universität freiburg

Video-Based Modelling Examples and Self-Explanation Prompts for Teaching a Complex Problem-Solving Strategy to Learners With Different Levels of Prior Knowledge

Inaugural-Dissertation
zur Erlangung der Doktorwürde
der Wirtschafts- und Verhaltenswissenschaftlichen Fakultät
der Albert-Ludwigs-Universität Freiburg. i. Br.

vorgelegt von
Julius Moritz Meier
aus Viersen

Sommersemester 2023

This dissertation was prepared within the project *Digitale Diagnostik und Intervention im Kfz-Wesen* (English translation: Digital diagnostics and intervention in the automotive sector) in the joint project ASCOT+ (Technology-based **A**ssessment of **S**kills and **Co**mpetences in VE**T**; the "+" stands for the transfer of results into training and examination practice), funded by the German Federal Ministry of Education and Research and the German Federal Institute for Vocational Education and Training.

Gutachter/innen	Prof. Dr. Alexander Renkl & Prof. Dr. Inga Glogger-Frey
Dekan	Prof. Dr. Matthias Nückles
Prüfungsdatum	29.09.2023

Table of Contents

Abstract

Learning from video-based modelling examples, as compared to learning by problem solving, is effective because it frees up working memory capacities. However, learners need to engage in generative learning activities such as self-explanation to use these freed-up capacities for learning. Such self-explanations can be elicited by prompts. Self-explanations prompts are usually directed backwards, that is, towards just studied solution steps (i.e., retrospective prompts). Forward-directed prompts, that is, towards a next step (i.e., anticipatory prompts) are presumably more demanding but – for higher prior knowledge learners – potentially also more beneficial for learning. In addition, self-explanation prompts are sometimes used to prompt learners to compare example cases. Such example comparisons, however, are difficult to implement for video-based modelling examples, as learners cannot watch two videos simultaneously. Instead, it might be useful to ask learners not to compare video examples directly but to ask them to compare non-transient representations of problem-solving processes that have been illustrated in the video-based modelling examples. Such comparative self-explanation prompts might be more demanding than sequentially studying and self-explaining example cases (or representations thereof) but – for higher prior knowledge learners – potentially also more beneficial for learning.

This dissertation includes three manuscripts describing two studies investigating the use of video-based modelling examples and retrospective versus anticipatory or sequential versus comparative representation-based self-explanation prompts for teaching a complex problem-solving strategy (i.e., the diagnosis of car malfunctions). Overall, results indicate that video-based modelling examples are useful for teaching problem-solving strategies in ill-structured domains. Furthermore, results indicate that anticipatory and comparative self-explanation prompts are more suitable for stronger learners. More research, especially on the

exact relationships between the use of worked or modelling examples, cognitive load and learning outcomes, is needed.

Zusammenfassung

Das Lernen mit videobasierten Modellierungsbeispiele ist im Vergleich mit Lernen durch Problemlösen effektiver, weil das Bearbeiten von Beispielen Arbeitsgedächtniskapazitäten schafft. Allerdings müssen die Lernenden sich mit generativen Lernaktivitäten wie Selbsterklärungen beschäftigen, um diese frei gewordenen kognitiven Kapazitäten aktiv für das Lernen zu nutzen. Solche Selbsterklärungen können mit Prompts gefördert werden. Selbsterklärungsprompts sind in der Regel rückwärts gerichtet, also auf gerade betrachtete Problemlöseschritte (d. h. retrospektive Prompts). Vorwärts gerichtete Prompts, die sich auf einen nächsten Problemlöseschritt beziehen (d. h. antizipatorische Prompts), sind vermutlich anspruchsvoller, aber – bei Lernenden mit höherem Vorwissen – möglicherweise auch lernförderlicher. Eine besondere Art von Selbsterklärungsprompts sind vergleichende Prompts, bei denen die Lernenden aufgefordert werden, Beispielfälle zu vergleichen. Ein solcher Vergleich von Beispielen ist jedoch bei videobasierten Modellierungsbeispielen schwierig umzusetzen, da Lernende nicht zwei Videos gleichzeitig ansehen können. Es könnte jedoch sinnvoll sein, die Lernenden aufzufordern, die Videobeispiele nicht direkt zu vergleichen, sondern sie aufzufordern, nicht-transiente Repräsentationen von Problemlöseprozessen zu vergleichen, die in den videobasierten Modellierungsbeispielen illustriert wurden. Solche vergleichenden Selbsterklärungsprompts könnten wiederum anspruchsvoller sein als das sequentielle Betrachten und Erklären von Beispielen (oder deren Repräsentationen), aber - für Lernende mit höherem Vorwissen - möglicherweise auch lernförderlicher.

Diese Dissertation umfasst drei Manuskripte, in denen zwei Studien beschrieben werden, in denen der Einsatz von videobasierten Modellierungsbeispielen und retrospektiven versus antizipativen bzw. sequentiellen versus vergleichenden repräsentationsbasierten Selbsterklärungsprompts für die Vermittlung einer komplexen Problemlösestrategie (d. h. die

Diagnose von Autofehlfunktionen) untersucht wurde. Insgesamt deuten die Ergebnisse darauf hin, dass videobasierte Modellierungsbeispiele für die Vermittlung von Problemlösestrategien in wenig strukturierten Domänen nützlich sind. Außerdem gibt es Hinweise darauf, dass antizipatorische und vergleichende Selbsterklärungsprompts für stärkere Lernende besser geeignet sind. Weitere Untersuchungen, insbesondere zu den genauen Zusammenhängen zwischen dem Einsatz von Beispielen, der kognitiven Belastung und den Lernergebnissen, sind erforderlich.

Introduction

Imagine you are an automotive mechatronic technician (AMT) and a customer reports an illuminated malfunction indicator light, possibly accompanied by other symptoms like a lack of power. It is now your job as an AMT to diagnose the cause of this malfunction and to eliminate this cause through repair. As in other domains, such as medicine (Elstein et al., 1978, 1990; Schmidt & Rikers, 2007), this diagnostic process includes formulating hypotheses, collecting and interpreting data, and eventually evaluating the hypotheses (Abele, 2018; Abele & von Davier, 2019). Strategies for conducting such diagnostic processes can thus be considered problem-solving strategies (van Merriënboer, 2013). For teaching problem-solving strategies, learning from examples is effective (Renkl, 2014; van Gog et al., 2019). Examples are effective both in well-structured domains (Atkinson et al., 2000; Renkl, 2014) and ill-structured domains (Renkl et al., 2009; Rourke & Sweller, 2009) and can take the form of text-based worked examples or video-based modelling examples with the latter being rather present in less well-structured domains (van Gog & Rummel, 2010). This worked or modelling example effect is usually explained via cognitive load theory (Sweller et al., 1998). Especially for novices (see expertise reversal effect; Kalyuga & Renkl, 2010), text-based worked examples or video-based modelling examples make the application of weak problem-solving strategies unnecessary, thereby reducing learning-irrelevant extraneous cognitive load. Consequently, the first aim of this dissertation is to investigate the use of video-based modelling examples for teaching AMT apprentices a strategy for diagnosing car malfunctions. This aim is pursued in all three manuscripts on which this dissertation is based.

If working memory capacities, which are freed up by examples, are used for learning-related activities such as self-explanation or comparison, learning is promoted. Usually, self-explanation prompts (Renkl et al., 1998) are directed backwards and ask learners to self-explain aspects of a problem-solving strategy that have just been illustrated in the corresponding example (retrospective prompts). Anticipatory prompts that refer to the next problem-solving step in an example could also be very effective, but they might also be more demanding. Presumably, only stronger learners with more prior knowledge can be expected to benefit from such prompts. Thus, the second aim of this dissertation is to investigate the effects of retrospective versus anticipatory self-explanation prompts for learners with different prior knowledge levels. Manuscript 2 focuses on this comparison.

Prompts can also be used to ask learners to compare several example cases (Alfieri et al., 2013; Rittle-Johnson & Star, 2011). Comparing examples is more demanding than studying them sequentially. Thus, again only learners with more prior knowledge can be expected to benefit from comparative self-explanation prompts (Rittle-Johnson et al., 2009). Example comparisons are difficult to implement for video-based modelling examples, as learners cannot watch two videos simultaneously. In manuscript 3 we, therefore, propose to not ask learners to compare transient video examples directly but to prompt them to compare non-transient representations of the problem-solving processes that are illustrated in a video-based modelling example. Thus, the third aim of this dissertation is to explore the use of sequential versus comparative representation-based self-explanation prompts for learners with different prior knowledge levels. Manuscript 3 focuses on this comparison.

In the following sections, I will first address text-based worked examples in well-structured domains and video-based modelling examples in ill-structured domains. I will further describe two qualifications for example-based learning and I will explain the beneficial effect of examples on learning as well as the qualifications for example-based learning via cognitive load theory. Eventually, I will introduce self-explanation and comparison as generative learning activities that are crucial for learning from examples. In the next section, I will then introduce the diagnosis of car malfunctions as a complex problem-solving process. I will also briefly explain which approaches to promoting diagnostic strategies exist so far in the automotive domain. Eventually, I present three manuscripts referring to two studies that address the aforementioned aims of this dissertation, followed by a general discussion.

Learning from Examples

Text-Based Worked Examples

For teaching problem-solving strategies, example-based learning is effective, especially for novices (Renkl, 2014; van Gog et al., 2019). In example-based learning, learners often first study an instruction on domain principles and/or the problem-solving strategy (VanLehn, 1996) and then receive examples of worked-out solutions where exemplary problems have been solved (Sweller, 2006). In research on example-based learning, the focus was first on text-based *worked examples* in well-structured domains, such as algebra. For example, in a series of experiments, Sweller and Cooper (1985) either provided students with solved examples of algebraic equations (i.e., worked example condition) or let students practise solving algebraic equations (i.e., conventional problem-solving condition). They found that for simple equations, students in the worked example condition completed the acquisition phase in less time while performing equally well in the posttest as students in the conventional problem-solving condition. Thus, students in the worked example condition learned more efficiently (Sweller & Cooper, 1985; experiment 2). In experiment 3, this finding was replicated for more complex algebraic equations. Moreover, students in the worked example condition also completed the posttest quicker, with fewer mathematical errors, and performed fewer unnecessary in-between steps than students in the conventional problem-solving condition (Sweller & Cooper, 1985; experiment 3). This *worked example effect* (Sweller, 2006), that is, the finding of worked examples being either more efficient, more effective, or both than practising to solve problems for well-structured domains such as algebra has been replicated numerous times since then (see Atkinson et al., 2000; Paas & van Gog, 2006; Renkl, 2014; VanLehn, 1996, for reviews).

In the early 2000s, research on example-based learning began to explore the application of worked examples for less well-structured problems. One example of such an ill-structured problem is mathematical proving. Mathematical proving – if it is done correctly, for example, by an expert – includes heuristic processes such as exploring and investigating the problem and producing conjectures for how to solve this problem (Boero, 1999). For such ill-structured problems, Reiss and Renkl (2002) presented the concept of *heuristic worked examples*. Unlike traditional (algorithmic) worked examples that illustrate clearly defined solution steps, heuristic worked examples rather illustrate an expert's cognitive processes (see cognitive modelling; Collins et al., 1988) while solving a heuristic problem (Reiss & Renkl, 2002). In further research, heuristic worked examples were found to be beneficial for teaching heuristic strategies for various ill-structured problems: Hilbert et al. (2008) used heuristic worked examples to teach mathematical proving in geometry (Boero, 1999). Roelle et al. (2012) used nonalgorithmic worked examples to promote students‘ knowledge about and application of cognitive and metacognitive strategies when writing learning journals. Hänze and Less (2022) developed heuristic worked examples to promote students students‘ mathematical modelling competencies, that is, a multi-step process of expressing a real-world task in mathematical terms (Blomhoj & Jensen, 2003). Taken together, text-based worked examples have proven beneficial for learning strategies for solving well- and ill-structured problems.

Video-Based Modelling Examples

In recent years there has been a shift in the format of examples, that are investigated. In early research, mainly text-based worked examples sometimes enriched by visual, were used (e.g., Catrambone, 1996; Renkl, 1997; Renkl et al., 1998; Sweller & Cooper, 1985; van

Gog et al., 2006). While text-based examples are, of course, still in use in research and educational practice (e.g., Safadi, 2022; Schalk et al., 2020), interest in video-based examples has increased since around the mid-2000s (e.g., McLaren et al., 2008; Rummel et al., 2009; Rummel & Spada, 2005). In their 2010 review article, van Gog and Rummel distinguished between the term *worked examples* for text-based examples the term *modelling examples* for video-based examples, which usually illustrate a person's actions and cognitive processes (van Gog & Rummel, 2010). With the term modelling examples, van Gog and Rummel (2010) followed previous theories and models such as Collins et al.'s (1988) cognitive apprenticeship and cognitive modelling and van Merriënboer's (1997) four-component instructional design model. However, this differentiation has not been fully accepted, as in some cases the term *video(-based) worked examples* is still used (e.g., Schmitz et al., 2017; Solé-Llussà et al., 2020). Therefore, I use the terms *text-based worked examples* and *video-based modelling examples* in the following to avoid any confusion about the format.

Video-based modelling examples are suited for illustrating a model's actions and cognitive processes in less well-structured and heuristic domains (van Gog & Rummel, 2010). The model's actions can be shown from different perspectives. Models can be shown in a third-person perspective, for example, in lecture-style modelling examples, in which the model is placed next to a screen on which steps of a problem-solving process are displayed (e.g., Hoogerheide et al., 2016, 2018; van Wermeskerken et al., 2018). For object manipulation tasks, such as assembly tasks, modelling examples from both a third-person perspective (e.g., van Gog et al., 2014) and a first-person perspective (e.g., Castro-Alonso et al., 2015; Marcus et al., 2013) have been investigated. Fiorella et al. (2017) found that for an assembly task of an electrical circuit, university students benefitted more from a first-person perspective than from a third-person perspective, especially for more complex tasks. A special form of first-person modelling examples are screencasts that show a model's actions on a computer. Such screencast examples are often used when the application of a problem-solving strategy in a computer simulation is illustrated. For example, Mulder et al. (2014) and Kant et al. (2017) used screencast examples to promote inquiry learning behaviour, that is, complying with the control-of-variables strategy (Chen & Klahr, 1999) when conducting simulated physics experiments. In summary, as text-based worked examples, video-based modelling examples are effective for teaching problem-solving strategies, especially in less well-structured and heuristic domains (van Gog & Rummel, 2010).

Irrespective of the perspective, so far video-based modelling examples were used to teach rather short problem-solving strategies. For example, Schmitz et al. (2017) applied four erroneous modelling examples lasting 30 to 51 seconds to teach nursing students to deliver bad news. Hoogerheide et al. (2014) used two modelling examples of less than 3 minutes each to teach students in pre-university education the procedure for calculating the probability of drawing balls from urns without replacement. Many other studies also applied modelling examples of similar length to teach rather simple problem-solving strategies (e.g., Fiorella et al., 2017: assembling an electrical circuit, 90 seconds; Hoogerheide, 2016; Hoogerheide et al., 2018: calculating current, voltage, and resistance, 240 seconds). Presumably, longer video-based modelling examples for more complex problem-solving strategies will also be beneficial in terms of learners‘ cognitive load and learning outcomes. However, such longer video-based modelling examples have hardly been investigated so far. Consequently, one of the goals of this dissertation is to replicate the worked or modelling example effect with longer video-based modelling examples for more complex problem-solving strategies. This goal is pursued in all three manuscripts on which this dissertation is based.

Boundary Conditions of Effective Example-Based Learning

Taken together, the beneficial effect of text-based worked examples and video-based modelling examples on learning has been found and replicated for well- and ill-structured problems in various domains and formats. However, research has identified various boundary conditions for the worked or modelling example effect. Two of those boundary conditions are the expertise reversal effect and the need for generative learning activities: First, according to the *expertise reversal effect* (Kalyuga & Renkl, 2010), examples are particularly effective for novices. For experts of a domain studying examples might be detrimental to learning. Experts rather benefit from practising the application of problem-solving strategy instead of studying exemplarily solved problems (e.g., Brunstein et al., 2009; Leppink et al., 2012). Second, examples alone might not fully exploit the full potential of example-based learning. Learners need to additionally engage in *generative learning activities* that might be stimulated by, for example, prompts to self-explain contents from the examples (Renkl & Eitel, 2019; Wylie & Chi, 2014) or prompts to compare or contrast several examples (Alfieri et al., 2013; Rittle-Johnson & Star, 2011).

Cognitive Load Theory

The worked or modelling example effect, as well as its boundary conditions are usually explained via *cognitive load theory* (Sweller et al., 1998). Cognitive load theory assumes that working memory is limited and that learning, as any other (non-automated) information processing, induces load on working memory. Three different types of cognitive load can be distinguished: *intrinsic* cognitive load (ICL), *germane* cognitive load (GCL), and *extraneous* cognitive load (ECL). If the sum of these load types exceeds available working memory capacities, learning is likely to fail. Recently, Sweller et al. (2019) presented updates to the theory that suggest that intrinsic and germane load can be classified as one type of load (Kalyuga, 2011), resulting in only two types of load (ICL/GCL vs. ECL) that can be distinguished. However, throughout this dissertation, I refer to the 1998 concept with three types of cognitive load (Sweller et al., 1998) as this concept is the basis for most of the research I refer to and because I had this original concept in mind when developing the learning materials and experimental design. Moreover, most cognitive load questionnaires assume a three-factor model (Krieglstein et al., 2022). Furthermore, a recent confirmatory factor analysis found more support for the three-factor model than for a two-factor model assuming no differences between ICL and GCL (Zavgorodniaia et al., 2020).

First, ICL is mainly defined by the learning material's complexity and the learner's prior knowledge. Complexity is defined as element interactivity. That means that the more elements a learner needs to consider at the same time during learning, the higher the ICL the learner experiences. However, learners with more prior knowledge are able to consider more elements at the same time (known as chunking; Sweller et al., 2011). They will therefore experience a lower ICL than novices. For a certain task and learners with a certain level of prior knowledge, ICL is considered fixed.

Second, GCL describes the load on working memory that is induced by generative learning activities. Generative learning activities are activities that support the *organisation* of information in a coherent mental representation in working memory and the *integration* of such representations with relevant prior knowledge structures in long-term memory (see SOI model; Fiorella & Mayer, 2016).

Third, ECL is induced by unproductive and learning-unrelated activities or cognitive processes. Learning materials that contain redundant repetitions or irrelevant information induce a higher ECL. According to cognitive load theory – considering ICL as fixed - ECL should be minimised to make sure that sufficient working memory resources are available for GCL (e.g., Mayer & Moreno, 2003).

Worked or Modelling Example Effect

When learning how to solve a problem, one possibility to reduce ECL (and potentially increase GCL) is to provide worked or modelling examples (Paas & van Gog, 2006; Sweller, 2006). When novices try to solve a problem, they often apply weak or superficial problem-solving strategies. An example for a superficial strategy is a copy-and-adapt strategy, that is, when trying to solve a problem, novices look for a presumably similar problem that has been already solved and copy and adapt the solution procedure. While novices might successfully solve a problem with this strategy in some cases, they are lacking understanding of basic domain principles to adapt the strategy to related but new problems. They may also fail to recognise that the original problem from which they are copying the strategy differs from the current problem in crucial aspects. Consequently, such a copy-and-adapt strategy is not suitable for a wide range of problems and must therefore be considered a superficial problem-solving strategy (Renkl, 2014). As the application of such superficial strategies is not conducive to learning, it can also be considered a learning-irrelevant activity inducing ECL (van Gog et al., 2019). However, studying solved problems (i.e., worked examples or modelling examples) makes the use of these weak problem-solving strategies unnecessary. Learners do not have to search for specific solution steps themselves. Instead, learners can focus on the problem-solving steps provided in the example. Consequently, ECL is reduced and working memory capacities become available that can be used for learning.

Besides the worked example effect, also the two boundary conditions can be explained via cognitive load theory: According to the expertise reversal effect (Kalyuga & Renkl, 2010), more experienced learners do not apply ineffective problem-solving strategies. Instead, those learners have already acquired a schema or a mental representation that can guide their problem-solving. Hence, studying worked-out steps in a worked or modelling example is an unproductive learning activity for them as these worked-out solution steps constitute redundant information inducing ECL (Kalyuga & Renkl, 2010). These more experienced learners rather benefit from practising a problem-solving strategy than from studying examples (e.g., Brunstein et al., 2009; Leppink et al., 2012). Regarding the second qualification: When novices learn a problem-solving strategy, worked and modelling examples are effective because they reduce ECL and thereby free working memory capacities. However, learners should ideally engage in generative learning activities to use these freed-up capacities for learning and to increase GCL. One example of such generative learning activities is self-explanation (Chi et al., 1989; Wylie & Chi, 2014). Such self-explanations often include comparisons of several worked or modelling examples (Alfieri et al., 2013; Gentner, 2010). In the following sections, I will focus on self-explanation prompts in general and on comparative self-explanation prompts in more detail.

Self-Explanation as Generative Learning Activities

In early studies on example-based learning in the late 1980s and 1990s, it was found that learners who are particularly successful at learning from examples explain the content of the examples to themselves (Chi et al., 1989; Renkl, 1997). Such self-explanation processes can also be elicited with *self-explanation prompts* (Renkl et al., 1998). The effectiveness of such prompts on self-explanation and subsequently on learning outcomes has been confirmed in many studies for both text-based worked examples and video-based modelling examples (see Bisra et al., 2018; Rittle-Johnson et al., 2017, for reviews). For example, Hilbert and Renkl (2009) first instructed students on a three-step process of concept mapping and then presented two text-based worked examples to illustrate this process. Although the use of worked examples alone was insufficient in promoting learning (Hilbert & Renkl, 2009; experiment 1), the combination of worked examples and self-explanation prompts proved to be advantageous for learning (Hilbert & Renkl, 2009; experiment 2). Similar effects were found for video-based modelling examples: Schworm and Renkl (2007) created video-based modelling examples for teaching student teachers about argumentations. Although the examples were effective in promoting declarative knowledge about argumentation, it was found that argumentation skills were only enhanced when self-explanation prompts were added to the examples (Schworm & Renkl, 2007). Similarly, in a series of experiments, Hefter et al. (2014, 2015, 2018) combined video examples and self-explanation prompts to promote knowledge about argumentative strategies and the application of such strategies (Hefter et al., 2014), to promote the disposition to apply these strategies (Hefter et al., 2015), and both (Hefter et al., 2018). In all three studies, Hefter et al. found that the learners‘ self-explanation quality mediated the video examples‘ beneficial effects on the respective outcome measures (2014, 2015, 2018).

When learners explain content from examples to themselves, they engage deeply with the underlying principles of the example, as they basically try to make sense of the given learning materials (Wylie & Chi, 2014). Thereby, referring to the SOI model (Fiorella & Mayer, 2016), the organisation of information into a mental representation is facilitated. Moreover, ideally, self-explanation prompts are designed in a way that learners draw on their prior knowledge to answer them (Van Merriënboer & Sweller, 2010), thereby promoting processes of integrating newly developed representations with relevant prior knowledge (Fiorella & Mayer, 2016). Thus, self-explanations promote GCL (Kalyuga, 2009; Renkl et al., 2009). However, this is only the case if students can generate high-quality self-explanations. If learners lack the necessary knowledge to produce solid self-explanations, asking them to self-explain might rather induce a higher ECL and therefore hinder learning (Paas & van Gog, 2006).

Retrospective Versus Anticipatory Self-Explanation Prompts

In most studies on self-explanation prompts, learners are asked to self-explain aspects of a problem-solving strategy that have just been illustrated in the corresponding example. In the aforementioned study on text-based worked examples by Hilbert & Renkl (2009), the self-explanation prompts read as 'To which phase of the concept mapping process can you assign what Carolin/Karsten just did? Why?' (Hilbert & Renkl, 2009, p. 271) with Carolin and Karsten being fictitious students in the text-based worked examples. Similarly, in the study by Schworm & Renkl (2007) on video-based modelling examples the self-explanation prompt read as 'Which argumentative elements does this sequence contain? How is it related to Kirsten's statement?' (Schworm & Renkl, 2007, p. 289) with Kirsten being the model student in the video-based modelling examples. Hence, in both studies learners answered the self-explanation prompts while reasoning on basis of the just completed step of the illustrated problem-solving strategy. For the remainder of the dissertation, I refer to such backwards-directed prompts as *retrospective self-explanation prompts*. When answering such retrospective prompts, learners only need to keep those elements active in working memory that are relevant for the just completed step of the problem-solving strategy. In terms of CLT, element interactivity and thus ICL is therefore rather low.

Another potentially successful type of self-explanation prompt is directed forward. In Renkl's study on individual learner characteristics when learning from worked examples (1997), learners who thought and reasoned about upcoming problem-solving steps were particularly successful. Renkl (1997) called these learners anticipative reasoners. Possibly, corresponding *anticipatory self-explanation prompts* that refer to the next problem-solving step in a worked or modelling examples could also be very effective. In terms of cognitive load theory, effects both on ICL and GCL can be expected. First, when learners reason about upcoming problem-solving steps, they must do so based on past steps. If learners were not aware of which problem-solving steps had already been completed, they would not be able to reason about upcoming steps. Consequently, anticipatory self-explanation prompts should induce a higher ICL than retrospective self-explanation prompts because of the higher element interactivity: relevant elements from two and not only one step of the problem-solving strategy need to be kept active in working memory. However, one could also argue that reasoning about upcoming problem-solving steps while keeping in mind past steps has beneficial effects on the organisation and integration of information into a coherent mental representation (Fiorella & Mayer, 2016). Consequently, anticipatory prompts might also induce a higher GCL than retrospective self-explanation prompts. However, as GCL can only take up working memory capacities that are left over from ICL and ECL (Sweller et al., 2011), presumably only learners with more prior knowledge can be expected to manage the increased demands by the anticipatory prompts without being overloaded by ICL. Consequently, also only these learners can be expected to experience a higher GCL and have better learning outcomes. However, anticipatory prompts have only seldom been investigated and have not been systematically compared with the usual retrospective self-explanation prompts (Bisra et al., 2018). Together with my co-authors, I address this gap in research in the second manuscript of this dissertation.

Comparative Self-Explanation Prompts

Prompts are also often used to ask learners to compare several examples (Alfieri et al., 2013; Rittle-Johnson & Star, 2011). This method is described by the *example comparison principle* (Renkl, 2014), which asserts that comparing multiple examples helps learners to develop abstract schemata and discover similarities and differences between them (Gentner, 2010). Comparisons, where several exemplary problems are solved with the same problem-solving strategy are called *within-category comparisons* (Renkl, 2014) or *problem comparisons* (Rittle-Johnson & Star, 2011). Instead of comparing how the same problem-solving strategy is applied to different problems, example comparisons can also be used to demonstrate how different problem-solving strategies are applied to the same problem. Both strategies could provide a correct solution to the problem (i.e., *correct method comparison*) or one of the strategies could be incorrect or provide a weaker solution to the problem than the other strategy (i.e., *incorrect method comparison*; Rittle-Johnson & Star, 2011). Both problem comparisons and method comparisons can be designed as so-called *critical feature comparisons* (Renkl, 2014) or *contrasting cases* (Glogger-Frey et al., 2017; Schwartz et al., 2011). These are sets of examples that share many features but differ in one or a few critical features to highlight the differences. They can, for example, be used to demonstrate how the same strategy is applied (possibly by different persons) in a more or less efficient manner (e.g., Glogger-Frey et al., 2015; Rittle-Johnson & Star, 2007).

For example, Rittle-Johnson and Star (2007) developed text-based worked examples for algebraic equations that were solved with varying degrees of efficiency. Seventh-grade students were paired up and presented with these worked examples either side-by-side with comparison prompts (i.e., comparison condition) or sequentially with self-explanation prompts that did not encourage comparisons (i.e., control condition). In the comparison condition, students demonstrated more improvement in procedural knowledge and procedural flexibility (i.e., the ability to select and apply the correct problem-solving strategy depending on certain features of the problem to be solved) and showed similar improvement in conceptual knowledge. The authors suggest that by comparing the worked examples, students were able to identify the most important features of the problem, explore different ways to solve it, and be better prepared for the summary lesson that was given to all students (Rittle-Johnson & Star, 2007).

Referring to CLT, the beneficial effect of example comparison can be explained as follows: In the study by Rittle-Johnson and Star (2007), for example, the comparison prompts asked learners to explain why two different methods for solving an algebraic equation yielded the same result (i.e., correct method comparison) or different results (i.e., incorrect method comparison). When learners are encouraged by comparison prompts to distinguish and judge the correct and incorrect use of a strategy they engage with the learning material in depth. This promotes the organisation of the information into a coherent and differentiated mental representation (Fiorella & Mayer, 2016), thereby promoting GCL.

However, the success of comparing example cases is heavily dependent on executive functions and induces heavy demands on working memory, as research on analogical reasoning shows (Holyoak, 2012). When contrasting side-by-side comparisons of example cases with studying the same examples sequentially, it becomes apparent that comparing examples side-by-side causes much higher element interactivity than when studying examples after another. Consequently, comparing examples side by side should induce a substantially higher ICL in comparison to subsequentially studying examples one by one. Only learners with sufficient prior knowledge can then be expected to manage the increased demands (i.e., the higher ICL) and, therefore, benefit from example comparison in terms of a higher GCL and better learning outcomes. For learners with less prior knowledge, comparisons with complex problems are likely to produce cognitive overload. For these learners, sequential study of examples might be more beneficial.

A study by Rittle-Johnson et al. (2009) confirms this assumption. Students‘ prior knowledge about solving linear algebraic equations was tested and then students studied pairs of worked examples including solved linear algebraic equations in one of three conditions: In a first condition, the worked examples included the same equations, but they were solved with different methods (i.e., method comparison). In a second condition, the worked example pairs included different equations that were solved with the same method (i.e., problem comparison). In a third condition, the worked examples were studied subsequentially without comparison (i.e., control condition). Students with more prior knowledge benefited most in terms of learning outcomes when they compared methods in the first condition. Students with less prior knowledge benefited most in the problem comparison or the control condition without comparison (Rittle-Johnson et al., 2009). This finding is an example of an expertise-reversal effect (Kalyuga et al., 2003), as the method that was most beneficial for learners with high prior knowledge was not beneficial for learners with little prior knowledge and vice versa.

Taken together, comparing text-based worked examples possibly promotes GCL and is, therefore, more beneficial for learning than subsequentially studying these examples – provided that the learners have sufficient prior knowledge to be able to manage the increased demands of the comparisons.

Comparison of Video-Based Modelling Examples

Interestingly, the effects of example comparisons have only been investigated with text-based worked examples, but not with video-based modelling examples. At least to my knowledge, there is no research published in peer-reviewed journals on comparing video-based modelling examples. This lack of studies is not very surprising. When comparing example cases, learners need to be able to consider the critical features of these examples side-by-side (see Glogger-Frey et al., 2015; Rittle-Johnson & Star, 2007). Side-by-side comparisons can be done easily with text-based worked examples, as they are static and non-transient in format. However, example comparisons seem to be less easy to implement for video-based modelling examples. One could display video-based modelling examples side-by-side, but learners would be required to either watch the videos at the same time or to pause the videos repeatedly. This makes direct comparison of critical features of the video-based modelling examples very difficult.

Against this background, I propose an alternative for how to implement example comparisons for video-based modelling examples. Instead of asking learners to directly compare video-based modelling examples with another, after watching (parts of) a video-based modelling example, learners are also provided with a static representation of the (so far) illustrated problem-solving process. For example, when a video-based modelling example shows how to conduct physics experiments in a computer simulation (e.g., Kant et al., 2017; Mulder et al., 2014), after watching an example, learners could be provided with a text-based or graphical summary (e.g., a table, a bullet-point summary, or a mindmap) that gives an overview of the most important steps the model has completed in the video-based modelling example. Such a static representation is a non-transient medium and thus better suited for comparison than a transient video. To allow for comparison, learners would then also be provided with an additional representation of the current state of an alternative solution to the same problem. This representation could be, for example, a summary of how the same problem was solved with a different strategy – possibly also resulting in a different (e.g. lower quality) result (i.e., method comparison; Rittle-Johnson et al., 2017). A *comparative representation-based self-explanation prompt* would then ask learners to compare the different representations and look for similarities and differences. Analogous to studies that have investigated example comparison with text-based examples (e.g., Rittle-Johnson et al., 2009; Rittle-Johnson & Star, 2007), a control group would receive the same representations, but would study and self-explain those representations sequentially (i.e., *sequential representation-based self-explanation prompts*). Together with my co-authors, I study these comparative versus sequential representation-based self-explanation prompts for video-based modelling examples in the third manuscript of this dissertation to investigate whether the example comparison principle (Renkl, 2014) also applies to video-based modelling examples.

Diagnosis of Car Malfunctions as an Example of Complex Problem-Solving

Studying worked or modelling examples is effective for learning problem-solving strategies. Problem-solving is about transforming a problem-state into a goal-state (van Merriënboer, 2013). Diagnoses, such as doctors trying to identify which disease a patient is suffering from on the basis of their symptoms (usually referred to as clinical reasoning; Elstein et al., 1978, 1990; Schmidt & Rikers, 2007), teachers evaluating their students' knowledge level in a certain subject or topic (Herppich et al., 2018), or car technicians trying to find the cause of a car malfunction (Abele, 2018; Abele & von Davier, 2019) are typical problem-solving processes. Irrespective of the domain, a diagnostic process usually begins with understanding the problem at hand, formulating hypotheses, and testing these hypotheses. A diagnosis is completed when all relevant information on a problem has been collected and classified and a countermeasure to the existing problem (i.e., a medical treatment, an educational intervention or a repair) can subsequently be developed.

The three manuscripts on which the present dissertation is based deal with promoting the diagnostic competence of apprentices being trained to become automotive mechatronics technicians (AMTs). Following Abele (2018), the diagnostic process of car malfunctions comprises four steps: (1) representing information, (2) formulating diagnostic hypotheses, (3) testing diagnostic hypotheses, and (4) evaluating diagnostic hypotheses: Firstly, AMTs mentally represent problem-related information such as an illuminated malfunction indicator light, possibly accompanied by other symptoms, such as a lack of power. This first step usually also includes reading the error memory of electronic control units that are found for various subsystems in modern cars, such as the engine control unit. Based on this information, in the second step AMTs formulate diagnostic hypotheses, that is, assumptions about potential but untested causes of the present malfunction. In the third step, these hypotheses then are tested by planning and executing corresponding measurements. In the final fourth step, the test results and with it the formulated hypotheses are evaluated. If an AMT concludes that he or she has identified the cause of a malfunction, the repair of this malfunction can be planned. This repair, however, is not part of the diagnostic process (Abele, 2018; Abele & von Davier, 2019)

Strategies for Diagnosing Car Malfunctions

When diagnosing car malfunctions, AMTs can use different diagnostic strategies. These diagnostic strategies differ primarily in the basis on which AMTs base their hypotheses. With the *case-based strategy*, AMTs draw on previous experience to formulate hypotheses: AMTs recognize a pattern of symptoms that they have successfully diagnosed in the past. They then hypothesize that the successful past diagnosis is also correct for the present malfunction. As such, the case-based strategy can be classified as a fast, automatic, effortless and non-analytical strategy including only little demanding problem-solving (Abele & von Davier, 2019; Norman et al., 2007)

A second diagnostic strategy is the *computer-based strategy*. As explained above, the first diagnostic step of representing information usually also includes reading the error memory of the relevant electronic control unit (e.g., the engine control unit) with a diagnostic device. Depending on the diagnostic software an AMT is using with his or her diagnostic device, a computer-based expert system (e.g., ESI[tronic] by Bosch) would also provide instructions on how to exactly diagnose and repair the present malfunction. As such, the computer-based strategy is not as automatic and effortless as the case-based strategy, as following the computer-based strategy requires domain-specific knowledge of how to handle the computer system and how to carry out the proposed measurements to diagnose the malfunction. However, this strategy can still be considered non-analytical, as AMTs solely follow the system's instructions and do not develop hypotheses themselves (Abele & von Davier, 2019).

