Methodological report on the first survey 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 first study that was conducted in the project *Digidin-Kfz,* sub-project Promotion of the Model-Based Diagnostic Strategy. We developed and evaluated an intervention to teach automotive apprentices a strategy for diagnosing car malfunctions. The data from this study and the insights gained from it have been used for two publications. Consequently, two separate data sets referring to these two publications are available and are described in this report. The first dataset *Digidin_DataSetPublication1.sav* refers to the publication by Meier et al. (2022). The second dataset *Digidin_DataSetPublication2.sav* includes the data that are reported in Meier et al. (2023). The data has been used for two publications, as the focus for the two publications was set very differently. While Meier et al. (2022) primarily describe the development of the intervention material and its evaluation, Meier et al. (2023) focus on the effects of the modelling examples and different self-explanation prompts. This report is therefore intended as a addendum to these two aforementioned publications. It is recommended to read these publications first. The study took place in September and October 2020 at the [redacted]

[redacted] (coded as School 1) and [redacted], Germany (coded as School 2).

Participants and Design

Originally, 78 apprentices participated in the experiment that comprised two sessions separated by approximately 10 days. 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 and organizational prompts. Then, the first group received modelling examples with retrospective self-explanation prompts and the second group learned with anticipatory self-explanation prompts. In the third condition (control), apprentices were given instructional videos but no modelling examples, and also no self-explanation prompts. For publication 1, only participants in the first two groups were considered, but the different self-explanation prompts were not taken into account for this publication. Hence, the corresponding cleaned data set (*Digidin_DataSetPublication1.sav*) includes data on 49 participants who learned with modelling examples. Participants are described in Meier et al. (2022). For publication 2, all three conditions were considered. Therefore, the corresponding cleaned data set (*Digidin_DataSetPublication2.sav*) includes data on 67 participants ($n_{Condition1} = 21$, $n_{Condition2} = 25$, $n_{Condition3} = 21$). Participants are described in Meier et al. (2023). Note that the differences of 49 participants in groups 1 and 2 in publication 1 versus 46 participants in groups 1 and 2 in publication 2 are due to the fact that in publication 1, participants could also be analysed on a dependent variable if they had missing values on another variable, whereas in publication 2, only complete data sets could be included in the analyses.

Procedure

The entire study took place on computers in the apprentices' schools. All learning and testing materials were presented in digital form via the page-based online survey tool LimeSurvey. PDF exports of these surveys can be found in the attachement (*Survey Export Session 1.pdf (this is an annotated version that allows a mapping to the different variables in the datasets); Survey Export Session 2_Condition1.pdf (this is an annotated version that allows a mapping to the different variables in the datasets); Survey Export Session 2_Condition2.pdf; Survey Export Session 2_Condition3.pdf*). 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 and avoided that few slow participants cause long waiting times for the other participants (see maximum duration until termination 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			
Phase 1	Introduction to study and computer simulation			30	45
Break				10	10
Phase 2	Assessment of motivation			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			
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	Assessment of motivation			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			5	
TOTAL SESSION 2				**250**	**275**

Learning Materials

In the intervention, apprentices learned a complex four-step diagnostic strategy with additional sub-steps that should help them to proceed in a structured way when diagnosing car malfunctions. The intervention comprised two learning phases. In the first learning phase, apprentices watched five animated instructional videos explaining the four-step diagnostic strategy (16:33 minutes). These instructional videos can be found in *Instruktionsvideos.zip.* Participants also completed four practice tasks, which served as *organizational prompt* (Roelle et al., 2017), and received the correct solution. Learning phase one took 35 minutes. Participants in all conditions received this first learning phase.

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 modelling example: 12 videos, total duration: 25:50 minutes, see *ModellingExample1.zip*; second modelling example: 10 videos, total duration: 19:37 minutes, see *ModellingExample2.zip*).

Both the diagnostic strategy and instructional videos and modelling examples were developed 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 either 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)" in the *retrospective self-explanation prompt* condition or "Which troubleshooting step **comes next**? Explain how you will proceed with this step and why it is important for troubleshooting (in general)" in the *anticipatory self-explanation prompt* condition. 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. Watching the video modelling examples one and two (incl. answering the prompts and reading the correct answers) took the participants about 50 minutes and 40 minutes, respectively. Thus, with the instructional videos in learning phase one taking 35 minutes and with a 5-minute break between learning phases, the total intervention took approximately 130 minutes.

Participants in the control condition did not receive the modelling examples and did not answer self-explanation prompts. Instead, these participants tried to diagnose the same malfunctions in the computer simulation that the expert in the modelling examples diagnosed. Hence, instead of studying examples, participants in the control condition practised 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. All tests can be studied in the PDF exports of the surveys. To assess *general prior knowledge* about car diagnoses, we used two different tests in session one only. To measure the apprentices' *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' *motivation* regarding making diagnoses was used in sessions one and two. Finally, a questionnaire aiming at the apprentices' *cognitive load* was given after the intervention took place in session 2. 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. These responses were completely in line with the diagnostic strategy taught. In addition, we also looked for alternative solutions in the responses of all participants that could be assessed as equivalent or similarly good from a domain perspective as the ideal responses based on the diagnostic strategy. 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

Prior knowledge measures were only relevant for research questions that were investigated in the second publication. Hence, these measures are not included in the data set regarding the first publication (*Digidin_DataSetPublication1.sav*) but only in the data set regarding the second publication (*Digidin_DataSetPublication2.sav*).

