multimedia Design of Avans University of Applied Sciences, Breda, the Netherlands.

Van Harsel, M. & Tillema, M. (2017). *Weten wat werkt en waarom [Knowing what works and why]*. Workshop at the Academie voor Industrie en Informatica (AI&I) of Avans University of Applied Sciences, Den Bosch, the Netherlands.

Van Harsel, M. & Van Peppen, L. M. (2018, November). *Effectief leren en studeren: Hoe doe je dat? [Learning and studying effectively: How to do that?]*. Workshop at Beroepshavo MBO College Hilversum, the Netherlands.

Dankwoord

(Acknowledgements in Dutch)



Het is een zonnige zondag in juni, de achterdeur staat open en ik zit in kleermakerszit op de bank met de laptop op mijn schoot. Tijd om mijn dankwoord te schrijven, na een reis van bijna zes jaar. Ik kijk door het raam naar buiten en laat mijn ogen afdwalen. Niet om de buitenwereld te bekijken, maar om naar binnen te kijken en te ervaren wat deze reis me heeft gebracht. Allerlei gedachten en emoties passeren de revue en ik realiseer me wat een bijzondere reis het was. Ik heb zowel persoonlijke als professionele groei doorgemaakt, ik ben gevallen en weer opgestaan en vreugde en verdriet hebben elkaar regelmatig afgewisseld. Ik heb nieuwe mensen leren kennen, van waaruit zelfs vriendschappen zijn ontstaan. Maar ik heb helaas ook afscheid moeten nemen van één van de allerliefste personen op deze aardbol, mijn lieve papa.

Er verschijnt een lach op mijn gezicht en tegelijkertijd voel ik een brok in mijn keel. Ik besef me hoe dankbaar ik ben voor alle lieve mensen om me heen. Ze hebben me tijdens deze reis geholpen, gesteund, gemotiveerd, geïnspireerd of op een andere manier bijgestaan. Een aantal van hen wil ik graag in het bijzonder te bedanken.

Allereerst wil ik graag mijn (co)promotoren, prof. dr. Tamara van Gog, dr. Peter Verkoeijen en dr. Vincent Hoogerheide bedanken voor de begeleiding tijdens mijn promotietraject. Zonder jullie was dit proefschrift er nooit gekomen! Door jullie fijne persoonlijkheden, jullie manier van begeleiden en natuurlijk al jullie kennis en kunde heb ik tijdens deze reis flinke afstanden kunnen afleggen.

Tamara, dank dat je mij de kans gaf om als buitenpromovendus binnen jouw onderzoeksteam aan de slag te gaan. Ik herinner me onze eerste kennismaking tijdens mijn eerste conferentie, namelijk de '*6th International Cognitive Load Theory*' in Toulouse. Na de presentatie over de resultaten van mijn masterthesis vroeg je me of ik overwoog om te gaan promoveren. Ik vertelde je dat ik het veel te hoog gegrepen vond en niet wist of ik het zou kunnen. Toch klopte ik een paar maanden later op de deur van jouw kamer bij de Erasmus Universiteit. We hadden een fijn eerste gesprek, waarbij mijn ideeën nog alle kanten uitgingen (*eye-tracking, example-based learning, critical thinking*... het was allemaal super interessant). Je hebt mij geholpen een prachtig voorstel te schrijven waardoor ik van start kon gaan. Wat was ik trots dat ik bij jou mocht gaan promoveren! Dank voor je vertrouwen in mij, je kritische blik, je betrokkenheid, je openheid, je enthousiasme voor mijn ideeën en je eerlijkheid. Dank ook voor de fijne persoonlijke gesprekken die we zo nu en dan hebben gevoerd. Ik waardeer je ontzettend als professional, maar nog meer als persoon.

Peter, dank dat je wilde optreden als copromotor in dit onderzoeksproject. Jouw vriendelijkheid, humor, motiverende feedback, kundigheid en persoonlijke interesse heb ik in dit traject als heel plezierig ervaren. Er is niemand die je *self-efficacy* zo kan verhogen als jij, dat gun ik echt iedereen. Ook dank ook voor je geduld als ik (meer dan eens) mailde over een (statistiek)vraag. Nooit voelde ik me teveel. Naast het feit dat je een fijn mens bent, vind ik je een fantastische lector en een voorbeeld van hoe je wetenschap

en praktijk met elkaar kunt verbinden. Ik voel me vereerd dat ik als senior onderzoeker in jouw kenniskring mag blijven participeren en ik kijk uit naar onze verdere samenwerking binnen het Leer- en Innovatiecentrum.