Eventually, there are situations where an AMT has no prior experience with a symptom pattern and/or a computer system does not provide (helpful) information for the diagnosis of malfunctions. Accordingly, neither the case-based strategy nor the computer-based strategy will lead to a successful diagnosis of the present malfunction. In such situations, AMTs need to follow the *model-based strategy*. This strategy implies that AMTs develop a mental model of the malfunctioning automotive (sub)system, for example, by using electrical circuit diagrams to understand how different sensors, actuators, and electronic control units are linked and influence each other. Based on that mental model, AMTs would then formulate hypotheses about the cause of a malfunction and test them subsequently. As such, the model-based strategy can be considered a systematic, analytical, slow, and effortful approach to the diagnosis of malfunctions and thus it represents real problem-solving (Abele & von Davier, 2019; van Merriënboer, 2013).

Promoting Strategies for Diagnosing Car Malfunctions

Diagnoses are a crucial aspect of an AMT's job profile, as diagnostic activities account for about 50 % of an AMT's working time (Spöttl et al., 2011). However, at the end of their three-year apprenticeship, the diagnostic skills of AMT apprentices are usually insufficient. Only about 15 % of the apprentices master the model-based strategy that is required to diagnose complex malfunctions (Abele & von Davier, 2019). However, at least to my knowledge, there has been no research conducted so far on how to promote AMTs diagnostic skills. There are providers of learning media and computer simulations that deal with teaching diagnostic skills. For example, Clark and Mayer (2016; chapter 18) describe a computer simulation that offers opportunities for AMTs to practice unusual diagnostic situations. However, this simulation mainly uses elements of guided discovery and has not been investigated systematically.

In other domains, however, several interventions have been developed to promote diagnostic skills. For example, Glogger et al. (2013) developed a computer-based learning environment to train teachers how to assess their students‘ use of learning strategies when writing learning journals. Within this learning environment, Glogger et al. (2013) provided teachers with worked examples of filled-in learning journals that represented the different learning strategies. Teachers were also prompted to self-explain and compare the worked examples. In the domain of medicine, Heitzmann et al. (2015) used text-based worked examples to promote medical students‘ diagnostic competence of heart failures. Typical of example-based learning, Heitzmann et al. (2015) combined these worked examples with self-explanation prompts.

Taken together, at the end of their apprenticeship, only 15 % of the AMT apprentices master the model-based strategy that is necessary to diagnose complex malfunctions in cars. As the diagnosis of car malfunctions can be considered problem-solving, an example-based learning environment seems promising for promoting such a model-based strategy in AMT apprentices. The development and evaluation of such a learning environment was one of the goals of the present dissertation and is dealt with in all three manuscripts.

Overview of Studies and Manuscripts

The overarching aim of this dissertation was to investigate the use of video-based modelling examples for teaching AMT apprentices a model-based strategy for diagnosing car malfunctions. Concerning video-based modelling examples, we also investigated the effects of retrospective versus anticipatory self-explanation prompts and explored the use of representation-based comparative self-explanation prompts for video examples. For these further research questions, the learners‘ prior knowledge was taken into account.

For this dissertation, two studies were conducted. Manuscript 1 and manuscript 2 describe the first study, and manuscript 3 refers to the second study. In study 1, in two conditions, the apprentices learned a model-based diagnostic strategy with modelling examples and received either retrospective or anticipatory prompts. In a third condition, apprentices did not receive modelling examples or prompts but the respective open problems. Manuscript 1 describes the development of the model-based strategy and the modelling examples. Following design-based research guidelines, we formatively evaluated the learning materials with expert judgments and a small study during the development of the learning environment. Finally, an evaluation study only considering the data of the two conditions in which apprentices learned with modelling examples showed that the learning environment promoted apprentices' knowledge about the diagnostic strategy. However, they could not transfer their knowledge to diagnostic problem-solving. Overall, the apprentices evaluated the learning environment positively, except it was considered too long and repetitive.

For manuscript 2, we re-analysed the data from study 1 for all three conditions (i.e., including the control condition) with a focus on the effects of retrospective versus anticipatory self-explanation prompts depending on the learners‘ prior knowledge. In comparison with the control condition, modelling examples did not promote learning. However, among the apprentices who learned with modelling examples, differential effects of the self-explanation prompts on diagnostic knowledge and germane cognitive load were found. For these outcomes, apprentices with low prior knowledge benefited from retrospective prompts, and apprentices with high prior knowledge benefited from anticipatory prompts. These findings suggest using different self-explanation prompts for learners with different levels of expertise.

The results of study 1 as described in manuscripts 1 and 2 indicated that apprentices found the intervention repetitive. To shorten the intervention we first adapted the strategy and then created new learning materials. This newly developed intervention is evaluated in study 2 as described in manuscript 3. Manuscript 3 also investigates the use of comparisons for video-based modelling examples: Similar to study 1, in two conditions, apprentices learned a (new) model-based strategy with modelling examples. These were accompanied by either comparative or sequential self-explanation prompts. In a third condition, apprentices did not receive modelling examples or prompts but the respective open problems. Modelling examples had beneficial effects on diagnostic knowledge and scaffolded diagnostic skills but not on independent problem-solving. In addition, there were no effects of examples and prompts on cognitive load. We assume that apprentices would have needed more practice opportunities. Moreover, the comparative prompts seem to be promising for stronger learners with more prior knowledge. We conclude that representation-based comparisons are useful for video-based modelling examples while comparative prompts seem promising for stronger learners. Further research, especially on the effects on cognitive load, is necessary.

Manuscript 1: Promoting Car Mechatronics Apprentices' Diagnostic Strategy with Modeling Examples: Development and Evaluation of a Simulation-based Learning Environment

Meier, J., Spliethoff, L., Hesse, P., Abele, S., Renkl, A., & Glogger-Frey, I. (2022). Promoting car mechatronics apprentices' diagnostic strategy with modeling examples: Development and evaluation of a simulation-based learning environment. *Studies in Educational Evaluation*, *72*, 101117. https://doi.org/10.1016/j.stueduc.2021.101117

My first authorship reflects the fact that I was responsible for planning the studies and design the material, for the data analyses, and for the preparation of the paper.

Abstract

Crucial for training automotive mechatronics technicians (AMTs) is enabling them to diagnose car malfunctions. AMTs are particularly successful when they base their diagnostic process on a mental model of the affected automotive system. Still, only few AMT apprentices master such diagnoses after their apprenticeship. Therefore, we created a simulation-based learning environment with modeling examples to teach AMT apprentices a diagnostic strategy that builds on mental models. Following design-based research guidelines, we formatively evaluated our learning and testing materials by expert judgments and a small study during the development of the learning environment. Finally, an evaluation study showed that the learning environment promoted apprentices' knowledge about the diagnostic strategy. However, they could not transfer their knowledge to diagnostic problem-solving. Overall, the apprentices evaluated the learning environment positively, except it was considered too long and repetitive. Reasons for the outcomes as well as possible further developments of the learning environment are discussed.

Keywords: Diagnosis of car malfunctions; mental models; modeling examples; simulation-based learning; learning environment

Introduction

An essential task for automotive mechatronics technicians (AMTs) is localizing car malfunctions. Therefore, AMTs must come up with hypothetical reasons for a malfunction, develop and execute test strategies, and evaluate test results to identify and ultimately repair the malfunction (i.e., AMTs *diagnose*). Diagnostic activities account for about half of an AMT's working time (Spöttl et al., 2011). However, at the end of their 3-year apprenticeship, only 15% of the AMT apprentices master the required strategies to diagnose complex malfunctions (Abele & von Davier, 2019). So far, the development of AMTs' diagnostic skills has been considered mainly descriptively (e.g., Nickolaus et al., 2012). Few attempts have been made to foster such processes (e.g., Clark & Mayer, 2016, chapter 18). However, several interventions have been developed to promote diagnostic skills in other domains, such as teacher education (e.g., Glogger et al., 2013) or medical education (e.g., Heitzmann et al., 2015).

This article describes the development and evaluation of a simulation-based learning environment that relies on example-based learning to promote AMT apprentices' diagnostic skills. Following design-based research guidelines, this development included formative evaluations of the learning material with subject matter experts and apprentices (Collins et al., 2004).

Strategies for Diagnosing Car Malfunctions

As in other domains, such as medicine (i.e., clinical reasoning, e.g., Klein et al., 2019) and education (i.e., teachers' assessments of students, e.g., Herppich et al., 2018), the process of diagnosing car malfunctions begins with comprehending the problem at hand (i.e., a disease, students' comprehension gaps, or a technical malfunction) as well as formulating hypotheses about possible causes and ends when the cause has been identified. When diagnosing car malfunctions, AMTs' diagnostic strategies differ on which basis the hypotheses are formulated (Abele & von Davier, 2019). AMTs often use a computer-based expert system that provides diagnostic hypotheses (*computer-based strategy*). AMTs also often rely on personal experience when they formulate hypotheses (*case-based strategy*). When applying a *model-based strategy,* AMTs base their diagnostic hypotheses on a mental model of the affected car system. This mental model contains information about the components in a car system, how they work, and how they are connected (Kluwe & Haider, 1990). AMTs should follow these model-based strategies when they do not receive sufficient guidance from the computer-based expert system or have little experience with the to-be-diagnosed malfunction. Thus, model-based strategies can be considered the most flexible and powerful diagnostic strategies. However, only 15% of the AMT apprentices master a model-based strategy at the end of their apprenticeship, presumably because apprentices experience situations too infrequently which require a model-based strategy (Abele & von Davier, 2019).

Socio-Cognitive Perspective on Teaching Diagnostic Strategies

Diagnosing can be considered problem-solving as it is about transforming a problem state (i.e., car malfunction) into a goal state (i.e., correct diagnosis for subsequent repair; van Merriënboer, 2013). Consequently, diagnostic strategies are problem-solving strategies, for which example-based learning is effective (Renkl, 2014). In example-based learning, learners usually first receive instructional explanations about the to-be-learned problem-solving strategy, for example in form of instructional videos. In these instructional phases, learners' organization of the learning content into mental representation can be promoted by providing learners with *organizational prompts* (Roelle et al., 2017). Following the instructional phase, learners receive examples, in which the application of the problem-solving strategy is shown. In less well-structured domains, often, *modeling examples* are used. Here, models (e.g., an expert) demonstrate how to solve a problem while verbalizing their thoughts. Modeling examples are therefore similar to the teaching method “modeling”, which is one of the core teaching methods of cognitive apprenticeship (Collins et al., 1988). Modeling examples may be provided face-to-face, as an animation or a video, or as a recording of the model's computer screen (van Gog & Rummel, 2010). After receiving examples, learners try to solve similar problems on their own, still receiving support, which fades over time (Merriënboer & Kirschner, 2018). Example-based learning has also been implemented successfully in simulation-based learning environments (see Chernikova et al., 2020 for a meta-analysis).

Designing and Evaluating Multimedia Learning Environments

When designing and evaluating multimedia learning environments, two critical factors should be considered: the learners' limited working memory and their motivation. The Cognitive Load research provides measures of whether working memory capacity is taken up by processes serving learning or by unproductive processes (Sweller et al., 2011). Three types of cognitive load that additively take up working memory resources can be distinguished: *intrinsic load*, *germane load,* and *extraneous load* (Kalyuga, 2011). Intrinsic load is mainly determined by the complexity of the learning content. Germane load refers to cognitive processes that are directly related to the comprehension of learning material. Extraneous load refers to unproductive cognitive processes and emerges typically from the way the learning content is presented. For example, visually poorly organized material induces high extraneous load, which then occupies working memory resources important for comprehension. Accordingly, it is crucial to minimize extraneous load by proper design of the learning materials (for guidelines see Mayer & Fiorella, 2014; or Mayer & Moreno, 2003). When designing and evaluating learning materials, it is, therefore, necessary to test the materials' effects on learners' cognitive load, especially their extraneous load.

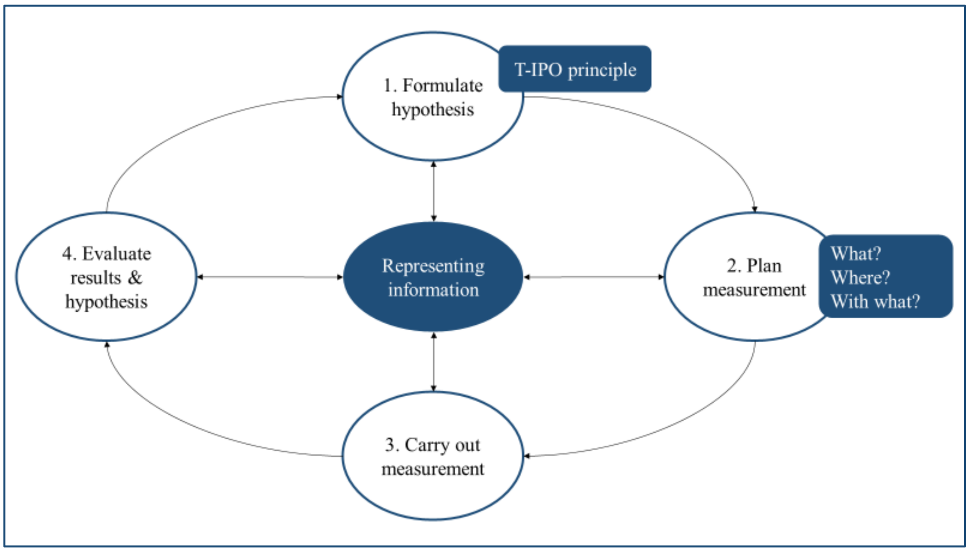
Additionally, motivational factors should be considered when designing learning materials to optimize learning outcomes. Following Vollmeyer and Rheinberg (1998, 2000), we consider four relevant factors when evaluating learning in a simulation-based environment. First, learners might or might not find learning material *interesting*. Second, learners differ in whether they perceive a (learning) task as a *challenge*. Third, learners might suffer from *incompetence fear*, that is, they might be afraid to perform poorly in a certain (learning) task. Finally, learners differ in their *mastery confidence*, that is how confident they are to master a new task, which is very similar to the construct of *self-efficacy* (Bandura, 1997). All these factors influence learning: Learners who find a topic interesting and perceive a learning task as a challenge will be more willing to invest effort during learning (Rheinberg et al., 2000; Schiefele, 1991). Moreover, self-efficacious learners with little incompetence fear are more successful as they are confident to also master challenging tasks and therefore do not avoid achievement situations (Bandura, 1997; Rheinberg et al., 2000). Consequently, in an evaluation of learning materials, the effects on learners' motivation should be monitored.

Definition of the Model-based Diagnostic Strategy

In this study, we developed a model-based strategy in collaboration with subject matter experts, building on a diagnostic problem-solving process in car mechatronics proposed by Abele (2018) and diagnostic strategies from other domains such as medicine (e.g., Elstein et al., 1990) and other technical professions (e.g., Schaafstal et al., 2000). This strategy consists of four steps (1) formulating a hypothesis, (2) planning a measurement, (3) carrying out the measurement, and (4) evaluating the measurement results and the hypothesis, where the mental model of the affected car system is mainly developed during steps 1 and 2 (Figure 1).

Figure 1

Overview of the four-step model-based strategy



To facilitate building a mental model in step 1, apprentices were taught the T-IPO principle, which describes the *transmission*, *input*, *processing*, and *output* of electrical parameters or signals. For example, an apprentice could ask herself, “Which component is responsible for *processing* a measured value?”. Apprentices were told to assign to each component in a car system its respective function according to the T-IPO principle. These functions of the individual components thus formed the basis for hypotheses, since each missing T-IPO function represents a potential source of malfunction.

For planning a measurement in step 2, apprentices were taught to follow the three W-questions, namely (1) “What do I need to measure?” (i.e., which physical entity?), (2) “Where do I need to measure?” (i.e., where is a component installed and which plugs have to be measured?), and (3) “With what do I need to measure?” (i.e., which tools are required?). Then, in step 3, apprentices carry out the measurement and evaluate its results in step 4. If the hypothesis formulated in step 1 needs to be rejected, apprentices need to formulate a new hypothesis and return to step 1.

Development of the Simulation-based Learning Environment

At the end of their apprenticeship, only 15% of AMT apprentices can correctly diagnose complex malfunctions that require a model-based strategy. Modeling examples that take into account cognitive load and learner motivation, and integrating these modeling examples into a simulation-based learning environment, seems promising for teaching apprentices a model-based diagnostic strategy.

Against this background, we have developed a simulation-based learning environment in which apprentices can diagnose the causes of car malfunctions. We modified a computer simulation by Gschwendtner and colleagues (2009) that has been repeatedly revised (e.g., Abele et al., 2014; Nickolaus et al., 2012). Additionally, we developed instructional videos to teach apprentices the model-based strategy already described. The instructional videos were accompanied by organizational prompts (Roelle et al., 2017). For example, right after the instructional video about how to plan a measurement, the corresponding organizational prompt asked the apprentices to plan a measurement by following the presented method.

Following design-based research guidelines, the development of these learning materials included two formative evaluations (Collins et al., 2004): First, we interviewed subject matter experts about the strategy and the learning materials. Second, in a pilot study, seven AMT apprentices worked on the learning materials and completed two newly developed diagnostic strategy tests. Following these formative evaluations, we then developed two modeling examples showing an expert applying the strategy in the simulation. The final intervention will be described in the Methods section of the evaluation study.

Expert Surveys

Five experts from the automotive vocational education field or workshop owners reviewed the strategy and the learning materials. Overall, these experts rated the strategy as applicable and the instructional videos as understandable and appropriate in language and complexity. However, the visual presentation of the instructional videos was regarded as suboptimal, so we added visual highlights and labels. Moreover, the organizational prompts while learning with the instructional videos were described as simplifying, but we did not revise them. Simplifying some aspects of automotive technology in the learning materials is critical to adhere to the theoretical background of example-based learning. The diagnostic strategy should be presented (and practiced) straightforwardly without overloading the learner's cognitive capacities with complex additional information. The later presented, more realistic modeling examples aimed at teaching apprentices that the strategy is also suitable for complex situations. The experts also assessed prototypes of the modeling examples and the self-explanation prompts. The latter were conceptualized as open-ended questions, which the experts found too demanding. Hence, we switched to a gap text format.

Pilot Study

Seven apprentices watched the instructional videos and answered the organizational prompts. They also completed two newly developed tests before and after watching the videos (strategy description test and strategy completion test; see Methods of the final evaluation study for a description). Eventually, apprentices evaluated the instructional videos and tests (open-ended and closed items). Here, we report only the most important insights. The appendix provides an overview of the procedure and results from this pilot study.

Regarding the strategy description test, three apprentices commented that the wording related to aspects of automotive technology was somewhat unclear, so we adapted the task.

After watching the instructional videos and working on the prompts, participants self-assessed their cognitive load on a seven-point Likert scale (Klepsch et al., 2017). Although the apprentices indicated a relatively high intrinsic load ($M = 4.86$, $SD = 0.69$), the videos were nevertheless rated as well designed (extraneous load: $M = 2.64$, $SD = 1.35$) and beneficial to learning (germane load: $M = 4.47$, $SD = 0.92$). Although the instruction was evaluated as lengthy, comprehensibility and structure were evaluated positively. The apprentices indicated that they would recommend the learning material to others. Therefore, we did not change the content of the instructional videos.

Overall, the learning material and the tests were evaluated positively. Once we had addressed the identified shortcomings, we planned a comprehensive evaluation study with the instructional videos, the organizational prompts, the tests, and two newly developed modeling examples. This study will be described below.

Evaluation Study

To eventually evaluate the learning environment, we investigated whether the learning environment promotes apprentices' diagnostic knowledge (i.e., knowledge about the diagnostic strategy) and their diagnostic skills (i.e., application of the diagnostic strategy). Moreover, we investigated whether the learning environment is perceived as motivating and whether it induces acceptable levels of cognitive load.

In this paper, we report a part of a more extensive study ($N = 78$), as the focus of this paper is the *development* and *evaluation* of the intervention. Accordingly, we report only apprentices who received instructional videos and modeling examples ($N = 49$). In the larger study, we also examined apprentices who did not receive modeling examples. Instead, directly after the instructional videos, these apprentices attempted to diagnose the problems presented in the modeling examples in the simulation (i.e., problem-solving instead of modeling examples). However, this is beyond the scope of this paper. Therefore, in the present paper we addressed the following four research questions (RQs):

- RQ 1: Does learning with the learning environment promote apprentices' knowledge of the diagnostic strategy and its application?
- RQ 2: What are the effects of the learning environment on apprentices' motivation?

- RQ 3: How do apprentices rate their cognitive load while working in the learning environment?
- RQ 4: How do apprentices subjectively evaluate the intervention regarding characteristics such as applicability, length, structure, narrator quality, comprehensibility, and satisfaction?

Methods

To answer these research questions, we assessed apprentices' diagnostic knowledge and skill on various diagnostic strategy tests and their motivation before *and* after working with the learning environment (RQs 1 and 2). Moreover, after working in the learning environment, we assessed apprentices' cognitive load while learning (RQ 3) and asked apprentices to subjectively evaluate the environment (RQ 4).

Participants and Design

Data collection took place in four classes at two schools in two separate sessions with a 10-day delay. We will refer to the classes from school one as classes 1a and 1b and to classes from school two as classes 2a and 2b. Pretests were conducted in the first session. During the second session, the intervention and posttest took place. Forty-nine apprentices in their fourth year and thus shortly before the end of their apprenticeship took part in both sessions. Descriptive data are available for forty-seven apprentices, of which forty-five were male. On average, the apprentices were $M = 21.47$ years ($SD = 3.26$) old. Forty apprentices stated that German was their only first language, seven stated at least one other first language.

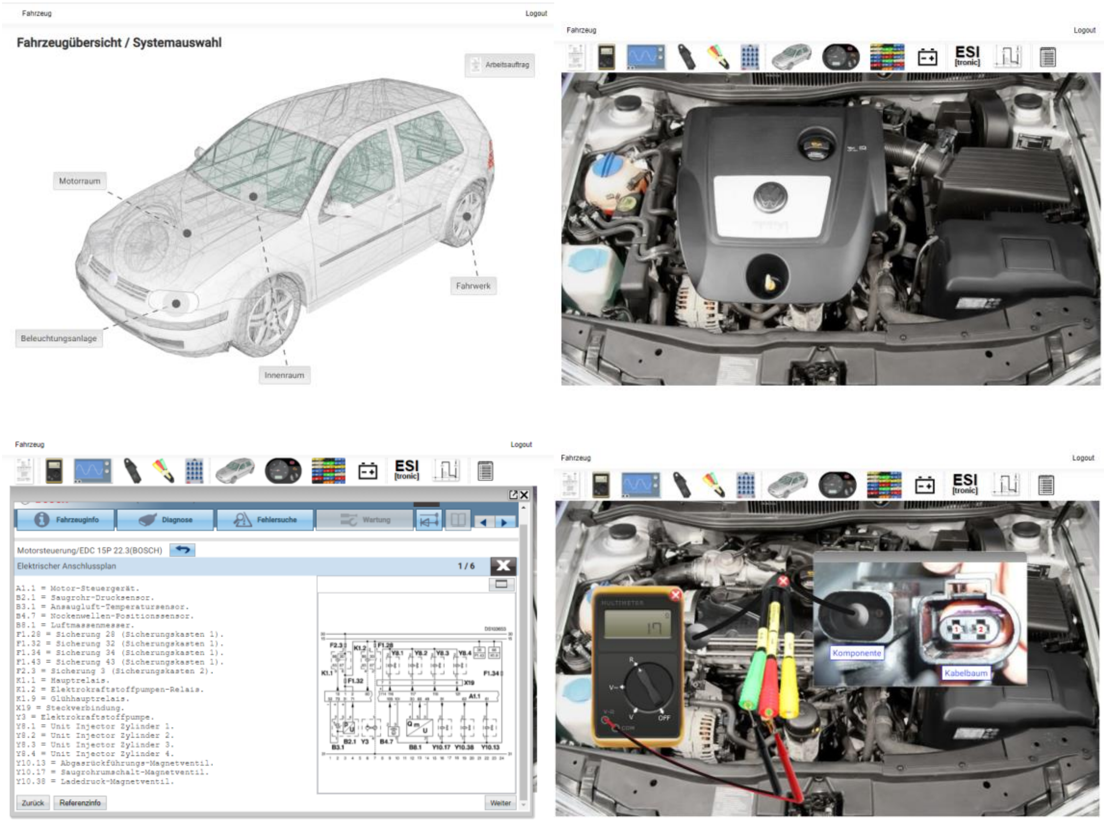
Final Intervention

Computer Simulation

The simulation is made up of authentic drawings, photos, and screenshots of the following parts of an AMT's work environment: (1) a selection of four relevant car systems, (2) a toolbox with various tools, and (3) relevant segments of the ESI[tronic], an internationally widespread computer-based expert system from Bosch (Figure 2). (1) The car systems are the engine compartment and engine control, the lighting system, the interior, and the chassis. We decided to focus only on malfunctions in the engine control, which are diagnosed in the engine compartment (bottom-right in Figure 2). Eleven components (e.g., sensors, cables, and fuses) are available for electronic measurements. These components have different numbers of measuring points (e.g., different numbers of plugs or terminals on these plugs). The car can be in four different operating states (ignition off & engine off, ignition on & engine off, ignition on & engine starting, ignition on & engine running). The components, connectors, and operating states allow for 3840 different measurements of voltages, resistances, and signals. (2) For these measurements, the AMT apprentices have access to various tools. Moreover, (3) the simulation covers relevant segments of the computer-based expert system ESI[tronic] by Bosch. It offers a great variety of information such as electrical circuit diagrams, installation plans, descriptions of components, troubleshooting instructions, or reference values for electrical measurements.

Figure 2

Screenshots of the computer simulation in German



Note. The top left picture shows the starting page of the simulation, giving an overview of the car systems; the top right picture gives an overview of the engine compartment with various tools in the top bar; in the bottom left, the computer-based expert system with a circuit diagram is depicted; the bottom right picture shows the measurement of the resistance of the exhaust gas recirculation valve with a multimeter.

A total of 54 malfunction scenarios are available as realistic work orders on which initial symptoms are listed. When starting a diagnosis in the simulation, an apprentice would usually first read the work order and then scan the fault memory of the engine control. Depending on the entry in the fault memory, the apprentices would then start looking for the cause of the malfunction. Every mouse click made in the simulation is recorded in log-files with a timestamp (Abele & von Davier, 2019).

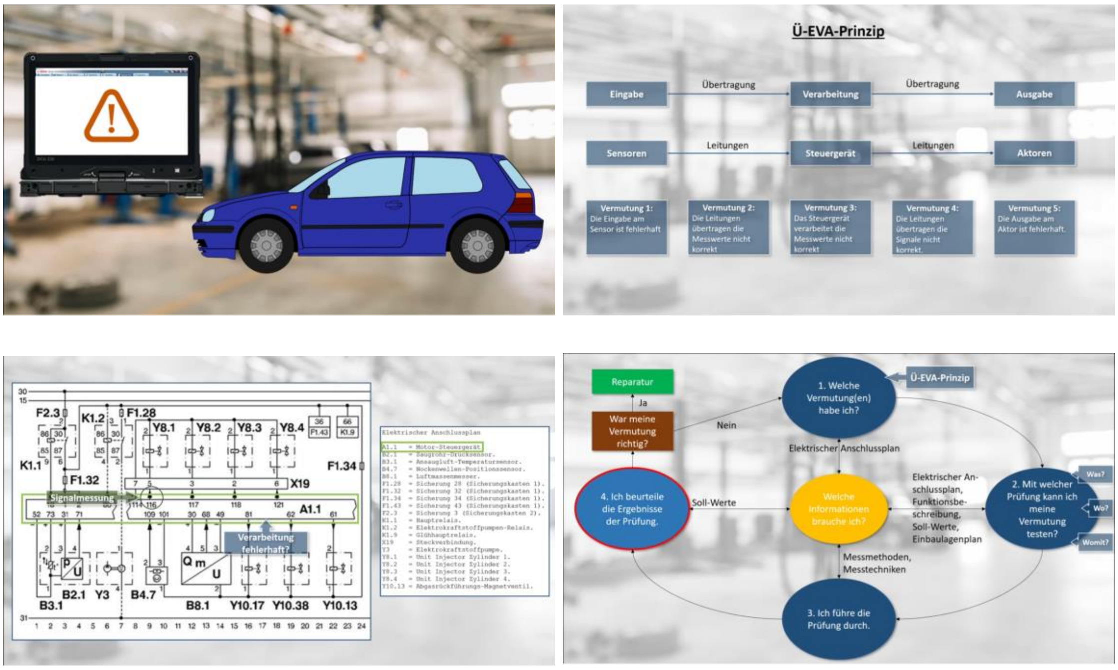
Learning Materials

The intervention consisted of two learning phases comprising instructional videos, organizational prompts, video-based modeling examples, and self-explanation prompts. The videos were integrated into a page-based online survey tool. Apprentices could proceed freely but not return to previous pages. Between the learning phases, apprentices had a break of 5 minutes.

Phase 1 comprised five instructional videos and four organizational prompts (see Table 1). In the introduction video, apprentices learned why the diagnostic strategy was essential. Then apprentices learned from four instructional videos about the four diagnostic steps with organizational prompts following each video. These videos were narrated animations (Figure 3).

Figure 3

Screenshots of the instructional videos in German



Note. The top left picture shows the introduction to diagnostic strategy; the top right picture explains the T-IPO principle in step 1; the bottom left picture explains how to plan a measurement with an electrical circuit diagram in step 2; the bottom right picture gives an overview of the complete diagnostic cycle in step 4.

When designing the videos, we considered multimedia principles to prevent unproductive cognitive load (Mayer & Moreno, 2003). First, some information was given visually on screen and some was presented by a narrator (i.e., *modality principle*). Second, following the *coherence principle*, we excluded extraneous elements. Third, we added visual cues to guide the apprentices' attention towards relevant elements (i.e., *signaling principle*, Figure 3, bottom left, green boxes). Thereby, we addressed one of the experts' concerns from the formative evaluation. Also, we chose an integrated representation for all visualizations, thereby avoiding a *split-attention effect* (Mayer & Moreno, 2003).

The four organizational prompts were similar to the *organizational prompts* that Roelle et al. (2017) recommend when a preceding instructional phase is included in example-based learning. These prompts support organizational processing of the learning content. After each organizational prompt, the apprentices received the correct solution.

Table 1

Instructional Elements, Contents, and Duration in Learning Phase 1

Element	Content	Duration in min
Introductory video	Introduction to diagnostic strategy: Why is the strategy important?	01:14
Instructional video	Step 1 of diagnostic strategy: Making an assumption with the T-IPO principle	06:52
Organizational prompt & solution	Select correct functions for five given components according to the T-IPO principle	01:30
Instructional video	Step 2 of diagnostic strategy: Planning a measurement with the three W-questions	05:14
Organizational prompt & solution	Plan a measurement by answering the W-questions for three given components	05:00
Instructional video	Step 3 of diagnostic strategy: Carrying out the measurement	00:27
Organizational prompt & solution	Recall the three steps of the diagnostic process that were presented so far	01:30
Instructional video	Step 4 of diagnostic strategy: Evaluating the measurement results and the assumption	02:46
Organizational prompt & solution	List other possible assumptions after a previously tested assumption could not be confirmed	06:00
Total video duration in learning phase 1		16:33
Estimated duration to complete learning phase 1		35:00

Learning phase 2 comprised two modeling examples of an expert applying the diagnostic strategy. When designing these modeling examples, we considered instructional principles of example-based learning (Renkl, 2014). First, in line with the *meaningful building blocks principle*, we presented the expert's diagnostic process with one video per diagnostic step (e.g., Schmidt-Weigand et al., 2009). Moreover, after each video, we provided a prompt that asked apprentices to self-explain the previous or the next diagnostic step (i.e., *self-explanation principle*, Renkl, 2014)). In the first round of the diagnostic process (i.e., when the *first* assumption was made, tested, and evaluated), apprentices answered the prompts by filling out cloze texts. In the following rounds (i.e., assumptions 2 and 3), apprentices received possible beginnings of sentences to support them in formulating their answers. Thereby, we addressed another concern of the experts from the formative evaluation while adhering to the *fading principle* (Renkl, 2014).

In the first modeling example, the engine power was too low. The cause was an electrical interruption in the signal line of the boost pressure valve. The model expert found this malfunction on the third attempt. That is, he first made and tested two other assumptions, which turned out to be wrong. The estimated time required to complete this modeling example comprising 12 videos and 10 self-explanation prompts was about 50 minutes, with the videos taking 25:50 min.

In the second modeling example, a car no longer started, although the starter motor was still turning. The cause for this malfunction was the interruption of a wire between the RPM sensor and the engine control unit. Here, the expert correctly diagnosed this cause for the malfunction in the second attempt. This modeling example consisted of 10 videos and 8 self-explanation prompts. The videos lasted 19:37 min. The estimated time for completion of this modeling example was 40 minutes. Taken together, the entire learning phase 2 took about 90 minutes. Hence, with learning phase 1 and a 5-minute break, the total intervention added up to approximately 130 minutes.

Testing Materials

This study investigated the apprentices' diagnostic knowledge and skills, motivation, and cognitive load while learning in the learning environment. In addition, the apprentices subjectively evaluated the learning environment.

Diagnostic Strategy Tests

We used three different diagnostic strategy tests. First, the *strategy description test* measured diagnostic knowledge and consisted of two questions that asked apprentices to (1) describe their diagnostic strategy in a situation in which they receive only minor support from a computer-based expert system, and (2) how they could narrow down which components might be responsible for a malfunction. The first question aimed at the four diagnostic steps. The second question referred to the T-IPO principle. The maximum score for this test was ten points.

Second, in the *strategy completion test*, apprentices performed or described (parts of) steps of the diagnostic strategy in four different scenarios. Hence, this test assessed scaffolded diagnostic skills. We developed two scenarios focusing on the first diagnostic step and one scenario for the second and fourth step. Within these scenarios, closed and open questions were used. The former dealt, for example, with which diagnostic step should be carried out next in the current scenario. In the open-ended questions, the apprentices, for example, studied a circuit diagram and described an appropriate measurement. In this test, participants could score a maximum of 47 points.

Third, to test diagnostic skills, participants performed a *diagnosis in the computer simulation*. They worked in a scenario in which a customer complained about a constantly burning engine control lamp. When apprentices had finished their diagnosis, they described the cause of the malfunction and how it could be repaired. This description was then scored with up to four points.

We also analyzed apprentices' diagnostic behavior by investigating their log-files. We defined critical information and test behavior for the scenario (Abele & von Davier, 2019). The critical *test behavior* included all meaningful measurements: (1) checking the power supply of the exhaust gas recirculation valve, (2) checking the resistance of the exhaust gas recirculation valve, (3) checking the signal that controls the component (this check could be done in two different ways, hence a division into 3 and 3b), and (4) checking the ground/signal line between the component and the engine control unit, which was the decisive measurement to find the cause of the malfunction. The critical *information behavior* included opening (1) the work schedule, (2) the electrical circuit diagram, (3) the installation plan, and (4) the reference values for measurements 1 to 3. For the decisive measurement 4, no reference values were available. If apprentices showed a specific behavior, they received a score of 1. It was not relevant when or how often a behavior was performed. Nine behaviors resulted in 9 dichotomous variables (behavior shown or not shown).

