CrossMark

Diagnostic Problem-Solving Process in Professional Contexts: Theory and Empirical Investigation in the Context of Car Mechatronics Using Computer-Generated Log-Files

Stephan Abele[1]

Abstract This article deals with a theory-based investigation of the diagnostic problem-solving process in professional contexts. To begin with, a theory of the diagnostic problem-solving process was developed drawing on findings from different professional contexts. The theory distinguishes between four sub-processes of the diagnostic problem-solving process and includes several hypotheses. According to the theory, the quality of the sub-processes "representing information" and "testing hypotheses" causally influences the diagnostic problem-solving success. Additionally, the theory suggests that the influence of "testing hypotheses" on the problem-solving success is higher than the influence of "representing information". Moreover, the theory assumes that the influence of the quality of "representing information" on the success is mediated by the quality of "testing hypotheses". These hypotheses were examined in the context of car mechatronics using diagnostic problems of the car sector, a computer-based assessment and a sample of car mechatronic apprentices ($N = 339$). To operationalize the sub-processes' quality, observable critical behavior was theoretically identified and extracted from computer-generated log-files. The empirical results were largely in line with the hypotheses and indicated a (very) first corroboration of the theory in the context of car mechatronics. The theory could be helpful to investigate and teach diagnostic problem solving in different professional contexts. Further studies, however, should scrutinize whether the theory applies to other studies and professional contexts.

✉ Stephan Abele
abele@bwt.uni-stuttgart.de

[1] Department of Vocational Education, Institute for Educational Science, University of Stuttgart, Geschwister-Scholl-Straße 24D, 70174 Stuttgart, Germany

🖄 Springer

Introduction

This article deals with a theory-based investigation of the diagnostic problem-solving process in professional contexts. In the first part of the article, a theory of the diagnostic problem-solving process is developed drawing on findings from different professional contexts. In the second part, this theory is empirically investigated in the context of car mechatronics. The emphasis is on a theory that relates the mental problem-solving process to observable problem-solving behavior and the observable diagnostic problem-solving success and provides hypotheses that can be examined empirically. Following Barrows and Feltovich (1987), the article concentrates on the overall process and does not strive to enhance our knowledge of specific aspects of the diagnostic problem-solving process.

Diagnostic problems refer to situations in which the cause(s) of an undesired state has/have to be detected. Diagnostic problems are relevant in diverse professional contexts: Teachers have to diagnose causes of learning difficulties, physicians must identify reasons for diseases, technicians and engineers have to detect causes of malfunctioning machines and so forth. Taking car mechatronics as an example, a diagnostic problem could reflect a situation in which the cause of a car's lighting system defect must be detected. This definition of a diagnostic problem is in line with Schaafstal et al. (2000, p. 75) but differs from other studies (e.g., Kassirer et al. 2010, p. 6) since it does *not* cover treatment options (e.g., repair). Diagnostic problems have the critical attributes of a problem (Jonassen 2000, p. 65): there is an unknown (e.g., cause of a lighting system defect) and it is worth finding this unknown (e.g., to satisfy a customer).[1]

The diagnostic problem-solving process is defined as the mental (latent, non-observable) activities that reveal the cause(s) of the undesired state and underlie the solution of a diagnostic problem (e.g., Durning et al. 2013, p. 444). Understanding this process means knowing critical differences between good and poor problem solvers and enhances our knowledge of how to foster diagnostic problem solving (Barrows and Feltovich 1987; p. 86). A theory of the diagnostic problem-solving process is useful for both assessment and didactical purposes.

There are a plethora of studies on the diagnostic problem-solving process in different professional contexts (technical context: e.g., Rasmussen 1993; Rouse 1983; Hoc and Amalberti 1995; medical context: e.g., Elstein et al. 1990; Norman 2005; Croskerry 2009; scientific context: Klahr and Dunbar 2000). These studies provide many valuable insights and process models. As Kassirer et al. (2010, p. 4) pointed out, studies covering the overall process are still reliant on normative principles which cannot be investigated empirically, due to a lack of information. Moreover, studies like the one of Coderre et al. (2003) that relate the diagnostic problem-solving process to the problem-solving success are rare.

In order to investigate theories on the diagnostic problem-solving process, process data is needed. The collection of such data (typically resulting from think-aloud

---

[1] According to this definition, each situation that is related to finding an unknown represents a problem, even if the unknown is found by applying a routine. The decisive point is that the situation does not reveal whether the routine will be successful; rather, the appropriateness of the routine must be tested.

protocols) is very often time-consuming and costly, which favors studies with small samples and low statistical power (Krems 1994, p. 60). In this context, log-files from computer-based assessment might be helpful. Log-files can be automatically generated during a computer-based assessment and contain behavioral actions of an individual (e.g., mouse clicks). Such behavioral process data can easily be gathered with large samples.

In the following sections, the theory of the diagnostic problem-solving process in professional contexts is presented. Following that, the theory is investigated using log-file data from a computer-based assessment for car mechatronics,[2] authentic diagnostic problems of the car sector and a sample of 339 apprentices. In light of the huge amount of research on diagnostic problem solving, the theory cannot cover all findings and demands further theoretical elaboration as well as empirical investigation. Nevertheless, I assume that the present study can contribute to research on the diagnostic problem-solving process in professional contexts by generating a theory that connects the diagnostic problem-solving process to observable behavior as well as the problem-solving success and that can be examined empirically.

Diagnostic Problem-Solving Process in Professional Contexts

The diagnostic problem-solving process has been investigated in medical and technical professional contexts. Many studies in the medical context do not use the term "diagnostic problem solving" but "clinical reasoning" (e.g., Kassirer et al. 2010; Barrows and Feltovich 1987). In the technical context, the term "troubleshooting" is very common (e.g., Schaafstal et al. 2000, p. 75). Frequently, both terms are referred to as covering diagnostic problem solving and treatment options. The emphasis, however, is often on diagnostic problem solving (Jonassen and Hung 2006, p. 79).

Schaafstal et al. (2000, p. 79) insinuated that the diagnostic problem-solving process consists of four sub-processes: formulate problem description, generate causes, test and evaluate. While the first sub-process refers, among other things, to understanding the problem and its symptoms, the second sub-process aims at generating hypotheses on the cause(s) of the defect. These diagnostic hypotheses are tested within the next sub-process. Finally, the entire process and its results have to be evaluated, leading to a statement on the cause(s) of the undesired state. Comparing different conceptions of the diagnostic problem-solving process, Jonassen and Hung (2006) concluded that the process starts with building a mental representation of the problem. This mental representation contains information on the problem (e.g., the undesired state) and relevant systems. Additionally, it often includes information coming from external information sources (e.g., circuit diagrams or health records). Drawing on the mental representation, the problem's cause(s) is/are examined: diagnostic hypotheses are

---

[2] The professional field of car mechatronics covers, among other things, troubleshooting, repair and maintenance of cars (Baethge and Arends 2009, p. 33–47). In Germany, car mechatronic apprentices usually attend a 3.5 years training programme including a school-based and workplace-based training ("dual apprenticeship system"). The training of car mechatronic technicians differs significantly from one country to the next (Baethge and Arends 2009, p. 34).

formulated and tested. Finally, the problem's solution is generated and evaluated. Barrows and Feltovich (1987) elaborated a synthesis of research on the clinical reasoning process. They identified the following sub-processes of clinical reasoning as being especially important: gathering information on the patient's problem, generating diagnostic hypotheses, examining these hypotheses and evaluating corresponding results. The examination of hypotheses sometimes include laboratory tests that cannot be applied in situ. Here, there is a difference to the technical context, where tests can usually be conducted at once and often are not as costly as in the medical context. In their book on clinical reasoning, Kassirer et al. (2010) stressed the role of gathering information on the problem (e.g., gender, age, appearance of the patient), hypotheses generation and testing. There are, however, several studies casting doubt on the key role of formulating hypotheses in diagnostic problem solving (e.g., Donner-Banzhoff et al. 2016). This criticism will be addressed in the final discussion.

Based on the studies cited, and inspired by others (medical context: Elstein et al. 1990; Durning et al. 2013; Patel et al. 1996; scientific context: Klahr and Dunbar 2000), a theory on the diagnostic problem-solving process in professional contexts was developed. In the next sections this theory will be presented. The theory differentiates between four sub-processes: representing information, as well as generating, testing and evaluating diagnostic hypotheses. Moreover, it assumes certain relations between the sub-processes as well as between these sub-processes and observable problem-solving behavior. The theory is a synthesis and further development of existing literature.

Representing Information

The aim of the sub-process "representing information" is to mentally represent the information required to solve the diagnostic problem. Here, there are three different types of information: The first type corresponds to information about the diagnostic problem: the undesired state (e.g., symptoms), the goal state (e.g., knowing the cause of the defect) and context information (e.g., brand of a car). The second type of information is related to testing diagnostic hypotheses, that is, test procedures and techniques, information necessary to conduct a test and so forth. In the technical context, testing hypotheses often have finding out the location of system components (e.g., a fuse) as a prerequisite. The third information type refers to evaluating diagnostic hypotheses: For example, testing the hypothesis of a broken fuse demands reference values; otherwise, whether the fuse is broken cannot be evaluated.