Vincent, dank dat ik jouw eerste promovendus mocht zijn! Ik had dit traject niet zonder jouw hulp, steun, sturing en gezelligheid kunnen voltooien. Ik moet tegelijkertijd Mark Zuckerberg bedanken voor het uitvinden van Facebook, want zonder hem had Facebook Messenger nooit bestaan en hadden we de begeleiding 'gewoon' en 'alleen' via Skype, Teams en e-mail moeten doen. Boring! Ik kan me overigens voorstellen dat jij soms een *mute button* miste als ik 's avonds om half elf weer eens een berichtje stuurde. Gelukkig leerde je mettertijd deze berichten te negeren en een reactie te sturen als het jou uitkwam. En zo hoort het! Dank ook voor je zorgzaamheid, als ik teveel werkte, als ik er even doorheen zat of als ik een te zware koffer mee had genomen naar een conferentie. Jij wilde wel van koffer wisselen tijdens onze treinreis (en vliegreis, ahum) naar Dijon. Ten slotte wil ik je bedanken voor jouw kennis en kunde, je punctualiteit (er is geen paper de deur uitgestuurd zonder een grondige check op komma's) en alle tijd en moeite die je in mij hebt gestoken. Ik hoop dat veel toekomstige promovendi jou als begeleider krijgen en ik gun jou een fantastische carrière in de wetenschap, want je bent een kei!

Also, I would like to thank the defense committee members, prof. dr. Anique de Bruin, prof. dr. Johan Jeuring, prof. dr. Liesbeth Kester, prof. dr. Sofie Loyens, and prof. dr. Alexander Renkl, for the time and effort invested in reading and assessing this dissertation and participating as an opponent in the defense committee. I feel honored to defend my dissertation to these experts.

Daarnaast wil ik de collega's bedanken die mij op wat voor manier dan ook hebben geholpen om de realisatie van dit proefschrift mogelijk te maken.

De directeur van het Leer- en Innovatiecentrum van Avans Hogeschool, Mirjam Woutersen, wil ik bedanken voor het vertrouwen en de faciliteiten om dit promotieonderzoek uit te kunnen voeren. Tevens wil ik mijn voorgaande teamcoördinatoren, Esther, Josje en Lydia, en mijn huidige teamcoördinator, Albert-Jan, bedanken voor de tijd, ruimte en steun om dit proefschrift te voltooien. Lydia, speciale dank gaat uit naar jou voor de extra tijd die je me hebt gegund om dit proefschrift af te kunnen ronden, je enthousiasme en je onvoorwaardelijke steun tijdens het overlijden van mijn vader. Ook dank ik de (voormalig) directeurs van de academies AI&I, AE&I en de PABO van Avans Hogeschool, André Gehring, Martin Rodenburg, Jan Reinhard en Dominique Majoor, voor de ruimte die ze mij hebben gegeven om het promotieonderzoek binnen hun academies uit te voeren. Ook Françoise Koole-Luteijn van de PABO Hogeschool Zeeland wil ik in dit kader bedanken voor haar medewerking aan dit proefschrift.

Tevens wil ik Jos van Weert, Bert Hoeks en Rob Müller heel erg bedanken voor hun hulp en tijd met het ontwikkelen van de taken. Jos, speciale dank gaat uit naar jou voor het bedenken en ontwerpen van de taken, je betrokkenheid bij het onderzoek en de fijne samenwerking de afgelopen jaren! Ook wil ik Fritschal Terheijden, Marieke de Jong, Sabine Hukema, Bianca Piek - van Wel, Dik Overdijk, Elly Cornelissen, Martijn Bogers, Menno Mandemaker, Onno Eerenberg, Frank Arnouts en Suzan van Brussel ontzettend bedanken voor hun hulp bij de organisatie van de experimenten en het enthousiasmeren van studenten om deel te nemen. Ook wil ik een aantal collega's hartelijk bedanken voor hun hulp bij het uitvoeren van de experimenten, namelijk Suzan van Brussel, Lottie Raaijmakers, Maud Müskens, Divna van Driel, Niek van Hoof, Rob Denevers, Mariette Vissers – Machielsen, Gabriëlle Peters, Gabby op de Weegh, Verine Vissers, Catharina Peekstok, Annette van Overeem, Astrid van de Wijer, Mirthe van Engelen, Patricia van Dongen, Shau-Sha Szeto, Marlies Otten, Yvette van den Bersselaar, Sanne Damsma, Sabine Hukema, Marian van der Lijke – Verheij, Marleen de Haan, Hans Slabbekoorn en Marianne Struijs. Bedankt! Suzan Ravensbergen, jou wil ik hartelijk bedanken voor je hulp bij het scoren van de data. Wat een werk heb je verzet! Daarnaast dank ik Ajda Ortac voor haar hulp bij het ontwerpen van de video's. Je bent een kei, Ajda!