Motivational Factors

Before apprentices diagnosed in the simulation, we assessed the apprentices' current motivation based on the four factors (cf. Vollmeyer & Rheinberg, 2000) with a 19-item questionnaire on a 7-point Likert-scale. Five items assessed the apprentices' self-efficacy regarding the following diagnosis in the simulation (Cronbach's $\alpha = 0.90$; Bandura, 2006). Five items related to the apprentices' interest in car diagnosis and diagnostic strategies (Cronbach's $\alpha = 0.86$). More specifically, three items related to emotion-related valences (i.e., whether an apprentice associates positive emotions with car diagnoses) and two items related to value-related valences (i.e., whether an apprentice ascribes personal significance to a topic) (Schiefele, 1991). Similar self-efficacy and interest items have been used successfully in previous studies (e.g., Glogger-Frey et al., 2015). Eventually, four items related to the extent to which the apprentices perceived the upcoming diagnosis in the simulation as a challenge (Cronbach's $\alpha = 0.87$) and five items related to whether they perceived incompetence fear (Cronbach's $\alpha = 0.93$).

Cognitive Load

We asked the apprentices to assess their cognitive load while learning on a seven-point Likert-scale. We used an instrument that distinguishes between intrinsic (two items), germane (two items), and extraneous cognitive load (three items; Klepsch et al., 2017). Reliability was acceptable (intrinsic load: Cronbach's $\alpha = 0.79$; germane load: Cronbach's α

= 0.84; extraneous load: Cronbach's α = 0.66).

Subjective Evaluation Items

At the end of Session 2, apprentices evaluated eight characteristics of the intervention and the strategy taught by answering fifteen closed questions on a 7-point Likert-scale (Table 2). For seven of these characteristics, two items were used. We tested whether scales could be formed from these item pairs, but as reliability was low for some of the pairs, we decided to report all items separately (applicability: Cronbach's α = 0.78; interestingness: Cronbach's α = 0.51; length: Cronbach's α = 0.38; structure: Cronbach's α = 0.29; narrator quality: Cronbach's α = 0.63; comprehensibility: Cronbach's α = 0.42; recommendation: Cronbach's α = 0.82).When developing these items, we followed other studies that dealt with the development and evaluation of computer-based learning environments (e.g., Glogger et al., 2013; Hilbert et al., 2008).

Table 2

Subjective Evaluation Items

Dimension	Question
Applicability	1. The content of the learning material will help me diagnose in practice.
Applicability	2. I can now put the diagnostic strategy presented into practice.
Interestingness	3. The learning material was interesting.
Interestingness	4. The learning material was boring.
Length	5. I would have liked to work on more learning tasks.
Length	6. The learning material was too long.
Structure	7. I was confused by the structure of the learning material.
Structure	8. I found the structure of the learning material useful.
Narrator quality	9. The narrator spoke too fast.
Narrator quality	10. I could follow the narrator well.
Comprehensibility	11. The content of the learning material was too hard for me.
Comprehensibility	12. I understood the content of the learning material well.
Recommendation	13. The diagnostic strategy should be taught to all apprentices.
Recommendation	14. I would recommend the learning material to other apprentices.
Satisfactions with learning progress	15. I am satisfied with my learning progress.

We also asked apprentices for further comments in two questions (1) “Were there things that you found difficult to deal with while learning?” and (2) “What ideas do you have on how the learning material could be further improved?”.

Procedure

Data were collected with a page-based online survey tool. We did not limit the time participants could spend on a page in the survey. However, we gave participants a time window for the upcoming phase at the beginning of each phase. After this time window, we told participants to move on, regardless of whether they had completed the phase or not (see Table 3 and 4 for the procedure of the two sessions).

Table 3

Procedure of Session 1

Phase	Content	Average duration in min	Maximum duration in min until termination
Phase 1	Introduction to study, informed consent, introduction to computer simulation, and measurement exercises	30	45
Break		10	
Phase 2	Questionnaire of motivation regarding diagnosis in simulation	5	50
	Strategy description test	10	
	Diagnosis in simulation (limited to 33 minutes)	25	
Break		10	
Phase 3	Partial competence test [a] (Abele et al., 2014)	40	85
	Test to assess the apprentices' diagnosis-related reading competencies [a] (Norwig et al., 2021)	10	
	Strategy completion test	25	
Total duration session 1			165 – 200

[a] Not a repeated measure. As explained, in this paper, we present only a part of a more extensive study. We focus the present analyses on repeated measures.

In session 2, apprentices completed the intervention as described in the description of the final intervention. This session was also divided into three phases (Table 4).

Table 4

Procedure of Session 2

Phase	Content	Average duration in min	Maximum duration in min until termination
Phase 1	Refresher on computer simulation	5	150
	Intervention (see description of final intervention)	130	
Break		10	
Phase 2 Posttest	Questionnaire of motivation and self-efficacy regarding diagnosis in simulation	5	75
	Strategy description test	10	
	First diagnosis in simulation (limited to 33 minutes)	25	
	Second diagnosis in simulation (limited to 33 minutes) [a]	25	
Break		10	
Phase 3 Posttest	Strategy completion test	25	30
	Subjective evaluation of the learning materials	5	
Total duration session 2			250 – 270

[a] Not a repeated measure. As explained, in this paper, we present only a part of a more extensive study. We focus the analyses on repeated measures.

Data Analysis

Scoring

The strategy description test, strategy completion test, and diagnosis in the simulation consisted of closed and open question items. The first author and a subject matter expert (i.e., the third author) developed a coding scheme for open questions. The first author and a student assistant then independently scored 25% of all answers and adjusted the coding schemes until an interrater reliability of Cohen's Kappa > 0.8 was achieved. Then the student assistant independently scored the remaining answers. The remaining testing materials, including the log-files, were scored automatically.

Analysis

Occasionally, responses were not saved in the survey tool. This resulted in $n = 45$ participants for the strategy completion test and the diagnosis in the simulation, $n = 46$ participants for the strategy description test and the motivation items, $n = 48$ participants for the subjective evaluation items, and $n = 49$ participants for the cognitive load items. Scores on the variables measured in sessions 1 and 2 (i.e., the three diagnostic strategy tests and the motivation items) were compared between sessions 1 and 2 using paired samples t-tests. Descriptive data will be reported for all other variables that were only measured in session 2 (i.e., cognitive load and subjective evaluation items).

The asymptotic McNemar test was performed for the log-file analyses, which is recommended for paired dichotomous variables (Fagerland et al., 2014). Here, only data from $n = 39$ apprentices were available since they generated their participant codes based on a predefined scheme. In several cases, these codes mismatched between the simulation and the survey tool.

All effects are reported as significant at $p < .05$. Cohen's d, where 0.2, 0.5, and 0.8 correspond to small, medium, and large effects, was used as effect size for the t-tests (Cohen, 1992).

We also analyzed whether apprentices watched the videos completely (timing data was available for $N = 41$ apprentices). We calculated the relative watching time as the time spent on a page of the survey divided by the length of the respective video. Rewatching or pausing parts of the videos could therefore result in values larger than 100%. While apprentices spent $M = 102\%$ ($SD = 34\%$) of the allotted time with the instructional videos and $M = 101\%$ ($SD = 23\%$) with the first modeling example, the relative proportion decreased to $M = 53\%$ ($SD = 29\%$) for the second modeling example. Twenty-five apprentices did at least use 80% of the allotted time. We conducted separate analyses for these 25 apprentices. However, as these analyses did not lead to different conclusions, these results will not be reported.

Results

Diagnostic Strategy Tests

Table 5 presents the descriptive statistics for the pretest and posttest on diagnostic knowledge and skill tests. The test scores were significantly higher in the posttest than in the pretest for the strategy description test, $t(45) = -3.91$, $p < .001$, Cohen's $d = 0.58$ (medium effect) and the strategy completion test, $t(44) = -9.53$, $p < .001$, Cohen's $d = 1.42$ (large effect). This was not the case for the diagnosis in the simulation, $t(44) = -0.11$, $p = .913$, Cohen's $d = 0.016$ (no statistically significant effect).

Table 5

Descriptive Statistics of the Three Strategy Tests in Pre- and Posttest

Measure	Pretest			Posttest	
	N	*M*	*SD*	*M*	*SD*
Strategy description test (0-10 points)	46	1.48	0.89	2.80	2.29
Strategy completion test (0-47 points)	45	16.44	5.86	25.24	4.75
Diagnosis in simulation (0-4 points)	45	1.22	1.22	1.24	1.25

Regarding the log-file analyses, Table 6 shows the frequency of the different information and test behaviors in the diagnosis in the simulation for pretest and posttest. The McNemar's chi-square test statistic in the right column shows significant differences between pretest and posttest only in the information behavior. More specifically, the number of apprentices researching reference values increased between pretest and posttest from 6 to 31 for measurement 1 and from 5 to 16 for measurement 2.

Table 6

Crosstabs of Absolute Scores on Critical Information and Test Behavior in Pretest and Posttest and McNemar's Chi-Square Test Statistics.

Problem-solving behavior					
Information behavior		Not shown in posttest	Shown in posttest	Cumulative *n*	McNemar's asymptotic *p*
Open work order	Not shown in pretest	1	3	4	
	Shown in pretest	0	35	35	
	Cumulative *n*	1	38	39	.248
Open electrical circuit diagram	Not shown in pretest	4	3	7	
	Shown in pretest	4	28	32	
	Cumulative *n*	8	31	39	1.00
Open installation plan	Not shown in pretest	1	5	6	
	Shown in pretest	5	28	33	
	Cumulative *n*	6	33	39	1.00
Research reference values for measurement 1	Not shown in pretest	17	16	33	
	Shown in pretest	2	4	6	
	Cumulative *n*	19	20	39	**.002**
Research reference values for measurement 2	Not shown in pretest	20	14	34	
	Shown in pretest	3	2	5	
	Cumulative *n*	23	16	39	**.015**
Research reference values for measurement 3	Not shown in pretest	27	7	34	
	Shown in pretest	3	2	5	
	Cumulative *n*	30	9	39	.343

Test behavior		Not shown in posttest	Shown in posttest	Cumulative *n*	McNemar's asymptotic *p*
Perform measurement 1	Not shown in pretest	20	6	26	
	Shown in pretest	8	5	13	
	Cumulative *n*	28	11	39	.789
Perform measurement 2	Not shown in pretest	14	4	18	
	Shown in pretest	9	12	21	
	Cumulative *n*	23	16	39	.267
Perform measurement 3	Not shown in pretest	23	8	31	
	Shown in pretest	6	2	8	
	Cumulative *n*	29	10	39	.789
Perform measurement 3b	Not shown in pretest	38	0	38	
	Shown in pretest	1	0	1	
	Cumulative *n*	39	0	39	1.00
Perform measurement 4 (decisive measurement)	Not shown in pretest	21	8	29	
	Shown in pretest	3	7	10	
	Cumulative *n*	24	15	39	.227

Motivational Factors

Table 7 shows apprentices' motivation and self-efficacy ratings regarding the diagnosis in the simulation. Paired samples t-tests revealed that the declines between pre- and posttest were significant for the apprentices' responses about their interest in the diagnosis, $t(45) = -3.58$, $p = .001$, Cohen's $d = 0.53$ (medium effect) and their perception of challenge, $t(45) = 5.45$, $p < .001$, Cohen's $d = 0.80$ (large effect), but not for self-efficacy, $t(45) = 0.25$, $p = .804$, or incompetence fear, $t(45) = 1.20$, $p = .236$.

Table 7

Descriptive Data on Apprentice Self-efficacy and Motivation

Measure	Pretest		Posttest	
	M	*SD*	*M*	*SD*
Self-efficacy regarding diagnosis	4.53	1.14	4.49	1.26
Interest in diagnosis	5.16	1.06	4.61	1.30
Perception of challenge	4.86	1.17	4.03	1.43
Incompetence fear	2.95	1.58	2.74	1.51

Cognitive Load

After learning phase 2, participants assessed their cognitive load during the learning phases. Intrinsic load was rated close the scale midpoint of 4 ($M = 3.93$, $SD = 1.64$). The germane load was rated slightly above ($M = 4.44$, $SD = 1.79$); the extraneous load was rated below the scale midpoint ($M = 3.36$, $SD = 1.36$).

Subjective Evaluation Items

Descriptive data for the subjective evaluation items are displayed in Table 8. Considering the scale midpoint of 4, one structure item (#7), both speaker quality items (#9, #10), and both comprehensibility items (#11, #12) were evaluated rather positively. However, the apprentices would not have liked to work on more learning tasks (#5). The scores of the remaining items are close to the scale midpoint.

Table 8

Descriptive Data on Apprentice Responses to Subjective Evaluation Questions

Dimension	Question	*M*	*SD*
Applicability	1. The content of the learning material will help me diagnose in practice.	4.35	1.68
Applicability	2. I can now put the diagnostic strategy presented into practice.	4.41	1.65
Interestingness	3. The learning material was interesting.	4.33	1.69
Interestingness	4. The learning material was boring. [a]	3.41	1.72
Length	5. I would have liked to work on more learning tasks.	3.04	1.56
Length	6. The learning material was too long.	4.15	1.70
Structure	7. I was confused by the structure of the learning material. [a]	2.76	1.61
Structure	8. I found the structure of the learning material useful.	4.13	1.54
Narrator quality	9. The narrator spoke too fast. [a]	2.13	1.57
Narrator quality	10. I could follow the narrator well.	5.46	1.50
Comprehensibility	11. The content of the learning material was too hard for me.[a]	2.43	1.34
Comprehensibility	12. I understood the content of the learning material well.	5.04	1.48
Recommendation	13. The diagnostic strategy should be taught to all apprentices.	4.80	1.77
Recommendation	14. I would recommend the learning material to other apprentices.	4.35	1.70
Satisfactions with learning progress	15. I am satisfied with my learning progress.	4.48	1.46

[a] Negatively formulated item. Low scores mean better evaluation.

Moreover, in responses to the first open-ended evaluation question about aspects that apprentices found difficult, seven apprentices (15%) mentioned technical problems. These were related to the embedding of the computer simulation within the online survey. Three apprentices (6%) complained about too many similar questions (i.e., practice tasks and self-explanation prompts). On the second question asking if things could be improved, again six (13%) apprentices mentioned technical problems. Fifteen apprentices (30%) again complained about too many repetitions in the learning materials.

Discussion

Although diagnosing car malfunctions is a crucial skill for AMTs, most apprentices experience problems with complex diagnoses, presumably because they experience such situations too rarely. To expose apprentices to such situations, we developed a simulation-based learning environment building on modeling examples. The present study aimed to evaluate the learning environment by investigating four research questions (RQs).

Regarding RQ 1, the findings are mixed. The final evaluation study showed that the intervention promoted apprentices' diagnostic knowledge (strategy description test) and scaffolded diagnostic skills (strategy completion test). However, apprentices could not transfer this knowledge to independent diagnosis problems. Concerning motivation (RQ 2), interest in the diagnosis and the perception of the diagnosis as a challenge decreased. With only minor deviations, all types of cognitive load were rated near the scale midpoint (RQ 3). Regarding apprentice subjective evaluations (RQ 4), the intervention was mainly rated neutral to slightly positive, except for its length. These findings will be discussed in the following paragraphs. Potentials for improvement will then be discussed in an integrated manner.

First (RQ 1), the intervention enabled apprentices to describe the steps of malfunction diagnoses (strategy description test) and to determine the next steps to be taken in various given scenarios (strategy completion test). This beneficial effect of the intervention corresponds with the analysis of the apprentices' self-reported cognitive load (RQ 3), as the apprentices reported a relatively low extraneous cognitive load, which is an essential factor for successful learning from multimedia (Mayer & Moreno, 2003). However, despite this increase in diagnostic knowledge and scaffolded diagnostic skills, the apprentices were not more successful when independently performing a diagnosis in the simulation. That is, the newly learned strategy did not pay off immediately. This is actually a common finding (Hübner et al., 2010): newly learned strategies need practice. One explanation is an *utilization deficiency* (Miller, 1994). Applying a new strategy requires so much cognitive capacity that only little capacity remains to process the new problem-solving case (i.e., the malfunction scenario in the simulation). Consequently, the independent application of the strategy to the novel context fails.

Regarding RQ 2, we found that participants' interest and their perception of diagnosis as a challenge decreased after working in the learning environment. A decrease of interest due to increasing knowledge is not uncommon. In three studies, Rotgans and Schmidt (2014) found that students' current interest in various topics covered in geography and history lessons decreased after learning. Probably, after learning learners do not perceive a knowledge gap about these topics anymore, thereby diminishing their interest in the topic. Besides, apprentices still rated their interest above the scale midpoint of four in the post-test session. Thus, it can be assumed that the apprentices' interest did not decrease to such an extent that it would have become detrimental to the apprentices' learning success.

We also found that apprentices perceived diagnosing car malfunctions as less challenging after the intervention. Multiple learning theories, such as Csikszentmihalyi's flow theory (1990) or Vygotsky's zone of proximal development (Schnotz et al., 2009) emphasize the importance of an appropriate level of challenge for learners. However, it must be remembered that, as in the case of interest, apprentices rated their perception of challenge above the midpoint of the scale in the posttest. Thus, despite the average decline, an appropriate level of challenge can still be assumed. In summary, neither the decrease in interest nor the decrease in the perception of challenge seems to be problematic. What must be noted, however, is that while on average interest and challenge have decreased, of 46 apprentices, 13 indicated an increased interest and 7 apprentices indicated an increased perception of challenge after the intervention. Why there is such a substantial variation in the development of these two motivational factors could be investigated in future studies.

Regarding RQ 4, most items seem to be evaluated mainly neutral to slightly positive. This means that the apprentices found the diagnostic strategy mostly applicable, the learning material interesting and comprehensible, they understood the structure of the learning material well, they could follow the speaker well, and they would recommend the diagnostic strategy and the learning material to other apprentices. However, since no scales could be constructed from the respective subjective evaluation item pairs due to poor reliabilities, these results must be interpreted cautiously.

The only item that scored lower than mid-scale was item #5, which asked whether the apprentices would have liked to work on more tasks. However, the apprentices completed the subjective evaluation at the very end of session 2, that is, after approximately 130 minutes of intervention and approximately 90 minutes of posttesting. So the rather poor evaluation of item #5 may be related to the extensive posttest, especially as item #5 asked if the apprentices would have liked to work on more learning tasks. Possibly, apprentices were not able to distinguish between learning tasks from the intervention and posttest tasks. Nevertheless, the interpretation that the apprentices might have found the intervention somewhat too long is supported by the watching time of the videos: While the apprentices watched the first modeling example (more or less) completely, they only used 53% of the allotted time for the second modeling example.

Moreover, in the open evaluation items, apprentices remarked that they considered the self-explanation prompts to be highly repetitive. To properly interpret these comments, it should be remembered how the self-explanation prompts were used. In total, apprentices were asked 18 times with the same prompt to self-explain a diagnostic step. Although an analysis of the responses showed that apprentices still gave mostly meaningful answers, answer quality decreased over time as it became worse for the second modeling example.

Potential for improvement

In summary, the evaluation study identified the following potential for improvement. First, apprentices could not apply their knowledge about diagnostic strategies when diagnosing a case in the simulation independently. Second, participants considered the intervention too long and repetitive. Third, together with the decreasing watching time, the decreasing answer quality to the self-explanation prompts in the second modeling example indicates that the apprentices became increasingly passive as the intervention progressed.

To address these potentials for improvement, we should first consider shortening the intervention. Discussions with experts on this topic resulted in the recommendation to focus the strategy on the first two steps, in which model building primarily takes place. Moreover, stretching the entire intervention over a longer period with several but shorter sessions might be beneficial. Apprentices would not only become less passive during a session but also get opportunities to practice the new strategy. For example, we could let the apprentices diagnose several malfunctions in the simulation between the modeling examples. Within these practices, we could provide the apprentices with scaffolds (e.g., hints which diagnostic step needs to be performed next), another core teaching method of cognitive apprenticeship besides modeling (Collins et al., 1988). These scaffolded practices could then allow apprentices to apply and practice only parts of the new strategy at once, thereby helping to overcome the utilization deficiency (Miller, 1994). Furthermore, with multiple sessions, we could also conduct a delayed posttest to be able to identify possible long-term effects on transfer (c.f., Hübner et al., 2010).

A final result of the evaluation study was that the apprentices found the self-explanation prompts repetitive and answered them less thorough over time. Thus, it might be helpful to vary the prompts by providing prompts that aim at the target learning processes but read very differently for learners (Nückles et al., 2020). Following an informed training approach, it might also be beneficial to explain why engaging in self-explanations would promote their learning (e.g., Hübner et al., 2010). In summary, these identified shortcomings concerning the learning environment could be addressed as follows. First, the learning contents could be reduced by condensing the diagnostic strategy. Second, an intervention over a longer period with shorter sessions would allow to provide apprentices with practice opportunities and conduct a delayed posttest. Finally, we could vary the self-explanation prompts.

Limitations

A few points about our research design should be mentioned. First, an investigation including a control group not receiving any elements of the present intervention (e.g., no instructional videos and no modeling examples) would be helpful to precisely quantify the effects of our developed learning environment on learning outcomes. Second, the sample size might appear small at first glance. Nevertheless, a post-hoc power analysis performed with Gpower 3.1 (Faul et al., 2007) indicated that for all t-tests that revealed significant results, the statistical power was at least 0.89. Third, classroom characteristics could affect the effectiveness of our intervention. For example, apprentices in some classes may not work concentrated. Such apprentices might then distract their classmates so that the entire class would work less concentrated than a class without disturbances. Accordingly, results from single classes might not be generalizable to all classes. We performed additional ANOVAs with class membership as the independent variable to determine possible effects of the apprentices' school class. We only found minor differences between two classes from one school regarding their subjective evaluations and cognitive load ratings. However, we did not find these differences in the diagnostic strategy tests. Thus, the learning effects do not differ across classes and seem to be thus of second-order importance.

Conclusion

In summary, our approach to developing a simulation-based learning environment with modeling examples to teach AMT apprentices a diagnostic strategy has taken first successful steps. Our structured approach to development, the multiple consultations of subject matter experts, and the piloting of large parts of the learning material are particularly noteworthy. Thus, this paper also provides possible guidance for others who want to develop a similar learning environment. The current version of the intervention already promotes apprentice performance on the strategy description test and strategy completion test. With the modifications described above, we expect that the intervention will also promote apprentice performance in the diagnosis in the simulation. Ultimately, we will provide vocational schools with an updated version of our intervention together with the simulation. This will be a valuable contribution to ensure that at the end of their apprenticeship, AMT apprentices will master strategies to independently diagnose more complex car malfunctions.

Acknowledgments

This work was supported by the Federal Ministry of Education and Research, Germany, and the Federal Institute for Vocational Education and Training, Germany, within the project “DigiDIn-Kfz – Digitale Diagnostik und Intervention im Kfz-Wesen” [Digital diagnostics and intervention in the automotive sector], which is a subproject of the research and transfer initiative “ASCOT+ - Technology-based Assessment of Skills and Competences in Vocational Education and Training” [grant number 21AP004C].

References

Abele, S. (2018). Diagnostic problem-solving process in professional contexts: Theory and empirical investigation in the context of car mechatronics using computer-generated log-files. *Vocations and Learning*, *11*(1), 133–159. https://doi.org/10.1007/s12186-017-9183-x

Abele, S., & von Davier, M. (2019). CDMs in vocational education: Assessment and usage of diagnostic problem-solving strategies in car mechatronics. In M. von Davier & Y.-S. Lee (Eds.), *Handbook of Diagnostic Classification Models*. Springer. https://doi.org/10.1007/978-3-030-05584-4_22

Abele, S., Walker, F., & Nickolaus, R. (2014). Zeitökonomische und reliable Diagnostik beruflicher Problemlösekompetenzen bei Auszubildenden zum Kfz-Mechatroniker [Time-saving and reliable diagnostics in measuring professional problem-solving competence in the domain of car mechatronics]. *Zeitschrift für Pädagogische Psychologie*, *28*(4), 167–179. https://doi.org/10.1024/1010-0652/a000138

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

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. C. Urdan (Eds.), *Self-Efficacy Beliefs of Adolescents* (pp. 307–337). IAP - Information Age Pub., Inc.

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Review of Educational Research*, *90*(4), 499–541. https://doi.org/10.3102/0034654320933544

Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (Fourth edition). Wiley.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Collins, A., Brown, J. S., & Newman, S. E. (1988). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. *Thinking: The Journal of Philosophy for Children*, *8*(1), 2–10. https://doi.org/10.5840/thinking19888129

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design Research: Theoretical and Methodological Issues. *Journal of the Learning Sciences*, *13*(1), 15–42. https://doi.org/10.1207/s15327809jls1301_2

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience* (1st ed). Harper & Row.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1990). Medical problem solving: A ten-year retrospective. *Evaluation & the Health Professions*, *13*(1), 5–36. https://doi.org/10.1177/016327879001300102

Fagerland, M. W., Lydersen, S., & Laake, P. (2014). Recommended tests and confidence intervals for paired binomial proportions. *Statistics in Medicine*, *33*(16), 2850–2875. https://doi.org/10.1002/sim.6148

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Glogger, I., Holzäpfel, L., Kappich, J., Schwonke, R., Nückles, M., & Renkl, A. (2013). Development and evaluation of a computer-based learning environment for teachers: Assessment of learning strategies in learning journals. *Education Research International*, *2013*, 1–12. https://doi.org/10.1155/2013/785065

Glogger-Frey, I., Fleischer, C., Grüny, L., Kappich, J., & Renkl, A. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct

instruction. *Learning and Instruction*, *39*, 72–87. https://doi.org/10.1016/j.learninstruc.2015.05.001

Gschwendtner, T., Abele, S., & Nickolaus, R. (2009). Computersimulierte Arbeitsproben: Eine Validierungsstudie am Beispiel der Fehlerdiagnoseleistungen von Kfz-Mechatronikern [Can troubleshooting skills of car mechatronic technicians validly be assessed using computer-based simulations of real work sample?]. *Zeitschrift Für Berufs- Und Wirtschaftspädagogik*, *105*(4), 557–578.

Heitzmann, N., Fischer, F., Kühne-Eversmann, L., & Fischer, M. R. (2015). Enhancing diagnostic competence with self-explanation prompts and adaptable feedback. *Medical Education*, *49*(10), 993–1003. https://doi.org/10.1111/medu.12778

Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Böhmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Böhmer, I., Karing, C., Kaiser, J., & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, *76*, 181–193. https://doi.org/10.1016/j.tate.2017.12.001

Hilbert, T. S., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning and Instruction*, *18*(1), 54–65. https://doi.org/10.1016/j.learninstruc.2006.10.008

Hübner, S., Nückles, M., & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. *Learning and Instruction*, *20*(1), 18–29. https://doi.org/10.1016/j.learninstruc.2008.12.001

Kalyuga, S. (2011). Cognitive Load Theory: How many types of load does it really need? *Educational Psychology Review*, *23*(1), 1–19. https://doi.org/10.1007/s10648-010-9150-7

Klein, M., Otto, B., Fischer, M. R., & Stark, R. (2019). Fostering medical students' clinical reasoning by learning from errors in clinical case vignettes: Effects and conditions of additional prompting procedures to foster self-explanations. *Advances in Health Sciences Education*, *24*(2), 331–351. https://doi.org/10.1007/s10459-018-09870-5

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*, 1997. https://doi.org/10.3389/fpsyg.2017.01997

Kluwe, R. H., & Haider, H. (1990). Modelle zur internen Repräsentation komplexer technischer Systeme [Models for the internal representation of complex technical systems]. *Sprache & Kognition*, *9*(4), 173–192.

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 279–315). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.015

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Merriënboer, J. J. G. van, & Kirschner, P. A. (2018). *Ten steps to complex learning: A systematic approach to four-component instructional design* (Third edition). Routledge.

Miller, P. H. (1994). Individual differences in children's strategic behaviors: Utilization deficiencies. *Learning and Individual Differences*, *6*(3), 285–307. https://doi.org/10.1016/1041-6080(94)90019-1

Nickolaus, R., Abele, S., Nitzschke, A., & Greiff, S. (2012). Fachspezifische Problemlösefähigkeit in gewerblich-technischen Ausbildungsberufen. Modellierung, erreichte Niveaus und relevante Einflussfaktoren [Domain-specific problem solving ability in technical apprenticed professions. Modeling, achieved level, and important determinants]. *Zeitschrift Für Berufs- Und Wirtschaftspädagogik*, *108*(2), 243–277. http://hdl.handle.net/10993/3188

Nückles, M., Roelle, J., Glogger-Frey, I., Waldeyer, J., & Renkl, A. (2020). The self-regulation-view in writing-to-learn: Using journal writing to optimize cognitive load in self-regulated learning. *Educational Psychology Review*, *32*(4), 1089–1126. https://doi.org/10.1007/s10648-020-09541-1

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, *38*(1), 1–37. https://doi.org/10.1111/cogs.12086

Rheinberg, F., Vollmeyer, R., & Burns, B. D. (2000). Motivation and Self-Regulated Learning. In J. Heckhausen (Ed.), *Motivational psychology of human development: Developing motivation and motivating development* (Vol. 131, pp. 81–108). Elsevier. https://doi.org/10.1016/S0166-4115(00)80007-2

Roelle, J., Hiller, S., Berthold, K., & Rumann, S. (2017). Example-based learning: The benefits of prompting organization before providing examples. *Learning and Instruction*, *49*, 1–12. https://doi.org/10.1016/j.learninstruc.2016.11.012

Rotgans, J. I., & Schmidt, H. G. (2014). Situational interest and learning: Thirst for knowledge. *Learning and Instruction*, *32*, 37–50. https://doi.org/10.1016/j.learninstruc.2014.01.002

Schaafstal, A., Schraagen, J. M., & van Berl, M. (2000). Cognitive task analysis and innovation of training: The case of structured troubleshooting. *Human Factors: The*

Journal of the Human Factors and Ergonomics Society, *42*(1), 75–86. https://doi.org/10.1518/001872000779656570

Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist*, *26*(3–4), 299–323. https://doi.org/10.1080/00461520.1991.9653136

Schmidt-Weigand, F., Hänze, M., & Wodzinski, R. (2009). Complex problem solving and worked examples: The role of prompting strategic behavior and fading-in solution steps. *Zeitschrift Für Pädagogische Psychologie*, *23*(2), 129–138. https://doi.org/10.1024/1010-0652.23.2.129

Schnotz, W., Fries, S., & Horz, H. (2009). Motivational aspects of cognitive load theory. In M. Wosnitza, S. A. Karabenick, A. Efklides, & P. Nenninger (Eds.), *Contemporary motivation research: From global to local perspectives* (pp. 66–69). Hogrefe & Huber Publishers.

Spöttl, G., Becker, M., & Musekamp, F. (2011). Anforderungen an Kfz-Mechatroniker und Implikationen für die Kompetenzerfassung [Requirements for automotive mechatronics technicians and implications for the assessment of competencies]. In R. Nickolaus (Ed.), *Lehr-Lernforschung in der gewerblich-technischen Berufsbildung (Zeitschrift für Berufs- und Wirtschaftspädagogik: Sonderbd. 25)* (pp. 37–53). Steiner.

Sweller, J., Ayres, P. L., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

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 Merriënboer, J. J. G. (2013). Perspectives on problem solving and instruction. *Computers & Education*, *64*, 153–160. https://doi.org/10.1016/j.compedu.2012.11.025

Vollmeyer, R., & Rheinberg, F. (1998). Motivationale Einflüsse auf Erwerb und Anwendung von Wissen in einem computersimulierten System [Motivational influences on the acquisition and application of knowledge in a simulated system]. *Zeitschrift Für Pädagogische Psychologie*, *12*(1), 11–23.

Vollmeyer, R., & Rheinberg, F. (2000). Does motivation affect performance via persistence? *Learning and Instruction*, *10*(4), 293–309. https://doi.org/10.1016/S0959-4752(99)00031-6

Appendix – Pilot Study

Procedure

The procedure of the pilot study is displayed in the following Table A.1.

Table A.1

Procedure of Pilot Study

Element	Content	Duration in min
Introduction	Demographical Data	03:30
Pretest	Strategy description test	07:30
	Evaluation of strategy description test	01:00
	Strategy completion test	34:00
	Evaluation of strategy completion test	02:30
Intervention	Instructional videos & practice tasks	32:00
	Self-assessment of cognitive load during learning	01:30
	Evaluation of learning materials	02:30
Posttest	Strategy description test	03:00
	Evaluation of strategy description test	00:30
	Strategy completion test	10:30
	Evaluation of strategy completion test	01:00
Total duration of pilot study		99:30

Evaluation Results

In the following paragraphs, we will describe the evaluations of the different tests and the learning materials.

Strategy Description Test

The strategy description test was evaluated with an open-ended item. Three of five participants answered this item in the pretest (i.e., before watching the instructional videos), their answers were:

- "with the last task, I was uncertain whether it was meant which defects lead exactly to errors, thus individual components, internal defect, or cable resistances too high, etc., or whether it was asked which systems produce an entry in the fault memory if they do not work correctly."
- "It would be easier to explain if, for example, a malfunction is given. Otherwise, you have to keep it pretty general and can't go into it in more detail."
- "Good and understandable. To solve the task, I would find it good if, for example, a circuit diagram would be available to refer to for the description."

None of the participants answered this evaluation item in the posttest (i.e., after watching the instructional videos).

Strategy Completion Test

The strategy completion test was evaluated with open and closed items. Table A.2 gives an overview of the apprentices' aggregated answers on the closed items (Likert-scale ranging from 1 = *absolutely not true* to 7 = *absolutely true*).

In the pretest, the two open items were answered as follows:

- "the answer options regarding the IPO principle drove me crazy, don't do such things at work at all"
- "Do not ask so many very similar questions"
- the first question on the T-IPO principle with the task. "Which of the five segments would you choose?" was not so easy to understand.

Only one apprentice answered one of the open items in the posttest. He stated, "Please do not ask the same questions so often".

Table A.2

Results for the Closed Evaluation Items Regarding the Strategy Completion Test

Evaluation dimension (aggregated)	Evaluation in Pretest		Evaluation in Posttest	
	M	*SD*	*M*	*SD*
Comprehensibility	4.57	1.10	4.52	1.30
Amount/length	4.24	1.47	3.52	0.98
Interestingness	5.36	1.11	4.00	1.38

Learning Materials

The learning materials (i.e., the instructional videos and the practice tasks) were also evaluated with closed and open evaluation items. Table A.3 gives an overview of the apprentices' aggregated answers on the closed items. Only one apprentice answered one of the open items. He stated, "Videos became rather boring after a while. Would be good to make them shorter."

Table A.3

Results for the Closed Evaluation Items Regarding the Strategy Completion Test

Evaluation dimension (aggregated)	Evaluation in Pretest	
	M	*SD*
Applicability	4.93	1.17
Interestingness	4.93	1.30
Length	3.36	0.80
Structure	5.14	0.75
Narrator	6.50	0.76
Comprehensibility	5.57	1.06
Recommendation	5.34	0.85
Satisfaction	4.71	1.70

Manuscript 2: Better Self-Explaining Backwards or Forwards? Video-Based Modelling Examples for Learning a Diagnostic Strategy and the Use of Different Self-Explanation Prompts

Meier, J., Hesse, P., Abele, S., Renkl, A., & Glogger-Frey, I. (2023). Better Self-Explaining Backwards or Forwards? Video-Based Modelling Examples for Learning a Diagnostic Strategy and the Use of Different Self-Explanation Prompts. *Manuscript Submitted for Publication.*

My first authorship reflects the fact that I was responsible for planning the studies and design the material, for the data analyses, and for the preparation of the paper.