Representing information is conceptualized as a mental sub-process that often relies on diagnostic knowledge and information resulting from external data. Diagnostic knowledge represents problem-solving experience and can be retrieved from memory. "External information" comes from interactions with the problem environment. To represent "external" information, car mechatronic technicians typically read problem descriptions, explore cars and their technical particularities and so forth. While retrieving knowledge is a mental (non-overt) activity, interactions with the environment become manifest in observable information behavior. Information behavior is both a consequence and prerequisite of the sub-process "representing information".

The research of Groves et al. (2003, p. 308) and Donner-Banzhoff et al. (2016) from the medical field provides evidence that "information representation" can be considered an independent sub-process and is critical to the problem-solving success (see also

Coderre et al. 2003; Johnson et al. 1995 and Nendaz et al. 2005). Johnson et al. (1995) summarized findings from the technical domain concluding that "information representation" influences the following problem-solving sub-processes. According to Elstein et al. (1990, p. 11), there is no correlation between thoroughness of information collection and the quality of information interpretation. This suggests that the diagnostic problem-solving success does not result from collecting a lot of information; rather, it results from gathering *critical* information, that is, information making sense from a substantive point of view and in terms of the present problem (see also Joseph and Patel 1990).

According to research on expertise (e.g., Boshuizen and Schmidt 2008, p. 115; Feltovich et al. 2006, p. 47), if a well-known diagnostic problem is worked on, case-specific knowledge is retrieved providing information on the diagnostic problem at hand. In this case, observable information behavior is replaced by retrieving knowledge. For this reason, experienced problem-solvers often need to collect relatively little critical information to solve a well-known problem (Groves et al. 2003, p. 308; Schmidt et al. 1990, p. 618).

Generating Diagnostic Hypotheses

The sub-process "generating diagnostic hypotheses" aims at formulating diagnostic hypotheses. Diagnostic hypotheses are defined as mental representations comprising potential causes(s) of an undesired state (e.g., a lighting system defect might be caused by a broken fuse) or, in other words: A diagnostic hypothesis contains a potential but untested problem solution. Diagnostic hypotheses are formulated drawing on information about the diagnostic problem. Thus, the quality of the "generating hypotheses" depends on the quality of "representing information".

In real-life problem-solving, diagnostic hypotheses are seldom explicated. Consequently, generating hypotheses is usually *not* associated with observable problem-solving behavior and interactions with the environment. In many professional contexts, it is a purely mental process, so that the relevance of this sub-process cannot be examined straightforwardly. Some studies meet this challenge by confronting their testees with artificial rather than authentic diagnostic problems, "forcing" them to explicate their diagnostic hypotheses (e.g., Krems and Bachmaier 1991; Mehle 1982). For example, given specific symptoms of a problem (i.e., some details of a diagnostic problem), testees are asked to explicate as many diagnostic hypotheses as possible.

It has been widely accepted that the generation of hypotheses is crucial to solving diagnostic problems (e.g., Elstein et al. 1990, p. 9; Morris and Rouse 1985, p. 508), even though this very plausible assumption has been tested rarely. Morris and Rouse (1985) reported findings showing that good problem-solvers formulate more hypotheses than poor ones do. Mehle (1982) found that, when administering diagnostic problems from the car sector, experts and novices generate a comparable number of hypotheses (see also Elstein et al. 1990, p. 9). In contrast, Krems and Bachmaier (1991) and Patel et al. (1996, p. 133) determined that experts formulate fewer hypotheses than novices do. The evidence appears to be inconsistent. Although there might be several reasons for that, the following seem to be of particular importance: When the role of "hypothesis generation" is investigated using somewhat artificial diagnostic problems,

the case/content specificity of diagnostic problem-solving is ignored (Schwartz and Elstein 2008, p. 225) and generating hypotheses is not systematically related to the problem-solving success. As Krems and Prechtl (1991) pointed out, it is not the number of hypotheses, but their quality, that makes the problem-solving success.

Testing Diagnostic Hypotheses

"Testing diagnostic hypotheses" is about gathering evidence to judge whether a diagnostic hypothesis is appropriate. It encompasses three main steps: (1) deducing observable events from a diagnostic hypothesis given the hypothesis is true, (2) planning how to test the occurrence of these events and (3) testing the events' occurrence and the mental representation of the test result. For example, testing the hypothesis of a broken fuse implies (1) deducing relevant events (e.g., an infinite resistance of the fuse), (2) planning how to test the resistance of the fuse (e.g., using a multimeter) and (3) measuring the resistance. "Hypothesis testing" is based on information coming from the problem environment and diagnostic knowledge, and requires application of this information to produce hypothesis-relevant evidence/information. For the latter reason, and *in contrast* to the present study, "hypothesis testing" is usually subsumed under the term "data collection", not differentiating between "information representation" and "hypothesis testing" in the *medical* field (e.g., Nendaz et al. 2005, p. 415; Schwartz and Elstein 2008, p. 224). The quality of "testing hypotheses" depends on the quality of "representing information" and "generating hypotheses".

The sub-process "testing hypotheses" is associated with observable test behavior. Whereas deduction and planning activities are usually not observable, the collection of evidence demands interactions with the environment. There might be cases in which diagnostic hypotheses can be tested without actively collecting external data; however, such cases should be (very) rare. The test behavior indicates the quality of the sub-process and influences the problem-solving process.

"Testing hypotheses" is considered an important sub-process of diagnostic problem-solving, although there is little empirical evidence supporting this assumption (e.g., Kassirer et al. 2010, p. 15). The study of Elstein et al. (1990, p. 11) showed that good diagnostic problem solvers have clearer concepts (i.e., knowledge) of how to test diagnostic hypotheses than poor ones do. Morris and Rouse (1985, p. 504) gave some empirical evidence proving the relevance of hypothesis tests to solving diagnostic problems. This is very plausible: When pure guessing is not acceptable, as in most professional contexts, the solution of a diagnostic problem should be grounded on evidence coming from hypothesis tests.

Evaluating Diagnostic Hypotheses

"Evaluating diagnostic hypotheses" aims to evaluate the evidence coming from hypothesis tests and decide whether a hypothesis is acceptable. The crucial point here is to interpret evidence in light of a diagnostic hypothesis and to conclude whether the evidence corroborates or refutes the hypothesis. When evaluating a diagnostic hypothesis, it might be necessary to consider several pieces of evidence and alternative hypotheses. "Evaluating hypotheses" is influenced by the foregoing sub-processes.

The evaluation sub-process is mental but leads to an observable problem solution. A diagnostic problem is solved when the correct cause (e.g., broken fuse) of an undesired state (e.g., lighting system defect) is given and proved by evidence (e.g., test result). The problem's solution (i.e., the problem-solving success) is the consequence of evaluating hypotheses.

It turned out that successful diagnostic problem-solvers were superior in interpreting evidence (Johnson et al. 1995, p. 10). This corresponds to Morris and Rouse (1985, p. 504), who additionally highlighted the fact that incorrect hypotheses are more quickly eliminated by successful than unsuccessful problem solvers. According to these findings, the quality of "hypothesis evaluation" varies among individuals and should affect the problem-solving success. Klahr and Dunbar (2000, 77ff.) showed that individuals can have serious problems with correctly interpreting data and give two reasons for that: confirmation bias, and a lack of alternative hypotheses. Another reason might be that individuals do not understand test results/evidence. For example, they cannot interpret the measurement value "OL", since they do not know that "OL" symbolizes an infinite resistance and/or they cannot apply this evidence to judge the appropriateness of a hypothesis. Table 1 gives an overview of the sub-processes.

Obviously, the accurate order of the problem-solving sub-processes can strongly differ between individuals and diagnostic problems: There might be situations in which "generating hypotheses" is followed by "representing information", "evaluating hypotheses" is followed by generating another hypothesis and so on. Against this background, the theory does *not* claim to reflect the "real" chronological sequence of the mental activities conducted during problem solving. The decisive point here is to have a theory that organizes mental problem-solving activities into sub-processes, gives the causal relationship between these sub-processes and causally explains the problem-solving success. Of course, causal relationship implies temporal precedence but, for instance, considering the influence of "representing information" on "testing hypotheses", it is beside the point whether the test information is represented before the hypothesis is formulated or afterwards.

Table 1 Overview and key aspects of the diagnostic problem-solving sub-processes

Mental sub-process	Aim	Associated manifest behavior
(1) Representing information	To mentally represent information relevant to the diagnostic problem	Retrieving and using external information material
(2) Generating diagnostic hypotheses	To generate diagnostic hypotheses providing potential but untested problem solutions	Usually none[a]
(3) Testing diagnostic hypotheses	To collect and mentally represent evidence required to judge whether a diagnostic hypothesis is appropriate	Conducting tests (e.g., resistance measurement or laboratory tests) to obtain relevant evidence
(4) Evaluating diagnostic hypotheses	Evaluating the evidence coming from tests to decide whether a hypothesis is acceptable	Giving the cause(s) of the diagnostic problem (i.e., the problem's solution)

[a] Please note that the focus here is on individual problem solving. When a team works on diagnostic problems, it might frequently be the case that diagnostic hypotheses are explicated (i.e., become manifest)

Operationalization of the Quality of "Representing Information" and "Testing Diagnostic Hypotheses"

It is assumed that the four sub-processes or, more specifically, their quality, causally influence the observable problem-solving success. In order to empirically investigate this influence, the quality of the non-observable sub-processes has to be operationalized. In this study, the focus is on the quality of "representing information" and "testing diagnostic hypotheses".

