Methodological report on the second study to promote a model-based car diagnosis strategy in the DigiDIn-Kfz project

Julius Meier & Peter Hesse

Information on the Project

This study was conducted 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.

Aim of the Report

This report describes the second study that was conducted in the project *Digidin-Kfz,* sub-project Promotion of the Model-Based Diagnostic Strategy. Based on the findings of study 1, as described in Meier et al. (2022) and Meier et al. (2023a) we adapted the diagnostic strategy for diagnosing car malfunctions and developed a new intervention to teach this strategy. The data from this study has been used for one publication: (Meier et al., 2023b). This report is intended as a addendum to this publications. It is recommended to read these publications first. The study took place in June and July 2022 at the [redacted], at the [redacted], and at the [redacted], Germany.

Participants and Design

We conducted a computer-based experiment in two sessions with three experimental conditions. Session one included the pretest. In session two, 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. In total, 135 apprentices participated in session one and 131 participated in session two. The corresponding cleaned data set (*Digidin_DataSetPublication3.sav*) includes data on these participants. Hoewever, only 118 apprentices participated in both sessions and can thus be included in the analyses. These participants are described in Meier et al. (2023b).

In session one, pre-tests were conducted. For the intervention and post-tests in session two, we randomly assigned the apprentices to the experimental conditions: First, apprentices in all conditions learned about the diagnostic strategy with instructional videos. Then, 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).

Procedure

The entire study took place on computers in the apprentices' schools. All learning and examination materials were presented in digital form in a specially programmed survey environment. Screenshots of these surveys can be found in the attachement (*Survey Screenshots Session 1.pptx (annotated version that explains which item corresponds to which variable in the data set); Survey Screenshots Session 2 Condition 2A.pptx (annotated version that explains which item corresponds to which variable in the data set; Survey Screenshots Session 2 Condition 2B.pptx; Survey Screenshots Session 2 Condition 2C.pptx*). Once apprentices 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 paper by Meier et al. (2023b) and is thus not included in the data file.

Intervention

Diagnostic Strategy

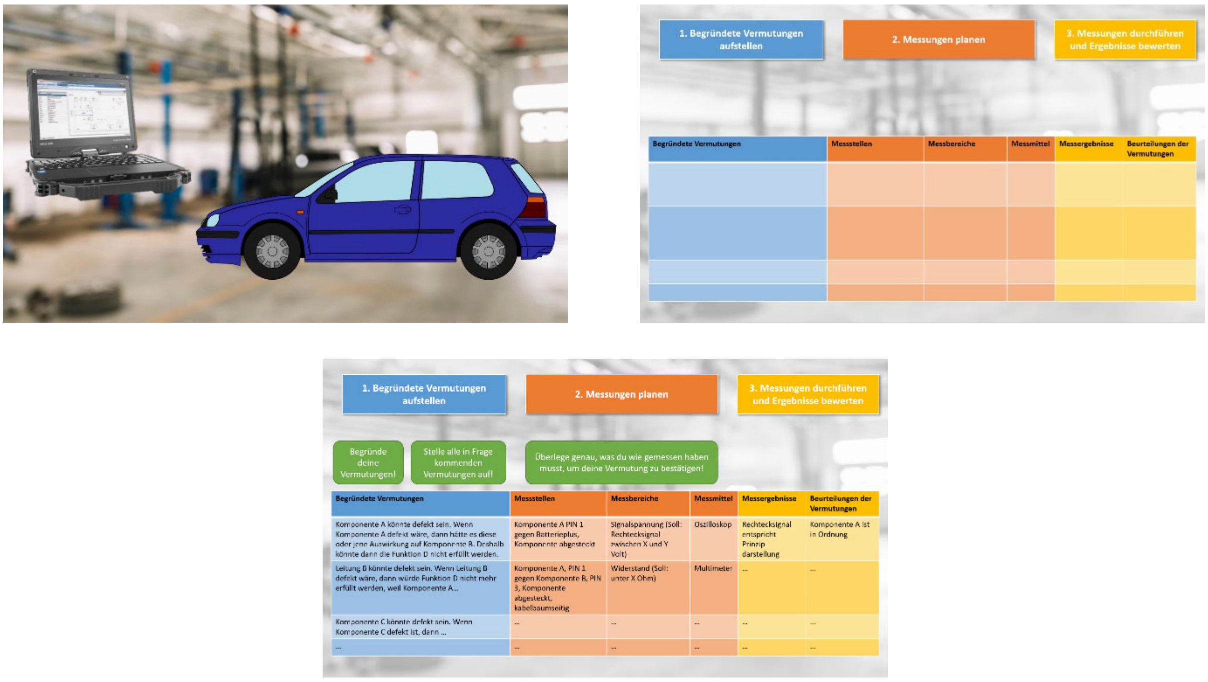
In collaboration with subject-matter experts and on basis of respective literature (e.g., Abele, 2014), we developed an intervention in which apprentices learned about a strategy to diagnose complex car malfunctions. This strategy is described in comprised three steps: (1) When diagnosing car malfunctions, 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 <u>all</u> 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