Abstract

Self-explanation prompts in example-based learning are usually directed backwards: Learners are required to self-explain problem-solving steps just presented (*retrospective* prompts). However, it might also help to self-explain upcoming steps (*anticipatory* prompts). The effects of the prompt type may differ for learners with various expertise levels, with anticipatory prompts being better for learners with more expertise. In an experiment, we employed extensive modelling examples and different types of self-explanations prompts to teach 78 automotive apprentices a complex and job-relevant problem-solving strategy, namely the diagnosis of car malfunctions. We tested the effects of these modelling examples and self-explanation prompts on problem-solving strategy knowledge and skill, self-efficacy, and cognitive load while learning. In two conditions, the apprentices learned with modelling examples and received either retrospective or anticipatory prompts. The third condition was a control condition receiving no modelling examples, but the respective open problems. In comparison with the control condition, modelling examples did not promote learning. However, we observed differential effects of the self-explanation prompts depending on the learner's prior knowledge level. Apprentices with higher prior knowledge learned more when learning with anticipatory prompts. Apprentices with less prior knowledge experienced a greater increase in self-efficacy and a higher germane cognitive load when learning with retrospective prompts. These findings suggest using different self-explanation prompts for learners possessing varying levels of expertise.

Keywords. Example-based learning, modelling examples, self-explanation prompts, complex problem-solving, diagnostic strategy

Introduction

When learning a problem-solving strategy, learners are often first instructed about the strategy and then study worked-out solutions of problems that have been solved with the instructed strategy (VanLehn, 1996). Such exemplary solved problems can be text-based worked examples (e.g., Najar & Mitrovic, 2013) or video-based modelling examples (e.g., screencasts showing a model's action on a computer; van Gog & Rummel, 2010). Studying worked or modelling examples frees up cognitive capacities and is thus more beneficial for learning than independently practising to apply the instructed strategy to solve problems (worked or modelling example effect; McLaren & Isotani, 2011; Renkl, 2014; Sweller, 2006; van Gog et al., 2019; van Gog & Rummel, 2010). Examples are especially beneficial for novices (see expertise-reversal effect; Kalyuga & Renkl, 2010). Regarding (video-based) modelling examples, research has focused on modelling examples illustrating rather brief and simple problem-solving strategies (e.g., Fiorella et al., 2017; Hoogerheide, 2016; Hoogerheide et al., 2014, 2018; Schmitz et al., 2017). How modelling examples can be used to teach more complicated problem-solving strategies, such as how to diagnose car malfunctions (Abele, 2018; Abele & von Davier, 2019), has seldom been studied.

Examples alone do not necessarily promote learning as learners might not actually use the freed-up cognitive capacities for learning. Generative learning activities stimulated by, for example, self-explanation prompts ensure that these capacities are used for learning (Renkl & Eitel, 2019). So far, prompts usually ask learners to explain an example's *previous* contents (i.e., *retrospective self-explanation prompts*). Prompts targeting the *upcoming* contents of an example have hardly been investigated (Bisra et al., 2018). Such *anticipatory self-explanation prompts* are probably more cognitively demanding, but potentially more conducive to learning. Presumably, the learners' prior knowledge is a crucial prerequisite of whether they can manage the more demanding anticipatory prompts

Consequently, the present paper pursued two goals. First, we investigated the effectiveness of more complex and, therefore, longer video-based modelling examples for teaching a complex problem-solving strategy. Furthermore, we compared the effects of retrospective and anticipatory self-explanation prompts for learners possessing different levels of prior knowledge.

Cognitive Load Theory and Example-based Learning

The effects of worked examples and self-explanation prompts can be explained via the *Cognitive Load Theory* (CLT; Sweller et al., 1998, 2011). In this paper, we refer to the still widely used conception of CLT from 1998[1]. CLT assumes that working memory capacity is limited and that learning induces three distinct types of cognitive load on working memory: germane cognitive load (GCL), intrinsic cognitive load (ICL), and extraneous cognitive load (ECL; Sweller et al., 1998). If the sum of these three load types exceeds available working memory capacities, learning fails. GCL describes the working memory load resulting from learning-related activities. Such activities include, for example, *organizing* and *integrating* new information with existing prior knowledge (see SOI model; Fiorella & Mayer, 2016). ICL is determined by the learning material's complexity and the learner's (prior) knowledge. That is, more complex learning materials (i.e., learning materials with higher element interactivity) induce higher ICL. However, the more prior knowledge learners have about a learning topic, the lower the ICL they experience. If learners have prior knowledge of a topic, they already have cognitive schemas enabling them to combine multiple elements from the learning material and handle those as a single element in their working memory. Element interactivity, and thus ICL, decreases. The third type of cognitive load is ECL, which is unproductive and learning-unrelated. Learning materials containing irrelevant information,

[1] Recently, Sweller and colleagues have presented new GCL concepts (2019). However, we refer to the 1998 concept in this paper, as it is the basis for most of the research we refer to, and because we had this original concept in mind when developing the learning materials and experimental design.

redundant repetition, or numerous references induce higher ECL. Given the same task (i.e., same element interactivity) and the same learners (i.e., same prior knowledge), ICL is considered fixed. Therefore, to ensure that sufficient working memory resources are available for GCL, ECL should be minimized (e.g., Mayer & Moreno, 2003).

When learning how to solve a problem, novices usually apply ineffective problem-solving strategies and thus experience a high ECL (van Gog et al., 2019). Learning from worked or modelling examples avoids such ineffective strategies and thus reduces ECL. If the freed-up capacities are used for learning-related activities, GCL increases and learning is promoted: a *worked* or *modelling example* effect occurs (Renkl et al., 2009).

Besides beneficial effects on cognitive load and learning, learning with (modelling) examples (in comparison to more open learning formats like inventing, or independent problem solving) is also known to promote self-efficacy (Glogger-Frey et al., 2015; Hoogerheide et al., 2014, 2018; van Harsel et al., 2019). Self-efficacy describes how confident learners are in performing a specific task (Bandura, 1997). Observing how a model successfully solves a task can strengthen learners' confidence that they can perform the task as well (Bandura, 1997; Schunk, 1995). For example, van Harsel et al. (2019) investigated how different sequences of studying examples and problem solving would affect various motivational aspects including self-efficacy. They found that studying examples only resulted in greater self-efficacy than mere problem solving (van Harsel et al., 2019). Finally, self-efficacy exerts a strong influence on learning outcomes, as it positively affects academic motivation and learning behaviour, such as learning perseverance (Bandura, 1997; Multon et al., 1991; Schunk, 1995).

Modelling examples are known to benefit learning in various domains and settings. However, in most cases, such problem-solving strategies were comparatively simple and could be taught with shorter modelling examples (e.g., Fiorella et al., 2017: assembling an electrical circuit, 90 seconds; Hoogerheide, 2016; Hoogerheide et al., 2018: calculating current, voltage, and resistance, 240 seconds). We can assume that substantially longer examples than in earlier studies, namely those illustrating more complex problem-solving strategies, also reveal beneficial effects on learners‘ cognitive load, learning outcomes, and self-efficacy. However, to our knowledge, this assumption has hardly been investigated so far. We, therefore, aimed to replicate the worked or modelling example effect with video-based modelling examples for more complex problem-solving strategies.

Self-explanation Prompts

By reducing ECL, examples liberate working memory capacities. To ensure that these capacities are used for learning (i.e., ensuring that GCL increases), learners should engage in self-explanations (Hilbert & Renkl, 2009) which can be elicited with *self-explanation prompts* (Atkinson et al., 2003; Renkl et al., 1998). With such prompts, learners are explicitly asked to relate the content in the illustrative example to the problem-solving strategy explained in an earlier instruction. For example, Hilbert and Renkl (2009) used two paper-based worked examples to teach students a circular, three-step-process of concept mapping that had already been introduced (Hilbert & Renkl, 2008). While worked examples alone failed to promote learning (Hilbert & Renkl, 2009; experiment 1), the combination of worked examples and self-explanation prompts proved beneficial for learning (Hilbert & Renkl, 2009; experiment 2). The self-explanation prompts used in experiment 2 asked students to explain: ‘To which phase of the concept mapping process can you assign what Carolin/Karsten just did? Why?’ (Hilbert & Renkl, 2009, p. 271) with Carolin and Karsten being fictitious students in the examples.

In this case and in most studies, self-explanation prompts refer to aspects already shown in corresponding examples (e.g., Berthold et al., 2009; Hilbert et al., 2008; Klein et al., 2019). We refer to such backwards-directed prompts as *retrospective self-explanation* *prompts*. Another potentially successful type of self-explanation prompts is directed *forward:* in Renkl (1997), successful learners were, inter alia, those who thought about a problem's *upcoming* solution steps (*anticipative reasoning*). Consequently, *anticipatory self-explanation prompts*, that is, prompts referring to upcoming problem-solving steps could also be useful. Referring to the study by Hilbert and Renkl (2009), such an anticipatory prompt could be 'Which step of concept mapping comes next and what will Carolin/Karsten have to do?'

Anticipatory and retrospective prompts presumably induce different cognitive processes: When answering retrospective self-explanation prompts, learners have to consider only previous steps in the illustrated problem-solving strategy. Conversely, when answering anticipatory prompts, learners have to represent the problem-solving strategy's next step. However, these mental processes can only take place by relying on already-completed problem-solving steps. Consequently, when learning with anticipatory prompts, more elements (i.e., the prior and subsequent step) must be considered overall, but more relevant information also has to be organized and integrated (Fiorella & Mayer, 2016).

In CLT terms, this could mean two things for learners' cognitive load: First, anticipatory prompts might induce higher GCL than regular retrospective prompts, as learners are prompted to organize and integrate more information. On the other hand, as more information to be considered results in greater element interactivity, anticipatory prompts will likely also induce higher ICL and might therefore be more demanding. Presumably, only those learners with greater prior knowledge will successfully manage the increased demands of such anticipatory prompts while remaining able to invest considerable amounts of GCL. Learners with lower prior knowledge, on the other hand, might be overwhelmed by the increased demands of the anticipatory prompts and will thus experience higher ICL (Gerjets et al., 2006; van Merriënboer et al., 2006). Consequently, in terms of learning outcomes, only learners with higher prior-knowledge levels can be expected to benefit from anticipatory prompts.

Self-explanation prompts affect both learners' cognitive processes and thus their cognitive load and learning outcomes, but they also influence learners' self-efficacy regarding the learning topic. For example, Crippen and Earl (2007) developed a web-based learning tool to teach undergraduate students problem solving skills in the domain of chemistry with quizzes. Students were allocated to one of three experimental conditions: in the control condition, students learned with the quizzes only. In the other two conditions, students were also provided with worked examples for each quiz item. Additionally, in one of these conditions, students were prompted to self-explain the worked examples. Regarding self-efficacy, these authors found that worked examples alone revealed no effects on self-efficacy, but worked examples provided together with self-explanation prompts did exert a positive effect on students' self-efficacy (Crippen & Earl, 2007). The question as to whether retrospective or anticipatory prompts reveal different effects on learners' self-efficacy cannot be answered based on existing research evidence.

Taken together, the potential positive effects of anticipatory prompts supposedly depend on whether learners can cope with the increased demands. Hence, the learners' prior knowledge likely plays an important role in the relationship between prompt type and cognitive load, learning outcomes, and self-efficacy. However, these theoretical considerations cannot be substantiated with empirical evidence, as anticipatory prompts have seldom been investigated (Bisra et al., 2018).

Present Study and Research Questions

The present study was conducted with automotive apprentices who were taught a diagnostic strategy to diagnose complex automotive malfunctions (Abele, 2018; Abele & von Davier, 2019). Although diagnoses of malfunctions are a crucial part of an automotive technicians day-to-day work (Spöttl et al., 2011), at the end of their 3-year apprenticeship, only 15 % of the apprentices master the required strategies to diagnose complex malfunctions; they can thus be considered novices (Abele & von Davier, 2019). We pursued two goals: First, we investigated the use and possible limitations of longer and more comprehensive modelling examples in a screencast video format for teaching a job-relevant and complex problem-solving strategy, namely diagnosing car malfunctions. Second, we compared the effects of anticipatory and retrospective self-explanation prompts for these modelling examples. For this comparison, we considered the apprentices' general prior knowledge of car diagnoses. We examined the effects of modelling examples and self-explanation prompts on apprentices' diagnostic strategy knowledge and skills (i.e., knowledge about and application of the instructed strategy), self-efficacy, and cognitive load. Diagnostic strategy knowledge and skills and self-efficacy were measured before and after the intervention. Cognitive load was measured only after the intervention.

We investigated the following hypotheses regarding modelling examples:

- H1: Following the worked or modelling example effect (Renkl, 2014; Sweller, 2006), we expected a greater increase in diagnostic strategy knowledge and skills from a pretest to a posttest when the apprentices learned with modelling examples than when apprentices practised applying the diagnostic strategy by solving open problems.
- H2: We expected a greater increase in self-efficacy among apprentices learning with modelling examples than among those practising applying the strategy (Crippen & Earl, 2007; Schunk, 1995).
- H3: Following the example-based learning literature (e.g., Renkl et al., 2009), we expected apprentices in the modelling example condition to perceive lower

extraneous and higher germane cognitive load while learning than apprentices practising to apply the strategy.

Moreover, we were interested in whether the effects of different self-explanation prompts depend on prior knowledge. However, since these effects have hardly been researched so far, we formulated no specific hypotheses. Instead, we posed these three open research questions:

- RQ1: Do anticipatory and retrospective self-explanation prompts reveal differential effects on the development of apprentices' diagnostic strategy knowledge and skills and does their prior knowledge moderate these effects?
- RQ2: Do anticipatory and retrospective prompts exert differential effects on the development of apprentices' self-efficacy and does their prior knowledge moderate these effects?
- RQ3: Third, do anticipatory and retrospective prompts demonstrate differential effects on apprentices' extraneous, intrinsic, and germane cognitive load while learning, and does their prior knowledge moderate these effects?

Methods

Participants

Originally, 78 apprentices participated in our experiment. Because of technical problems with the survey software, only 67 complete data sets could be analysed. Apprentices were 20.85 years old (*SD* = 2.74), 65 were male, and two were female. German was the first language of 57 apprentices, and 10 reported an additional first language. Seven apprentices had a university entrance qualification (*Abitur*), 55 apprentices had a secondary school leaving certificate (*Mittlere Reife*), and five apprentices had a lower secondary school leaving certificate (*Hauptschulabschluss*).

To determine the required sample sizes, we conducted two a-priori power analyses with Gpower 3.1 (Faul et al., 2007). We aimed for a power of .80. Based on previous studies on the worked example effect (e.g., Nievelstein et al., 2013; Schwonke et al., 2009; van Gog et al., 2011) and self-explanation prompts (e.g., Atkinson et al., 2003; Hilbert & Renkl, 2009), we expected medium effect sizes (e.g., Cohen's f > .25 or $\eta^2 > .06$; Cohen, 1988). For the analyses regarding hypotheses H1 and H2 and research questions RQ1 and RQ2 (i.e., repeated measures analyses of variance, RM-ANOVAs), the required sample size was $N = 34$ (about half of the collected sample). For the analyses regarding hypothesis H3 and research question RQ3 (i.e., analyses of variance, ANOVAs), the required sample size was $N = 128$. As we had to stop collecting data at an early stage because of school closures during the COVID-19 pandemic, the required sample size for the ANOVAs could not be realized. A larger sample may have enabled us to demonstrate additional effects. However, the effects we did discover can still be interpreted.

Design and Procedure

The experiment comprised two sessions separated by approximately 10 days. Table 1 shows the detailed procedure. In session two, first, apprentices in all conditions learned about the diagnostic strategy with instructional videos and organizational prompts. Then, they learned according to their randomly assigned experimental condition: Two groups received modelling examples, one ($n = 21$) with retrospective self-explanation prompts and the other ($n = 25$) with anticipatory self-explanation prompts. The third group (control, $n = 21$) received no modelling examples and no self-explanation prompts.

The entire study took place on computers in the apprentices' schools. All learning and testing materials, which can be requested from the first author, were presented in digital form via the page-based online survey tool LimeSurvey. Once apprentices left a page, they could not go back. We told participants when we expected them to have completed a phase and to proceed with the next phase. Thereby we ensured an equal time on task within and between conditions (see maximum durations in Table 1).

Table 1

Procedures in Sessions 1 and 2

Phase	Content	Average duration in min	Maximum duration in min until termination
	SESSION 1: Pretest		
Phase 1	Introduction to study and computer simulation	30	45
Break		10	10
Phase 2	Self-efficacy rating[a]	5	50
	Diagnostic strategy knowledge and skills test: Strategy description test	10	
	Diagnostic strategy knowledge and skills test: First diagnosis in simulation	25	
Break		10	10
Phase 3	General prior knowledge test: Partial skills test	40	85
	General prior knowledge test: Diagnosis-relevant reception competence test	10	
	Diagnostic strategy knowledge and skills test: Strategy completion test	25	
TOTAL SESSION 1		**165**	**200**
	SESSION 2: Intervention and posttest		
Phase 1	Refresher on computer simulation	5	150
	Learning phase 1: Instructional videos and organizational prompts	35	
	Break	5	
	Learning phase 2: **Content depending on experimental condition:**	90	
	Modelling examples and *retrospective* prompts / Modelling examples and *anticipatory* prompts / *Control* group: no modelling examples or prompts but independent diagnosis in simulation		
Break		10	10
Phase 2	Self-efficacy rating[a]	5	75
	Diagnostic strategy knowledge and skills test: Strategy description test	10	
	Diagnostic strategy knowledge and skills test: First diagnosis in simulation	25	
	Diagnostic strategy knowledge and skills test: Second diagnosis in simulation	25	
Break		10	10
Phase 3	Diagnostic strategy knowledge and skills test: Strategy completion test	25	30
	Subjective evaluation of the learning materials[b]	5	
TOTAL SESSION 2		**250**	**275**

[a] In addition to self-efficacy, some motivational items were administered to evaluate the learning materials. These are reported in Meier et al. (2022)
[b] Results of these evaluations are also reported in Meier et al. (2022)

Learning Materials

In the experimentally varied intervention, apprentices learned a complex diagnostic strategy that should help them to diagnose car malfunctions in a structured way. The intervention comprised two learning phases. In the first learning phase, apprentices in *all* conditions watched five animated instructional videos explaining the strategy (16:33 minutes). All participants also completed four practice tasks during this phase, which served as *organizational prompts* (Roelle et al., 2017), and received the correct solution. Learning phase one took 35 minutes.

In learning phase two, participants in the two modelling example conditions received two video-based modelling examples showing an expert applying the diagnostic strategy in a computer simulation (Gschwendtner et al., 2009; Meier et al., 2022). Both modelling examples consisted of several videos (first example: 12 videos; 25:50 minutes; second example: 10 videos;19:37 minutes).

We developed the diagnostic strategy, the instructional videos, and the modelling examples in close collaboration with subject-matter experts. The development and evaluation of this content are described in detail by Meier et al. (2022).

After each video of the modelling examples, participants answered the same self-explanation prompt in writing. Depending on the condition, the prompt differed: the *retrospective self-explanation prompt* read as “Which troubleshooting step **was just completed**? Explain how you will proceed with this step and why it is important for troubleshooting (in general)”. The *anticipatory self-explanation prompt* read “Which troubleshooting step **comes next**? Explain how you will proceed with this step and why it is important for troubleshooting (in general)”. For the first four prompts, participants were supported in their answers by answering fill-in-the-blank self-explanation prompts (i.e., *assisting* self-explanation prompts; Berthold et al., 2009). For all following prompts, participants received suggestions for how to start their answers' first sentences. Exemplary responses to the self-explanation prompts are displayed in Table A-1. Participants did not receive individual feedback but the correct answer for each prompt after answering it, that is, they received an example of how the respecting prompt could have been answered correctly.

Participants in the control condition did not receive the modelling examples but tried to diagnose the same diagnostic problems in the computer simulation (Gschwendtner et al., 2009; Meier et al., 2022) that the expert in the modelling examples solved. Hence, instead of studying the worked-out solutions to the two diagnostic problems in the modelling examples, participants in the control condition were required to solve the problems independently, that is, to practise applying the diagnostic strategy on their own.

Testing Materials

To investigate the effects of modelling examples and different self-explanation prompts depending on the learners' general prior knowledge on diagnostic strategy knowledge and skills, self-efficacy, and cognitive load, different tests were used: To assess *general prior knowledge* about car diagnoses, we used two different tests in session one. For *diagnostic strategy knowledge and skills* (i.e., knowledge about and application of the instructed diagnostic strategy) three tests were given in both sessions one (i.e., before the intervention) and two (i.e., after the intervention). Likewise, a questionnaire assessing the apprentices' *self-efficacy* in performing diagnoses was used in sessions one and two. Finally, a questionnaire aiming at the apprentices' *cognitive load* was given after the intervention in session two. All these tests are described below. Closed and open items were used in most of them. Closed items were scored automatically. For all open items, the first author and a subject matter expert (i.e., the second author) developed a coding scheme. We developed these schemes based on ideal responses to the different tests. Ideal means that these responses were perfectly in line with the taught diagnostic strategy. In addition, we also looked for alternative solutions in the responses of all participants that could be assessed as similarly good from a subject matter perspective. Then, a student assistant and the first author scored 25% of all answers and adjusted the coding schemes until achieving an interrater reliability of Cohen's Kappa > 0.8. Then the student assistant independently scored the remaining answers.

General Prior Knowledge Tests

As a first measure of general prior knowledge about car diagnoses, we selected five out of 24 items in the *diagnosis-relevant reception competence* (DRC) test by Norwig et al. (2021). This competence describes the ability to read various documents relevant to the diagnosis (e.g., electrical circuit diagrams) and can thus be seen as prerequisite knowledge for car diagnoses. For example, we gave participants a schematic diagram and a photo of an engine compartment and asked them to use the schematic diagram to locate a particular component in the realistic photo. We selected items with a midrange solution rate (ranging from 32% to 71% in Norwig et al., 2021) to prevent floor and ceiling effects and with the highest item-total correlation (> 0.43 for all 5 items).

Second, we selected three of seven items of a partial skills test by Abele (2014) with a high item-total correlation (between .48 and .60 in Abele, 2014). In these items, participants were instructed to perform specific measurements in the simulation and to evaluate whether the measurement results indicated a malfunction or not.

Diagnostic Strategy Knowledge and Skills Tests

We administered three different tests to measure the apprentices' diagnostic strategy knowledge and skills in the pretest (i.e., in session one) *as well as* in the posttest (i.e., after the intervention in session two). First, the *strategy description test* measured conceptual knowledge and comprised two questions asking participants (1) to describe their troubleshooting procedure in a situation where they are given little assistance from a computer-based expert system (i.e., complex diagnostic problems), and (2) how they would narrow down which components might be responsible for a malfunction.

Second, in the *strategy completion test*, apprentices carried out or described (parts of) steps of the diagnostic strategy in four different scenarios. Within these scenarios, closed and open questions were used. The former dealt, for example, with which diagnostic step should be taken next in the current scenario. In the open-ended questions, the apprentices, for example, studied a circuit diagram and described an appropriate measurement.

Third, to test *diagnostic skills*, participants performed diagnoses in the computer simulation. They were provided with a description of the malfunction and then diagnosed it. Eventually, participants described the cause of the malfunction and how it could be repaired. Participants made their *first diagnosis* in both the pretest in session one and the posttest in session two, and one additional *second diagnosis* in the posttest only.

Self-efficacy and Cognitive Load

Both in the pretest and posttest and before performing the first diagnosis in the computer simulation, participants rated their *self-efficacy* regarding this diagnosis with five items on a seven-point Likert scale (Cronbach's $\alpha = 0.89$; Bandura, 2006). After the intervention, participants rated their *intrinsic* (two items), *germane* (two items) and *extraneous cognitive load* (three items) on a seven-point Likert-scale (Klepsch et al., 2017; Klepsch & Seufert, 2020, 2021). Reliability was acceptable (intrinsic load: Cronbach's $\alpha = 0.74$; germane load: Cronbach's $\alpha = 0.84$; extraneous load: Cronbach's $\alpha = 0.60$).

Results

To test the effects of modelling examples and self-explanation prompts on variables measured in the pretest (i.e., session 1) and posttest (i.e., session 2), we ran repeated measures analyses of variance (RM-ANOVAs). For variables only measured once in session 2, regular analyses of variance (ANOVAs) were conducted. Both analyses were run separately for the modelling examples' effects (H1 to H3; modelling examples yes versus no) and the effects of self-explanation prompts (RQ1 to RQ3; retrospective versus anticipatory prompts). For the latter, the mean-centered test scores on both prior knowledge tests were included as additional continuous factors (Schneider et al., 2015). A significance level of .05 applies to all analyses. As effect size we used $\eta^2_{partial}$ with .01, .06, and .14 corresponding to a small, medium, and large effect, respectively (Cohen, 1988; Lakens, 2013). Analyses were conducted with IBM SPSS Statistics 27.

Apart from two exceptions, there were no differences in demographic variables, prior knowledge tests, or first measures of repeated measures between the two example conditions or two prompt conditions (all $p > .05$). The first exception was the age between the two example conditions, $F(1, 65) = 4.227$.; $p = .044$. However, as age did not correlate with any of the dependent variables (or with the development of the dependent variables with repeated measures), this difference is negligible. Second, in the pretest, apprentices in the modelling example condition scored significantly higher in the first diagnosis in the simulation than apprentices did in the control condition (see Table 2), $F(1, 65) = 7.727$.; $p = .007$. This difference must be considered in the later interpretation of our results. In the following section, we first report the effects of modelling examples. In the second section, we report the effects of retrospective versus anticipatory self-explanation prompts.

Effects of Modelling Examples

To analyse variables measured in the pretest and posttest, we performed an RM-ANOVA with *example condition* (i.e., modelling examples yes versus no) as between-subjects variable and *timepoint* (pretest versus posttest) as within-subjects variable. For variables measured only in the posttest, we conducted a regular ANOVA. Table 2 illustrates descriptive data. Table 3 shows the results of the statistical tests.

Table 2

Descriptive Data of Dependent Variables for the Control Condition (i.e., no Modelling Examples) and Modelling Examples Condition

	No modelling examples (*n* =21)				Modelling examples (*n* =46)			
	Pretest		Posttest		Pretest		Posttest	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
DRC Test Score[a]	3.29	0.27	-	-	3.20	0.18	-	-
Partial Skills Test Score[b]	3.33	0.41	-	-	4.09	0.39	-	-
Strategy Description Test Score[c]	1.10	0.77	2.38	2.27	1.48	0.89	2.80	2.29
Strategy Completion Test Score[d]	15.14	5.84	22.00	6.47	16.52	5.82	25.33	4.73
First Diagnosis Score[e]	0.38	1.07	1.43	1.29	1.24	1.21	1.22	1.25
Second Diagnosis Score[f]	-	-	0.48	0.81	-	-	0.83	1.06
Self-efficacy[g]	4.29	1.19	4.15	1.37	4.53	1.14	4.49	1.26
Intrinsic Load[g]	-	-	3.98	1.07	-	-	3.99	1.62
Germane Load[g]	-	-	4.71	1.62	-	-	4.46	1.72
Extraneous Load[g]	-	-	3.27	1.10	-	-	3.41	1.35

[a]0-5 points; this test was a prior knowledge measure but not relevant for the analyses of the effects of the modelling examples
[b]0-9 points; this test was a prior knowledge measure but not relevant for the analyses of the effects of the modelling examples
[c]0-10 points
[d]0-47 points
[e]0-4 points
[f]0-6 points
[g]7-point Likert-scale ranging from 1 = *absolutely not true* to 7 = *absolutely true*

Table 3

Main and Interaction Effects of the Example Condition and Timepoint on Dependent Variables

Analysis	Hypothesis	Independent Variable(s)	Dependent Variables	Statistical test results			
				df	F	p	$\eta^2_{partial}$
RM-ANOVA	-	Example Condition	Strategy Description Test Score	1, 65	1.470	.230	.022
	-		Strategy Completion Test Score	1, 65	3.629	.061	.053
	-		First Diagnosis Score	1, 65	1.578	.214	.024
	-		Self-efficacy	1, 65	1.072	.304	.016
	-	Timepoint	Strategy Description Test Score	1, 65	17.938	< **.001**	.216
	-		Strategy Completion Test Score	1, 65	96.430	< **.001**	.597
	-		First Diagnosis Score	1, 65	7.256	**.009**	.100
	-		Self-efficacy	1, 65	0.333	.566	.005
	H1	Timepoint*Example Condition	Strategy Description Test Score	1, 65	.004	.948	.000
			Strategy Completion Test Score	1, 65	1.491	.227	.022
			First Diagnosis Score	1, 65	7.884	**.007**	.108
	H2		Self-efficacy	1, 65	.086	.770	.001
ANOVA	H1	Example Condition	Second Diagnosis Score	1, 65	1.797	.185	.027
	H3	Example Condition	Intrinsic load	1, 65	.001	.973	.000
			Germane load	1, 65	.337	.563	.005
			Extraneous load	1, 65	.181	.672	.003

Tables 2 and Table 3 indicate significant main effects of timepoint on strategy description test score and strategy completion test score with large effects. In both tests, participants in both conditions scored significantly higher in the posttest than in the pretest. However, as there were no interaction effects of example condition and timepoint on these test scores, this improvement was not larger in participants in the modelling example condition. Nor were there any effects on self-efficacy, either as main effects by timepoint or example condition or as interaction effects.

We observed a significant main effect of timepoint and an interaction effect of example condition and timepoint on the score in the first diagnosis. However, both effects arise from a difference in the first measurement of the score on the first diagnosis, which we already pointed out at the beginning of the results section. Thus, these effects should be disregarded.

Regarding the variables measured only once, we detected no effects of modelling examples on participants' score on the second diagnosis, nor any effects on participants' ICL, GCL, or ECL.

Taken together, we were unable to confirm hypotheses H1 to H3: modelling examples did not lead to a greater increase in diagnostic strategy knowledge and skills (H1), or in self-efficacy (H2), and participants in the modelling examples did not perceive lower ECL and higher GCL (H3).

Effects of Retrospective versus Anticipatory Self-Explanation Prompts

Our open research questions regarding the effects of the various self-explanation prompts also included participants' prior knowledge. To analyse variables measured in the pretest and posttest, we ran an RM-ANOVA with *self-explanation prompt condition* (i.e., retrospective versus anticipatory prompts) as between-subjects variable and *timepoint* (pretest versus posttest) as within-subjects variable. For variables measured only in the posttest, we ran another ANOVA. To test for moderating effects of prior knowledge, we included *diagnosis-relevant reception competence (DRC) test score* and *partial skill test score* as additional continuous factors for both analyses. For those analyses, we grand mean-centered these prior-knowledge factors (Schneider et al., 2015). Table 4 shows the descriptive data of the various dependent variables for the two self-explanation prompt conditions. Table 5 shows the results of the statistical tests of the effects of the different self-explanation prompts.

Table 4

Descriptive Data of Moderation Variables and Dependent Variables for the Retrospective Prompt Condition and Anticipatory Prompt Condition

	Retrosp. SE-Prompts (*n* =21)				Anticipat. SE-Prompts (*n* =25)			
	Pretest		Posttest		Pretest		Posttest	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
DRC Test Score[a]	3.14	1.06	-	-	3.24	1.30	-	-
Partial Skills Test Score[b]	4.05	2.29	-	-	4.12	2.91	-	-
Strategy Description Test Score[c]	1.48	0.93	2.62	2.27	1.48	0.87	2.96	2.34
Strategy Completion Test Score[d]	17.67	4.56	24.90	4.06	15.56	6.63	25.68	5.29
First Diagnosis Score[e]	1.24	1.34	1.33	1.20	1.24	1.13	1.12	1.30
Second Diagnosis Score[f]	-	-	0.57	0.75	-	-	1.04	1.24
Self-efficacy[g]	4.50	1.15	4.51	1.13	4.56	1.15	4.47	1.38
Intrinsic Load[g]	-	-	4.05	1.68	-	-	3.94	1.60
Germane Load[g]	-	-	4.55	1.73	-	-	4.38	1.73
Extraneous Load[g]	-	-	3.89	1.43	-	-	3.01	1.16

[a]0-5 points; this test was a prior knowledge measure and used as a moderation variable for the analyses of the effects of the different types of self-explanation prompts
[b]0-9 points; this test was a prior knowledge measure and used as a moderation variable for the analyses of the effects of the different types of self-explanation prompts
[c]0-10 points
[d]0-47 points
[e]0-4 points
[f]0-6 points
[g]7-point Likert-scale ranging from 1 = *absolutely not true* to 7 = *absolutely true*

Table 5

Main and Interaction Effects of the Prompt Condition, the Moderation Variables, and Timepoint on Dependent Variables

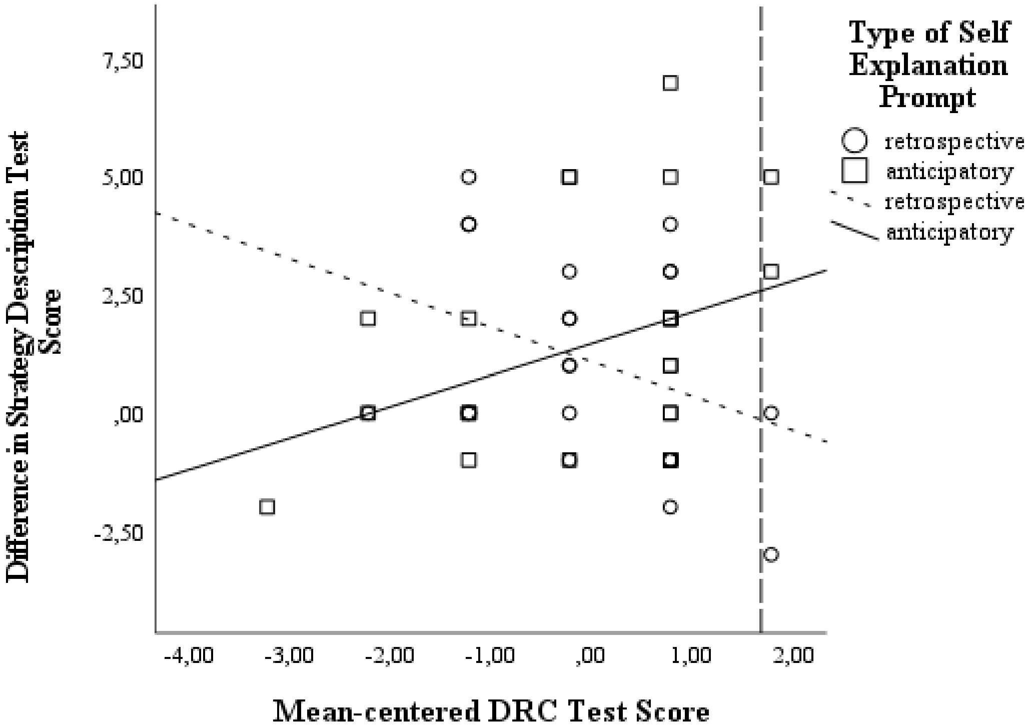
Analysis	Research Question	Independent Variable(s)	Dependent Variables	Statistical test results			
				df	F	p	$\eta^2_{partial}$
RM-ANOVA	-	Timepoint	Strategy Description Test Score	1, 40	16.724	< **.001**	.295
	-		Strategy Completion Test Score	1, 40	94.246	< **.001**	.702
	-		First Diagnosis Score	1, 40	.001	.978	.000
	-		Self-efficacy	1, 40	.381	.541	.009
	RQ1	Timepoint*Prompt Condition	Strategy Description Test Score	1, 40	.192	.663	.005
	RQ1		Strategy Completion Test Score	1, 40	2.368	.132	.056
	RQ1		First Diagnosis Score	1, 40	.339	.564	.008
	RQ2		Self-efficacy	1, 40	.045	.834	.001
	RQ1	Timepoint*Prompt Condition*DRC Test Score	Strategy Description Test Score	1, 40	4.368	**.043**	.098
	RQ1		Strategy Completion Test Score	1, 40	.308	.582	.008
	RQ1		First Diagnosis Score	1, 40	.674	.416	.017
	RQ2		Self-efficacy	1, 40	8.968	**.005**	.183
	RQ1	Timepoint*Prompt Condition*Partial Skills Test Score	Strategy Description Test Score	1, 40	2.422	.127	.057
	RQ1		Strategy Completion Test Score	1, 40	.743	.394	.018
	RQ1		First Diagnosis Score	1, 40	.193	.663	.005
	RQ2		Self-efficacy	1, 40	.007	.936	.000
ANOVA	RQ1	Prompt Condition	Second Diagnosis Score	1, 40	2.121	.153	.050
	RQ3		Intrinsic Load	1, 40	.010	.921	.000
	RQ3		Germane Load	1, 40	.178	.675	.004
	RQ3		Extraneous Load	1, 40	5.394	**.025**	.119
	RQ1	Prompt Condition*DRC Test Score	Second Diagnosis Score	1, 40	.476	.494	.012
	RQ3		Intrinsic Load	1, 40	3.490	.069	.080
	RQ3		Germane Load	1, 40	8.813	**.005**	.181
	RQ3		Extraneous Load	1, 40	2.244	.142	.053
	RQ1	Prompt Condition*Partial Skills Test Score	Second Diagnosis Score	1, 40	.063	.803	.002
	RQ3		Intrinsic Load	1, 40	.004	.950	.000
	RQ3		Germane Load	1, 40	2.180	.148	.052
	RQ3		Extraneous Load	1, 40	.009	.924	.000

We noted significant main effects of timepoint on the strategy description test score and strategy completion test score with large effects. These effects correspond to the effects we had already observed when comparing the two example conditions.