Indicators of the sub-Processes' Quality: Critical Information Behavior and Critical Test Behavior

The quality of "representing information" can be operationalized using critical information behavior. With regard to "representing information", a high quality is associated with mentally representing a lot of critical information. Critical information refers to information that is relevant to solve the diagnostic problem. For example, it is necessary to know the undesired state of the problem (e.g., the car defect) to solve a diagnostic problem. To represent such critical information, critical information behavior is required. For example, in order to learn the undesired state of a problem, the problem description must be selected. This critical information behavior is triggered by the sub-process "representing information" to make available and to mentally represent critical information. Generally speaking, critical information behavior is both a causal result and a prerequisite of the quality of the sub-process. In this vein, critical information behavior can be interpreted as a quality indicator of "representing information": If problem solvers exhibit critical information behavior, a higher quality of "representing information" is assumed than if they do *not* show such behavior.

The quality of "testing diagnostic hypotheses" can be operationalized using critical test behavior. In terms of "testing hypotheses", a high quality is associated with mental activities initiating critical tests. Critical tests refer to collecting evidence that is needed to investigate a diagnostic hypothesis. Critical tests require critical test behavior (e.g., measuring a fuse's resistance). Critical test behavior demonstrates reasonable mental activities (i.e., deduction, planning and application of tests) and provides relevant evidence. Accordingly, critical test behavior can be considered a quality indicator of "testing hypotheses".

Identification of Critical Information Behavior and Critical Test Behavior

Critical information behavior and critical test behavior (in short: critical behavior) can be theoretically identified based on critical diagnostic hypotheses. Critical diagnostic hypotheses are defined as assumptions that relate to a specific diagnostic problem, provide potential causes of the undesired state of the problem and make sense from a substantive point of view. For instance, it makes sense to suppose that a broken fuse is the reason for a lighting system defect and, consequently, this assumption is considered a critical diagnostic hypothesis. In contrast, it is unreasonable to assume that the defect is caused by an empty fuel tank. Diagnostic problems usually allow for several critical diagnostic hypotheses: A lighting system defect might be caused by a broken fuse, lamp and so forth.

Based on the critical diagnostic hypotheses, the critical behavior can be identified applying domain-specific expertise. For example, in order to examine a critical diagnostic hypothesis (e.g., the fuse causes the lighting system's defect) critical information and critical tests are needed. Critical information comes from specific critical information behavior (e.g., selecting the fuse card to check the fuse's location), critical tests are associated with specific critical test behavior (e.g., measuring the fuse's resistance). In the following empirical study, the theoretically identified critical behavior is used to determine individuals' quality of "representing information" and "testing diagnostic hypotheses" (see the section on measures and scoring).

## Aim and Hypotheses of the Empirical Study

The aim of the empirical study is to investigate the theory of the diagnostic problem-solving process. Fig. 1 summarizes the theory graphically.

The theory includes the *theoretical* hypothesis that the quality of the sub-processes "representing information" and "testing diagnostic hypotheses" have – mediated by the other sub-processes – a causal influence on the problem-solving success. As mentioned before, the quality of these latent sub-processes are operationalized using critical behavior. From this and the theoretical hypothesis results the following *empirical* hypothesis: The critical information behavior and the critical test behavior have an effect on the diagnostic problem-solving success (RH1).

The theory advises that, in comparison to "representing information", "testing hypotheses" draws upon more critical mental activities and, therefore, should have a higher influence on the problem-solving success. Against this background, it is assumed that the effect of the critical test behavior on the success is stronger than the effect of the critical information behavior (RH2).

Experience (i.e., problem-related knowledge) can be retrieved from memory and can lead to a high quality of "representing information", although no observable critical information behavior is exhibited. This suggests that the effect of the critical information behavior on the problem-solving success is moderated by problem-related experience (RH3).

Diagnostic problem-solving process
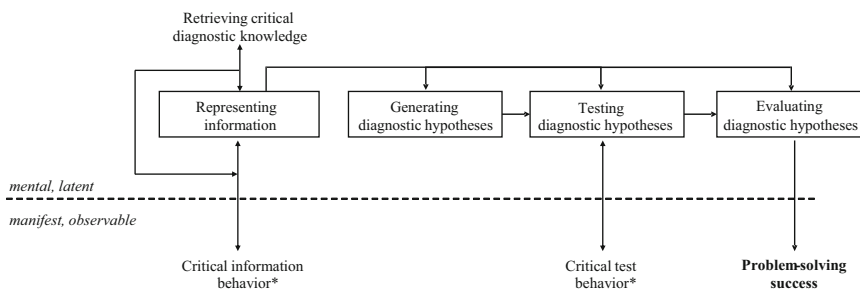
* Critical behavior is identified by applying domain-specific expertise to a specific diagnostic problem and based on critical diagnostic hypotheses.

Fig. 1 Theory of the diagnostic problem-solving process in professional contexts

The quality of "representing information" affects the quality of "testing hypotheses" which, in turn, influences the problem-solving success. Consequently, the influence of the critical information behavior on the diagnostic problem-solving success should be mediated by the critical test behavior (RH4).

The theory has domain-general and domain-specific aspects. Drawing on findings from different professional contexts, it assumes that the distinction of four sub-processes, their causal relationship and empirical consequences apply to different professional contexts. The critical behavior, however, can only be identified by applying domain-specific expertise to a specific diagnostic problem. Thus, the theory can be investigated only in specific professional contexts. In this study, the empirical hypotheses were investigated in the context of car mechatronics.

With regard to RH1 and RH4, contradictory empirical results have logical implications for the theory. For example, if no effect of the critical test behavior on the critical information behavior was found, the results would seriously question the hypothesis that the quality of "representing hypotheses" causally influences the quality of "testing hypotheses". Contradictory results might suggest that the theory, or at least parts of it, do not apply to the context of car mechatronics. They might also cast doubt on the operationalization of the sub-processes' quality. This would be a matter of further examination. Importantly, if the empirical results are in line with RH1 and RH4, there is no logical ground to conclude that the theory is confirmed or even verified. From a logical viewpoint, it is impossible to verify empirical hypotheses and related theories; however, it is possible to investigate whether empirical data falsify the empirical hypotheses and the related theories (Popper 2005). If the investigation fails to falsify the hypotheses, the theory is *corroborated* but not confirmed or even verified. As the theory presented here applies to different professional contexts, it has to be investigated in different professional contexts. The following study focuses on a specific professional context (car mechatronics) and resembles a (very) first empirical investigation of the theory.

Please note that RH2 and RH3 are not logically derivable from, but in line with, the theory. Regarding these empirical hypotheses, the empirical results do not have logical implications for the theory but they provide evidence to evaluate its empirical plausibility.

Material and Methods

The Computer-Simulation-Based Assessment

The diagnostic problem-solving behavior (i.e., the critical information behavior and critical test behavior) and success were assessed in the context of car mechatronics using a computer simulation. The computer simulation uses authentic graphic material (pictures, screenshots, etc.) and represents the following parts of the work environment of car mechatronics: (1) a selection of car systems, (2) a toolbox and (3) a computer-based expert system (Fig. 2).

(1) The simulation covers four systems of a VW Golf, which were identified to be of high practical relevance by experienced car mechatronic technicians, teachers/trainers of car mechatronic apprentices and relevant scientists (Baethge and Arends 2009, p. 16). Here, the system "electronic engine management" is relevant. In this system, 17
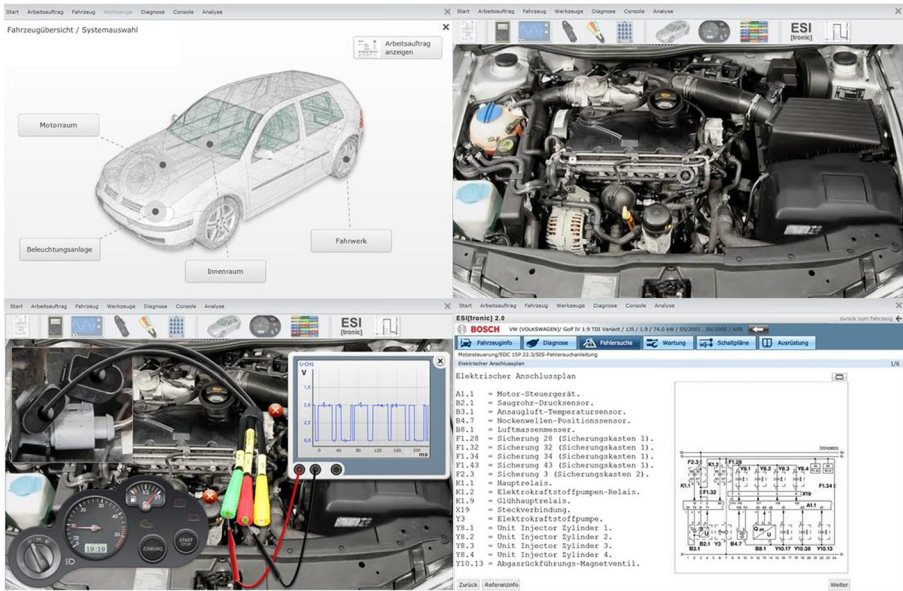
Fig. 2 Screenshots of the computer simulation in German (top left: start page giving an overview of the car systems; top right: the upper part shows the icons of the toolbox, below the motor compartment referring to the system "electronic engine management" is shown; bottom left: measurement of a signal using the oscilloscope, cockpit and adapter; bottom right: circuit diagram retrieved from the computer-based expert system)

components (plugs of actuators and sensors, the battery, etc.) are available. (2) The toolbox contains icons representing different work equipment (e.g., problem description, multimeter, fuse box, computer-based expert system). (3) Computer-based expert systems are an integral part of the car mechatronic technicians' work environment. The simulation covers relevant segments of the ESI[tronic] from Bosch, which is an internationally widespread system and applicable to a broad selection of car brands. It offers a great variety of relevant information.