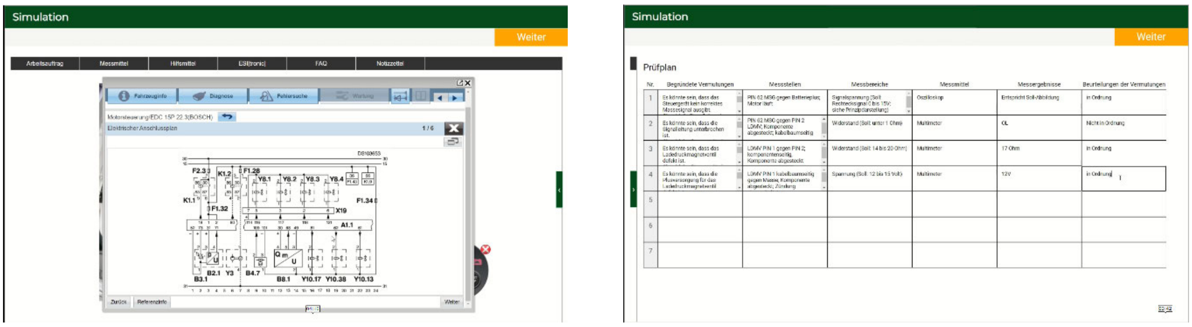
Note. The screenshots are from the instructional videos explaining the three diagnostic steps using the diagnosis plan. The three green boxes in the bottom screenshot contain the rules underlying the first two diagnostic steps.

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 (*DRCTestQuestion1_V1A2* to *DRCTestQuestion5_V2A17*) 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. The total score in this test is represented by the variable *DRCTestScore* in *Digidin_DataSetPublication3.sav.*

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; *StrategyDescriptionASteps*). 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 κ = .864) and session 2 (Cohen's κ = .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 (*StrategyDescriptionBModell*). 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. The total score in this test is represented by the variables *StrategyDescriptionPre_Scored* for session one and *StrategyDescriptionPost_Scored* for session two in *Digidin_DataSetPublication3.sav.*

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 (*StrategyCompletionPre_gestörteFunktion_Aufgabe1_1_1 to StrategyCompletionPre_Schritt_Aufgabe3_2*). 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. The total score in this test is represented by the variables *StrategyCompletionPre_Scored* for session one and *StrategyCompletionPost_Scored* for session two in *Digidin_DataSetPublication3.sav.*

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 (*DiagnosisFM02* and *DiagnosisFM02*). 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. The total score in this test is represented by the variables *DiagnosisCombinedPre_Scored* for session one and *DiagnosisCombinedPost_Scored* for session 2 in *Digidin_DataSetPublication3.sav.* 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 (*Selfefficacy1* to *Selfefficacy5*), 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 $\alpha = 0.998$). The mean score of these items is represented by the variables *SelfefficacyPre* for session one and *SelfefficacyPost* for session 2. Five items (*Interest1* to *Interest5*) assessed the apprentices' *interest* in car diagnoses (Schiefele, 1991). Reliability was again good (Session 1: Cronbach's $\alpha = 0.847$; Session 2: Cronbach's $\alpha = 0.853$). The mean score of these items is represented by the variables *InterestPre* for session one and *InterestPost* for session 2. With four items (*Challenge1* to *Challenge4*) 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 $\alpha = 0.997$). The mean score of these items is represented by the variables *ChallengePre* for session one and *ChallengePost* for session 2. Five items (*IncompetenceFear1* to *IncompetenceFear5*) assessed the apprentices' *incompetence fear* (Session 1: Cronbach's $\alpha = 0.903$; Session 2: Cronbach's $\alpha = 0.998$). The mean score of these items is represented by the variables *IncompetenceFearPre* for session one and *IncompetenceFearPost* for session 2.

Cognitive Load

After the intervention, we assessed the apprentices' *intrinsic* (two items; *CL_Intrinsic1, CL_Intrinsic2*), *germane* (two items; *CL_Germane1, CL_Germane2*), and *extraneous cognitive load* (three items; *CL_Extr1, CL_Extr2, CL_Extr3*) 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 $\alpha = 0.61$). The mean score of these items is represented by the variables *CL_Intrinsic_Mean, CL_Extraneous_Mean,* and *CL_Germane_Mean.*

References

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