Regarding research questions RQ1 and RQ2, we detected no interaction effects of timepoint and prompt condition on any of the dependent variables. When also considering participants' prior knowledge, however, our results revealed two significant three-way interactions of timepoint, prompt condition, and the DRC test score on the strategy description test score (RQ1) and self-efficacy (RQ2). Figure 1 indicates that concerning strategy description test scores, participants who got a lower DRC test score benefitted from retrospective prompts, while those with higher DRC test scores benefitted more from anticipatory prompts. To further explore this interaction effect, we used the Johnson-Neyman procedure (Hayes & Matthes, 2009; Montoya, 2019) by performing a moderation analysis using the PROCESS macro by Hayes (2022). We tested where in the distribution of mean-centered DRC test scores the condition (i.e., retrospective versus anticipatory prompts) had a statistically significant effect on the difference of strategy description test scores, calculated as posttest score minus pretest score. We found that for learners with mean-centered DRC test scores larger than 1.68, that is, the higher prior knowledge participants, retrospective prompts had detrimental effects and anticipatory prompts had beneficial effects on the difference of strategy description test scores, $t(42) = 2.02$, $p = .05$.

Figure 1

Scatter Plot of Grand Mean-centered DRC Test Scores Against Difference in Strategy Description Test Scores for the Retrospective Prompt Condition and Anticipatory Prompt Condition

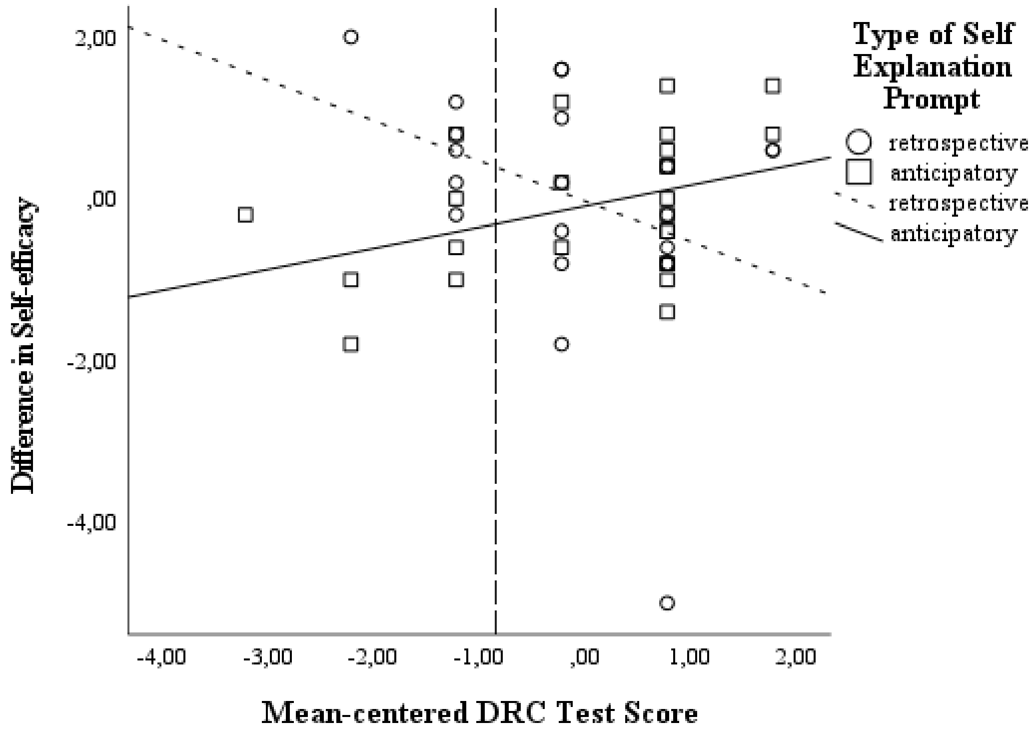


Note. The differential effect of retrospective and anticipatory prompts on the difference in strategy description test scores is significant right of the vertical longer dashed line (DRC test scores > 1.68).

Regarding self-efficacy (RQ2): Figure 2 indicates that participants with low DRC test scores rather benefited from retrospective prompts while participants with higher DRC test scores rather benefited from anticipatory prompts. We again used the Johnson-Neyman technique to fully explicate the nature of this interaction effect: We found that for participants with mean-centered DRC test scores lower than -0.85, that is, the lower prior knowledge participants, retrospective prompts had beneficial effects and anticipatory prompts had detrimental effects in terms of difference in self-efficacy, $t(42) = -2.02$, $p = .05$.

Figure 2

Scatter Plot of Grand Mean-centered DRC Test Scores Against Difference in Self-Efficacy for the Retrospective Prompt Condition and Anticipatory Prompt Condition

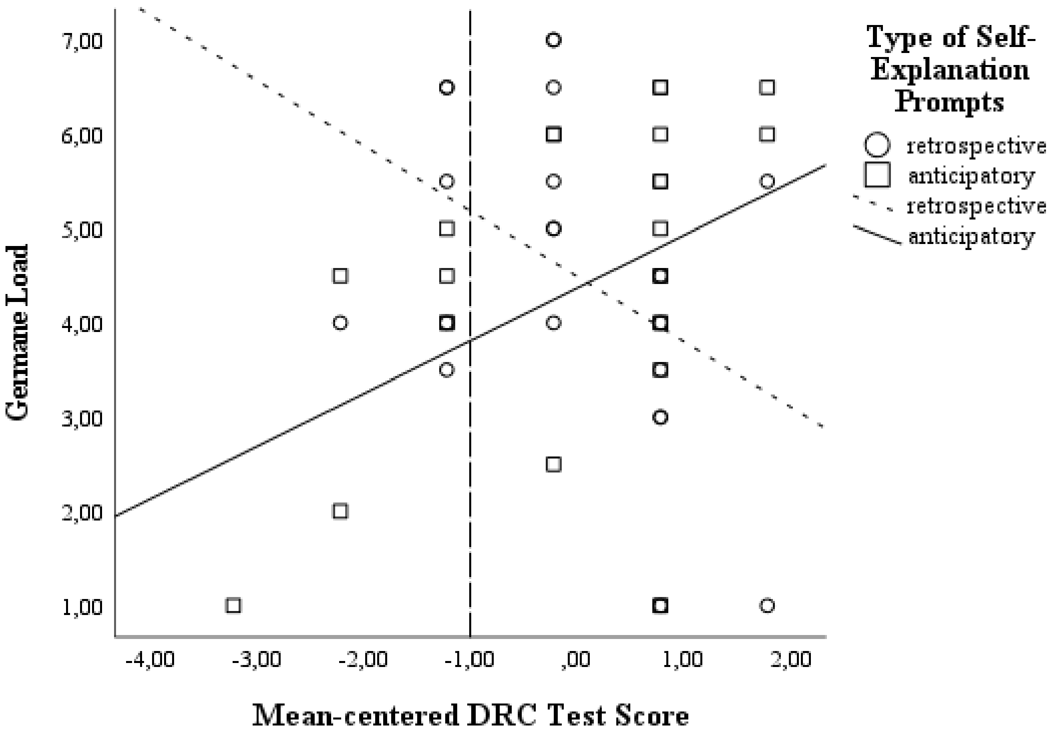


Note. The differential effect of retrospective and anticipatory prompts on the difference in self-efficacy is significant left of the vertical longer dashed line (DRC test scores < -0.85).

Eventually, regarding RQ3, we found that ECL was lower in the anticipatory prompt group. Moreover, we identified a significant two-way interaction of prompt condition and DRC test score on GCL. Figure 3 indicates that apprentices with lower DRC test scores experienced higher GCL when learning with retrospective prompts, while those with higher DRC test scores experienced higher GCL when learning with anticipatory prompts. Following the Johnson-Neyman procedure, we found that for lower prior knowledge participants (mean-centered DRC test scores < -0.98) retrospective prompts induced a higher GCL than anticipatory prompts, $t(42) = -2.02$, $p = .05$.

Figure 3

Scatter Plot of Grand Mean-centered DRC Test Scores Against Germane Load for the Retrospective Prompt Condition and Anticipatory Prompt Condition



Note. The differential effect of retrospective and anticipatory prompts on GCL is significant left of the vertical longer dashed line (DRC test scores < -0.98).

Taken together, RQ1 to RQ3 cannot be answered unambiguously, but there is a tendency that indicates that apprentices with more prior knowledge learned more when learning with anticipatory prompts, while apprentices with less prior knowledge experienced a greater increase in self-efficacy and a higher GCL when learning with retrospective prompts.

Discussion

So far, research on modelling examples has tended to focus on brief modelling examples teaching quite simple problem-solving strategies. Moreover, self-explanation prompts asking learners to explain past problem-solving steps illustrated in an example have mainly been used. Thus, the present study had two objectives: First, we investigated the effects of modelling examples when teaching longer problem-solving strategies, such as diagnosing car malfunctions, on diagnostic strategy knowledge and skills (H1), self-efficacy (H2), and extraneous and germane cognitive load during learning (H3). Second, while taking into account the apprentices' prior knowledge, we compared the effects of retrospective and anticipatory self-explanation prompts on the development of diagnostic strategy knowledge and skills (RQ1), self-efficacy (RQ2), and cognitive load during learning (RQ3).

Effects of Modelling Examples

Contrary to H3, we observed that the modelling examples exerted no effects on the apprentices' extraneous (ECL) or germane cognitive load (GCL). Since example-based learning's positive effect on learning outcomes relies on reducing ECL and increasing GCL (Sweller, 2006), we would not expect the modelling examples to reveal any positive effect on learning outcomes. Accordingly, and in contrast to our H1, we detected no such effect. One interpretation of this finding is that longer modelling examples are less suitable for teaching complex problem-solving strategies. However, both text-based worked examples (e.g., Heitzmann et al., 2015; Schalk et al., 2020) and video-based modelling examples (Fiorella et al., 2017; Hoogerheide, 2016; Hoogerheide et al., 2014; Schmitz et al., 2017; van Harsel et al., 2019) have proven to be conducive to learning. We assume that we were unable to detect beneficial effects of the modelling examples because of the long instruction phase in which the diagnostic strategy was initially explained to all apprentices, that is, also in the control condition. Learning phase one took 35 minutes and comprised five instructional videos and four practice tasks that presumably supported knowledge organisation well. This 35-minute instruction phase is substantially longer than in other studies. Schmitz et al. (2017) tested an instruction lasting only 17 minutes. Some studies used no instruction at all (e.g., Hoogerheide et al., 2014). Therefore, our extensive instruction may have provided all apprentices, that is, also those in the control condition, with sufficient knowledge to begin independent problem solving (i.e., independent diagnosis in the simulation).

Regarding self-efficacy, we expected the modelling examples to promote self-efficacy, as learners would be able to see how the diagnostic process is completed (H2; Glogger-Frey et al., 2015; Schunk, 1995). However, that is not what we observed. One reason for this might be that the inexperienced apprentices could not identify with the model, who was an experienced expert. According to the model-observer similarity principle (Renkl, 2014; van Gog et al., 2019), the model should also have been an apprentice so that the learners could have identified better with it.

We, therefore, recommend future studies investigating the use of longer modelling examples for complex problem-solving strategies to use a shorter instruction phase and a model with which learners can better identify.

Effects of Retrospective and Anticipatory Self-explanation Prompts

Between the two prompt conditions we noted different effects on learning outcomes (RQ1), self-efficacy (RQ2), and cognitive load (RQ3) depending on the apprentices' prior knowledge: We found a greater increase in declarative knowledge (i.e., in the strategy description test) among the stronger apprentices when learning with the anticipatory prompts. Regarding self-efficacy, the weaker apprentices' self-efficacy was better supported by the retrospective prompts. Similarly, in terms of cognitive load, apprentices with less prior knowledge reported a higher GCL when learning with retrospective prompts. Overall, these effects suggest that anticipatory prompts are more beneficial for learners with more prior knowledge, whereas learners with less prior knowledge profit more from retrospective prompts. Consequently, we recommend researchers and practitioners designing example-based learning scenarios to adapt their self-explanation prompts to their learners' prior knowledge. By doing so, detrimental effects, such as the expertise reversal effect (Kalyuga &

Renkl, 2010), can possibly be avoided. Moreover, to better understand the specific mechanisms of the anticipatory prompts for learners with different prior knowledge levels, a replication study in a laboratory setting with fewer interfering environmental influences and, possibly, think-aloud protocols may be helpful.

Besides the interactions, we also found that the anticipatory prompts induced a lower ECL, that is, a lower learning-irrelevant load. This finding is rather surprising. We expected anticipatory prompts to mainly influence element interactivity and thus ICL. The ECL items mainly addressed the learning material's (visual) appearance, that is, whether the content was easy to process. However, the two prompt conditions did not differ in their learning materials' (visual) design. For example, Klepsch and Seufert (2020, study 1), who developed the cognitive load instruments used in the present study, found differences in learners' ECL ratings when element interactivity had been manipulated and argued that participants sometimes struggle to distinguish between ICL and ECL, which resulted in effects on both scales. However, we could only refer to this argumentation if the anticipatory prompts had caused an increase and not a decrease in ECL. To make sure that this finding is not a false positive, future studies should try to replicate it.

Limitations and Implications for Future Research

Above, we already gave several recommendations for future research: First, in future studies on longer modelling examples, a shorter instruction phase and a model with which learners can better identify should be used. Second, the recurrent pattern of anticipatory prompts being more beneficial for higher prior knowledge learners and retrospective prompts being more beneficial for lower prior knowledge learners needs to be further investigated – possibly in a laboratory setting with think-aloud protocols. Third, the effect of anticipatory prompts inducing a lower ECL should also be further investigated.

Our study had one limitation that should be addressed in future research: We found that the self-explanation prompts were answered sub-optimally. That is, participants in both prompt conditions answered only about 50% of the prompts in a meaningful way. The other half of the prompts were often not answered meaningfully, with participants either entering only single letters or blanks or making entries without any reference to car diagnoses. Table A-1 gives exemplary responses to the self-explanation prompts. There are two potential reasons for this. First, the self-explanation prompts were not very specific, as they simply asked learners to name and explain the previous or subsequent diagnostic step. Accordingly, they provided little guidance to the learners. For example, Glogger et al. (2009) showed that for ninth graders prompted to apply learning strategies, specific prompts were superior to general prompts. Second, in the present paper, after each diagnostic step in the modelling example, the apprentices answered exactly the same self-explanation prompt. These prompts may have been perceived as too repetitive. The differential effects of retrospective and anticipatory prompts depending on prior knowledge may be even stronger with more specific and more engaging prompts. This possibility should be investigated in the future.

Conclusion

Even if modelling examples did not yield the desired effects in the present study, anticipatory self-explanation prompts seem to function differently from retrospective self-explanation prompts and are a promising alternative for stronger learners. When designing modelling examples, educational practitioners should thus consider using various types of self-explanation prompts for learners possessing different levels of prior knowledge. Our results indicate the potential of anticipatory prompts that should be explored in future research.

References

Abele, S. (2018). Diagnostic problem-solving process in professional contexts: Theory and empirical investigation in the context of car mechatronics using computer-generated log-files. *Vocations and Learning*, *11*(1), 133–159. https://doi.org/10.1007/s12186-017-9183-x

Abele, S., & von Davier, M. (2019). CDMs in vocational education: Assessment and usage of diagnostic problem-solving strategies in car mechatronics. In M. von Davier & Y.-S. Lee (Eds.), *Handbook of Diagnostic Classification Models*. Springer. https://doi.org/10.1007/978-3-030-05584-4_22

Abele, S., Walker, F., & Nickolaus, R. (2014). Zeitökonomische und reliable Diagnostik beruflicher Problemlösekompetenzen bei Auszubildenden zum Kfz-Mechatroniker [Time-saving and reliable diagnostics in measuring professional problem-solving competence in the domain of car mechatronics]. *Zeitschrift für Pädagogische Psychologie*, *28*(4), 167–179. https://doi.org/10.1024/1010-0652/a000138

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. *Journal of Educational Psychology*, *95*(4), 774–783. https://doi.org/10.1037/0022-0663.95.4.774

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

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. C. Urdan (Eds.), *Self-Efficacy Beliefs of Adolescents* (pp. 307–337). IAP - Information Age Pub., Inc.

Berthold, K., Eysink, T. H. S., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science*, *37*(4), 345–363. https://doi.org/10.1007/s11251-008-9051-z

Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing Self-Explanation: A Meta-Analysis. *Educational Psychology Review*, *30*(3), 703–725. https://doi.org/10.1007/s10648-018-9434-x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.

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

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fiorella, L., & Mayer, R. E. (2016). Eight Ways to Promote Generative Learning. *Educational Psychology Review*, *28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9

Fiorella, L., van Gog, T., Hoogerheide, V., & Mayer, R. E. (2017). It's all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task. *Journal of Educational Psychology*, *109*(5), 653–665. https://doi.org/10.1037/edu0000161

Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction*, *16*(2), 104–121. https://doi.org/10.1016/j.learninstruc.2006.02.007

Glogger-Frey, I., Fleischer, C., Grüny, L., Kappich, J., & Renkl, A. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct

instruction. *Learning and Instruction*, *39*, 72–87. https://doi.org/10.1016/j.learninstruc.2015.05.001

Gschwendtner, T., Abele, S., & Nickolaus, R. (2009). Computersimulierte Arbeitsproben: Eine Validierungsstudie am Beispiel der Fehlerdiagnoseleistungen von Kfz-Mechatronikern [Can troubleshooting skills of car mechatronic technicians validly be assessed using computer-based simulations of real work sample?]. *Zeitschrift Für Berufs- Und Wirtschaftspädagogik*, *105*(4), 557–578.

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (Third edition). The Guilford Press.

Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavior Research Methods*, *41*(3), 924–936. https://doi.org/10.3758/BRM.41.3.924

Heitzmann, N., Fischer, F., Kühne-Eversmann, L., & Fischer, M. R. (2015). Enhancing diagnostic competence with self-explanation prompts and adaptable feedback. *Medical Education*, *49*(10), 993–1003. https://doi.org/10.1111/medu.12778

Hilbert, T. S., & Renkl, A. (2008). Concept mapping as a follow-up strategy to learning from texts: What characterizes good and poor mappers? *Instructional Science*, *36*(1), 53–73. https://doi.org/10.1007/s11251-007-9022-9

Hilbert, T. S., & Renkl, A. (2009). Learning how to use a computer-based concept-mapping tool: Self-explaining examples helps. *Computers in Human Behavior*, *25*(2), 267–274. https://doi.org/10.1016/j.chb.2008.12.006

Hilbert, T. S., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning and Instruction*, *18*(1), 54–65. https://doi.org/10.1016/j.learninstruc.2006.10.008

Hoogerheide, V. (2016). *Effects of Observing and Creating Video Modeling Examples on Cognitive and Motivational Aspects of Learning*.

Hoogerheide, V., Loyens, S. M. 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., 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

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

Klein, M., Otto, B., Fischer, M. R., & Stark, R. (2019). Fostering medical students' clinical reasoning by learning from errors in clinical case vignettes: Effects and conditions of additional prompting procedures to foster self-explanations. *Advances in Health Sciences Education*, *24*(2), 331–351. https://doi.org/10.1007/s10459-018-09870-5

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*, 1997. https://doi.org/10.3389/fpsyg.2017.01997

Klepsch, M., & Seufert, T. (2020). Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. *Instructional Science*, *48*(1), 45–77. https://doi.org/10.1007/s11251-020-09502-9

Klepsch, M., & Seufert, T. (2021). Making an Effort Versus Experiencing Load. *Frontiers in Education*, *6*, 645284. https://doi.org/10.3389/feduc.2021.645284

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

McLaren, B. M., & Isotani, S. (2011). When Is It Best to Learn with All Worked Examples? In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Artificial Intelligence in Education* (Vol. 6738, pp. 222–229). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21869-9_30

Meier, J., Spliethoff, L., Hesse, P., Abele, S., Renkl, A., & Glogger-Frey, I. (2022). Promoting car mechatronics apprentices' diagnostic strategy with modeling examples: Development and evaluation of a simulation-based learning environment. *Studies in Educational Evaluation*, *72*, 101117. https://doi.org/10.1016/j.stueduc.2021.101117

Montoya, A. K. (2019). Moderation analysis in two-instance repeated measures designs: Probing methods and multiple moderator models. *Behavior Research Methods*, *51*(1), 61–82. https://doi.org/10.3758/s13428-018-1088-6

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology*, *38*(1), 30–38. https://doi.org/10.1037/0022-0167.38.1.30

Najar, A. S., & Mitrovic, A. (2013). Examples and Tutored Problems: How Can Self-Explanation Make a Difference to Learning? In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial Intelligence in Education: 16th International Conference, AIED 2013, Memphis, TN, USA, July 9-13, 2013. Proceedings* (Vol. 7926, pp. 339–348). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39112-5

Nievelstein, F., van Gog, T., van Dijck, G., & Boshuizen, H. P. A. (2013). The worked example and expertise reversal effect in less structured tasks: Learning to reason about legal cases. *Contemporary Educational Psychology*, *38*(2), 118–125. https://doi.org/10.1016/j.cedpsych.2012.12.004

Norwig, K., Güzel, E., Hartmann, S., & Gschwendtner, T. (2021). „Tools to tap into the content of human minds" – Think-Aloud-Interviews und Cognitive Labs als zentrale Bausteine zur Identifikation von Barrieren in Fehlerdiagnoseprozessen bei Auszubildenden des Kfz-Handwerks und zur Entwicklung addressantespezifischer Lehr-/Lernarrangements ["Tools to tap into the content of human minds" – The use of Think-Aloud and Cognitive Laboratory Interviews for Identifying Cognitive Barriers of Car Mechatronics Apprentices Diagnosing Electronic Car Systems and for Developing Educational Interventions Tailored to the Apprentices' Needs.]. *Zeitschrift Für Berufs- Und Wirtschaftspädagogik*, *117*, 658–693. https://doi.org/10.25162/zbw-2021-0025

Renkl, A. (1997). Learning from Worked-Out Examples: A Study on Individual Differences. *Cognitive Science*, *21*(1), 1–29. https://doi.org/10.1207/s15516709cog2101_1

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, *38*(1), 1–37. https://doi.org/10.1111/cogs.12086

Renkl, A., & Eitel, A. (2019). Self-Explaining: Learning About Principles and Their Application. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge Handbook of Cognition and Education* (1st ed., pp. 528–549). Cambridge University Press. https://doi.org/10.1017/9781108235631.022

Renkl, A., Hilbert, T., & Schworm, S. (2009). Example-Based Learning in Heuristic Domains: A Cognitive Load Theory Account. *Educational Psychology Review*, *21*(1), 67–78. https://doi.org/10.1007/s10648-008-9093-4

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from Worked-Out Examples: The Effects of Example Variability and Elicited Self-Explanations. *Contemporary Educational Psychology*, *23*(1), 90–108. https://doi.org/10.1006/ceps.1997.0959

Roelle, J., Hiller, S., Berthold, K., & Rumann, S. (2017). Example-based learning: The benefits of prompting organization before providing examples. *Learning and Instruction*, *49*, 1–12. https://doi.org/10.1016/j.learninstruc.2016.11.012

Schalk, L., Roelle, J., Saalbach, H., Berthold, K., Stern, E., & Renkl, A. (2020). Providing worked examples for learning multiple principles. *Applied Cognitive Psychology*, *34*(4), 813–824. https://doi.org/10.1002/acp.3653

Schmitz, F. M., Schnabel, K. P., Stricker, D., Fischer, M. R., & Guttormsen, S. (2017). Learning communication from erroneous video-based examples: A double-blind randomised controlled trial. *Patient Education and Counseling*, *100*(6), 1203–1212. https://doi.org/10.1016/j.pec.2017.01.016

Schneider, B. A., Avivi-Reich, M., & Mozuraitis, M. (2015). A cautionary note on the use of the Analysis of Covariance (ANCOVA) in classification designs with and without within-subject factors. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00474

Schunk, D. H. (1995). Self-efficacy, motivation, and performance. *Journal of Applied Sport Psychology*, *7*(2), 112–137. https://doi.org/10.1080/10413209508406961

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. (2009). The worked-example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior*, *25*(2), 258–266. https://doi.org/10.1016/j.chb.2008.12.011

Spöttl, G., Becker, M., & Musekamp, F. (2011). Anforderungen an Kfz-Mechatroniker und Implikationen für die Kompetenzerfassung [Requirements for automotive mechatronics technicians and implications for the assessment of competencies]. In R.

Nickolaus (Ed.), *Lehr-Lernforschung in der gewerblich-technischen Berufsbildung (Zeitschrift für Berufs- und Wirtschaftspädagogik: Sonderbd. 25)* (pp. 37–53). Steiner.

Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction*, *16*(2), 165–169. https://doi.org/10.1016/j.learninstruc.2006.02.005

Sweller, J., Ayres, P. L., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

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., & 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. A. Rawson (Eds.), *The Cambridge Handbook of Cognition and Education* (1st ed., pp. 183–208). Cambridge University Press. https://doi.org/10.1017/9781108235631.009

van Harsel, M., Hoogerheide, V., Verkoeijen, P., & 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 Merriënboer, J. J. G., Kester, L., & Paas, F. (2006). Teaching complex rather than simple tasks: Balancing intrinsic and germane load to enhance transfer of learning. *Applied Cognitive Psychology*, *20*(3), 343–352. https://doi.org/10.1002/acp.1250

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, *47*(1), 513–539. https://doi.org/10.1146/annurev.psych.47.1.513

Appendix A

Answers to Self-Explanation Prompts

The following Table A-1 gives an overview of examples of apprentices' answers to retrospective and anticipatory self-explanation prompts. The table also includes a column explaining whether a prompt was answered meaningfully. As such, correctly answered prompts were coded as being answered meaningfully. Sometimes participants provided explanations that were correct in content but referred to the wrong step of the diagnostic strategy. For example, participants explained step 2 when being prompted to explain step 1. However, we still coded such responses as meaningful as it was apparent that the participants had at least attempted to refer to the diagnostic strategy in their answer. Finally, when participants answered the prompts by entering only single letters or blanks, or when they made entries without any reference to car diagnoses, these responses were coded as not meaningful.

Table A-1

Overview of Exemplary Apprentices' Answers on Retrospective and Anticipatory Self-Explanation Prompts

Type of Prompt	Exemplary answer	Coded as
Retrospective	The second step has just been completed. This step is about making a plan. To do this step, I take notes so that I can compare my measurements. This step is important so that I can proceed in a structured way.	Meaningful
Anticipatory	next is step 2 - planning a test for the assumption. to do this step, I answer the three questions. this step is important so that I know how to measure quickly and correctly.	
Retrospective	XXXX	Not meaningful
Anticipatory	I don't care	

Manuscript 3: Video-Based Modelling Examples and Comparative Self-Explanation Prompts for Teaching a Complex Problem-Solving Strategy

Meier, J., Hesse, P., Abele, S., Renkl, A., & Glogger-Frey, I. (2023). Video-Based Modelling Examples and Comparative Self-Explanation Prompts for Teaching a Complex Problem-Solving Strategy. *Manuscript Submitted for Publication.*

My first authorship reflects the fact that I was responsible for planning the studies and design the material, for the data analyses, and for the preparation of the paper.

Abstract

Background

In example-based learning, examples are often combined with generative activities, such as self-explanation aiming at comparing example cases. Such comparisons can easily be prompted for static text-based worked examples. For video-based modelling examples that are transient in format, however, side-by-side comparisons are hard to implement as two videos cannot be watched and processed simultaneously.

Objectives

To allow for such comparisons, we propose to combine video-based modelling examples with a static representation (e.g., a summarizing table) of the observed problem-solving process. Such a representation is a non-transient medium and thus better suited for comparison than a video. Moreover, learners are provided with an additional representation of an alternative solution approach to the same problem. A comparative self-explanation prompt then asks learners to compare the different solution approaches by comparing the different representations.

Methods

In an experiment, we taught 118 automotive apprentices a complex strategy for diagnosing car malfunctions. Apprentices were assigned to one of three conditions: apprentices learned with modelling examples and (1) <u>comparative</u> self-explanation prompts (i.e., static representations were provided side-by-side), (2) or <u>sequential</u> self-explanation prompts (i.e., static representations were provided subsequently), (3) or with neither modelling examples nor prompts. Diagnostic strategy knowledge and skills were assessed before and after the intervention, cognitive load was retrospectively assessed after the intervention.

Results and Conclusions

Modelling examples had beneficial effects on diagnostic knowledge but not on diagnostic skills. In addition, there were no effects of examples and prompts on cognitive load. We assume that apprentices would have needed more practice opportunities. Moreover, the comparative prompts seem to be promising for stronger learners with more prior knowledge.

Takeaways

The combination with static representations of observed problem-solving strategies is useful for video-based modelling examples while comparative prompts seem promising for stringer learners. Further research, especially on the effects on cognitive load, is necessary.

Keywords. Example-based learning, modelling examples, comparative self-explanation prompts, contrasting cases, complex problem-solving, diagnostic strategy

Practitioner Notes

What is already known about the topic?

- Text and video examples are widely used in education.
- Text examples often include self-explanation prompts that ask learners to compare several examples.
- For video examples, such comparison prompts have seldom been investigated, because comparisons are difficult to implement for transient videos.

What does this paper add?

- This paper investigates how to combine video examples with static summaries of processes that have been shown in the video example to allow for comparisons.

Implications for practice

- Combining static summaries of processes with video examples allows for comparisons
- Direct side-by-side comparisons of such summaries seem to be more promising for stronger learners, but further research is needed.

Introduction

When learning how to solve a problem with a specific problem-solving strategy, learners are often first instructed about the strategy and then receive examples that illustrate how the strategy is applied to solve a problem. Such examples can take the form of text-based worked examples or video-based modelling examples. In comparison to practising how to apply the instructed problem-solving strategy, example study induces less learning-irrelevant cognitive load, that is, extraneous cognitive load. Consequently, more cognitive capacities are available for learning (Renkl, 2014; Sweller, 2006; van Gog et al., 2019). To benefit from the lower extraneous cognitive load, learners need to increase their germane cognitive load, that is, engage in learning-related activities, such as self-explanations (Renkl & Eitel, 2019; Wylie & Chi, 2014). Such self-explanation activities may occur spontaneously, but they can also be prompted (Atkinson et al., 2003; Renkl et al., 1998). Another learning activity that is often applied in example-based learning to promote germane cognitive load is comparing several examples (Alfieri et al., 2013; Rittle-Johnson & Star, 2011). Comparing examples allows learners to discover similarities and differences between these examples (Gentner, 2010). Example comparison usually includes self-explanation (Sidney et al., 2015) and can thus be prompted. However, comparing examples also induces heavy demands on working memory in general (Holyoak, 2012), especially in terms of intrinsic cognitive load. Consequently, for very complex examples, only stronger learners who can manage the increased demands can be expected to benefit from comparisons or *comparative self-explanation prompts* while weaker learners might rather benefit from studying examples sequentially (i.e., *sequential self-explanation prompts*) instead of comparing them (Rittle-Johnson et al., 2009).

Moreover, to allow for the comparison of examples, learners need to be able to look at these examples simultaneously. This is possible with static and non-transient text-based worked examples, that can be studied side by side at the same time (e.g., Rittle-Johnson &

Star, 2007). But when learning with dynamic and transient video-based modelling examples, comparing examples is more difficult to implement, as learners cannot watch two video examples at the same time and thus a direct comparison of distinctive features of the examples is difficult. One possibility to allow for comparisons of video-based modelling examples is asking learners not to compare and self-explain the dynamic videos directly but rather have them compare a *static and non-transient representation* of the problem-solving process that was demonstrated in the video, for example, a table-based summary comprising the most important steps of the problem-solving process.

Consequently, the present study aimed to compare the effects of *comparative and sequential self-explanation prompts* on cognitive load and learning outcomes for learners with differing levels of prior knowledge when learning a complex problem-solving strategy with a combination of video-based modelling examples and static representation of observed problem-solving processes.

Example-based Learning

When learning how to solve a problem, learning from examples is very effective and superior to just practising how to solve problems (e.g., Renkl, 2014; Sweller, 2006; van Gog et al., 2019). This is true not only for well-structured domains, such as mathematics, where mostly text-based worked examples are used (e.g., Schalk et al., 2020). Examples are also beneficial for learning in ill-structured domains, where often video-based modelling examples are used (van Gog et al., 2019; van Gog & Rummel, 2010). For example, Schmitz and colleagues (2017) successfully used short (i.e., 30 s – 51 s) erroneous video-based examples to teach healthcare students a strategy for delivering bad news. Frerejean and colleagues (2018) showed that also longer (i.e., 10 minutes) video-based modelling examples are effective for teaching information problem-solving skills (i.e., researching and critically assessing information).