The computer simulation provides numerous authentic diagnostic problem-solving steps: In the system "electronic engine management" alone, there are more than a thousand possibilities to measure voltage, resistance and signals. A guiding principle of the simulation's development was to allow interactions that largely correspond with the professional reality of car mechatronics. The computer-simulation-based assessment proved to produce valid test score interpretations, that is, measures indicating authentic diagnostic problem-solving skills (Gschwendtner et al. 2009).

Measures and Scoring

In the assessment, two diagnostic problems were administered, both referring to the fuel temperature sensor. P1 and P2 were similar in terms of their symptoms, but differed in terms of the symptoms' causes and their difficulty. In previous studies, P1 was solved by 85% (Gschwendtner et al. 2009, p. 573) and P2 by 25% (Abele et al. 2014, p. 174) of the testees. To detect the problems' causes, electrotechnical measurements had to be conducted. The problems allowed for using the computer-based expert system to retrieve location diagrams, circuit diagrams and test instructions as well as to

read out the error-storage. Test instructions contained information on electrotechnical measurements useful to solving the diagnostic problem.

The *problem-solving success* was determined by analyzing handwritten documentations. A problem was considered solved if the correct cause had been given, documented and proved by appropriate measurements. The scoring was conducted by two independent raters applying a coding manual to the documentations and produced dichotomous data (correct solution: 1, incorrect solution: 0; no partial credits). In very rare cases of diverging scores, content-oriented discussions produced a consensual scoring. Additionally, the interrater reliability Cohen's Kappa was calculated for a selection of the sample ($N = 67$): κ = .95 (P1) and .97 (P2).

To determine the *critical behavior*, the problems were analyzed and the critical diagnostic hypotheses were identified (see first line of Table 2). Since both problems referred to the fuel temperature sensor and comparable symptoms, the diagnostic hypotheses were identical. The critical information behavior and critical test behavior were derived from the critical hypotheses. While some information behavior is related to each critical hypothesis (e.g., PD), each test behavior is linked to a specific hypothesis: T1 stands for the critical test behavior relevant to test C1, T2 for the behavior relevant to test C2 and so forth. The critical behavior was dichotomously scored (behavior not shown: 0; behavior shown: 1; no partial credits) and extracted from computer-generated log-files (Fig. 3).

Please note that C2 represents the diagnostic hypothesis that contains the correct cause of P1; C3 contains the correct cause of P2. To solve the problem, the correct cause of the problem must be given and proved by relevant evidence. That is, the appropriateness of a diagnostic hypothesis had to be interpreted in the light of evidence. As mentioned before, the interpretation of evidence could be (very) challenging. Consequently, collecting relevant evidence (i.e., conducting relevant tests; e.g., T2) does not logically imply solving the diagnostic problem (e.g., P1).

For exploratory purposes, the overall number of behavioral actions and the following time measures were included, too: the time spent on a diagnostic problem (time on problem) and the time used for critical behavior (time on critical behavior). To determine the time on critical behavior, the theoretical assumptions of Table 2 were applied. That is, the time period for P1 was calculated adding the portions of time spent to conduct T2, I2, I3, ES and LD. PD was not considered, since there was almost no between-individual variation (Table 3). While the time spent on a problem is presumably rather a rough indicator of the process quality (Greiff et al. 2016), the time on critical behavior should be much more relevant, because it directly relates to critical behavior and the process theory.

Sample and Design

In order to test the research hypotheses, 339 car mechatronic apprentices nearing the end of the third year of training were sampled. Overall, three German federal states (Baden-Württemberg, Bavaria, Hesse) and 25 classes of vocational schools were included. The apprentices were 17 to 41 years old (M = 20.8) and, as could be expected in this profession, almost all of them were male (97%).

The problems were administered in a computer lab and within a large project. The relevant problems here refer to a testing time of 45 min (P1: 20 min and P2: 25 min). To

Table 2 Critical information and test behavior of P1 and P2

Critical behavior	C1: Broken cable	C2: Broken sensor (P1)	C3: Broken positive cable (P2)	C4: Broken negative cable	C5: Control unit defect
Critical test behavior	T1: Measuring voltage 1	T2: Measuring resistance 1	T3: Measuring resistance 2	T4: Measuring resistance 3	T5: Measuring voltage 2
Critical information behavior	I1: Selecting instruction 1	I2: Selecting instruction 2 I3: Selecting instruction 3	CD: Selecting the circuit diagram		PD: Selecting the problem description ES: Reading out the error-storage LD: Selecting the localization diagram

C1-C5: Critical diagnostic hypotheses; T1-T5: specific Tests; I1-I3: selecting Instruction 1–3 of the computer-based expert system; CD: selecting the Circuit Diagram; PD: selecting the Problem Description; ES: reading out the Error-Storage; LD: selecting the Localization Diagram

```xml
<?xml version="1.0"?>
- <activityLog>
    <activity lightState="00" motorState="000" timeStamp="12:15:19" id="1">lesson loaded: Fehler 6</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:00:02" id="2">menu open: SystemSelection</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:01:52" id="3">menu open: Document</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:01:52" id="4">Arbeitsauftrag page: 1</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:02:29" id="5">button close: Arbeitsauftrag</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:02:39" id="6">system loaded: Motorraum</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:02:42" id="7">system loaded: Cockpit</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:02:44" id="8">ignition on: true</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:02:45" id="9">shortCut open: DiagnoseSoftware</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:02:45" id="10">ESITRONIC screen: 00_01</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:02:47" id="11">ESITRONIC screen: 00_02</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:02:56" id="12">ESITRONIC screen: g4_00_01</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:02:58" id="13">ESITRONIC screen: g4_03_01_01</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:03:03" id="14">ESITRONIC screen: g4_03_01_13_01</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:03:10" id="15">ESITRONIC screen: g4_03_01_13_m01</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:03:13" id="16">ESITRONIC screen: g4_03_01_13_m01_m08_m02_04_01</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:03:18" id="17">ESITRONIC Fehlerspeicher lesen:</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:03:22" id="18">ESITRONIC screen: g4_03_01_13_m01_m08_m02_04_02</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:03:57" id="19">ESITRONIC screen: g4_44A3_01</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:04:12" id="20">button close: null</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:04:21" id="21">module open: Motorabdeckung</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:04:23" id="22">module zoom in: Kraftstofftemperatursensor</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:04:25" id="23">module open: Kraftstofftemperatursensor</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:04:27" id="24">module close: Kraftstofftemperatursensor</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:04:29" id="25">module open: Kraftstofftemperatursensor</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:04:34" id="26">shortCut open: Multimeter</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:04:42" id="27">Messspitze aufgesetzt: red/01-04-03-01</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:04:45" id="28">Messspitze aufgesetzt: black/01-04-03-02</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:04:45" id="29">Multimeter measure: 01-04-03-01/01-04-03-02 (5 Udc)</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:05:20" id="30">button close: Multimeter</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:05:22" id="31">shortCut open: DiagnoseSoftware</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:05:27" id="32">ESITRONIC screen: g4_44A3_02</activity>
    <activity lightState="100" motorState="000" timeStamp="+00:05:38" id="33">button close: null</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:05:41" id="34">ignition on: false</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:05:45" id="35">shortCut open: Multimeter</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:05:54" id="36">Messspitze aufgesetzt: black/01-04-03-04</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:05:59" id="37">Messspitze aufgesetzt: red/01-04-03-03</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:05:59" id="38">Multimeter measure: 01-04-03-03/01-04-03-04 (OL R)</activity>                T2
    <activity lightState="00" motorState="000" timeStamp="+00:06:39" id="39">button close: Multimeter</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:06:40" id="40">module close: Kraftstofftemperatursensor</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:06:42" id="41">module open: Kraftstofftemperatursensor</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:06:43" id="42">module close: Kraftstofftemperatursensor</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:06:44" id="43">module zoom out: Kraftstofftemperatursensor</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:06:46" id="44">button close: null</activity>
    <activity lightState="00" motorState="000" timeStamp="+00:06:49" id="45">lesson FINISHED</activity>
</activityLog>
```

Fig. 3 Example of a log-file. T2: test referring to the critical diagnostic hypotheses C2

control for position, that is exhaustion effects, a *Latin square design* was used (Frey et al. 2009, p. 45). Since both problems are similar, this design allowed analysis of the moderator effect of experience (RH3): For each problem, there is one group with no experience and one group with experience in diagnosing the fuel temperature sensor.

The standardized instruction for the assessment took 30 min. Initially, the instructor demonstrated the handling of the simulation by means of a video projector. Afterwards, the apprentices individually worked on standardized tasks concerning the handling of the simulation. In very rare cases, apprentices could not complete a task. Then, the instructor gave explanations in front of the class using the video projector. Finally, the apprentices were acquainted with how to prepare the handwritten documentation.