De (oud)collega's van de afdeling Educatie aan de Universiteit Utrecht wil ik bedanken voor hun interesse in mijn onderzoek, de feedback tijdens presentaties (op congressen) en de openheid en hartelijkheid die ik als buitenpromovendus heb ervaren als ik een bezoekje bracht aan de universiteit. Speciale dank gaat uit naar Margot, Eva, Tim en Steven voor hun adviezen en gezelligheid.

I would also like to thank the (former) coordinators of *EARLI Special Interest Group 6 & 7* for the collaboration, especially during the preparation of our members meetings in 2018 and 2020. Margot, Anne, Andreas, Martin, Jean-Michel, Laurie, Moritz, and Alberto, you're all great researchers and lovely persons! Margot, *I'm so happy we were in this together!*

Ook alle (oud)collega's van AI&I, team Onderwijs en team Kennismanagement en andere geledingen binnen Avans wil ik bedanken voor hun interesse in mijn promotieonderzoek, hun collegialiteit en gezelligheid. Floor, Maud, Mirthe, Lydia, Marjolein, Lottie en Sanne, dank voor alle gesprekken, borrels, lunches, etentjes en andere fijne momenten. Jullie zijn fantastische meiden! Lottie, Han, Verine en Irma, dank voor alle interessante discussies en fijne gesprekken in de Expertgroep Brein en Leren. Ik ben heel trots dat ik onderdeel heb mogen uitmaken van deze geweldige groep van experts! Nies, Teun en Ronald, wat heb ik prettig met jullie samengewerkt en veel van jullie geleerd tijdens het ontwikkelen van de nieuwe Avans-visie op praktijkgericht onderzoek. Ingeborg, wat fijn dat jij recentelijk mijn collega bent geworden en wat werk ik fijn met je samen! Ann, Anita, Teun & Mariëtte, ik ben blij dat ik samen met jullie kan sparren over hoe we onderzoek en onderwijs nog sterker met elkaar kunnen verbinden. Ik geniet van onze discussies en leer iedere keer weer bij! Lotte, Corinne, Geert en Marion, dank voor de fijne jaren bij AI&I en de vriendschap die hieruit is ontstaan!

De (oud)collega's van het Lectoraat Brein en Leren, Anita, Anton, Eva, Han, Hans, Ilse, Janneke, Lara, Lottie, Marion, Marloes, Michael, Peter en Stefan, wil ik bedanken voor de fijne donderdagen, goede discussies, gezelligheid en gedeelde passie voor onderzoek en onderwijs. Ilse, wat heb ik genoten van onze samenwerking tijdens het opzetten van ons allereerst praktijkonderzoek naar kritisch denken bij studenten in de lab opleidingen, nog voordat het Lectoraat Brein en Leren startte. Marloes, Suzan, Lara en Eva, heel veel dank voor jullie adviezen, relativering en gezelligheid tijdens onze gesprekken over PhD-life. Jullie zijn alle vier kanjers! Hans, dank dat je mij als PABO-docent ooit het vertrouwen en de moed hebt gegeven om een master Onderwijskunde te gaan volgen. Esther, ik kijk ontzettend uit naar onze samenwerking binnen het kenniscentrum en Lectoraat Digitale Didactiek! Anita, jou wil ik in het bijzonder bedanken. In november 2011 heb jij mij aangenomen als stagiaire binnen het Leer- en Innovatiecentrum. Wat heb ik veel van jou mogen leren en wat heb ik door jou veel mogen en kunnen ervaren binnen en buiten de muren van Avans. Je hebt me tevens aangespoord om dit promotieonderzoek te gaan doen. Mijn passie voor leren, onderwijs en onderzoek is door jou alleen maar sterker geworden. Dank voor alle mooie persoonlijke en werk gerelateerde gesprekken, onze gezellige conferentiebezoekjes, je adviezen, en je geloof in mij en mijn kunnen. Jij bent één van de mooie voorbeelden in mijn professionele carrière die ik niet snel zal vergeten!