Usually, the effectiveness of examples is explained with *cognitive load theory* (Paas & van Gog, 2006; Sweller, 2006): Cognitive load theory (CLT) distinguishes three types of cognitive load that together load on a working memory that is limited in capacity (Sweller et al., 1998): In this paper, we refer to the still widely used conception of CLT from 1998[2]. First, *extraneous* cognitive load (ECL) is considered learning-irrelevant and unproductive load that is often induced by the (sub-optimal) design of learning materials. Second, *germane* cognitive load (GCL) results from learning-related activities. Ideally, most of a learner's working memory capacity would be taken by GCL. Finally, *intrinsic* cognitive load (ICL) is mainly determined by the complexity of the learning material (i.e., element interactivity) and the learner's prior knowledge. The more elements learners need to consider simultaneously during learning, the higher the ICL they experience. However, learners with more prior knowledge are able to consider more elements in the learning material at the same time or as one element and will thus experience a lower ICL as they (known as chunking; Endres et al., 2022; Sweller et al., 2011). Given the same task (i.e., same element interactivity) and the same learners (i.e., same prior knowledge), ICL is considered fixed. Therefore, to ensure that sufficient working memory resources are available for the GCL, the ECL must be minimised. Accordingly, most guidelines for the design of learning materials in terms of CLT refer to reducing ECL (e.g., Mayer & Moreno, 2003).

When learning how to solve a problem, one possibility to reduce ECL (and potentially increase GCL) is to provide worked or modelling examples (Paas & van Gog, 2006; Sweller, 2006). When novices try to solve a problem, they often apply weak problem-solving

[2] Recently, Sweller et al. (2019) presented updates to the theory mainly with innovations to the concept of germane cognitive load: Most importantly, these updates suggest that intrinsic and germane load can be classified as one type of load, resulting in only two types of load that can be distinguished. However, we refer to the 1998 concept with three types of cognitive load as this concept is the basis for most of the research we refer to and because we had this original concept in mind when developing the learning materials and experimental design. Moreover, most cognitive load questionnaires assume a three-factor model (Krieglstein et al., 2022). Furthermore, a recent confirmatory factor analysis found stronger support for the three-factor model than for the two-factor model (Zavgorodniaia et al., 2020).

strategies (Renkl, 2014). As the application of such weak strategies is hardly conducive to learning in terms of schema construction, it can also be considered a learning-irrelevant activity inducing ECL (van Gog et al., 2019). However, studying worked or modelling examples makes the use of weak problem-solving strategies unnecessary. Learners can focus on the problem-solving steps provided in the example. Consequently, ECL is reduced. These freed cognitive capacities can then be used for learning (i.e., for GCL), which explains the possible beneficial effect for learning outcomes (Renkl, 2014; Sweller, 2006). However, for learners to benefit from the reduced ECL, they must actively use these freed-up cognitive capacities for learning, that is, they must engage in generative learning-related activities to produce GCL. Examples of such activities are self-explaining or comparing examples (Renkl, 2014).

Self-explanations

An effective generative learning activity is *self-explaining* (e.g., Bisra et al., 2018; Rittle-Johnson et al., 2017). Self-explanations can occur spontaneously (Chi et al., 1989), but learners can also be *prompted to self-explain* (Renkl & Eitel, 2019). For example, Schworm and Renkl (2007) developed video-based examples that modelled argumentative principles to teach student teachers declarative knowledge about argumentation as well as argumentation skills. While the examples in general fostered declarative knowledge about argumentation, argumentation skills were only promoted when examples were combined with self-explanation prompts. Similarly, Hefter et al. (2014) investigated the effects of a web-based training intervention consisting of video examples and self-explanation prompts to promote knowledge about and application of argumentative strategies (Hefter et al., 2014), to promote the disposition to apply these strategies (Hefter et al., 2015), and both (Hefter et al., 2018). In all three studies, the self-explanation quality mediated the intervention's beneficial effect on the respective outcome measures.

When learners explain content from examples to themselves, this can make them engage more deeply with the underlying principles of the example, as they basically try to make sense of the given learning materials (Wylie & Chi, 2014). Thus, self-explanations promote GCL (Renkl et al., 2009).

Example Comparison

Another effective generative learning activity is asking learners to compare examples (Alfieri et al., 2013; Rittle-Johnson & Star, 2011). According to the *example comparison principle* (Renkl, 2014), comparing several examples benefits the development of abstract schemata and allows learners to discover similarities and differences between these examples (Gentner, 2010). A specific type of example comparisons are *critical feature comparisons* (Renkl, 2014) or *contrasting cases* (e.g., Glogger-Frey et al., 2017; Schwartz et al., 2011): These are sets of examples, where examples share many features but differ only in one or few critical features so that this difference stands out. Such example sets could be used to demonstrate how different problem-solving strategies are applied to the same problem or how the same strategy is applied in a more and less efficient way (e.g., Glogger-Frey et al., 2015; Rittle-Johnson & Star, 2007). For example, Rittle-Johnson and Star (2007) designed text-based worked examples for algebraic equations that were solved with more and less efficient solution methods. Students in seventh grade worked in pairs and received these worked examples either side-by-side with self-explanation prompts that encouraged comparisons (i.e., comparison condition) or sequentially with prompts that did not encourage comparisons (i.e., control condition). Students in the comparison condition showed more improvement in procedural knowledge and procedural flexibility (i.e., the ability to select and apply the correct problem-solving strategy depending on certain features of the problem to be solved) and demonstrated similar improvement in conceptual knowledge. When learners are encouraged by comparison prompts to distinguish and judge the correct and incorrect use of a strategy they engage with the learning material in depth (e.g., Rittle-Johnson & Star, 2007). Thereby, the development of a differentiated mental representation of the problem-solving strategy to be learned is promoted. From a CLT perspective, one could thus argue that the beneficial effects of comparisons are due to an increase in GCL.

However, the effects of (different types of) comparisons on learning depend on the learners' prior knowledge: Rittle-Johnson et al. (2009) tested seventh- and eighth-grade students' prior knowledge of algebra and then provided them with pairs of worked examples of solved linear algebraic equations in three conditions: a method comparison condition (i.e., worked example pairs included the same equations but were solved with different methods), a problem comparison condition (i.e., worked example pairs included different equations that were solved with the same method), and a sequential condition without comparison. In terms of learning outcomes, students with little prior knowledge benefited most in the problem comparison condition or the sequential condition. Students with more prior knowledge benefited the most when they compared methods (Rittle-Johnson et al., 2009). The authors argue that this finding is an example of an expertise-reversal effect (Kalyuga et al., 2003): The instructional approach that was most beneficial for novice learners with little prior knowledge was not beneficial for learners with higher prior knowledge and vice versa.

This effect can again be explained with the CLT: Comparison processes in general induce heavy demands on working memory (Holyoak, 2012). Thus, comparing examples side by side should induce a substantially higher ICL in comparison to subsequentially studying examples one by one, as element interactivity is much higher. Consequently, especially for complex problems that are high in ICL, only learners with more prior knowledge who can manage the increased demands (i.e., the higher ICL), can be expected to benefit from comparisons, that is, to perceive a higher GCL and have better learning outcomes. For learners with less prior knowledge, comparisons with complex problems are likely to produce cognitive overload. For these learners, sequential study of examples might be more beneficial (e.g., Rittle-Johnson et al., 2009).

Static Representations for Comparisons of Video-based Modelling Examples

Example comparisons can be easily implemented for text-based worked examples in well-structured domains. The static and non-transient format of these text-based examples allows comparisons of the critical features of the examples side by side (see Glogger-Frey et al., 2015; Rittle-Johnson & Star, 2007). However, example comparisons seem to be less suitable for ill-structured domains using video-based modelling examples illustrating complex multiple-step strategies. Comparing such video examples would require learners to either watch two videos at the same time or pause the videos repeatedly. Hence, it is not surprising, that (at least to our knowledge) there is no research published in peer-reviewed journals on comparing video examples.

Against this background, we propose that after watching (parts of) a video example, learners are provided with a static representation of the (so far) observed problem-solving process as the basis for comparing and explaining. Such a representation could be realized, for example, by a text-based or graphical summary (e.g., a table, a bullet-point summary, or a mindmap) of the problem-solving process so far, highlighting the critical steps, or it might be an overview of a product created in the problem-solving process. Such a representation consisting of text and/or image is a non-transient medium and thus better suited for comparison than a transient video. To allow for comparison, learners are also provided with an additional representation of the current state of an alternative solution to the same problem. This could be, for example, a summary of how the same problem was solved with a different strategy – possibly also resulting in a different (e.g. lower quality) result (i.e., method comparison; Rittle-Johnson et al., 2017).

Present Study and Research Questions

The present study was conducted with automotive apprentices learning a complex problem-solving strategy, namely a strategy for diagnosing complex car malfunctions. Such complex malfunctions cannot be easily diagnosed using the usually available computer-based expert system (Abele, 2018; Abele & von Davier, 2019). Applying the instructed diagnostic strategy included filling in and executing a so-called diagnosis plan. This plan is a table in which the apprentices (or other mechanics who apply this strategy) list possible causes, diagnostic measurement methods and their requirements, measurement results, as well as conclusions drawn therefrom for the present malfunction. Thus, a filled-in diagnosis plan represents a summary of a diagnostic process. We developed modelling examples in a screencast format showing an expert applying the diagnostic strategy and filling out a diagnosis plan in a computer simulation (Gschwendtner et al., 2009; Meier et al., 2022, 2023). While studying the modelling examples, apprentices received self-explanation prompts that referred to the filled-in diagnosis plans that served as static representations of the problem-solving process. For each modelling example, we designed two versions of diagnosis plans: the expert plan from the modelling examples and a novice plan providing an overview of an alternative but worse solution to the same problem. In a first condition, apprentices received these two diagnosis plans side by side and answered *comparative self-explanation prompts*, that is, they were asked to explain and compare how well the expert and the novice had filled out their diagnosis plans, respectively. In a second condition, apprentices answered *sequential self-explanation prompts*, that is, first for the expert plan and then for the novice plan apprentices self-explained how well first the expert and then the novice had filled out their respective diagnosis plans without comparing them directly. In a third condition (*control*), which can be regarded as a learning by problem-solving condition, the apprentices did not receive any modelling examples and therefore no self-explanation prompts or diagnosis plans, but tried to diagnose the respecting car malfunctions on their own.

We investigated the effects of modelling examples and the different self-explanation prompts on apprentices' diagnostic knowledge and skills and cognitive load. Diagnostic knowledge and skills were assessed before and after the intervention. Cognitive load was assessed once after the intervention. Building on cognitive load theory (e.g., Paas & van Gog, 2006) and the worked or modelling example effect (Renkl, 2014; van Gog et al., 2019), we investigated the following hypotheses:

- H1: We expected apprentices learning with modelling examples to experience a lower extraneous and a higher germane cognitive load than apprentices trying to solve the respective problems on their own.
- H2: We expected a greater increase in diagnostic knowledge and skills from a pretest to a posttest for apprentices learning with the modelling examples in comparison with apprentices attempting to solve the respective problems on their own.

We assumed that, due to the higher element interactivity, the comparative self-explanation prompts would be more demanding than the sequential self-explanation prompts. Following research on example comparison and the expertise reversal effect (e.g., Kalyuga et al., 2003; Renkl, 2014; Rittle-Johnson et al., 2009, 2017; Rittle-Johnson & Star, 2007), regarding the comparison of comparative and sequential self-explanation prompts, we hypothesized the following:

- H3: We expected the more demanding comparative self-explanation prompts to induce a higher intrinsic cognitive load than the sequential self-explanation prompts.

- H4: We expected apprentices with low prior knowledge to perceive a higher germane cognitive load when learning with sequential self-explanation prompts than when learning with comparative self-explanation prompts. In contrast, apprentices with high prior knowledge were expected to experience a higher germane cognitive load when learning with comparative self-explanation prompts than when learning with sequential self-explanation prompts.
- H5: For participants with low prior knowledge, we expected a larger increase in diagnostic knowledge and skills when learning with sequential self-explanation prompts than when learning with comparative self-explanation prompts. In contrast, for participants with high prior knowledge, we expected a larger increase in diagnostic knowledge and skills when learning with comparative self-explanation prompts than when learning with sequential self-explanation prompts.

Methods

Participants and Design

We conducted a computer-based experiment in two sessions with three experimental conditions at German vocational schools. Session 1 included the pretest. In session 2, the intervention and the posttest took place. The sessions were conducted during school hours in the apprentices' classrooms. All material was presented in digital form.

We conducted two a-priori power analyses with Gpower 3.1 (Faul et al., 2007) to calculate the required sample sizes. We aimed for a power of .80. Based on previous studies on the worked example effect (e.g., Nievelstein et al., 2013; Schwonke et al., 2009; van Gog et al., 2011) and self-explanation prompts (e.g., Atkinson et al., 2003; Hilbert & Renkl, 2009), we expected medium effect sizes (e.g., Cohen's $f > .25$ or $\eta^2 > .06$; Cohen, 1988). For the analyses regarding H1, H3, and H4 (i.e., analyses of variance, ANOVAs), the required sample size was $N = 128$. For the analyses regarding H2 and H5 (i.e., repeated measures analyses of variance, RM-ANOVAs), the required sample size was $N = 34$. In total, 136 apprentices participated in session one and 132 participated in session two. However, only 118 apprentices participated in both sessions and can thus be included in the analyses. Thus, the sample size falls a little short of the sample size ($N = 128$) required for the ANOVAs, but is sufficient for the RM-ANOVAs. As data collection took place shortly before the summer holidays (6 weeks), a secondary data collection to achieve the required sample size was not feasible. A larger sample may have enabled us to demonstrate additional (smaller) effects. However, the effects we did discover can still be interpreted.

Apprentices were 20.08 years old ($SD = 2.04$), 114 were male, and 4 were female. For most apprentices ($n = 103$), German was their only first language, 13 reported another first language, and two apprentices reported that German was not their first language. Although these two non-native speakers might be at a disadvantage because of the German test and learning material, we did not exclude them from the analyses because they did not show extreme values (i.e. values more than 3 standard deviations above or below the mean) on any of the variables in either session. Regarding general school education, 13 apprentices had a university entrance qualification (*Abitur*), 96 apprentices had a secondary school leaving certificate (*Mittlere Reife*), and nine apprentices had a lower secondary school leaving certificate (*Hauptschulabschluss*). At the beginning of the first session, we assessed the apprentices' general prior knowledge about car diagnoses with a prior knowledge test. Then the apprentices completed the pretest on all repeated measures variables (i.e., diagnostic knowledge and skills). In the second session, the apprentices received the intervention, rated their cognitive load, and completed the posttest on the repeated measures variables. For the intervention, apprentices were assigned to one of three experimental conditions. In a first condition, apprentices learned with modelling examples and *comparative self-explanation prompts* ($n = 42$). In a second condition, apprentices received modelling examples and *sequential self-explanation prompts* ($n = 39$). In a third condition (*control*), apprentices received no modelling examples and, thus, no self-explanation prompts ($n = 37$). Instead, these apprentices tried to diagnose the malfunctions that were illustrated in the modelling examples themselves in the computer simulation.

Intervention

Diagnostic Strategy

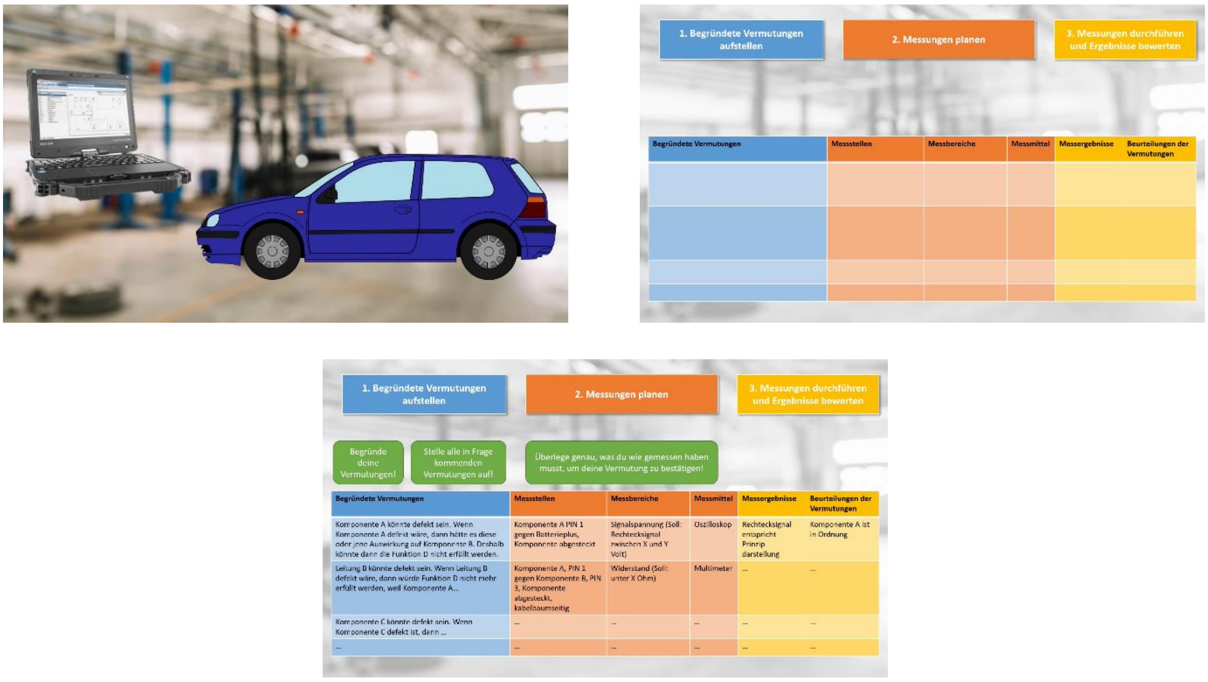
In collaboration with subject-matter experts and on basis of respective literature (e.g., Abele, 2014), we developed an intervention in which apprentices learned about a strategy to diagnose complex car malfunctions. This strategy comprised three steps: (1) When diagnosing car malfunctions, apprentices should first formulate hypotheses about possible causes for the present malfunctions. These hypotheses should be reasoned, that is, based on the functional relationships of different relevant components in an automotive system. To formulate these reasoned hypotheses, apprentices learned about two underlying rules, namely the *reasoning rule* (i.e., 'formulate what function is probably impaired, what components are relevant to accomplishing that function, and how those components typically work together to accomplish the function'), and the *rule of completeness* (i.e., 'formulate all possible hypotheses and do not just rely on your first idea'). (2) The second diagnostic strategy step comprises the planning of (electrotechnical) measurements to verify the hypotheses. The planning includes collecting information on (2a) measuring points, (2b) measuring range, and (2c) measuring equipment for each of the hypotheses. We emphasised the importance of these three points with the so-called *carefulness rule* (i.e., 'think carefully about what and how you must have measured to confirm your hypothesis'). (3) In the third and last diagnostic step the planned measurements are executed and the measurement results and with it the respective hypotheses are evaluated. Proceeding through these three steps of the diagnostic strategy was supported by a diagnosis plan. This diagnosis plan was a six-column table with the six columns corresponding to (1) reasoned hypotheses, (2a) measuring points, (2b) measuring ranges, (2c) measuring equipment, (3a) measuring results, and (3b) evaluations of the hypotheses. To teach apprentices this strategy as well as how to fill out the diagnosis plan, we developed an intervention consisting of instructional videos, two modelling examples, and three self-explanation prompts for each of the modelling examples. These learning materials are described below. Note that for the first modelling example, the instructional videos and the modelling example were presented in an *interleaved format*. This means that the instructional videos explaining the strategy initially and the first modelling example, which consisted of one video per step illustrating the application of the strategy, were shown in alternation. A detailed explanation and rationale for this format can be found in the appendix.

Instructional Videos

Six instructional videos (see Figure 1) briefly explained the three diagnostic steps with the three underlying rules and how to fill out the diagnosis plan along these steps (overall duration: 10:38 minutes). Participants from all three conditions received these instructional videos and thus learned about the diagnostic strategy.

Figure 1

Screenshots of the Instructional Videos



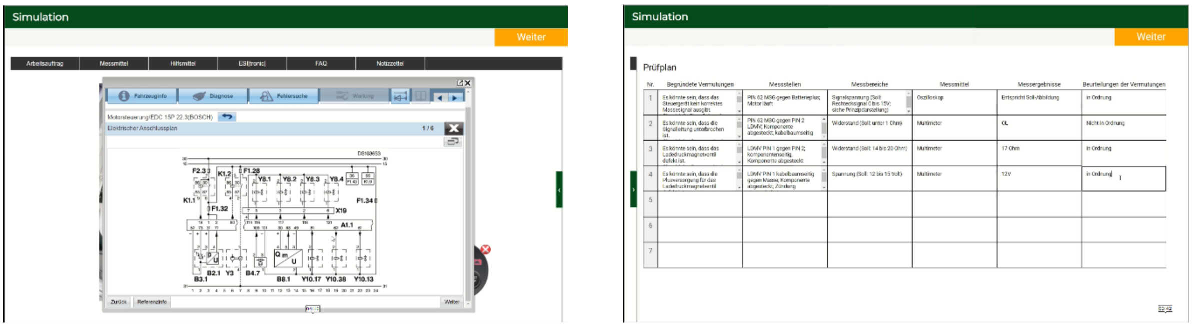
Note. The screenshots are from the instructional videos explaining the three diagnostic steps using the diagnosis plan. The three green boxes in the bottom screenshot contain the rules underlying the first two diagnostic steps.

Modelling Examples (First Experimental Variation)

The modelling examples showed an expert diagnosing a malfunction by applying the steps of the diagnostic strategy in the computer simulation while also filling out a diagnosis plan (see Figure 2). The expert verbalised his cognitive processes. Corresponding to the three diagnostic steps, both modelling examples consisted of three videos. The three videos of the first modelling example took 20:12 minutes, the second modelling example took 13:50 minutes.

Figure 2

Screenshots of the First Modelling Example



Note. The left screenshot shows how the expert uses the computer-based expert system to open an electrical circuit diagram. These diagrams illustrate the interrelationships between electrotechnical components and are thus an important resource for formulating hypotheses. The right screenshot shows how the expert fills in the diagnosis plan.

The modelling examples constituted the first experimental variation as – dependent on the experimental condition – apprentices either learned with modelling examples or tried to solve the respective problem on their own, that is, they tried to diagnose the malfunction on their own.

Self-Explanation Prompts (Second Experimental Variation)

Three self-explanation prompts were given after diagnostic steps 1 and 2 in the modelling examples that asked learners to explain how well the three underlying rules in these diagnostic steps, namely the reasoning rule, the rule of completeness, and the carefulness rule were applied in the example. The prompts had an open-book format, that is, the respective rule was displayed at the top of the page (Hiller et al., 2020). Besides the rule, apprentices were provided with (a relevant section of) the diagnosis plan as it had been filled out by the expert in the modelling examples (i.e., expert solution). The apprentices also received a novice solution of the same diagnostic step for the same problem, namely they were provided with (a section of) a diagnosis plan as it had been filled out by a less experienced hobby mechanic.

The format of the self-explanation prompts constituted the second experimental variation: In the *comparative self-explanation prompt condition*, the apprentices received the expert solution and the novice solution at the same time side by side and were instructed to compare the solutions, to look for similarities and differences, and to explain how differently well the expert and the hobby mechanic applied the respecting rule. After each prompt, apprentices were provided with a solution: In a written text it was explained and demonstrated that, for example regarding the rule of completeness, the expert had formulated all possible hypotheses while the hobby mechanic's diagnosis plan was not complete. In the *sequential self-explanation prompt condition*, the apprentices received the expert solution and the novice solution successively. For both the expert and the novice solution the apprentices were asked to explain how well the expert or the novice had applied the respecting rule. After providing an answer, apprentices received the corresponding solution. Note that apprentices in the control condition did not receive modelling examples and thus also no prompts. Instead, these apprentices tried to diagnose the malfunctions that were illustrated in the modelling examples themselves in the computer simulation.

Testing Materials

We used different tests to investigate the effects of modelling examples and comparative versus sequential self-explanation prompts: Only in the pretest in session 1, we assessed apprentices' *diagnosis relevant reception competence* (i.e., prerequisite knowledge for car diagnoses). Both in the pretest and posttest, various tests were administered to measure the apprentices' development in *diagnostic knowledge* and *diagnostic skills*. In the posttest only, we assessed the participants' *cognitive load* during learning. These tests are described below. In most tests, closed and open question items were used. Closed items were scored automatically. For open question items, the first author and a subject matter expert (i.e., the second author) developed a coding scheme. Then, a student assistant and the first author scored 25% of all answers and adjusted the coding schemes until achieving an interrater reliability of Cohen's $\kappa > 0.6$. Then the student assistant independently scored the remaining answers. For some items, coding required very detailed automotive diagnostic expertise and no sufficient reliability could be established in the codings of the student assistant and the first author. In these cases, the first author coded the answers. For items where this applies, this is noted separately in the detailed description below.

Prior Knowledge

As a measure of general prior knowledge, we assessed the apprentices' *diagnosis-relevant reception competence*. This competence includes the ability to read different diagnosis-relevant documents, such as electrical circuit diagrams, and is thus required for successful diagnoses of automotive malfunctions. For this test, we used a selection of five out of 24 items from the diagnosis-relevant reception competence (DRC) test by Norwig and colleagues (2021), as in our previous study (Meier et al., 2022, 2023). To prevent floor and ceiling effects, we selected items for their midrange solution range (ranging from 32% to 71% in Norwig et al., 2021) and with the highest item-total correlation (> 0.43 for all 5 items in Norwig et al., 2021). Apprentices could achieve up to five points on this test.

Diagnostic Knowledge and Skills

We applied several tests to measure the apprentices' diagnostic knowledge and skills both in sessions 1 and 2. In the *strategy description test*, apprentices were asked two questions: First, they were asked to describe by which steps they would proceed in a diagnosis when there is only little assistance from a computer-based expert system (i.e., complex diagnosis). Apprentices could achieve six points for this question. The interrater reliability between the student assistant and the first author was acceptable both for session 1 (Cohen's $\kappa = .864$) and session 2 (Cohen's $\kappa = .689$). In the second question of the strategy description test apprentices described what would go through their minds when reading the error memory of a car and thinking about why the component/subsystem named in the error memory entry might be malfunctioning. Apprentices could achieve six points for this question. This second question required extensive knowledge of electrotechnical car systems and was thus coded by the first author. Taken together, the maximum achievable score for the strategy description test was nine points.

Second, in the *strategy completion test*, apprentices were successively provided with three diagnostic scenarios – one scenario for each of the diagnostic steps. For each scenario, apprentices answered different open and closed questions to describe or carry out (parts of) and thereby complete the three diagnostic steps. For example, in the scenario regarding the second step, after reading the respective diagnostic scenario, apprentices studied a circuit diagram and described an appropriate measurement, thereby completing the second diagnostic step, that is, planning measurements. Hence, this test assessed scaffolded diagnostic skills. All open questions in the strategy completion test were scored by the first author and not by the student assistant. Apprentices could achieve up to 47 points on this test.

Eventually, to test independent diagnostic skills, in the *strategy application test* participants performed two diagnoses in the computer simulation both in sessions 1 and 2. For these independent diagnoses, apprentices first read a description of the malfunction and then diagnosed it. Eventually, apprentices were asked to describe the malfunction and how it could be repaired. Apprentices had 30 minutes to complete one diagnosis. The maximum score for each diagnosis was four points, resulting in a maximum score of eight points for the strategy application test. Interrater agreement was acceptable (first diagnosis, session 1: Cohen's $\kappa = .625$; second diagnosis, session 1: Cohen's $\kappa = .756$; first diagnosis, session 2: Cohen's $\kappa = .657$; second diagnosis, session 2: Cohen's $\kappa = .681$).

Motivation

In our previous study (Meier et al., 2022, 2023) we assessed the apprentices' motivation (i.e., self-efficacy, interest, perception of challenge, and incompetence fear). In this study, we assessed the apprentices' motivation with the same items on a seven-point Likert-scale to ensure that neither the modelling examples nor the different prompts had negative effects on the apprentices' motivation. However, we did not have any hypotheses regarding the effects of conditions on the apprentices' motivation. Both in the pretest and posttest, before performing the first diagnosis in the computer simulation, we assessed the apprentices' current motivation (Vollmeyer & Rheinberg, 2000) with a 19-item questionnaire on a 7-point Likert-scale. With five items, we measured the apprentices' *self-efficacy* regarding the subsequent diagnosis (Bandura, 2006). Reliability was good (Session 1: Cronbach's $\alpha = 0.885$; Session 2: Cronbach's $\alpha = 0.998$). Five items assessed the apprentices' *interest* in car diagnoses (Schiefele, 1991). Reliability was again good (Session 1: Cronbach's $\alpha = 0.847$; Session 2: Cronbach's $\alpha = 0.853$). With four items we examined the extent to which the apprentices perceived the upcoming diagnosis in the simulation as a *challenge* (Session 1: Cronbach's $\alpha = 0.654$; Session 2: Cronbach's $\alpha = 0.997$) and five items assessed the apprentices' *incompetence fear* (Session 1: Cronbach's $\alpha = 0.903$; Session 2: Cronbach's $\alpha = 0.998$).

Cognitive Load

After the intervention, we assessed the apprentices' *intrinsic* (two items), *germane* (two items), and *extraneous cognitive load* (three items) while learning on a seven-point Likert-scale (Klepsch et al., 2017; Klepsch & Seufert, 2020, 2021). Reliability was acceptable (intrinsic load: Cronbach's $\alpha = 0.62$; germane load: Cronbach's $\alpha = 0.64$; extraneous load: Cronbach's $\alpha = 0.61$).

Procedure

The procedures in sessions 1 and 2 are displayed in Table 1. All material was presented on computers in digital form in a page-based learning environment. Once participants left a page, they could not go back.

Table 1

Procedures in Sessions 1 and 2

Phase	Content	Planned duration in min	Actual duration in min
	Session 1		
Phase 1	Introduction to study and computer simulation, demographics	35	31
	Assessment of motivation	5	4
	Strategy description test	10	4
Break		15	22
Phase 2	Strategy application test: First diagnosis in simulation	30	28
	Strategy application test: Second diagnosis in simulation	30	22
	Strategy completion test	20	25
Break		15	23
Phase 3	Diagnosis-relevant reception competence test	10	7
	Expertise of car technology test [a]	50	50
TOTAL SESSION 1		**220**	**216**
	Session 2		
Phase 1	Refresher on computer simulation	5	4
	Instructional videos and modelling example 1 in interleaved format	55	44
	Modelling example 2	30	24
	Cognitive load rating	5	1
Break		15	27
Phase 2	Assessment of motivation	5	2
	Strategy description test	10	3
	Strategy application test: First diagnosis in simulation	30	20
	Strategy application test: Second diagnosis in simulation	30	17
Break		15	28
Phase 3	Strategy completion test	20	14
TOTAL SESSION 2		**220**	**184**

[a] This expertise test on different automotive systems was not related to research questions investigated in the present paper and is thus not presented or analysed here.

Results

Exploratory analyses revealed participants with scores on dependent variables in the pretest or posttest that were more than 3 standard deviations below or above the grand mean. These participants were removed as outliers (n = 6). Exploratory analyses also showed a large variance in the quality of responses to the self-explanation prompts, as some apprentices gave meaningless answers (e.g., only single letters). Apprentices who answered less than 80% of the prompts meaningfully were consequently excluded from further analyses (n = 14). The final sample included in the analyses thus consisted of N = 99 apprentices with n = 38 apprentices learning with modelling examples and comparative self-explanation prompts, n = 27 apprentices learning with examples and sequential self-explanation prompts, and n = 34 apprentices in the control condition.

Effects of Modelling Examples

To analyze the effects of modelling examples versus independent problem-solving on cognitive load, we conducted an analysis of variance (ANOVA) with *example condition* (i.e., modelling examples yes vs. no) as between-subjects variable. To analyze the effects on diagnostic knowledge and skills and motivation, we performed a repeated measures analysis of variance (RM-ANOVA) with *example condition* as between-subjects variable and *timepoint* (pretest vs. posttest) as within-subjects variable. Table 2 displays descriptive data. Table 3 shows the results of the test of statistical significance. A significance level of .05 applies to all analyses. As effect size we used $\eta^2_{partial}$ with .01, .06, and .14 corresponding to a small, medium, and large effect, respectively (Lakens, 2013).

Table 2

Descriptive Data of Dependent Variables for the Control Condition (i.e., no Modelling Examples) and Modelling Examples Condition

	No modelling examples (*n* = 34)				Modelling examples (*n* = 65)			
	Pretest		Posttest		Pretest		Posttest	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Intrinsic Load[a]	-	-	4.54	1.14	-	-	4.15	1.40
Germane Load[a]	-	-	3.33	0.99	-	-	3.24	1.25
Extraneous Load[a]	-	-	5.01	1.29	-	-	5.28	1.41
Strategy Description Test Score[b]	0.97	0.87	1.41	1.58	0.86	0.98	1.95	1.61
Strategy Completion Test Score[c]	14.78	5.31	16.65	6.41	16.58	4.95	21.09	7.22
Strategy Application Test Score[d]	0.85	1.42	2.29	1.90	0.78	1.28	1.58	1.58
Self-efficacy[a]	4.24	1.39	4.29	1.32	4.40	1.19	4.27	1.12
Interest[a]	5.00	1.31	4.79	1.18	5.33	1.05	4.88	1.26
Perception of Challenge[a]	4.86	0.89	4.50	1.11	4.95	1.04	4.61	1.24
Fear of Failure[a]	3.15	1.54	3.15	1.60	2.90	1.50	3.06	1.51

[a]7-point Likert-scale ranging from 1 = *absolutely not true* to 7 = *absolutely true*
[b]0-9 points
[c]0-47 points
[d]0-8 points

Table 3

Main and Interaction Effects of the Example Condition and Timepoint on Dependent Variables

Analysis	Hypothesis	Independent Variable(s)	Dependent Variables	Statistical test results			
				df	F	p	$\eta^2_{partial}$
ANOVA		Example Condition	Intrinsic load	1, 97	1.961	.165	.020
	H1		Germane load	1, 97	.139	.710	.001
	H1		Extraneous load	1, 97	.817	.368	.008
RM-ANOVA		Timepoint	Strategy Description Test Score	1, 97	22.350	**<.001**	.187
			Strategy Completion Test Score	1, 97	28.460	**<.001**	.227
			Strategy Application Test Score	1, 97	41.389	**<.001**	.299
			Self-efficacy	1, 97	.165	.686	.002
			Interest	1, 97	9.076	**.003**	.086
			Perception of Challenge	1, 97	12.380	**<.001**	.113
			Fear of Failure	1, 97	.426	.515	.004
	H2	Timepoint* Example Condition	Strategy Description Test Score	1, 97	4.030	**.047**	.040
	H2		Strategy Completion Test Score	1, 97	4.880	**.030**	.048
			Strategy Application Test Score	1, 97	3.388	.069	.034
			Self-efficacy	1, 97	.897	.346	.009
			Interest	1, 97	1.171	.282	.012
			Perception of Challenge	1, 97	.008	.928	.000
			Fear of Failure	1, 97	.495	.483	.005

Regarding cognitive load, which was measured only once after the learning phase, the ANOVA indicated no effects of modelling examples as compared to independent problem-solving on participants' intrinsic, germane, or extraneous cognitive load. Regarding variables measured both in pre- and posttest, the RM-ANOVA indicated two-way interaction effects of timepoint and example condition on the participants' scores in the strategy description test (small to medium effect) and the strategy completion test (small to medium effect). Figures 3 and 4 illustrate these interactions. For the strategy description test and the strategy completion test, apprentices who learned with modelling examples had a higher increase in test scores from the pretest to the posttest than apprentices in the control group who did not learn with the modelling examples.