Statistical Analyses

Prior to the test of the research hypotheses, descriptive statistics and correlations of the process data with the problem-solving success were calculated using SPSS 23. For two dichotomous measures, the phi correlation coefficient was used; for a dichotomous and continuous measure, the point biserial correlation was calculated.

The hypotheses were tested using Mplus 7 and probit regression models. Such models can handle binary mediator and outcome variables. The mediator and outcome variables were statistically modeled as normal distributed latent response variables underlying the observed responses (Muthén 2011, pp. 19). In the case of binary *predictors*, identical to linear regression, observed responses (0, 1) were considered. The estimation of the parameters drew on the weighted least square estimator (WLSMV). Due to the sampling of school classes, that is, the dependence of observations, the Mplus option "TYPE = COMPLEX" was applied to get correct standard errors.

Table 3 Descriptive statistics and correlations of the critical behavior with the problem-solving success

Measure		Scale	*M*	*SD*	r (success)	*M*	*SD*	r (success)
			Problem 1			Problem 2		
Critical information behavior	PD	0/1	.99		.09	1.00		.02
	ES	0/1	.94		.12*	.96		.09
	I1	0/1	.70		−.03	.72		−.09
	I2	0/1	.42		.20***	.45		−.06
	I3	0/1	.32		.15**	.37		.00
	LD	0/1	.34		.22***	.34		.18**
	CD	0/1	.13		.01	.36		.43***
Critical test behavior	T1	0/1	.49		.28***	.62		.13*
	T2	0/1	.62		.70***	.51		.17**
	T3	0/1	.03		.12*	.15		.89***
	T4	0/1	.03		.15**	.12		.82***
	T5	0/1	.02		.00	.11		.29***
Behavioral actions			103.4	70.9	−.11*	147.6	95.4	.11*
Time on problem			581.9	303.9	−.16**	775.2	403.1	.13*
Time on critical behavior			93.21	93.2	.24***	79.97	80.7	.35***
M		0/1	.57			.16		

P1: $N = 336$; P2: $N = 339$; PD: selecting the Problem Description; ES: reading out the Error Storage; I1-I3: selecting Instruction 1–3; LD: selecting the Localization Diagram; CD: selecting the Circuit Diagram; T1-T5: specific Tests

* $p < .05$. ** $p < .01$. *** $p < .001$

To investigate the influence of the critical behavior on the problem-solving success (RH1), regression analyses for each problem were carried out and the following effects were examined: non-standardized regression coefficients (*b*), their standard errors (SE *b*) and statistical significance (*p*), the standardized coefficients (*β*) and the variance of the success explained by the critical behavior (R^2).

Whether the influence of the critical test behavior is stronger than the influence of the critical information behavior (RH2), was examined by comparing the R^2 of the critical information behavior to the R^2 of the critical test behavior. In order to examine the moderator effect of experience (RH3), two-group analyses were done where one group represented test takers with no experience and the other group represented test takers with experience. Speaking about Problem 1, the inexperienced group consisted of apprentices starting with Problem 1, and the experienced group was made up of apprentices starting with Problem 2 and working on problem 1 afterwards. So, when working on Problem 1, the experienced group already had experience with diagnosing the fuel temperature sensor. The moderator effect was evaluated using a two-step procedure: In the first step, the model parameters of both groups were freely estimated; in the second step, the effects (*b*) of the problem-solving behavior were equated across the groups. Whether both models differed significantly was analyzed by comparing the

Satorra-Bentler scaled chi-square values of the models (Mplus option "DIFFTEST"). Significant differences between the models indicate the moderation effect of experience.

Whether the influence of the critical information behavior is mediated by the critical test behavior (RH4), was investigated by means of mediation analyses. As suggested by MacKinnon (2008, pp. 334-335), the bias-corrected bootstrap method (1000 samples) was employed to evaluate mediation (i.e., indirect) effects, and the 95% confidence intervals were inspected. Mplus does not allow for bootstrapping and controlling for dependent observations simultaneously. Thus, the estimations were run twice; the results showed no mentionable differences. The mediation models were regarded as having a good overall fit if they met the following criteria: insignificant χ^2 value, ratio of the χ^2 value and degrees of freedom ≤3, root mean square error of approximation (RMSEA) ≤ .08, weighted root mean square residual (WRMR) ≤ 1 and comparative fit index (CFI) ≥ .95 (Moosbrugger and Schermelleh-Engel 2008, p. 319; for WRMR see Wang and Wang 2012, p. 70). Results were considered significant if p was less than .05 in all statistical analyses.

Results

Descriptive Statistics

As expected, both diagnostic problems proved to be different in difficulty: Problem 1 was solved by 57% and P2 by 16% of the testees (bottom line of Table 3). Considering *Problem 1*, Table 3 illustrates a great variation between the actions taken during the test: Whereas practically every testee selected the problem description (99%), very few of them conducted the tests "T3", "T4" and "T5" (3%, 3% and 2%). There were four significant and low-to-moderate correlations between the *critical information behavior* and the problem-solving success. In terms of PD and ES, the low correlations were to be expected due to little variation between the individuals. Most of the correlations of the *critical test behavior* and the success were significant, too. T2 appeared to have the highest correlation compared to the other coefficients. It should be noted that PD, T3, T4 and T5 will be dropped in further analyses of P1 because some expected values of the 2 × 2 contingency table underlying the correlations were below 5 (Field 2013, p. 735).

Analyzing *P2*, two significant correlations between the *critical information behavior* and the success were found, and the correlation of CD was substantial. The correlations of the *critical test behavior* and the problem-solving success were also significant, and the magnitude of the coefficient "T3" was salient. Again, the inspection of the expected frequencies of the contingency tables suggested excluding some behavior from further analyses: PD and ES. Complementary to these descriptive statistics, the contingency tables of the critical behavior and the problem-solving success are given in Appendix A, Table 8.

Comparing the results of the *other process data* revealed expectable differences: The difficult problem induced more behavioral actions and time on a problem than the

easier one. The correlations between these measures and the success were low-to-moderate, and the time on critical behavior turned out to be the best predictor of success. It is also worth mentioning that the direction of the correlations related to time on problem and behavioral actions changed depending on the problem, whereas the direction of the relationship of time on critical behavior with success was stable.

Effects of the Critical Behavior (RH1)

I2 and I3 turned out to correlate perfectly meaning that when instruction 2 was selected, instruction 3 was selected too. Due to this perfect correlation, the following analyses do not include I3. Expectedly, sensitivity analyses showed that the results did not depend on whether I2 or I3 was included.

Starting with *Problem 1*, Table 4 shows that the *critical information behavior* had a considerable overall effect on the problem-solving success (bottom line of model 1). Here, and in the following analyses, I1 were, surprisingly, revealed to have a negative effect, which will be dealt with in the discussion. Regarding the *critical test behavior*, the overall effect was higher than in model 1. The integration of the test and information behavior in a common regression model slightly increased the effect.

Turning to *P2*, the previous findings were largely confirmed (Table 5): The *information behavior* and *test behavior* had remarkable effects on the success. In comparison to P1, the effects were even higher. Due to the high prognostic value of T3, which is also documented by the high correlation of T3 with the success ($\Phi = .89$, Table 3), the other test behavior was dropped in model 2 and 3. Here, and in contrast to the following analyses, I2 turned out to have a negative effect, which proved to be irrelevant: If I2 had been excluded, the R^2 would have changed from .360 to .359.

Table 4 Effects of the critical information behavior and critical test behavior on success in solving Problem 1

Critical behavior	Model 1: Information behavior			Model 2: Test behavior			Model 3: Problem-solving behavior		
	b	SE b	β	b	SE b	β	b	SE b	β
ES	0.73*	0.32	.16				0.03	0.49	.01
I1	-0.81***	0.20	−.34				−0.69**	0.29	−.21
I2	0.73***	0.18	.33				0.43*	0.21	.14
LD	0.55*	0.17	.24				0.24	0.19	.08
CD	-0.11	0.22	−.03				0.09	0.27	.02
T1				-0.03	0.18	−.01	−0.02	0.20	−.01
T2				2.17***	0.19	.73	2.13***	0.20	.69
Constant	0.43	0.28		1.16***	0.18		0.97*	0.41	
R^2	.18			.52			.56		

N = 336; ES: reading out the Error Storage; I1-I2: selecting Instruction 1–2; LD: selecting the Localization Diagram; CD: selecting the Circuit Diagram; T1-T5: specific Tests

* $p < .05$. ** $p < .01$. *** $p < .001$

Table 5 Effects of the critical information behavior and critical test behavior on the success in solving Problem 2

Critical behavior	Model 1: Information behavior			Model 2: Test behavior			Model 3: Problem-solving behavior		
	b	SE b	β	b	SE b	β	b	SE b	β
I1	−0.45*	0.18	−.16				−0.18	1.56	−.05
I2	-0.22***	0.06	−.09				0.14	1.64	.04
LD	0.17	0.12	.06				0.37	0.33	.11
CD	1.46***	0.04	.57				0.15	0.28	.05
T3				3.46***	0.34	.78	3.37***	.35	.74
Constant	1.43***	0.31		2.03**	0.19		2.17***	0.35	
R^2	.36			.61			.63		

$N = 339$; ES: reading out the Error Storage; I1-I2: selecting Instruction 1–2; LD: selecting the Localization Diagram; CD: selecting the Circuit Diagram; T1-T5: specific Tests

* $p < .05$. ** $p < .01$. *** $p < .001$

Comparing the Effects of the Critical Behavior (RH2)

Considering Table 4 and *Problem 1*, the clear difference in R^2 documented a (much) higher effect of the test behavior on the success than of the information behavior on the success. In model 3, the probability of solving P1 would increase from 16.6% to 87.7% if only T2 changed from 0 to 1 (i.e., from not conducting to conducting T2). If the scoring had been based on T2, 85.4% of the sample (287 testees) would have been scored correctly.

