



# Problemlösekompetenzen in beruflichen Kontexten

Resultate aus Lehr-Lern-Prozessen sichtbar machen



# **Problemlösekompetenzen in beruflichen Kontexten**

Resultate aus Lehr-Lern-Prozessen sichtbar machen

Stephan Abele

## „Wirtschaft – Beruf – Ethik“

Herausgegeben von:

**Prof.in Dr.in Birgit Ziegler**, Arbeitsbereich Berufspädagogik und Bildungsforschung  
an der Technischen Universität Darmstadt

**Prof. Dr. Gerhard Minnameier**, Lehrstuhl für Wirtschaftsethik und Wirtschaftspädagogik  
an der Goethe-Universität Frankfurt am Main

Die Reihe „Wirtschaft – Beruf – Ethik“ widmet sich Fragen der ökonomischen Bildung, der beruflichen Aus- und Weiterbildung sowie der Berufs-, Unternehmens- und Wirtschaftsethik im Kontext lokaler und globaler Entwicklungen. Sie umfasst theoretische, empirische, systematische und historische Arbeiten, die disziplinär in der Berufs- und Wirtschaftspädagogik sowie der Wirtschaftsethik verankert sind.

Ulrich Pleiß gründete 1982 die Reihe „Wirtschaftsdidaktik, Berufsbildung und Konsumentenerziehung“, sie wurde 2015 umbenannt in „Wirtschaft – Beruf – Ethik“.

Die Reihe wird gefördert durch die Käthe und Ulrich Pleiß-Stiftung.

### Publikationen in der Reihe „Wirtschaft – Beruf – Ethik“:

Band 32

Fritz Oser; Margarete Landenberger; Klaus Beck (Hg.)

#### **Technologiebasierte Kompetenzmessung in der beruflichen Bildung**

Ergebnisse aus der BMBF-Förderinitiative ASCOT  
Bielefeld 2016, ISBN: 978-3-7639-5459-9

Band 33

Gerhard Minnameier (Hg.)

#### **Ethik und Beruf**

Interdisziplinäre Zugänge  
Bielefeld 2016, ISBN: 978-3-7639-5461-2

Band 34

Maxi Deppe

#### **Fehler als Stationen im Lernprozess**

Eine kognitionswissenschaftliche Untersuchung  
am Beispiel Rechnungswesen  
Bielefeld: 2017, ISBN: 978-3-7639-5463-6

Band 35

Jürgen Seifried; Klaus Beck; Bernd-Joachim Ertelt;  
Andreas Frey (Hg.)

#### **Beruf, Beruflichkeit, Employability**

Bielefeld 2019, ISBN: 978-3-7639-5465-0

Band 36

Karin Heinrichs; Hannes Reinke (Hg.)

#### **Heterogenität in der beruflichen Bildung**

Im Spannungsfeld von Erziehung, Förderung und  
Fachausbildung  
Bielefeld 2019, ISBN: 978-3-7639-6003-3

Band 37

Juliana Schlicht

#### **Kommunikation und Kooperation in Geschäftsprozessen**

Modellierung aus pädagogischer, ökonomischer und  
informationstechnischer Perspektive  
Bielefeld: 2019, ISBN: 978-3-7639-6005-7

Band 38

Georg Hans Neuweg; Rico Hermkes; Tim Bonowski (Hg.)

#### **Implizites Wissen**

Berufs- und wirtschaftspädagogische Annäherungen  
Bielefeld 2020, ISBN: 978-3-7639-6007-1

Band 39

Christian Michaelis, Florian Berding

#### **Berufsbildung für nachhaltige Entwicklung**

Umsetzungsbarrieren und interdisziplinäre Forschungsfragen  
Bielefeld 2022, ISBN: 978-7639-7009-4

Band 40

Elisabeth Riebenbauer

#### **Kompetenzentwicklung im Masterstudium Wirtschaftspädagogik**

Längsschnittstudie zur Unterrichtsplanung im Rechnungswesen  
Bielefeld 2022, ISBN: 978-7639-7016-2

Band 41

Stephan Schumann, Susan Seeber, Stephan Abele (Hg.)