Figure 3

Interaction Effect of Timepoint and Example Condition on Strategy Description Test Score

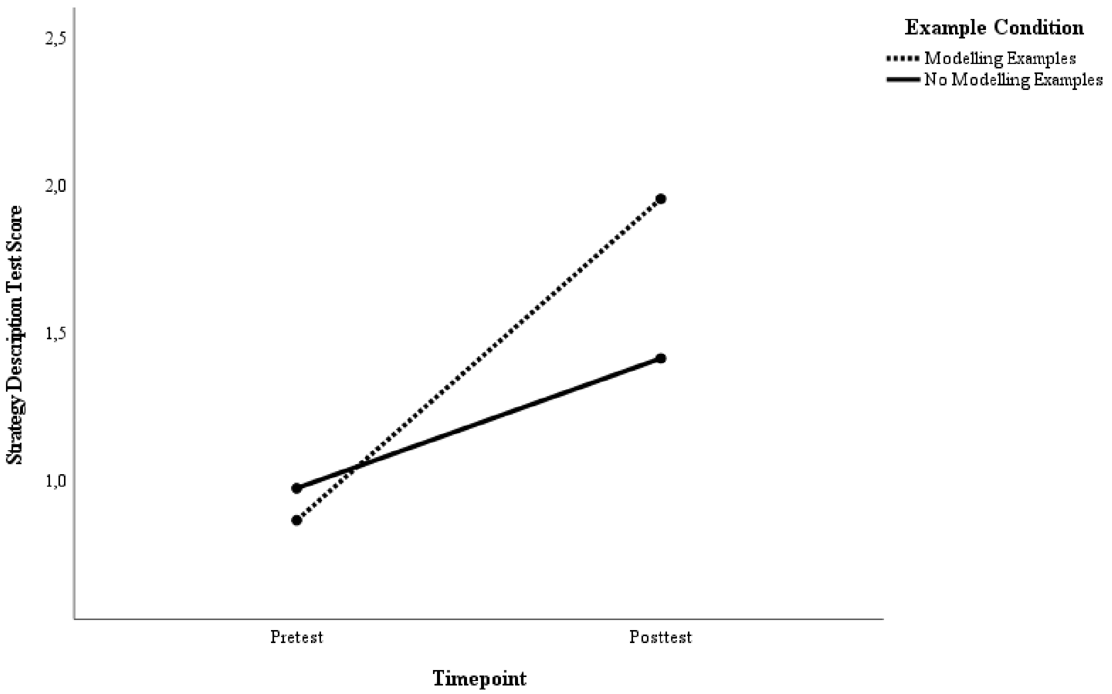
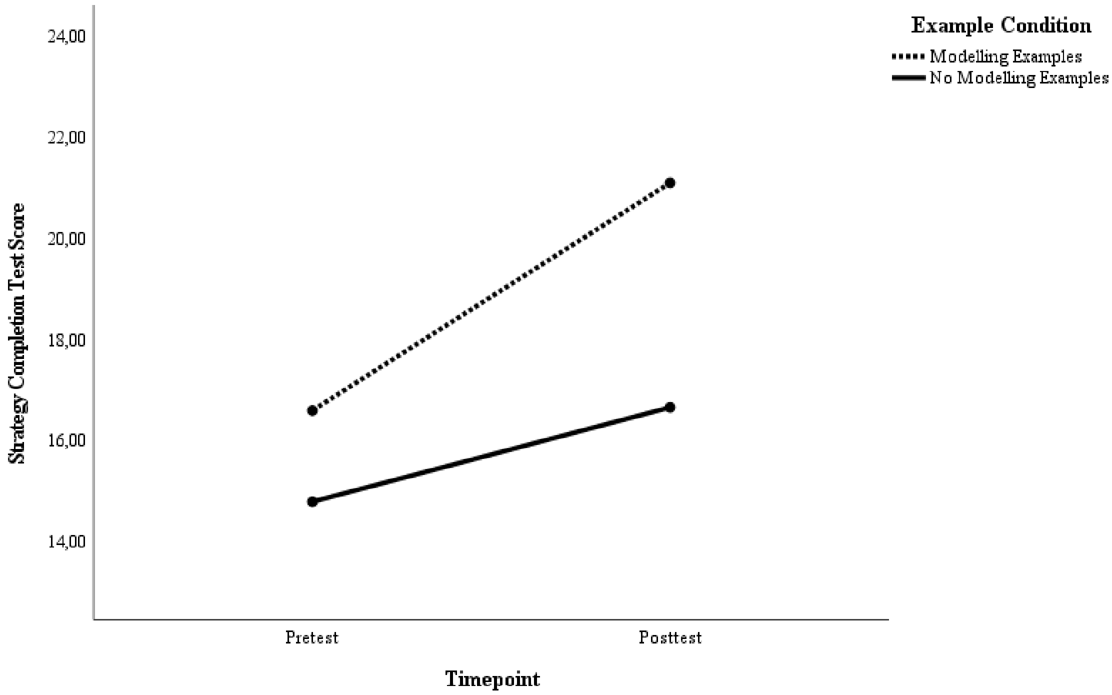


Figure 4

Interaction Effect of Timepoint and Example Condition on Strategy Completion Test Score



Besides these interaction effects, we found three main effects of the timepoint on dependent variables that were not affected by interaction effects, namely on the participants‘ strategy application test score (large effect), their interest in diagnoses (medium effect), and their perception of diagnoses as a challenge (medium to large effect). Descriptive data in Table 2 indicate that the participants‘ scores in the strategy application test increased from the pretest to the posttest. Participants‘ interest and their perception of diagnoses as a challenge decreased from the pretest to the posttest.

Taken together, hypothesis H1 is not supported and must therefore be rejected, as participants in the modelling example condition did not perceive a lower extraneous and higher germane cognitive load. Hypothesis H2 is partially supported, as participants in the modelling example condition showed a greater increase in diagnostic knowledge (i.e., strategy description test) and scaffolded diagnostic skills (i.e, strategy completion test) but not in independent diagnostic skills (i.e., strategy application test).

Effects of Comparative versus Sequential Self-explanation Prompts

To analyze the effects of comparative versus sequential self-explanation prompts on cognitive load in the posttest, we conducted an ANOVA with *prompt condition* (i.e., comparative vs. sequential prompts) as between-subjects variable. To analyze the effects on diagnostic knowledge and skills and motivation, we performed an RM-ANOVA with *prompt condition* as between-subjects variable and *timepoint* (pretest vs. posttest) as within-subjects variable. To test for moderating effects of prior knowledge, we included the *diagnosis-relevant reception competence (DRC) test score* as additional continuous factors for both analyses. This factor was grand mean-centered for these analyses (Schneider et al., 2015). Table 4 shows descriptive data. Table 5 provides the results of the tests on statistical significance.

Table 4

Descriptive Data of Dependent Variables and the Moderator Variable for the Sequential and Comparative Self-explanation Prompt Condition

	Sequential SE-Prompts (*n* = 27)				Comparative SE-Prompts (*n* = 38)			
	Pretest		Posttest		Pretest		Posttest	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
DRC Test Score[a]	3.44	0.97	-	-	3.55	1.13	-	--
Intrinsic Load[b]	-	-	4.26	1.62	-	-	4.08	1.23
Germane Load[b]	-	-	5.17	1.39	-	-	5.36	1.43
Extraneous Load[b]	-	-	3.44	1.38	-	-	3.10	1.15
Strategy Description Test Score[c]	0.78	0.97	1.44	1.37	0.92	1.00	2.32	1.68
Strategy Completion Test Score[d]	15.67	5.41	19.80	7.26	17.24	4.56	22.01	7.14
Strategy Application Test Score[e]	0.81	1.33	1.63	1.64	0.76	1.26	1.55	1.55
Self-efficacy[b]	4.40	1.12	4.40	1.24	4.41	1.26	4.18	1.02
Interest[b]	5.21	1.04	4.79	1.17	5.41	1.07	4.94	1.32
Perception of Challenge[b]	5.13	1.04	4.70	1.32	4.82	1.04	4.54	1.19
Fear of Failure[b]	2.77	1.54	2.93	1.63	3.00	1.48	3.15	1.43

[a]0-5 points; this test was a prior knowledge measure and used as a moderation variable
[b]7-point Likert-scale ranging from 1 = *absolutely not true* to 7 = *absolutely true*
[c]0-9 points
[d]0-47 points
[e]0-8 points

Table 5

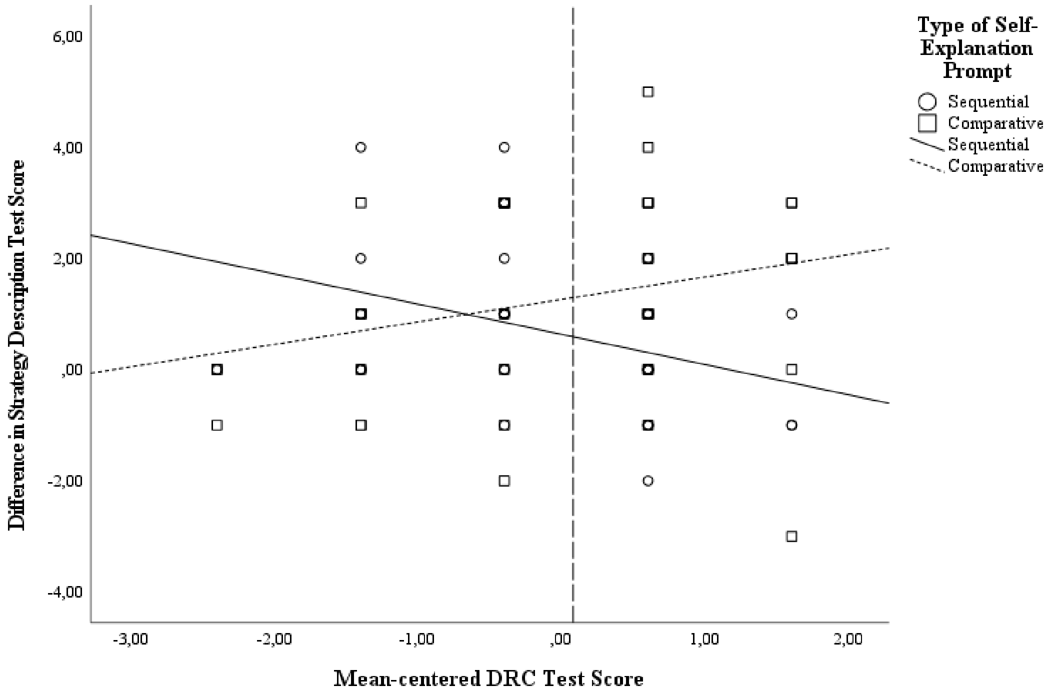
Main and Interaction Effects of the Prompt Condition, the Moderation Variables, and Timepoint on Dependent Variables

Analysis	Research Question	Independent Variable(s)	Dependent Variables	Statistical test results			
				df	F	p	$\eta^2_{partial}$
ANOVA	H3	Prompt Condition	Intrinsic Load	1, 61	.358	.552	.006
			Germane Load	1, 61	.252	.617	.004
			Extraneous Load	1, 61	1.078	.303	.017
		Prompt Condition*DRC Test Score	Intrinsic Load	1, 61	.717	.492	.023
	H4		Germane Load	1, 61	.611	.546	.020
			Extraneous Load	1, 61	.672	.515	.022
RM-ANOVA		Timepoint	Strategy Description Test Score	1, 61	26.244	**<.001**	.301
			Strategy Completion Test Score	1, 61	40.710	**<.001**	.400
			Strategy Application Test Score	1, 61	20,876	**<.001**	.255
			Self-efficacy	1, 61	1.085	.302	.017
			Interest	1, 61	11.505	**.001**	.159
			Perception of Challenge	1, 61	7.856	**.007**	.114
			Fear of Failure	1, 61	1.994	.163	.032
		Timepoint*Prompt Condition	Strategy Description Test Score	1, 61	2.739	.103	.043
			Strategy Completion Test Score	1, 61	.114	.737	.002
			Strategy Application Test Score	1, 61	.004	.947	.000
			Self-efficacy	1, 61	1.027	.315	.017
			Interest	1, 61	.023	.879	.000
			Perception of Challenge	1, 61	.239	.627	.004
			Fear of Failure	1, 61	.011	.918	.000
	H5	Timepoint*Prompt Condition*DRC Test Score	Strategy Description Test Score	1, 61	4.171	**.045**	.064
	H5		Strategy Completion Test Score	1, 61	1.296	.259	.021
	H5		Strategy Application Test Score	1, 61	.422	.518	.007
			Self-efficacy	1, 61	.029	.864	.000
			Interest	1, 61	1.131	.292	.018
			Perception of Challenge	1, 61	.740	.393	.012
			Fear of Failure	1, 61	.461	.500	.008

Regarding cognitive load, which was measured only once in the posttest, the ANOVA indicated no effects of comparative versus sequential self-explanation prompts (i.e., neither main effects nor interaction effects with the DRC test score) on participants‘ intrinsic, germane, or extraneous cognitive load. Regarding variables measured in the pretest and posttest, the RM-ANOVA indicated one significant three-way interaction of timepoint, prompt condition and mean-centered DRC test score on strategy description test scores (medium effect). Figure 5 indicates that participants with low DRC test scores benefitted (in terms of a higher increase in strategy description test scores) from sequential self-explanation prompts while participants with higher DRC test scores rather benefitted from comparative self-explanation prompts. To further explore this interaction, the Johnson-Neyman procedure (Hayes & Matthes, 2009; Montoya, 2019) was applied by using the SPSS-macro PROCESS by Hayes (2022). We tested where in the distribution of mean-centered DRC test scores the condition (i.e., comparative versus sequential prompts) had a statistically significant effect on the difference of strategy description test scores, calculated as posttest score minus pretest score. We found that the interaction effect was significant for learners with mean-centered DRC test scores larger than 0.08, which is essentially the half of participants with higher prior knowledge, $t(74) = 1.99$, $p = .05$.

Figure 5

Scatter Plot of Grand Mean-centered DRC Test Scores Against the Difference in Strategy Description Test Scores for Both Prompt Conditions



Note. The depicted effect is only significant for mean-centered DRC test scores > 0.08, i.e., right of the vertical longer dashed line.

No two-way interactions of timepoint and prompt condition were found. We found significant main effects of timepoint on the strategy description test score (large effect), strategy completion test score (large effect), strategy application test score (large effect), interest (large effect), and perception of challenge (medium to large effect). These effects correspond to the effects we had already observed when comparing the two example conditions.

Taken together, hypotheses H3 and H4 need to be rejected: Apprentices learning with comparative self-explanation prompts did not experience a higher intrinsic cognitive load than apprentices learning with sequential self-explanation prompts (H3). Moreover, apprentices‘ prior knowledge, as measured with the DRC test, did not moderate the effects of the different self-explanation prompts on germane cognitive load (H4). Hypothesis H5 is partially supported: regarding the strategy description test, apprentices with higher prior knowledge benefitted more from comparative self-explanation prompts.

Discussion

The present paper aimed to compare the effects of comparative and sequential self-explanation prompts in combination with static representations of problem-solving processes on cognitive load and learning outcomes when learning a complex problem-solving strategy with video-based modelling examples for learners with different levels of prior knowledge. In the following paragraphs, we will first discuss findings regarding cognitive load (i.e., hypotheses H1, H3, and H4). In the second section, we will discuss findings regarding learning outcomes (i.e., hypotheses H2 and H5). Eventually, limitations of the study and implications for future research will be discussed.

Effects on Cognitive Load

Regarding modelling examples we expected that – irrespective of their prior knowledge – apprentices learning with modelling examples would experience a lower extraneous cognitive load (ECL) and a higher germane cognitive load (GCL) than apprentices in the control group trying to solve the respective problems on their own (H1). Contrary to H1, we detected no effects of modelling examples on ECL or GCL. Regarding the comparison of comparative and sequential self-explanation prompts, we expected the comparative prompts to induce a higher intrinsic cognitive load (ICL) than the sequential self-explanation prompts (H3). Moreover, low prior knowledge apprentices were expected to experience a higher GCL when learning with sequential self-explanation prompts, and higher prior knowledge apprentices were expected to experience a higher GCL when learning with comparative self-explanation prompts (H4). However, contrary to H3 and H4, no effects of prompt condition on ICL and no effects of prompt condition on GCL depending on participants' prior knowledge could be found.

These findings fail to support the widely shared assumption that worked examples reduce ECL and potentially increase GCL as they make the use of ineffective problem-solving strategies unnecessary and free up working memory capacities for learning (Renkl, 2014; Sweller, 2006). Interestingly, many of the research contributions that have argued in this direction in the past have either used only one general measure of cognitive load: For example, Spanjers et al. (2012) and Van Gog et al. (2008) used the single-item mental effort scale by Paas (1992). Other publications have not investigated the effects of worked or modelling examples on cognitive load at all, but have only analysed the effects on learning outcomes but still argued with CLT (e.g., Rourke & Sweller, 2009). One could therefore argue that the assumption of working examples reducing ECL is much less well covered by evidence than is widely supposed. Therefore, we suggest that researchers investigating the effects of worked or modelling examples on learning should also investigate the effects on cognitive load to close this gap (Sweller, 2017). Moreover, it would be important for researchers to agree on the same instruments to measure cognitive load to allow for overarching conclusions.

An exception and one first step in this direction is study 3 by Klepsch and Seufert (2020), who were involved in the development of the cognitive load instruments we also used in the present study. In their study, participants learned to solve complex math problems: In a learning phase, participants received such a math problem as well as the correct final result. In an example condition, participants additionally received the worked-out solution (i.e., a text-based worked example), while participants in the control or problem-solving condition had to work out the solution path themselves. The participants in the example condition reported a lower ECL and had better learning outcomes. However, Klepsch and Seufert (2020) found no effects of the examples on GCL. Although we did use the same instruments to measure cognitive load, we did not find any effects of the modelling examples on ECL.

One explanation for this might be that reliability was considerably worse in our study (i.e., while we obtained reliability estimates of Cronbach's α between 0.61 und 0.64, Klepsch & Seufert (2020) report reliability estimates of McDonald's ω between 0.74 and 0.83 for study 3). Besides reliability, there are also important differences between our study and the study by Klepsch and Seufert (2020): First, in the study by Klepsch and Seufert, the learning phase comprising a short introduction and study of two text-based worked examples took around ten minutes. In the present study, the learning phase both for the modelling example condition and the control condition took more than 60 minutes. The question arises if one single retrospective rating of cognitive load can be informative for a study phase of more than 60 minutes or whether a repeated measurement of cognitive load during this time would be preferable (Schmeck et al., 2015; van Gog et al., 2012). Second, while Klepsch and Seufert (2020) used rather simple text-based worked examples for a well-structured domain (i.e., mathematics), we used video-based modelling examples in a complex and ill-defined domain that were additionally combined with comparative or sequential self-explanation prompts. It seems reasonable to assume that subjective rating processes of cognitive load become more complex as the learning material gets more complex.

Effects on Learning Outcomes

Regarding the effects on learning outcomes, it can be noted that over all conditions apprentices benefitted from the intervention substantially, as there were large increases in scores on the diagnostic knowledge tests as well as when apprentices performed diagnoses in the computer simulation themselves. Regarding hypothesized effects, we expected a greater increase in diagnostic knowledge and skills for the apprentices learning with modelling examples (H2). Moreover, we expected low prior knowledge apprentices to benefit more from learning with modelling examples when learning with sequential self-explanation prompts, and higher prior knowledge apprentices were expected to benefit more in terms of learning outcomes when learning with comparative self-explanation prompts (H5).

Concerning these hypotheses, although we could not confirm the expected effects on cognitive load, both hypotheses regarding learning outcomes could be partially confirmed: Apprentices learning with modelling examples had a higher increase in diagnostic knowledge and they were also more proficient in applying the strategy when they worked on scaffolded tasks. However, these learning gains did not transfer to independent diagnoses in the computer simulation (H2). One possible explanation might be that apprentices experienced a utilization deficiency (Hübner et al., 2010; Miller, 1994). This term describes that using a newly learned strategy, which is not yet automated, requires so much cognitive capacity that only little capacity remains to process the new problem that is to be solved (i.e., the malfunction in the simulation to be diagnosed). Consequently, the application of the strategy to the novel context fails. Such a utilization deficiency can be overcome by practice, that is, apprentices would have needed more time and opportunity to practice the diagnostic strategy (possibly with additional support) before performing diagnoses on their own.

Eventually, regarding the effects of the different self-explanation prompts on learning outcomes depending on the learners' prior knowledge, for the strategy description test asking for declarative knowledge about the diagnostic strategy, the hypothesised effect could be partially confirmed: participants with higher prior knowledge rather benefitted from comparative self-explanation prompts. This finding is consistent with Rittle-Johnson et al. (2009) who found that in the domain of mathematics students with more prior knowledge learned more when they compared worked-examples.

Strengths and Limitations

The study makes an important contribution to improving vocational education for automotive apprentices: Although the diagnosis of car malfunctions is a crucial aspect of the work of car mechatronics technicians (Spöttl et al., 2011), only about 15 % of the apprentices master strategies for the diagnosis of complex malfunctions at the end of their apprenticeship (Abele & von Davier, 2019). Over all conditions, there were large increases in scores on the diagnostic knowledge tests as well as in the two independent diagnoses in the simulation. Accordingly, the developed intervention might be helpful to teach more apprentices a diagnosis for complex car malfunctions. Future research could investigate whether these positive effects persist over a longer period of time and whether they could also be confirmed in diagnoses on real vehicles.

Moreover, with the present study, we make a first proposal on how the positive effect of comparing examples can also be used for video-based modelling examples by combining them with static representations of the problem-solving process. Future studies on modelling examples could follow this direction and could investigate whether findings regarding the effects of comparisons of text-based worked examples also hold true for such representation-based comparisons of video-based modelling examples.

One limitation of the study concerns the procedure in the control group. While the instructional videos and the first modelling example were presented in an interleaved format for the modelling example conditions, such an interleaved format was not feasible for the control group (see Appendix A). For the apprentices in the control group, it would have meant first learning about only the first step of the strategy in the instructional videos. Then, analogous to the modelling examples, they would have had to carry out only the first step of the diagnosis, that is, formulating hypotheses in the simulation. Then they would have learned about the second step of the diagnostic strategy in the instructional videos and would have again only carried out the second step in the simulation themselves. Such a switch between instructional videos and the simulation would have required the simulation to save and reload intermediate states of the diagnostic process. This was technically not feasible.

Hence, apprentices in the control group did first watch all six instructional videos on the diagnostic strategy and then tried to solve the problem in the computer simulation.

Conclusion

The intervention in general had beneficial effects on both diagnostic knowledge and skills and could therefore be implemented in the vocational education of automotive mechatronics apprentices. Modelling examples had a beneficial effect on diagnostic knowledge but not on diagnostic skills. We assume that apprentices would have needed more practice opportunities to apply the diagnostic strategy. The comparative prompts seem promising for stronger learners when learning with video-based modelling examples. Finally, none of our hypotheses regarding cognitive load could be confirmed. We argue that future research focusing on the effects of examples on learning should also investigate the mediating effects of cognitive load and that, furthermore, similar items should be used to measure cognitive load across studies, possibly in different versions depending on the example conditions being studied.

References

Abele, S. (2014). Modellierung und Entwicklung berufsfachlicher Kompetenz in der gewerblich-technischen Ausbildung [Modelling and development of vocational competence in vocational-technical education and training]. Steiner.

Abele, S. (2018). Diagnostic problem-solving process in professional contexts: Theory and empirical investigation in the context of car mechatronics using computer-generated log-files. Vocations and Learning, 11(1), 133–159. https://doi.org/10.1007/s12186-017-9183-x

Abele, S., & von Davier, M. (2019). CDMs in vocational education: Assessment and usage of diagnostic problem-solving strategies in car mechatronics. In M. von Davier & Y.-S. Lee (Eds.), Handbook of Diagnostic Classification Models. Springer. https://doi.org/10.1007/978-3-030-05584-4_22

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning Through Case Comparisons: A Meta-Analytic Review. Educational Psychologist, 48(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

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. Journal of Educational Psychology, 95(4), 774–783. https://doi.org/10.1037/0022-0663.95.4.774

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. C. Urdan (Eds.), Self-Efficacy Beliefs of Adolescents (pp. 307–337). IAP - Information Age Pub., Inc.

Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing Self-Explanation: A Meta-Analysis. Educational Psychology Review, 30(3), 703–725. https://doi.org/10.1007/s10648-018-9434-x

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. Cognitive Science, 13(2), 145–182. https://doi.org/10.1207/s15516709cog1302_1

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed). L. Erlbaum Associates.

Endres, T., Lovell, O., Morkunas, D., Rieß, W., & Renkl, A. (2022). Can prior knowledge increase task complexity? – Cases in which higher prior knowledge leads to higher intrinsic cognitive load. British Journal of Educational Psychology, bjep.12563. https://doi.org/10.1111/bjep.12563

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175–191. https://doi.org/10.3758/BF03193146

Frerejean, J., van Strien, J. L. H., Kirschner, P. A., & Brand-Gruwel, S. (2018). Effects of a modelling example for teaching information problem solving skills. Journal of Computer Assisted Learning, 34(6), 688–700. https://doi.org/10.1111/jcal.12276

Gentner, D. (2010). Bootstrapping the Mind: Analogical Processes and Symbol Systems. Cognitive Science, 34(5), 752–775. https://doi.org/10.1111/j.1551-6709.2010.01114.x

Glogger-Frey, I., Fleischer, C., Grüny, L., Kappich, J., & Renkl, A. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. Learning and Instruction, 39, 72–87. https://doi.org/10.1016/j.learninstruc.2015.05.001

Glogger-Frey, I., Gaus, K., & Renkl, A. (2017). Learning from direct instruction: Best prepared by several self-regulated or guided invention activities? Learning and Instruction, 51, 26–35. https://doi.org/10.1016/j.learninstruc.2016.11.002

Gschwendtner, T., Abele, S., & Nickolaus, R. (2009). Computersimulierte Arbeitsproben: Eine Validierungsstudie am Beispiel der Fehlerdiagnoseleistungen von Kfz-Mechatronikern [Can troubleshooting skills of car mechatronic technicians validly be assessed using computer-based simulations of real work sample?]. Zeitschrift Für Berufs- Und Wirtschaftspädagogik, 105(4), 557–578.

Hayes, A. F. (2022). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (Third edition). The Guilford Press.

Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. Behavior Research Methods, 41(3), 924–936. https://doi.org/10.3758/BRM.41.3.924

Hefter, M. H., Berthold, K., Renkl, A., Riess, W., Schmid, S., & Fries, S. (2014). Effects of a training intervention to foster argumentation skills while processing conflicting scientific positions. Instructional Science, 42(6), 929–947. https://doi.org/10.1007/s11251-014-9320-y

Hefter, M. H., Renkl, A., Riess, W., Schmid, S., Fries, S., & Berthold, K. (2015). Effects of a training intervention to foster precursors of evaluativist epistemological understanding and intellectual values. Learning and Instruction, 39, 11–22. https://doi.org/10.1016/j.learninstruc.2015.05.002

Hefter, M. H., Renkl, A., Riess, W., Schmid, S., Fries, S., & Berthold, K. (2018). Training Interventions to Foster Skill and Will of Argumentative Thinking. The Journal of Experimental Education, 86(3), 325–343. https://doi.org/10.1080/00220973.2017.1363689

Hilbert, T. S., & Renkl, A. (2009). Learning how to use a computer-based concept-mapping tool: Self-explaining examples helps. Computers in Human Behavior, 25(2), 267–274. https://doi.org/10.1016/j.chb.2008.12.006

Hiller, S., Rumann, S., Berthold, K., & Roelle, J. (2020). Example-based learning: Should learners receive closed-book or open-book self-explanation prompts? Instructional Science, 48(6), 623–649. https://doi.org/10.1007/s11251-020-09523-4

Holyoak, K. J. (2012). Analogy and Relational Reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), The Oxford Handbook of Thinking and Reasoning (1st ed., pp. 234–259). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199734689.013.0013

Hübner, S., Nückles, M., & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. Learning and Instruction, 20(1), 18–29. https://doi.org/10.1016/j.learninstruc.2008.12.001

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. Educational Psychologist, 38(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. Frontiers in Psychology, 8, 1997. https://doi.org/10.3389/fpsyg.2017.01997

Klepsch, M., & Seufert, T. (2020). Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. Instructional Science, 48(1), 45–77. https://doi.org/10.1007/s11251-020-09502-9

Klepsch, M., & Seufert, T. (2021). Making an Effort Versus Experiencing Load. Frontiers in Education, 6, 645284. https://doi.org/10.3389/feduc.2021.645284

Krieglstein, F., Beege, M., Rey, G. D., Ginns, P., Krell, M., & Schneider, S. (2022). A Systematic Meta-analysis of the Reliability and Validity of Subjective Cognitive Load Questionnaires in Experimental Multimedia Learning Research. Educational Psychology Review, 34(4), 2485–2541. https://doi.org/10.1007/s10648-022-09683-4

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Frontiers in Psychology, 4. https://doi.org/10.3389/fpsyg.2013.00863

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. Educational Psychologist, 38(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Meier, J., Hesse, P., Abele, S., Renkl, A., & Glogger-Frey, I. (2023). Better Self-Explaining Backwards or Forwards? Video-Based Modelling Examples for Learning a Diagnostic Strategy and the Use of Different Self-Explanation Prompts. Manuscript Submitted for Publication.

Meier, J., Spliethoff, L., Hesse, P., Abele, S., Renkl, A., & Glogger-Frey, I. (2022). Promoting car mechatronics apprentices' diagnostic strategy with modeling examples: Development and evaluation of a simulation-based learning environment. Studies in Educational Evaluation, 72, 101117. https://doi.org/10.1016/j.stueduc.2021.101117

Miller, P. H. (1994). Individual differences in children's strategic behaviors: Utilization deficiencies. Learning and Individual Differences, 6(3), 285–307. https://doi.org/10.1016/1041-6080(94)90019-1

Montoya, A. K. (2019). Moderation analysis in two-instance repeated measures designs: Probing methods and multiple moderator models. Behavior Research Methods, 51(1), 61–82. https://doi.org/10.3758/s13428-018-1088-6

Nievelstein, F., van Gog, T., van Dijck, G., & Boshuizen, H. P. A. (2013). The worked example and expertise reversal effect in less structured tasks: Learning to reason about legal cases. Contemporary Educational Psychology, 38(2), 118–125. https://doi.org/10.1016/j.cedpsych.2012.12.004

Norwig, K., Güzel, E., Hartmann, S., & Gschwendtner, T. (2021). „Tools to tap into the content of human minds“ – Think-Aloud-Interviews und Cognitive Labs als zentrale Bausteine zur Identifikation von Barrieren in Fehlerdiagnoseprozessen bei Auszubildenden des Kfz-Handwerks und zur Entwicklung addressantespezifischer Lehr-/Lernarrangements ["Tools to tap into the content of human minds" – The use of Think-Aloud and Cognitive Laboratory Interviews for Identifying Cognitive Barriers of Car Mechatronics Apprentices Diagnosing Electronic Car Systems and for Developing Educational Interventions Tailored to the Apprentices' Needs.]. Zeitschrift Für Berufs- Und Wirtschaftspädagogik, 117, 658–693. https://doi.org/10.25162/zbw-2021-0025

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., & van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. Learning and Instruction, 16(2), 87–91. https://doi.org/10.1016/j.learninstruc.2006.02.004

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. Cognitive Science, 38(1), 1–37. https://doi.org/10.1111/cogs.12086

Renkl, A., & Eitel, A. (2019). Self-Explaining: Learning About Principles and Their Application. In J. Dunlosky & K. A. Rawson (Eds.), The Cambridge Handbook of Cognition and Education (1st ed., pp. 528–549). Cambridge University Press. https://doi.org/10.1017/9781108235631.022

Renkl, A., Hilbert, T., & Schworm, S. (2009). Example-Based Learning in Heuristic Domains: A Cognitive Load Theory Account. Educational Psychology Review, 21(1), 67–78. https://doi.org/10.1007/s10648-008-9093-4

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from Worked-Out Examples: The Effects of Example Variability and Elicited Self-Explanations. Contemporary Educational Psychology, 23(1), 90–108. https://doi.org/10.1006/ceps.1997.0959

Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. ZDM, 49(4), 599–611. https://doi.org/10.1007/s11858-017-0834-z

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. Journal of Educational Psychology, 99(3), 561–574. https://doi.org/10.1037/0022-0663.99.3.561

Rittle-Johnson, B., & Star, J. R. (2011). The Power of Comparison in Learning and Instruction: Learning Outcomes Supported by Different Types of Comparisons. In Psychology of Learning and Motivation (Vol. 55, pp. 199–225). Elsevier. https://doi.org/10.1016/B978-0-12-387691-1.00007-7

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. Journal of Educational Psychology, 101(4), 836–852. https://doi.org/10.1037/a0016026

Roelle, J., Hiller, S., Berthold, K., & Rumann, S. (2017). Example-based learning: The benefits of prompting organization before providing examples. Learning and Instruction, 49, 1–12. https://doi.org/10.1016/j.learninstruc.2016.11.012

Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognise designers' styles. Learning and Instruction, 19(2), 185–199. https://doi.org/10.1016/j.learninstruc.2008.03.006

Schalk, L., Roelle, J., Saalbach, H., Berthold, K., Stern, E., & Renkl, A. (2020). Providing worked examples for learning multiple principles. Applied Cognitive Psychology, 34(4), 813–824. https://doi.org/10.1002/acp.3653

Schiefele, U. (1991). Interest, learning, and motivation. Educational Psychologist, 26(3–4), 299–323. https://doi.org/10.1080/00461520.1991.9653136

Schmeck, A., Opfermann, M., van Gog, T., Paas, F., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: Differences between immediate and delayed ratings. Instructional Science, 43(1), 93–114. https://doi.org/10.1007/s11251-014-9328-3

Schmitz, F. M., Schnabel, K. P., Stricker, D., Fischer, M. R., & Guttormsen, S. (2017). Learning communication from erroneous video-based examples: A double-blind randomised controlled trial. Patient Education and Counseling, 100(6), 1203–1212. https://doi.org/10.1016/j.pec.2017.01.016

Schneider, B. A., Avivi-Reich, M., & Mozuraitis, M. (2015). A cautionary note on the use of the Analysis of Covariance (ANCOVA) in classification designs with and without within-subject factors. Frontiers in Psychology, 6. https://doi.org/10.3389/fpsyg.2015.00474

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. Journal of Educational Psychology, 103(4), 759–775. https://doi.org/10.1037/a0025140

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V., & Salden, R. (2009). The worked-example effect: Not an artefact of lousy control conditions. Computers in Human Behavior, 25(2), 258–266. https://doi.org/10.1016/j.chb.2008.12.011

Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. Journal of Educational Psychology, 99(2), 285–296. https://doi.org/10.1037/0022-0663.99.2.285

Sidney, P. G., Hattikudur, S., & Alibali, M. W. (2015). How do contrasting cases and self-explanation promote learning? Evidence from fraction division. Learning and Instruction, 40, 29–38. https://doi.org/10.1016/j.learninstruc.2015.07.006

Spanjers, I. A. E., Gog, T., & Merriënboer, J. J. G. (2012). Segmentation of Worked Examples: Effects on Cognitive Load and Learning: Segmentation of Worked Examples. Applied Cognitive Psychology, 26(3), 352–358. https://doi.org/10.1002/acp.1832

Spöttl, G., Becker, M., & Musekamp, F. (2011). Anforderungen an Kfz-Mechatroniker und Implikationen für die Kompetenzerfassung [Requirements for automotive mechatronics technicians and implications for the assessment of competencies]. In R. Nickolaus (Ed.), Lehr-Lernforschung in der gewerblich-technischen Berufsbildung (Zeitschrift für Berufs- und Wirtschaftspädagogik: Sonderbd. 25) (pp. 37–53). Steiner.