With regard to Table 5, the test behavior also appeared to have a higher effect than the information behavior. Assuming that only CD and T3 changed from 0 to 1, the probability of the problem-solving success would increase from 7.6% to 51.2% (model 1) and 2.1% to 92.4% (model 2). Drawing on T3, an automatic scoring of P2 would have led to a correct scoring of 97% of the sample (329 testees).

Moderation Effect of Experience (RH3)

Table 6 gives the results of the two-group analysis for Problem 1 and the moderator effect of experience. In case of ES, splitting the sample into two groups entailed categories having too few observations, which imposed to exclude ES from the analysis.

Comparing the groups, the effects of LD differed substantially: In the inexperienced group, the effect of LD was high and significant; in the experienced group, it was insignificant. LD refers to selecting the location diagram to find out the location of the fuel temperature sensor. Obviously, this piece of information is easy to memorize. When the location of the sensor had already been figured out (while working on P2), the testees tended to not conduct LD once again. In this vein, LD was selected by 56.4% of the inexperienced testees and only 13.2% of the experienced ones.

Table 6 Moderation effect of experience analyzing Problem 1

Critical information behavior	Model 1: Inexperienced group			Model 2: Experienced group		
	b	SE b	β	b	SE b	β
I1	−0.61*	0.30	−.21	−.62*	0.30	−.28
I2	0.54*	0.24	.22	.80***	0.22	.11
LD	1.32***	0.24	.53	−0.27	0.46	−.08
CD	-0.31	0.27	−.09	0.41	0.24	.12
Constant	0.41*	0.20		−0.41	0.23	
R^2	.35			.11		

Model 1: $N = 176$; model 2: $N = 157$; I1-I2: selecting Instruction 1–2; LD: selecting the Localization Diagram; CD: selecting the Circuit Diagram

* $p < .05$. *** $p < .001$

Equating the effects of both groups, and comparing the chi-square values of the resulting model to the value of the two-group model with freely estimated parameters of Table 6, significant model differences ($\Delta\chi^2(4) = 13.42$, $p = .009$) were found, supporting the moderation effect of experience. The decline in R^2 from .35 (model 1) to .11 (model 2) is another piece of evidence for this effect.

Table 7 gives the results for P2. Please remember that CD refers to selecting the circuit diagram required to test T3 and that experience could make a difference here. The influence of CD, however, was quite stable across the groups. In this line, the difference of the chi-square values of the freely estimated and the "equated" two-group model was not significant ($\Delta\chi^2 (4) = 2.18$, $p = .70$). The fact that the change in R^2 (from .47 to .32) gave some evidence for the moderation effect should, however, not be ignored.

Mediation Effect of the Critical Test Behavior (RH4)

To investigate the mediation effect of the test behavior, only the information behavior that correlated with the problem-solving success was considered (Field 2013, p. 410). In *Problem 1*, this holds for ES, I2 and LD.[3] The left part of Fig. 4 illustrates that T2 explained 82% of the variance of the success in solving P1. In comparison to Table 4, this R^2 is much higher. This is due to different statistical approaches: Whereas T3 is a binary predictor in Table 4, here it is a mediator variable (see the section on statistical analyses). The information behavior appeared to have significant direct effects on T2 and, more importantly here, considerable indirect effects.

The right part of Fig. 4 gives the mediator model for P2 and a very high R^2 caused by T3. CD turned out to have a significant and substantial indirect effect. It should be stressed, however, that this mediator model might be biased because of cells

---

[3] As mentioned before, I3 was not considered due to a perfect correlation with I2.

Table 7 Moderation effect of experience problem analyzing Problem 2

Critical information behavior	Model 1: Inexperienced group			Model 2: Experienced group		
	b	SE b	β	b	SE b	β
I1	−0.26	0.38	−.09	−.64	0.35	−.23
I2	-0.21	0.35	−.08	−.25	0.29	−.10
LD	0.52	0.45	.19	0.06	0.45	.02
CD	1.61***	0.40	.57	1.42***	0.34	.56
Constant	2.02***	0.55		1.15**	0.35	
R^2	.47			.32		

Model 1: $N = 165$; model 2: $N = 174$; I1-I2: selecting Instruction 1–2; LD: selecting the Localization Diagram; CD: selecting the Circuit Diagram

* $p < .05$. ** $p < .01$. *** $p < .001$

comprising near-zero cases in the contingency table of T3 and PS2 (Byrne 2012, p. 131) due to an almost perfect correlation between these measures. In comparison to the previous analyses, this issue resulted from a different statistical approach. Although this very high correlation is in line with the theory on the diagnostic problem-solving process, the findings of this mediator analysis should be interpreted cautiously.
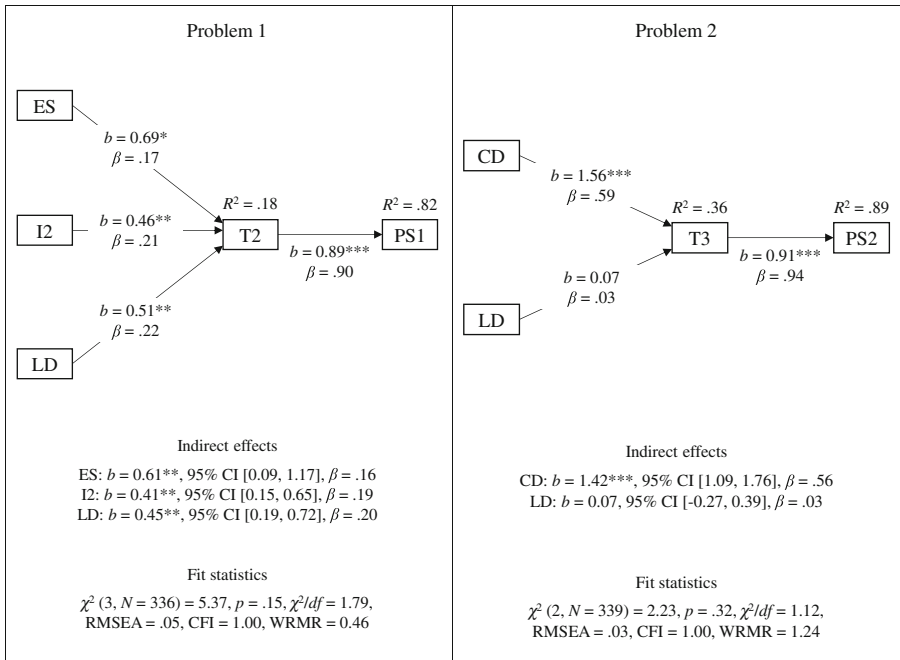

Fig. 4 Mediator effects of the critical test behavior. ES: reading out the Error-Storage; I1-I2: selecting Instruction 1–2; LD: selecting the Localization Diagram; CD: selecting the Circuit Diagram; T2-T3: specific Tests; PS: Problem-solving Success; CI: Confidence Interval. * $p < .05$. ** $p < .01$. *** $p < .001$

Discussion

Summary

The aim of the empirical study was to investigate the presented theory of the diagnostic problem-solving process in professional contexts. The theory distinguishes between four sub-processes of the diagnostic problem-solving process and includes several hypotheses. In this study, four hypotheses were examined. According to the theory, the quality of the sub-processes "representing information" and "testing diagnostic hypotheses" causally influences the problem-solving success (RH1). Secondly, the theory suggests that the influence of the quality of "testing hypotheses" on the problem-solving success is higher than the influence of "representing information" (RH2). Thirdly, the theory advises that the influence of the critical information behavior on the success depends on problem-related experience (RH3). Fourthly, the theory assumes that the influence of the quality of "representing information" on the problem-solving success is mediated by the quality of "testing hypotheses" (RH4). While RH1 and RH4 logically follow from the theory, RH2 and RH3 are in line with, but not logically derivable from, the theory. The quality of the sub-processes was operationalized by critical information behavior and critical test behavior, respectively. This critical behavior follows from critical diagnostic hypotheses and problem-specific as well as substantive considerations. Accordingly, the theory can only be investigated in specific professional contexts. In this study, the theory was examined in the context of car mechatronics with diagnostic problems of the car sector, a computer-based assessment and a sample of car mechatronic apprentices. The critical behavior was extracted from computer-generated log-files.

In accordance with the theory, critical information behavior and critical test behavior substantially affected the problem-solving success (RH1). Some critical behavior, however, could not be included in the statistical analyses due to (almost) perfect correlations (e.g., I2 and I3) or (very) little variance (e.g., PD). These findings did not refute the empirical hypotheses and are irrelevant to the theory. Results indicated that the effects of some critical behavior were strong, whereas other critical behavior had only weak or even no effect.