As the first measure of general prior knowledge about car diagnoses, the participants' *diagnosis-relevant reception competence* was assessed. 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 five out of 24 items (*DRCTestQuestion1_V1A2* to *DRCTestQuestion5_V2A17*) in the diagnosis-relevant reception competence (DRC) test by Norwig et al. (2021). First, we selected the items for their midrange solution rate (ranging from 32% to 71% in Norwig et al., 2021) to prevent floor and ceiling effects. Furthermore, we selected the items with the highest item-total correlation (> 0.43 for all 5 items). The total score in the diagnosis-relevant reception competence test is included in *Digidin_DataSetPublication2.sav* as *DRCTestScore*.

Second, we selected three of seven items of a *partial skills* test by Abele (2014), in which participants were instructed to perform specific measurements in the simulation and to evaluate whether the measurement results indicated a malfunction or not. Again, we selected items with a high item-total correlation (between .48 and .60 in Abele, 2014) and made sure that the items did not overlap in content with other tests used. Total score in the partial skills test is included in *Digidin_DataSetPublication2.sav* as *PartialSkillsTestScore*.

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; *StrategyDescriptionQuestionA*), and (2) how they would narrow down which components might be responsible for a malfunction (*StrategyDescriptionQuestionB*). Answers to these two questions were coded as explained in *CodingSchemeStrategyDescription.docx*. The total score is included in the variable *StrategyDescriptionPre* (for session one) and *StrategyDescriptionPost* (for session two).

Second, in the *strategy completion test*, apprentices carried out or described (parts of) steps of the diagnostic strategy in four different scenarios. Hence, this test assessed scaffolded diagnostic skills. 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. Answers to these questions are included in the variables *StrategyCompletionTaks_1a* to *StrategyCompletionTask_4d*. Answers to these questions were coded as explained in *CodingSchemeStrategyCompletion.docx*. The total score is included in the variable *StrategyCompletionPre* (for session one) and *StrategyCompletionPost* (for session two).

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 on their own. Eventually, participants described the cause of the malfunction and how it could be repaired. These answers were coded as described in *CodingSchemeDiagnosis.docx*. 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. The dataset for the first publication *Digidin_DataSetPublication1.sav* only includes the score in the first diagnosis in session 1 and 2 as *DiagnosisFM13Pre* and *DiagnosisFM13Post*. The dataset for the second publication *Digidin_DataSetPublication2.sav* includes this data but also the score in the second diagnosis in session 2 *DiagnosisFM47Post*.

Motivation and Cognitive Load

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 (*Selfefficacy1* to *Selfefficacy5*) assessed the apprentices' self-efficacy regarding the following diagnosis in the simulation (Cronbach's $\alpha = 0.90$; Bandura, 2006). For a mean score of these items, see *SelfefficacyPre* and *SelfefficacyPost* in both datasets *Digidin_DataSetPublication1.sav* and *Digidin_DataSetPublication2.sav*. Five items (*Interest1* to *Interest5*) related to the apprentices' interest in car diagnosis and diagnostic strategies (Cronbach's $\alpha = 0.86$). For a mean score of these items, see *InterestPre* and *InterestPost* in both datasets *Digidin_DataSetPublication1.sav* and *Digidin_DataSetPublication2.sav*. Moreover, four items (*Challenge1* to *Challenge4*) related to the extent to which the apprentices perceived the upcoming diagnosis in the simulationas a challenge (Cronbach's α = 0.87). For a mean score of these items, see *ChallengePre* and *ChallengePost* in both datasets *Digidin_DataSetPublication1.sav* and *Digidin_DataSetPublication2.sav*. Eventually, five items (IncompetenceFear1 to IncompetenceFear5) related to whether they perceived incompetence fear (Cronbach's α = 0.93). For a mean score of these items, see *IncompetenceFearPre* and *IncompetenceFearPost* in both datasets *Digidin_DataSetPublication1.sav* and *Digidin_DataSetPublication2.sav*.

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: *CL_Intrinsic1* and *CL_Intrinsic2*), germane (two items; *CL_Germane1* and *CL_Germane2*), and extraneous cognitive load (three items; *CL_Extr1*, *CL_Extr2*, and *CL_Extr3*; Klepsch et al., 2017). Reliability was acceptable (intrinsic load: Cronbach's α = 0.79; germane load: Cronbach's α = 0.84; extraneous load: Cronbach's α = 0.66). Data on cognitive load is included in both datasets *Digidin_DataSetPublication1.sav* and *Digidin_DataSetPublication2.sav* as *CL_Intrinsic*, *CL_Germane*, and *CL_Extraneous.*

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. 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). This data is included only in the dataset regarding the first publication *Digidin_DataSetPublication1.sav*. These variables are *EvaluationApplicability1,* *EvaluationApplicability2, EvaluationInterestingness1, EvaluationInterestingness2_negative, EvaluationLength1, EvaluationLength2_negative, EvaluationStructure1_negative, EvaluationStructure2, EvaluationNarratorQuality1, EvaluationNarratorQuality2, EvaluationComprehensibility1_negative, EvaluationComprehensibility2, EvaluationRecommendation1, EvaluationRecommendation2 ,EvaluationSatisfaction*

References

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

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

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.

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

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

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

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