Sweller, J. (2006). The worked example effect and human cognition. Learning and Instruction, 16(2), 165–169. https://doi.org/10.1016/j.learninstruc.2006.02.005

Sweller, J. (2017). The Role of Independent Measures of Load in Cognitive Load Theory. In R. Zheng (Ed.), Cognitive load measurement and application: A theoretical framework for meaningful research and practice. Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315296258-1/role-independent-measures-load-cognitive-load-theory-john-sweller

Sweller, J., Ayres, P. L., & Kalyuga, S. (2011). Cognitive load theory. Springer. https://doi.org/10.1007/978-1-4419-8126-4

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. Educational Psychology Review, 31(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10(3), 251–296. https://doi.org/10.1023/A:1022193728205

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., Kirschner, F., Kester, L., & Paas, F. (2012). Timing and Frequency of Mental Effort Measurement: Evidence in Favour of Repeated Measures: Measuring mental effort. Applied Cognitive Psychology, 26(6), 833–839. https://doi.org/10.1002/acp.2883

van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. Learning and Instruction, 18(3), 211–222. https://doi.org/10.1016/j.learninstruc.2007.03.003

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. A. Rawson (Eds.), The Cambridge Handbook of Cognition and Education (1st ed., pp. 183–208). Cambridge University Press. https://doi.org/10.1017/9781108235631.009

Vollmeyer, R., & Rheinberg, F. (2000). Does motivation affect performance via persistence? Learning and Instruction, 10(4), 293–309. https://doi.org/10.1016/S0959-4752(99)00031-6

Wylie, R., & Chi, M. T. H. (2014). The Self-Explanation Principle in Multimedia Learning. In R. E. Mayer (Ed.), The Cambridge Handbook of Multimedia Learning (2nd ed., pp. 413–432). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.021

Zavgorodniaia, A., Duran, R., Hellas, A., Seppala, O., & Sorva, J. (2020). Measuring the Cognitive Load of Learning to Program: A Replication Study. United Kingdom & Ireland Computing Education Research Conference., 3–9. https://doi.org/10.1145/3416465.3416468

Appendix A: Interleaved Presentation Format of Instructional Videos and Modelling Examples

For the first modelling example, instructional videos and the modelling example were presented in an *interleaved format*: After a general introduction to the diagnostic strategy, only the first diagnostic step was explained in the instructional videos. This video was followed by the first part of the first modelling example which showed the expert only performing the first step in the simulation. After the corresponding self-explanation prompts, apprentices watched the instructional video explaining the second diagnostic step followed by the second part of the modelling example and so on. We chose this interleaved format for the following reason: In a previous study on the same learning topic, we found no beneficial effects of the modelling examples on participants' learning outcomes (Meier et al., 2023). In this study, the first learning phase took 35 minutes and comprised five instructional videos that already included partial examples (16:33 minutes) and four practice tasks that presumably supported knowledge organisation well (Roelle et al., 2017). This extensive instruction may have provided all apprentices with enough knowledge of the respective strategy rendering the modelling examples useless. In the present study, to omit the partial examples within the instructional videos, we decided on the interleaved format in learning phase one, in which only parts of the diagnostic strategy are instructed and then immediately illustrated by parts of the subsequent modelling examples. For technical reasons, this interleaved format was not viable for the control group. Apprentices in this group did first watch all six instructional videos on the diagnostic strategy and then tried to solve the problem in the simulation.

When learning with the second modelling example, apprentices had already been introduced to the diagnostic strategy. Hence, here apprentices only watched the three videos of the second modelling examples and worked on the three self-explanation prompts, but we did not provide apprentices with additional instructional videos.

General Discussion

The overarching aim of this dissertation was to investigate the use of video-based modelling examples and different types of self-explanation prompts for teaching AMT apprentices a model-based strategy for diagnosing car malfunctions. In the following sections, I will first briefly summarize the findings of the two studies the dissertation is based on. I will then discuss the effects of the video-based modelling examples and the different types of self-explanation prompts. In a separate paragraph, I will discuss findings regarding cognitive load. Eventually, the practical implications of our findings for the vocational education practice will be presented and I will give suggestions for future research.

Summary of Results

In study 1, irrespective of the experimental condition, the intervention promoted apprentices‘ diagnostic knowledge and their scaffolded diagnostic skills but not their independent diagnostic skills. When considering the experimental conditions, we did not find a modelling example effect, as modelling examples did not affect extraneous (ECL) or germane cognitive load (GCL) nor were there any additional beneficial effects of modelling examples on learning outcomes. Regarding the effects of retrospective versus anticipatory self-explanation prompts, we found that apprentices with less prior knowledge reported a higher GCL when learning with retrospective prompts. Moreover, we found a greater increase in diagnostic knowledge among the stronger apprentices when learning with the anticipatory prompts.

In study 2, we could replicate the general beneficial effect of the intervention on learning outcomes as the intervention promoted apprentices‘ diagnostic knowledge and scaffolded diagnostic skills. Unlike in study 1, in study 2 apprentices could also transfer this increase in knowledge to independent problem-solving in the computer simulation. Considering the experimental conditions in study 2, modelling examples again had no effects on ECL or GCL. However, we found a modelling example effect on apprentices‘ diagnostic knowledge and their scaffolded diagnostic skills. This modelling example effect did not extend to the independent diagnosis in the simulation. Regarding the representation-based self-explanation prompts, sequential versus comparative self-explanation prompts did not have any effects on ICL or GCL regardless of whether the prior knowledge was considered or not. Nevertheless, the prompt condition affected learning outcomes. With respect to diagnostic knowledge, participants with higher prior knowledge benefitted from comparative self-explanation prompts.

Modelling Examples for Teaching a Strategy for Diagnosing Car Malfunctions

Interestingly, neither in study 1 nor in study 2, modelling examples had effects on cognitive load. This finding is surprising as the beneficial effect of worked or modelling examples on learning outcomes is usually explained with a decrease in ECL which results in more working memory capacities (i.e., GCL) that can be used for learning (Renkl, 2014; Sweller, 2006). These findings regarding cognitive load will be discussed in a later section of this discussion in an integrated manner together with other unexpected findings regarding cognitive load.

Nevertheless, the interventions in general (i.e., irrespective of modelling examples) are effective for promoting a complex problem-solving strategy such as the diagnosis of car malfunctions. After eliminating the shortcomings of the first intervention as evaluated in study 1 (i.e., especially the long instruction phase and the repetitive prompts), we found a beneficial effect of modelling examples on diagnostic knowledge and scaffolded diagnostic skills in study 2. Hence, we can conclude that also longer video-based modelling examples are useful for teaching problem-solving strategies in ill-structured domains, thus providing a positive answer to one of the underlying questions of the dissertation.

In both studies, however, we found indications of a possible utilisation deficiency (Hübner et al., 2010; Miller, 1994): While in study 1 the overall positive effect of the intervention extended only to diagnostic knowledge and scaffolded diagnostic skills, we also found the modelling example effect in study 2 only for these outcome measures. In both cases, the respective effect did not occur for independent problem solving, that is, the independent diagnosis in the computer simulation. A utilisation deficiency occurs when a newly learned strategy is not yet automated. In such situations, trying to apply the strategy requires so much cognitive capacity that the new problem that needs to be solved cannot be processed. Consequently, the application of the strategy to the novel situation (Hübner et al., 2010; Miller, 1994). One explanation might be that in our interventions apprentices first watched the instructional videos that introduced the diagnostic strategy and then studied two modelling examples. They did, however, not get any opportunity to practise the newly learned diagnostic strategy. Hence, a possible solution to the utilisation deficiency could be interleaving example study with independent problem-solving in later phases of skill acquisition (Renkl, 2014; VanLehn, 1996). Such interleaving could be implemented by mixing complete modelling examples with problems that need to be solved completely. How precisely such full (modelling) examples and problems should be paired (e.g., example first versus problem first) as well as the exact effects of different pairings on cognitive (Kant et al., 2017; Leppink et al., 2014; van Gog, 2011; van Gog et al., 2011) and motivational outcome measures (van Harsel et al., 2019) is still under investigation. Another option would be to fade out solution steps in the modelling examples by making them more and more incomplete and asking the apprentices to complete the problem-solving processes shown in the modelling examples until only the problem formulation is left, that is, a problem to be solved (Atkinson et al., 2003). Ideally, such a fading procedure would be adaptive, by fading out solution steps only if a learner has shown progress with the respect to the to be faded steps (e.g., Kalyuga & Sweller, 2004). Fading out solution steps from examples has been investigated for well-structured worked examples, but how to sensibly fade out solution steps in video-based modelling examples is an open question (Renkl, 2014).

Self-Explanation and Comparisons for Video-Based Modelling Examples

Modelling examples are effective because they reduce ECL and free up working memory capacities that can be used for generative learning activities such as self-explanation (Bisra et al., 2018; Rittle-Johnson et al., 2017). In most studies with self-explanation prompts, learners are asked to self-explain aspects of a problem-solving strategy that have just been illustrated in the corresponding example (i.e., retrospective prompts). Anticipatory prompts that refer to the next problem-solving step have not been investigated systematically (Bisra et al., 2018). Thus, in study 1 we compared the effects of retrospective versus anticipatory self-explanation prompts on cognitive load and learning. Another generative learning activity is comparison (Alfieri et al., 2013; Rittle-Johnson & Star, 2011), which is difficult to implement for video-based modelling examples. Hence, in study 2 we explored the use of representation-based comparisons for video-based modelling examples and investigated the effects of comparative (i.e., side-by-side) self-explanation versus sequential self-explanation of such representations. In both studies, we expected differential effects of the different types of self-explanation prompts depending on the apprentices' prior knowledge. We expected anticipatory prompts and comparative prompts to be more demanding and, therefore, to induce a higher ICL. However, we also expected that learners with more prior knowledge would be able to manage these increased demands and would experience a higher GCL and would show better learning outcomes.

Retrospective Versus Anticipatory Self-Explanation Prompts

In study 1, we found that the anticipatory prompts induced a lower ECL. This finding is rather surprising and will be discussed in a later section of this discussion in an integrated manner together with other unexpected findings regarding cognitive load. In addition, apprentices with less prior knowledge reported a higher GCL when learning with retrospective prompts. These apprentices were challenged by the rather less demanding retrospective prompts in an appropriate way so that a high GCL could emerge. Moreover, we found a greater increase in strategy knowledge among the stronger apprentices when learning with the anticipatory prompts. Taken together, these effects suggest that, in line with our hypotheses, anticipatory prompts are more beneficial for learners with higher prior knowledge. Consequently, educational practitioners using example-based learning should adapt their self-explanation prompts to their learners' prior knowledge (Kalyuga & Renkl, 2010), for example, by using anticipatory prompts for more experienced learners. Hence, together with my co-authors, I have taken a first step towards investigating the use of anticipatory self-explanation prompts. Nevertheless, as already described in detail in the discussion of manuscript 2, further research should be conducted to understand more precisely how anticipatory prompts work, for example by using think-aloud protocols.

Representation-Based Comparisons for Video-Based Modelling Examples

Comparing examples is another effective generative learning activity (Alfieri et al., 2013; Rittle-Johnson & Star, 2011), but it is difficult to implement for video-based modelling examples as learners would need to watch two videos at the same time or pause the videos repeatedly. Hence, we proposed to base comparisons not on the (transient) video-based modelling examples but on a (static) representation of the problem-solving process that has been illustrated in a video-based modelling example. In manuscript 3 we used diagnosis plans of experts and novices as representations of diagnostic processes that were completed with different degrees of quality. We found no evidence that comparing or sequentially self-explaining these representations led to any complications. The comparison of (differently solved) problem-solving processes seems to be as successful for learners using such representations as when they compare examples directly. Therefore, future research can now use such representation-based comparisons and start investigating whether the various effects of comparison in example-based learning that have been studied with text-based examples (Alfieri et al., 2013; Rittle-Johnson & Star, 2011) also apply to video-based modelling examples.

Sequential Versus Comparative Self-Explanation Prompts

We made a first step in this direction by investigating whether the example comparison principle also applies to video-based modelling examples (Renkl, 2014). We expected comparative prompts to be more demanding and, therefore, to induce a higher ICL for all learners. We also expected learners with more prior knowledge to experience a higher GCL when learning with comparative self-explanation prompts. However, regarding the cognitive load, we found no effects of sequential versus comparative self-explanation prompts. These unexpected findings regarding cognitive load will be discussed in a later section of this discussion in an integrated manner together with other unexpected findings regarding cognitive load. Regarding learning outcomes, however, we found that for diagnostic strategy knowledge participants with higher prior knowledge rather benefitted from comparative self-explanation prompts. This finding is consistent with Rittle-Johnson et al. (2009) who found that in the domain of mathematics students with more prior knowledge learned more when they compared worked-examples instead of studying them sequentially. Hence, we conclude that the example comparison principle also applies to representation-based comparisons of video-based modelling examples.

Unexpected Effects on Cognitive Load

In both studies we found surprising effects on cognitive load: First, in both studies, modelling examples had no effects on cognitive load. However, the usual explanation for the beneficial effect of modelling examples on learning is that examples decrease ECL and free up cognitive capacities for learning (Renkl, 2014; Sweller, 2006). As we found a modelling example effect in study 2, we also should have found a decrease in ECL.

Second, anticipatory prompts induced less ECL than retrospective prompts in study 1. We expected the anticipatory prompts to induce a higher ICL and – for learners with higher prior knowledge – a higher GCL. The ECL items referred to the visual characteristics of the learning materials. However, the visual design of the learning materials did not differ between the two prompt conditions. It has been argued that learners sometimes struggle to distinguish between ICL and ECL (Klepsch & Seufert, 2020; study 1). Nevertheless, anticipatory prompts did not induce a higher but a lower ECL than retrospective prompts. To make sure that this finding is not a false positive, future studies should try to replicate it.

Third, in study 2 we expected the comparative prompts to induce a higher ICL and – for learners with higher prior knowledge – a higher GCL. However, no effects of sequential versus comparative prompts on cognitive load could be found. Taken together, these three surprising findings (or the absence thereof) on cognitive load raise doubts about the relationships between the use of worked or modelling examples, cognitive load, and learning outcomes. As argued in the discussion of manuscript 3 in detail, most research on worked or modelling examples was conducted within a cognitive load framework, that is, researchers assumed that examples decrease ECL, free up cognitive capacities and – if generative learning is encouraged – thereby promote GCL and learning. However, many of these studies have used only one general measure of cognitive load (e.g., Spanjers et al., 2012; van Gog et al., 2008) or they have not assessed the learners‘ cognitive load at all (e.g., Rourke & Sweller, 2009). Accordingly, these studies cannot provide any insight into the specific relationships between the use of worked or modelling examples, cognitive load, and learning outcomes. Therefore, future research on example-based learning should also assess the effects on cognitive load (Sweller, 2017). Moreover, even if studies used measures that differentiated between ICL, ECL, and GCL, the use of many different questionnaires impedes drawing overarching conclusions. Ideally, the research community would agree on using the same instruments to measure cognitive load, such as the instruments developed by Klepsch, Seufert and colleagues (Klepsch et al., 2017; Klepsch & Seufert, 2020, 2021). Furthermore, differences between interventions or studies evaluating interventions need to be accounted for more carefully. Although we used instruments that detected the effects of examples on cognitive load successfully in the past (Klepsch & Seufert, 2020; experiment 3), we could not replicate these effects. As discussed in manuscript 3 in more detail, in our studies the learning phases were very extensive. Probably, repeated measurements of cognitive load to also account for changes in cognitive load over time would have been more appropriate (Schmeck et al., 2015; van Gog et al., 2012). Furthermore, the question arises whether instruments for measuring cognitive load when learning from text-based worked examples are equally suitable for measuring cognitive load when learning from video-based modelling examples. In video-based modelling examples, usually longer and more complex problem-solving processes are illustrated. It can be assumed that the retrospective self-assessment of the cognitive load becomes more complex when the learning content and the learning materials become more complex. It may therefore be necessary to develop different instruments for measuring cognitive load for different types of learning material.

Conclusion

With this dissertation, I contributed to the research on example-based learning. First, I have shown that longer video-based modelling examples are also useful for teaching problem-solving strategies in ill-structured domains. Second, I have described anticipatory self-explanation prompts as a novel type of prompt and compared them for the first time with the usual retrospective prompts. These results suggest that anticipatory prompts are more suitable for stronger learners. Third, with the representation-based comparisons, I have presented a proposal on how example comparisons can be used for video-based modelling examples. I have also been able to show that the example comparison principle also applies to video-based modelling examples. I conclude with practical implications and suggestions for future research:

Practical Implications

The results of the two studies are promising concerning the training of automotive mechatronics technicians. Currently, only 15 % of apprentices have mastered a diagnostic strategy to diagnose complex malfunctions at the end of their apprenticeship (Abele & von Davier, 2019). The interventions are helpful in teaching such a diagnostic strategy. However, as presumably utilisation deficiencies occurred (Hübner et al., 2010; Miller, 1994), when the intervention is used in vocational schools or other vocational education sites, it should be combined with further practice opportunities for the apprentices (Renkl, 2014; VanLehn, 1996). In further studies, it would then be interesting to find out whether the positive effects of the intervention persist over a longer period and whether they also transfer to real vehicles.

Suggestions for Future Research

The following recommendations for future research result from this dissertation: First, anticipatory self-explanation prompts should be investigated further and in more detail. Here, laboratory studies without interfering environmental influences using think-aloud protocols would be helpful to find out how exactly learners process anticipatory prompts. Second, we showed that representation-based comparisons can be used to implement comparisons for video-based modelling examples. Future research can now investigate whether the various effects that have been found for text-based worked examples also account for video-based modelling examples. Finally, the specific relationships between the use of worked or modelling examples, cognitive load, and learning outcomes need to be investigated in more detail.

References

Abele, S. (2018). Diagnostic problem-solving process in professional contexts: Theory and empirical investigation in the context of car mechatronics using computer-generated log-files. *Vocations and Learning*, *11*(1), 133–159. https://doi.org/10.1007/s12186-017-9183-x

Abele, S., & von Davier, M. (2019). CDMs in vocational education: Assessment and usage of diagnostic problem-solving strategies in car mechatronics. In M. von Davier & Y.-S. Lee (Eds.), *Handbook of Diagnostic Classification Models*. Springer. https://doi.org/10.1007/978-3-030-05584-4_22

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning Through Case Comparisons: A Meta-Analytic Review. *Educational Psychologist*, *48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

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., & Merrill, M. M. (2003). Transitioning From Studying Examples to Solving Problems: Effects of Self-Explanation Prompts and Fading Worked-Out Steps. *Journal of Educational Psychology*, *95*(4), 774–783. https://doi.org/10.1037/0022-0663.95.4.774

Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing Self-Explanation: A Meta-Analysis. *Educational Psychology Review*, *30*(3), 703–725. https://doi.org/10.1007/s10648-018-9434-x

Blomhoj, M., & Jensen, T. H. (2003). Developing mathematical modelling competence: Conceptual clarification and educational planning. *Teaching Mathematics and Its Applications*, *22*(3), 123–139. https://doi.org/10.1093/teamat/22.3.123

Boero, P. (1999). Argumentation and mathematical proof: A complex, productive, unavoidable relationship in mathematics and mathematics education. *International Newsletter on the Teaching and Learning of Mathematical Proof*, *7*(8). http://www.lettredelapreuve.org/OldPreuve/Newsletter/990708Theme/990708ThemeUK.html

Brunstein, A., Betts, S., & Anderson, J. R. (2009). Practice enables successful learning under minimal guidance. *Journal of Educational Psychology*, *101*(4), 790–802. https://doi.org/10.1037/a0016656

Castro-Alonso, J. C., Ayres, P., & Paas, F. (2015). Animations showing Lego manipulative tasks: Three potential moderators of effectiveness. *Computers & Education*, *85*, 1–13. https://doi.org/10.1016/j.compedu.2014.12.022

Catrambone, R. (1996). Generalizing Solution Procedures Learned From Examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(4), 1020–1031. https://doi.org/10.1037/0278-7393.22.4.1020

Chen, Z., & Klahr, D. (1999). All Other Things Being Equal: Acquisition and Transfer of the Control of Variables Strategy. *Child Development*, *70*(5), 1098–1120. https://doi.org/10.1111/1467-8624.00081

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science*, *13*(2), 145–182. https://doi.org/10.1207/s15516709cog1302_1

Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (Fourth edition). Wiley.

Collins, A., Brown, J. S., & Newman, S. E. (1988). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. *Thinking: The Journal of Philosophy for Children*, *8*(1), 2–10. https://doi.org/10.5840/thinking19888129

Elstein, A. S., Allal, L., Gordon, M., Hilliard, J., Jordan, R. D., Kagan, N., Loupe, M. J., Shulman, L. S., & Sprafka, S. A. (1978). *Medical Problem Solving: An Analysis of Clinical Reasoning*. Harvard University Press.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1990). Medical problem solving: A ten-year retrospective. *Evaluation & the Health Professions*, *13*(1), 5–36. https://doi.org/10.1177/016327879001300102

Fiorella, L., & Mayer, R. E. (2016). Eight Ways to Promote Generative Learning. *Educational Psychology Review*, *28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9

Fiorella, L., van Gog, T., Hoogerheide, V., & Mayer, R. E. (2017). It's all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task. *Journal of Educational Psychology*, *109*(5), 653–665. https://doi.org/10.1037/edu0000161

Gentner, D. (2010). Bootstrapping the Mind: Analogical Processes and Symbol Systems. *Cognitive Science*, *34*(5), 752–775. https://doi.org/10.1111/j.1551-6709.2010.01114.x

Glogger, I., Holzäpfel, L., Kappich, J., Schwonke, R., Nückles, M., & Renkl, A. (2013). Development and evaluation of a computer-based learning environment for teachers: Assessment of learning strategies in learning journals. *Education Research International*, *2013*, 1–12. https://doi.org/10.1155/2013/785065

Glogger-Frey, I., Fleischer, C., Grüny, L., Kappich, J., & Renkl, A. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct

instruction. *Learning and Instruction*, *39*, 72–87. https://doi.org/10.1016/j.learninstruc.2015.05.001

Glogger-Frey, I., Gaus, K., & Renkl, A. (2017). Learning from direct instruction: Best prepared by several self-regulated or guided invention activities? *Learning and Instruction*, *51*, 26–35. https://doi.org/10.1016/j.learninstruc.2016.11.002

Hänze, M., & Leiss, D. (2022). Using heuristic worked examples to promote solving of reality-based tasks in mathematics in lower secondary school. *Instructional Science*, *50*(4), 529–549. https://doi.org/10.1007/s11251-022-09583-8

Hefter, M. H., Berthold, K., Renkl, A., Riess, W., Schmid, S., & Fries, S. (2014). Effects of a training intervention to foster argumentation skills while processing conflicting scientific positions. *Instructional Science*, *42*(6), 929–947. https://doi.org/10.1007/s11251-014-9320-y

Hefter, M. H., Renkl, A., Riess, W., Schmid, S., Fries, S., & Berthold, K. (2015). Effects of a training intervention to foster precursors of evaluativist epistemological understanding and intellectual values. *Learning and Instruction*, *39*, 11–22. https://doi.org/10.1016/j.learninstruc.2015.05.002

Hefter, M. H., Renkl, A., Riess, W., Schmid, S., Fries, S., & Berthold, K. (2018). Training Interventions to Foster Skill and Will of Argumentative Thinking. *The Journal of Experimental Education*, *86*(3), 325–343. https://doi.org/10.1080/00220973.2017.1363689

Heitzmann, N., Fischer, F., Kühne-Eversmann, L., & Fischer, M. R. (2015). Enhancing diagnostic competence with self-explanation prompts and adaptable feedback. *Medical Education*, *49*(10), 993–1003. https://doi.org/10.1111/medu.12778

Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Böhmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Böhmer, I., Karing, C., Kaiser, J., & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, *76*, 181–193. https://doi.org/10.1016/j.tate.2017.12.001

Hilbert, T. S., & Renkl, A. (2009). Learning how to use a computer-based concept-mapping tool: Self-explaining examples helps. *Computers in Human Behavior*, *25*(2), 267–274. https://doi.org/10.1016/j.chb.2008.12.006

Hilbert, T. S., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning and Instruction*, *18*(1), 54–65. https://doi.org/10.1016/j.learninstruc.2006.10.008

Holyoak, K. J. (2012). Analogy and Relational Reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning* (1st ed., pp. 234–259). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199734689.013.0013

Hoogerheide, V. (2016). *Effects of Observing and Creating Video Modeling Examples on Cognitive and Motivational Aspects of Learning*.

Hoogerheide, V., Loyens, S. M. 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., van Wermeskerken, M., Loyens, S. M. M., & van Gog, T. (2016). Learning from video modeling examples: Content kept equal, adults are more effective models than peers. *Learning and Instruction*, *44*, 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

Hübner, S., Nückles, M., & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. *Learning and Instruction*, *20*(1), 18–29. https://doi.org/10.1016/j.learninstruc.2008.12.001

Kalyuga, S. (2009). Instructional designs for the development of transferable knowledge and skills: A cognitive load perspective. *Computers in Human Behavior*, *25*(2), 332–338. https://doi.org/10.1016/j.chb.2008.12.019

Kalyuga, S. (2011). Cognitive Load Theory: How many types of load does it really need? *Educational Psychology Review*, *23*(1), 1–19. https://doi.org/10.1007/s10648-010-9150-7

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist*, *38*(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4

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

Kant, J. M., 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

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*, 1997. https://doi.org/10.3389/fpsyg.2017.01997

Klepsch, M., & Seufert, T. (2020). Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. *Instructional Science*, *48*(1), 45–77. https://doi.org/10.1007/s11251-020-09502-9

Klepsch, M., & Seufert, T. (2021). Making an Effort Versus Experiencing Load. *Frontiers in Education*, *6*, 645284. https://doi.org/10.3389/feduc.2021.645284

Krieglstein, F., Beege, M., Rey, G. D., Ginns, P., Krell, M., & Schneider, S. (2022). A Systematic Meta-analysis of the Reliability and Validity of Subjective Cognitive Load Questionnaires in Experimental Multimedia Learning Research. *Educational Psychology Review*, *34*(4), 2485–2541. https://doi.org/10.1007/s10648-022-09683-4

Leppink, J., Broers, N. J., Imbos, T., van der Vleuten, C. P. M., & Berger, M. P. F. (2012). Self-explanation in the domain of statistics: An expertise reversal effect. *Higher Education*, *63*(6), 771–785. https://doi.org/10.1007/s10734-011-9476-1

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

Marcus, N., Cleary, B., Wong, A., & Ayres, P. (2013). Should hand actions be observed when learning hand motor skills from instructional animations? *Computers in Human Behavior*, *29*(6), 2172–2178. https://doi.org/10.1016/j.chb.2013.04.035

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

McLaren, B. M., Lim, S.-J., & 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. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 2176–2181.

Miller, P. H. (1994). Individual differences in children's strategic behaviors: Utilization deficiencies. *Learning and Individual Differences*, *6*(3), 285–307. https://doi.org/10.1016/1041-6080(94)90019-1

Mulder, Y. G., Lazonder, A. W., & de Jong, T. (2014). Using heuristic worked examples to promote inquiry-based learning. *Learning and Instruction*, *29*, 56–64. https://doi.org/10.1016/j.learninstruc.2013.08.001

Norman, G., Young, M., & Brooks, L. (2007). Non-analytical models of clinical reasoning: The role of experience. *Medical Education*, *0*(0), 071116225013001-??? https://doi.org/10.1111/j.1365-2923.2007.02914.x

Paas, F., & van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction*, *16*(2), 87–91. https://doi.org/10.1016/j.learninstruc.2006.02.004

Reiss, K., & Renkl, A. (2002). Learning to prove: The idea of heuristic examples. *Zentralblatt Für Didaktik Der Mathematik*, *34*(1), 29–35. https://doi.org/10.1007/BF02655690

Renkl, A. (1997). Learning from Worked-Out Examples: A Study on Individual Differences. *Cognitive Science*, *21*(1), 1–29. https://doi.org/10.1207/s15516709cog2101_1

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, *38*(1), 1–37. https://doi.org/10.1111/cogs.12086

Renkl, A., & Eitel, A. (2019). Self-Explaining: Learning About Principles and Their Application. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge Handbook of Cognition and Education* (1st ed., pp. 528–549). Cambridge University Press. https://doi.org/10.1017/9781108235631.022

Renkl, A., Hilbert, T., & Schworm, S. (2009). Example-Based Learning in Heuristic Domains: A Cognitive Load Theory Account. *Educational Psychology Review*, *21*(1), 67–78. https://doi.org/10.1007/s10648-008-9093-4

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from Worked-Out Examples: The Effects of Example Variability and Elicited Self-Explanations. *Contemporary Educational Psychology*, *23*(1), 90–108. https://doi.org/10.1006/ceps.1997.0959

Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM*, *49*(4), 599–611. https://doi.org/10.1007/s11858-017-0834-z

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, *99*(3), 561–574. https://doi.org/10.1037/0022-0663.99.3.561

Rittle-Johnson, B., & Star, J. R. (2011). The Power of Comparison in Learning and Instruction: Learning Outcomes Supported by Different Types of Comparisons. In *Psychology of Learning and Motivation* (Vol. 55, pp. 199–225). Elsevier. https://doi.org/10.1016/B978-0-12-387691-1.00007-7

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology*, *101*(4), 836–852. https://doi.org/10.1037/a0016026

Roelle, J., Krüger, S., Jansen, C., & Berthold, K. (2012). The Use of Solved Example Problems for Fostering Strategies of Self-Regulated Learning in Journal Writing. *Education Research International*, *2012*, 1–14. https://doi.org/10.1155/2012/751625

Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognise designers' styles. *Learning and Instruction*, *19*(2), 185–199. https://doi.org/10.1016/j.learninstruc.2008.03.006

Rummel, N., & Spada, H. (2005). Learning to Collaborate: An Instructional Approach to Promoting Collaborative Problem Solving in Computer-Mediated Settings. *Journal of the Learning Sciences*, *14*(2), 201–241. https://doi.org/10.1207/s15327809jls1402_2

Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning*, *4*(1), 69–92. https://doi.org/10.1007/s11412-008-9054-4

Safadi, R. (2022). Supporting student learning from diagnosing erroneous examples when contrasting them with worked examples in the physics classroom. *International Journal of Science Education*, 1–26. https://doi.org/10.1080/09500693.2021.2023834

Schalk, L., Roelle, J., Saalbach, H., Berthold, K., Stern, E., & Renkl, A. (2020). Providing worked examples for learning multiple principles. *Applied Cognitive Psychology*, *34*(4), 813–824. https://doi.org/10.1002/acp.3653

Schmeck, A., Opfermann, M., van Gog, T., Paas, F., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: Differences between immediate and delayed ratings. *Instructional Science*, *43*(1), 93–114. https://doi.org/10.1007/s11251-014-9328-3

Schmidt, H. G., & Rikers, R. M. J. P. (2007). How expertise develops in medicine: Knowledge encapsulation and illness script formation. *Medical Education*, *0*(0), 071116225013002-??? https://doi.org/10.1111/j.1365-2923.2007.02915.x

Schmitz, F. M., Schnabel, K. P., Stricker, D., Fischer, M. R., & Guttormsen, S. (2017). Learning communication from erroneous video-based examples: A double-blind randomised controlled trial. *Patient Education and Counseling*, *100*(6), 1203–1212. https://doi.org/10.1016/j.pec.2017.01.016

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, *103*(4), 759–775. https://doi.org/10.1037/a0025140

Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology*, *99*(2), 285–296. https://doi.org/10.1037/0022-0663.99.2.285

Solé-Llussà, A., Aguilar, D., & Ibáñez, M. (2020). Video-worked examples to support the development of elementary students' science process skills: A case study in an inquiry activity on electrical circuits. *Research in Science & Technological Education*, 1–21. https://doi.org/10.1080/02635143.2020.1786361

Spanjers, I. A. E., Gog, T., & Merriënboer, J. J. G. (2012). Segmentation of Worked Examples: Effects on Cognitive Load and Learning: Segmentation of Worked Examples. *Applied Cognitive Psychology*, *26*(3), 352–358. https://doi.org/10.1002/acp.1832

Spöttl, G., Becker, M., & Musekamp, F. (2011). Anforderungen an Kfz-Mechatroniker und Implikationen für die Kompetenzerfassung [Requirements for automotive mechatronics technicians and implications for the assessment of competencies]. In R. Nickolaus (Ed.), *Lehr-Lernforschung in der gewerblich-technischen Berufsbildung (Zeitschrift für Berufs- und Wirtschaftspädagogik: Sonderbd. 25)* (pp. 37–53). Steiner.

Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction*, *16*(2), 165–169. https://doi.org/10.1016/j.learninstruc.2006.02.005

Sweller, J. (2017). The Role of Independent Measures of Load in Cognitive Load Theory. In R. Zheng (Ed.), *Cognitive load measurement and application: A theoretical framework for meaningful research and practice*. Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315296258-1/role-independent-measures-load-cognitive-load-theory-john-sweller

Sweller, J., Ayres, P. L., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

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/s1532690xci0201_3

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

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., 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., Kirschner, F., Kester, L., & Paas, F. (2012). Timing and Frequency of Mental Effort Measurement: Evidence in Favour of Repeated Measures: Measuring mental effort. *Applied Cognitive Psychology*, *26*(6), 833–839. https://doi.org/10.1002/acp.2883

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., Paas, F., & van Merriënboer, J. J. G. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction*, *18*(3), 211–222. https://doi.org/10.1016/j.learninstruc.2007.03.003

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. A. Rawson (Eds.), *The Cambridge Handbook of Cognition and Education* (1st ed., pp. 183–208). Cambridge University Press. https://doi.org/10.1017/9781108235631.009

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., Janssen, E., Verkoeijen, P., & van Gog, T. (2022). How do higher education students regulate their learning with video modeling examples, worked examples, and practice problems? *Instructional Science*, *50*(5), 703–728. https://doi.org/10.1007/s11251-022-09589-2

van Harsel, M., Hoogerheide, V., Verkoeijen, P., & 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 Merriënboer, J. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Educational Technology Publications.

van Merriënboer, J. J. G. (2013). Perspectives on problem solving and instruction. *Computers & Education*, *64*, 153–160. https://doi.org/10.1016/j.compedu.2012.11.025

Van Merriënboer, J., & Sweller, J. (2010). Cognitive load theory in health professional education: Design principles and strategies: Cognitive load theory. *Medical Education*, *44*(1), 85–93. https://doi.org/10.1111/j.1365-2923.2009.03498.x

van Wermeskerken, M., Ravensbergen, S., & van Gog, T. (2018). Effects of instructor presence in video modeling examples on attention and learning. *Computers in Human Behavior*, *89*, 430–438. https://doi.org/10.1016/j.chb.2017.11.038

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, *47*(1), 513–539. https://doi.org/10.1146/annurev.psych.47.1.513

Wylie, R., & Chi, M. T. H. (2014). The Self-Explanation Principle in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 413–432). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.021

Zavgorodniaia, A., Duran, R., Hellas, A., Seppala, O., & Sorva, J. (2020). Measuring the Cognitive Load of Learning to Program: A Replication Study. *United Kingdom & Ireland Computing Education Research Conference.*, 3–9. https://doi.org/10.1145/3416465.3416468