In line with RH2, results showed that the effect of the critical test behavior was higher than the effect of the critical information behavior. Drawing on the theory, this finding was explained by the fact that the quality of "testing hypotheses" draws upon more critical mental activities than the quality of "representing information".

Unexpectedly, the influence of the critical behavior, related to the correct diagnostic hypothesis, turned out to be stronger than the influence of behavior related to other critical hypotheses: With respect to Problem 1, T2, I2, ES and LD had remarkable effects on the success (Fig. 4). As Table 2 illustrates, this behavior is connected to the correct diagnostic hypothesis of Problem 1. Notably, the test behavior related to the correct diagnostic hypothesis of the problems (T2 and T3, Table 2) proved to have particular strong effects on the success. In contrast to other findings (Klahr and Dunbar 2000, p. 77), this suggests that many testees did not have difficulties with interpreting evidence coming from the tests.

The moderation effect of experience (RH3) was supported by the results on Problem 1; respective results on Problem 2 demand further discussion (see below). From a theoretical viewpoint, experience (i.e., problem-related knowledge) can be retrieved from memory and could lead to a high quality of "representing information", although no observable critical information behavior is exhibited.

With regard to RH4, findings documented that the effects of the critical information behavior on the problem-solving success were mediated by the critical test behavior. Lastly, the exploratory investigation of other process data showed that a time measure created by means of the theory had more predictive power than the rather crude quality indicators "time on problem" and "number of behavioral actions".

In terms of the moderation effect of experience (RH3) in Problem 2 (P2), the effects of the critical information behavior did not prove to differ significantly depending on experience. A remarkable effect difference was anticipated, especially with respect to selecting the circuit diagram (CD). This unexpected finding might be explained by the following reasons: Firstly, several pieces of information of the circuit diagram had to be known to solve P2, increasing the probability of selecting the circuit diagram even if it had been looked up before (while working on Problem 1). Secondly, relatively few testees had selected the circuit diagram while working on Problem 1 (13%, Table 3) meaning that they had not had the opportunity to memorize relevant information beforehand. Both reasons indicate that many testees did not have relevant experience and could not retrieve relevant knowledge. Thus, with Problem 2, the moderation effect of experience was probably not found because many testees did not have relevant experience.

Surprisingly, in some regression models, a specific information behavior, I1, had a negative effect on the problem-solving success, although no correlation was found between I1 and the success. More detailed analyses showed that I1 and other predictors were clearly associated, and entering I1 into the regression models increased the amount of explained variance: I1 turned out to be a classical suppressor variable (Paulhus et al. 2004, p. 306). So there were unsuccessful apprentices that had conducted I1 and other information behavior; controlling for these apprentices increased the explained variance. In this light, the negative effect of I1 is irrelevant to the theory.

Against this background, it seems defensible to conclude that the empirical results did not contradict the empirical hypotheses. It should be stressed, however, that the results do not confirm or even verify the theory of the diagnostic problem-solving process in professional contexts. They solely document that the empirical investigation failed to disprove the theory in the context of car mechatronics and indicate a (very) first empirical corroboration of the theory. Results on RH1 did not refute the theory but suggest a specification: Based on these findings, it could be speculated that the critical behavior, related to the correct diagnostic hypothesis, is especially suitable to operationalize the quality of the sub-processes.

Implications and Limitations of the Study

The theory takes a domain-specific and domain-general perspective: On the one hand, the quality of the sub-processes has to be operationalized by applying domain-specific expertise to identify critical behavior. On the other hand, it is supposed that the distinction of four sub-processes, their causal relationship and their influence on the diagnostic problem-solving process generalize to different professional contexts.

It is very important to take into account that the theory has only been empirically investigated so far in the context of car mechatronics, and with only two diagnostic problems. Moreover, the requirements of diagnostic problems and professional contexts could be (very) different. For example, whereas diagnostic problems of the

technical context usually refer to technical systems, diagnostic problems of the medical context relate to individuals. When dealing with individuals, the side effects of problem-solving behavior (e.g., painful tests) are especially important. These side effects might influence the problem-solving process. Moreover, it should be stressed that the theory concentrates on specific aspects and does not cover other process aspects (e.g., the two-system view, Schwartz and Elstein 2008, p. 229).

Methodically, it should be borne in mind that the statistical analyses might be biased by categories comprising too few observations. It is questionable whether this problem can be solved by increasing the sample size, since near-to-zero categories could be in line with the theory. For example, there might be diagnostic situations in which the "correct" critical test behavior and the success correlate (very) closely, implying near-to-zero categories in contingency tables. Another limitation of the study is that treatment options (e.g., repair) were excluded by definition. Including treatment options might seriously affect the diagnostic problem-solving process (Holmboe and Durning 2014, p. 114).

Critical diagnostic hypotheses are the key component of the presented theory. Donner-Banzhoff et al. (2016) still argue that experienced physicians often do not formulate diagnostic hypotheses; instead, they extensively explore the patient's situation and sometimes even unconsciously collect information. In this context, it should be underlined that the process theory does not imply statements on strategies used by a diagnostician and her/his level of awareness. Diagnostic problems might be solved based on routines, deliberate approaches, or a mixture of both. Diagnostic hypotheses might be formulated early or late in the process, consciously or unconsciously. Such considerations are beyond the scope of the theory. A crucial point of the theory, however, is that successful diagnostic problem solving relies on diagnostic hypotheses. In the end, information collection is completely useless as long as there is not at least one piece of information (i.e., evidence) that corroborates a specific hypothesis and refutes others. Even if a hypothesis is not actively and consciously formulated, it is pivotal: From a theoretical viewpoint, a well-founded solution of a diagnostic problem is nothing but an evidence-based diagnostic hypothesis. This point was also made by Barrows and Feltovich (1987, p. 89).

Overall, the theory could be useful to empirically investigate the problem-solving process in different professional contexts, but it needs further empirical investigations in the context of car mechatronics *and* other professional contexts. It is an open question whether the presented approach to operationalize the sub-processes' quality, that is, to identify critical behavior, would apply to other professional contexts. In future studies, the theory could also be examined using verbal reports on thought processes. In this case, the idea that verbalization "may lead to rationalizations that do not accurately explain actual cognitive processes" (Durning et al. 2015, p. 128) should be considered.

From a didactical perspective, the theory gives helpful orientation. It advises ensuring that individuals learn how to gather problem-relevant information and have the knowledge as well as the skills to formulate, test and evaluate critical hypotheses. Following van Merrienboer (2013), it seems wise to teach problem-solving skills providing "whole" rather than "pieces" of diagnostic problems, starting with simple problems and gradually increasing the difficulty. When teaching diagnostic problem solving, the theory helps to clarify the content of the lessons, to identify the areas where students need support and to figure out relevant learning difficulties. On top of that, it emphasizes the relevance of experience and diagnostic knowledge.

Finally, I would like to stress the role of research on the professional problem-solving process for vocational education and training. Undoubtedly, teaching professional problem solving is essential to prepare students for their professional lives (van Merrienboer 2013). Learning to solve diagnostic problems means acquiring the skills necessary to successfully organize a *process*. In some fields, we can reliably and validly assess whether someone can solve a professional problem (e.g., Rausch et al. 2016; Walker et al. 2016) but we still know very little about the process which underlies the problem's solution or failure. In short, we do not know a lot about the process which we want to foster. Given an elaborated process theory, log-file data provide an immense potential for examining this process, but have been rarely used to date. This is particularly unsatisfactory, since such data can help to understand what students have learned and still have to learn (Greiff et al. 2015, p. 103), and it can be smoothly combined with other process data. For example, interesting insights might be gained when considering whether a relevant information sheet was selected *and* whether relevant areas of the sheet were focused on for a certain time. From this perspective, combining log-file and eye-tracking data makes a "promising match" that can further enhance our understanding of the professional problem-solving process.

Appendix 1

Table 8 Contingency tables: Numbers of testees that solved the problems and showed the critical behavior

Critical behavior		Problem 1			Problem 2		
		0: Unsuccessful	1: Successful	Total	0: Unsuccessful	1: Successful	Total
Information behavior							
PD	0: Not shown	2	0	2	1	0	1
	1: Shown	144	190	334	284	54	338
	Total	146	190	336	285	54	339
ES	0: Not shown	14	7	21	13	0	13
	1: Shown	132	183	315	272	54	326
I1	0: Not shown	42	60	102	75	20	95
	1: Shown	104	130	234	210	34	244
I2	0: Not shown	101	94	195	152	33	185
	1: Shown	45	96	141	133	21	154
I3	0: Not shown	111	118	229	180	34	214
	1: Shown	35	72	107	105	20	125
LD	0: Not shown	113	108	221	198	25	223
	1: Shown	33	82	115	87	29	116

Table 8 (continued)

Critical behavior		Problem 1			Problem 2		
		0: Unsuccessful	1: Successful	Total	0: Unsuccessful	1: Successful	Total
CD	0: Not shown	127	164	291	209	9	218
	1: Shown	19	26	45	76	45	121
Test behavior							
T1	0: Not shown	97	73	170	117	13	130
	1: Shown	49	117	166	168	41	209
T2	0: Not shown	112	15	127	150	16	166
	1: Shown	34	175	209	135	38	173
T3	0: Not shown	145	181	326	281	6	287
	1: Shown	1	9	10	4	48	52
T4	0: Not shown	146	181	327	283	14	297
	1: Shown	0	9	9	2	40	42
T5	0: Not shown	144	187	331	265	37	302
	1: Shown	2	3	5	20	17	37

The total of the unsuccessful/successful testees is identical for each contingency table and given in the first table; PD = selecting the Problem Description; ES = reading out the Error Storage; I1-I3: selecting Instruction 1–3; LD: selecting the Localization Diagram; CD: selecting the Circuit Diagram; T1-T5: specific Tests

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]. *Zeitschrift für Pädagogische Psychologie, 28*(4), 167–179.