Mijn lieve vrienden en vriendinnen ben ik ook zeer dankbaar voor de nodige steun en ontspanning tijdens dit traject. Zij hebben mij regelmatig doen beseffen dat werk niet het belangrijkste is in het leven, al deed ik dat soms wel zo voorkomen. Vickie, Mandy, Sanne, Laura, Corinne, Lotte, Marion, Noortje, Dominique, Eef, Tamara, Tessa en Anouk, dank voor alle gezellige, mooie en hilarische momenten die we de afgelopen jaren hebben meegemaakt. Ook veel dank voor jullie interesse, lieve kaartjes, begrip en geduld (als ik weer eens een weekend moest doorwerken om iets af te krijgen). En dank voor jullie onvoorwaardelijke steun voor, tijdens en na het overlijden van papa. Ik ben dol op jullie allemaal!

Grote dank gaat ook uit naar mijn lieve paranimfen en vriendinnen, Dominique en Marion. Wat ben ik blij dat jullie er zijn en dat jullie mijn paranimfen willen zijn!

Do, ik ken niemand die zo lief, warm en zorgzaam is als jij. Jij bent er altijd, waar je ook bent. Ik word altijd zo ontzettend blij van onze (minimaal) anderhalf uur durende telefoongesprekken, van onze gezellige (*home-cooked* of *chique*) etentjes en van onze spontane bezoekjes aan de Efteling. Dankjewel voor alles wat ik op persoonlijk en professioneel vlak (al) van je heb mogen leren en voor alle prachtige gesprekken die we tot nu toe hebben gevoerd, over van alles in het leven. *You're one in a million!*

Tillie, herinner je nog ons eerste gesprek in het noodgebouw van Avans? Wat een klik was daar! Vanaf dat moment spreken we elkaar bijna iedere dag (lang leve Whatsapp) en zien we elkaar regelmatig in Roosendaal, Tilburg of ergens daartussen, om te wandelen, te lunchen, te chillen of spelletjes te spelen met jouw fantastische jongens, Wouter en Pieter. Dank voor je vriendschap, openheid en gezelligheid. Fijn dat ik bij jou helemaal mezelf kan zijn!

Tot slot wil ik mijn lieve (schoon)familie bedanken voor hun liefde en steun.

Els & Leen, dank voor jullie interesse en zorgzaamheid. Fijn dat Bart en ik altijd bij jullie kunnen aanschuiven als we het weer eens druk hebben.

Mam en pap★, ik wil jullie bedanken voor jullie onvoorwaardelijke liefde en steun. Wat voel ik me bevoorrecht dat ik ben opgegroeid in zo'n warm en liefdevol gezin, waar ik altijd mezelf kon zijn en waar er altijd ruimte was voor het delen van gedachten en gevoelens. Pap★, jij bent en blijft mijn grote voorbeeld als het gaat om zorgzaamheid, wilskracht en eerlijkheid. Ik weet hoe graag je de afronding van mijn promotie had willen meemaken, net zoals nog heel veel andere gebeurtenissen in mijn leven. Ik mis je verschrikkelijk, maar ik koester alle mooie herinneringen die ik heb aan jou. Mam, jij bent mijn grote voorbeeld als het gaat om doorzettingsvermogen, relativering en zelfbewustzijn. Ik ben ontzettend dankbaar voor alles wat je me hebt geleerd (en nog steeds leert) en ik ben trots op hoe je in het leven staat. Een lieve, krachtige en warme vrouw. Je bent mijn kanjer!

Bart, wat hebben we in de afgelopen vijf jaar al veel meegemaakt samen, zowel mooie als verdrietige gebeurtenissen. Ik bewonder je integriteit, leergierigheid, analytische blik, gevoeligheid en grote inlevingsvermogen. Dank voor je geduld met mij, voor je luisterend oor, voor je eerlijkheid en vooral voor je warme en liefdevolle knuffels als ik het even niet meer zag zitten. Ik kijk uit naar de toekomst met jou! ❤️

*Sail on silver girl
Sail on by
Your time has come to shine
All your dreams are on their way
See how they shine
Oh, if you need a friend
I'm sailing right behind*

Simon & Garfunkel