Baethge, M., & Arends, L. (2009). Feasibility study VET-LSA: A comparative analysis of occupational profiles and VET programmes in 8 European countries - international report. In *Vocational training research volume 8*. Bielefeld: Bertelsmann.

Barrows, H. S., & Feltovich, P. J. (1987). The clinical reasoning process. *Medical Education, 21*, 86–91.

Boshuizen, H. P., & Schmidt, H. G. (2008). The development of clinical reasoning expertise. In J. Higgs, M. A. Jones, S. Loftus, & N. Christensen (Eds.), *Clinical reasoning in the health professions* (3rd ed., pp. 113–121). Oxford: Elsevier Ltd..

Byrne, M. B. (2012). *Structural equation modeling with Mplus*. New York: Routledge.

Coderre, S., Mandin, H., Harasym, P. H., & Fick, G. H. (2003). Diagnostic reasoning strategies and diagnostic success. *Medical Education, 37*, 695–703.

Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic Medicine, 84*(8), 1022–1028.

Donner-Banzhoff, N., Seidel, J., Sikeler, A. M., Bösner, S., Vogelmeier, M., Westram, A., et al. (2016). The phenomenology of the diagnostic process: A primary care-based survey. *Medical Decision Making*. doi:10.1177/0272989X16653401.

Durning, S. J., Artino, A., R., Schuwirth, L., & van der Vleuten, C. (2013). Clarifying assumptions to enhance our understanding and assessment of clinical reasoning. Academic Medicine, 88(4), 442–448.

Durning, S. J., Rencic, J. & Schuwirth, L. (2015). Assessing clinical reasoning. In L. N. Pangaro & W. C. McGaghie (Eds.), *Handbook on medical student evaluation and assessment* (pp. 127–146). North Syracuse, NY: Gegensatz 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.

Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 41–67). Cambridge: Cambridge University Press.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: Sage Publications.

Frey, A., Hartig, J., & Rupp, A. A. (2009). An NCME instructional module on booklet designs in large-scale assessments of student achievement: Theory and practice. *Educational Measurement: Issues and Practice, 28*(3), 39–53.

Greiff, S., Wüstenberg, S., & Avvisati, F. (2015). Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving. *Computers & Education, 91*, 92–105.

Greiff, S., Niepel, C., Scherer, R., & Martin, R. (2016). Understanding student' sperformance in a computer-based assessment of complex problem-solving: An analysis of behavioral data from computer generated log files, *61*, 36–46.

Groves, M., O'Rourke, P., & Alexander, H. (2003). The clinical reasoning characteristics of diagnostic experts. *Medical Teacher, 25*(3), 308–313.

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*, 557–578.

Hoc, J.-M., & Amalberti, R. (1995). Diagnosis: Some theoretical questions raised by applied research. *Current Psychology of Cognition, 14*(1), 73–101.

Holmboe, E. S., & Durning, S. J. (2014). Assessing clinical reasoning: Moving from in vitro to in vivo. *Diagnosis, 1*(1), 11–117.

Johnson, S. D., Flesher, J. W., & Chung, S.-P. (1995). *Understanding troubleshooting styles to improve training methods* (Annual meeting of the American Vocational Association). Denver, Col.

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development, 48*(4), 63–85.

Jonassen, D. H., & Hung, W. (2006). Learning to troubleshoot: A new theory-based design architecture. *Educational Psychology Review, 18*(1), 77–114.

Joseph, G.-M., & Patel, V. L. (1990). Domain knowledge and hypothesis generation in diagnostic reasoning. *Medical Decision Making, 10*(1), 31–44.

Kassirer, J., Wong, J., & Kopelman, R. (2010). *Learning clinical reasoning* (3rd ed.). Baltimore: Lippincott Williams & Wilkins.

Klahr, D., & Dunbar, K. (2000). Coordinating dual search: The role of evidence. In D. Klahr (Ed.), *Exploring science: The cognition and development of discovery processes* (pp. 61–82). MIT Press: Cambridge.

Krems, J. F. (1994). *Wissensbasierte Urteilsbildung: Diagnostisches Problemlösen durch Experten und Expertensysteme [knowledge-based decision making: Diagnostic problem-solving of experts and expert systems]*. Bern: Hans Huber.

Krems, J., & Bachmaier, M. (1991). Hypothesenbildung und Strategieauswahl in Abhängigkeit vom Expertisegrad [hypothesis generation and strategy-selection dependent on the level of expertise]. *Zeitschrift für Experimentelle und Angewandte Psychologie, 38*(3), 394–410.

Krems, J., & Prechtl, C. (1991). Urteilsbildung und Berufserfahrung: Eine experimentelle Untersuchung zur Generierung, evaluation und Modifikation diagnostischer Hypothesen. [decision making and professional experience: An experimental study on the generation, evaluation and modification of diagnostic hypotheses]. *Zeitschrift für Experimentelle und Angewandte Psychologie, 38*(2), 248–263.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis* (2nd ed.). New York: Taylor & Francis.

Mehle, T. (1982). Hypothesis generation in an automobile malfunction inference task. *Acta Psychologica, 52*(1–2), 87–106. doi:10.1016/0001-6918(82)90028-2.

van Merrienboer, J. (2013). Perspectives on problem solving and instruction. *Computers & Education, 64*, 153–160.

Moosbrugger, H., & Schermelleh-Engel, K. (2008). Exploratorische (EFA) und konfirmatorische Faktorenanalyse (CFA) [exploratory and confirmatory factor analysis]. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 308–324). Berlin: Springer.

Morris, N. M., & Rouse, W. B. (1985). Review and evaluation of empirical research in troubleshooting. *Human Factors, 27*, 503–530.

Muthén, B. O. (2011). *Applications of causally defined direct and indirect effects in mediation analysis using SEM in Mplus* (Technical report). Los Angeles.

Nendaz, M. R., Gut, A. M., Perrier, A., Louis-Simonet, M., Reuille, O., Junod, A. F., et al. (2005). Common strategies in clinical data collection displayed by experienced clinician-teachers in internal medicine. *Medical Teacher, 27*(5), 415–421.

Norman, G. R. (2005). Research in clinical reasoning: Past history and current trends. *Medical Education, 39*, 418–427.

Patel, V. L., Kaufman, D. R., & Magder, S. A. (1996). The acquisition of medical expertise in complex dynamic environments. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 127–165). Mahwah: Lawrence Erlbaum Associates.

Paulhus, D. L., Robins, R. W., Trzesniewski, K. H., & Tracy, J. L. (2004). Two replicable suppressor situations in personality research. *Multivariate Behavioral Research, 39*(2), 301–326.

Popper, R. (2005). *Logik der Forschung [the logic of scientific discovery]*. Tübingen: Mohr Siebeck.

Rasmussen, J. (1993). Diagnostic reasoning in action. *Systems, Man and Cybernetics, IEEE Transactions on, 23*(4), 981–992. doi:10.1109/21.247883.

Rausch, A., Seifried, J., Wuttke, E., Kögler, K., & Brandt, S. (2016). Reliability and validity of a computer-based assessment of cognitive and non-cognitive facets of problem-solving competence in the business domain. *Empirical Research in Vocational Education and Training, 8*(1), 1–23. doi:10.1186/s40461-016-0035-y.

Rouse, W. B. (1983). Models of human problem solving: Detection, diagnosis, and compensation for system failures. *Automatica, 19*(6), 613–625. doi:10.1016/0005-1098(83)90025-0.

Schaafstal, A., Schraagen, J. M., & van Berlo, M. (2000). Cognitive task analysis and innovation of training: the case of structured troubleshooting. *Human Factors,* 42, 75–86.

Schmidt, H. G., Norman, G., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine, 65*(10), 611–621.

Schwartz, A., & Elstein, A. S. (2008). Clinical reasoning in medicine. In J. Higgs, M. A. Jones, S. Loftus, & N. Christensen (Eds.), *Clinical reasoning in the health professions* (3rd ed., pp. 223–234). Oxford: Elsevier Ltd..

Walker, F., Link, N., & Nickolaus, R. (2016). A multidimensional structure of domain-specific problem-solving competencies of electronics technicians for automation technology. *Empirical Research in Vocational Education and Training, 8*(1), 1–16. doi:10.1186/s40461-016-0034-z.

Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. Chichester: Wiley.

Dr. Stephan Abele is a post-doc researcher and research project leader in the Department of Vocational and Technical Education at Stuttgart University. He received his diploma (FH) in mechanical engineering from Reutlingen University (Germany) and his diploma as well as his Ph.D. in Technical Education and Training from Stuttgart University (Germany). Currently, he is working on his postdoctoral qualification ("Habilitation") dealing with diagnostic problem-solving in professional contexts. His main research interests are: teaching/learning in industrial-technical professions, measurement, modeling and development of professional competencies, computer-based assessment of domain-specific problem-solving competences and vocational teacher education.