#### **Digitale Transformation in der Berufsbildung**

Konzepte, Befunde und Herausforderungen  
Bielefeld 2022, ISBN: 978-7639-7137-4



Weitere Informationen finden  
Sie auf [wbv.de/wbe](http://wbv.de/wbe)

Stephan Abele

# Problemlösekompetenzen in beruflichen Kontexten

Resultate aus Lehr-Lern-Prozessen sichtbar machen



Die Produktionskosten für diese Publikation wurden vom Open Access Publikationsfonds der Sächsischen Landesbibliothek, Staats- und Universitätsbibliothek (SLUB Dresden) gefördert.

2023 wbv Publikation  
ein Geschäftsbereich der  
wbv Media GmbH & Co. KG, Bielefeld

Gesamtherstellung:  
wbv Media GmbH & Co. KG, Bielefeld  
**wbv.de**

Umschlagillustration:  
Shutterstock.com/Kev Draws

ISBN (Print): 978-3-7639-73668  
ISBN (E-Book): 978-3-7639-73675  
DOI 10.3278/9783763973675

Printed in Germany

Diese Publikation ist frei verfügbar zum Download  
unter [wbv-open-access.de](http://wbv-open-access.de)

Diese Publikation ist unter folgender  
Creative-Commons-Lizenz veröffentlicht:  
[creativecommons.org/licenses/by-sa/4.0/deed.de](http://creativecommons.org/licenses/by-sa/4.0/deed.de)



Für alle in diesem Werk verwendeten Warennamen  
sowie Firmen- und Markenbezeichnungen können  
Schutzrechte bestehen, auch wenn diese nicht als solche  
gekennzeichnet sind. Deren Verwendung in diesem  
Werk berechtigt nicht zu der Annahme, dass diese frei  
verfügbar seien.

---

### **Bibliografische Information der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;  
detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

---

# Inhalt

Vorwort der Herausgeber .....	7
-------------------------------	---

*Stephan Abele*

<b>Berufliche Problemlösekompetenz: Modellierung und Befunde .....</b>	<b>9</b>
1    Forschungsthema und Gang der Untersuchung .....	9
1.1  Erfassung diagnostischer Problemlösekompetenzen als Forschungsproblem .....	9
1.2  Systematische Einordnung des Forschungsproblems .....	11
1.3  Bedeutung des Forschungsproblems .....	12
1.4  Aufbau und Funktion des vorliegenden Texts .....	14
2    Erfassung diagnostischer Problemlösekompetenzen .....	14
2.1  Modellierungsprozess .....	15
2.2  Konsequenzen .....	17
3    Konkretisierung des Sprachmodells diagnostischer Problemlösekompetenz .....	18
3.1  Wie wird das Sprachmodell konkretisiert? .....	18
3.2  Festlegung konkretisierungsbedürftiger Sprachmodellaspekte .....	19
3.3  Konkretisierung der Sprachmodellaspekte .....	22
3.4  Zusammenfassung und Forschungsdesiderate .....	31
4    Konkretisierung des Erhebungsmodells diagnostischer Problemlösekompetenzen .....	32
4.1  Konkretisierungsziel und Konkretisierungskriterien .....	33
4.2  Konkretisierung des Erhebungsmodells unter den Validitätskriterien ....	35
4.3  Zusammenfassung und Forschungsdesiderate .....	44
5    Forschungsprogramm und Forschungsfragen .....	46
5.1  Auf das Sprachmodell bezogene Forschungsfragen .....	46
5.2  Auf das Erhebungsmodell bezogene Forschungsfragen .....	46
6    Ergebnisse der Einzelstudien .....	47
6.1  Studien zum Sprachmodell .....	47
6.2  Studien zum Erhebungsmodell .....	58
7    Diskussion .....	63
7.1  Zusammenfassung der Erkenntnisse .....	64
7.2  Implikationen der Ergebnisse für die Erfassung diagnostischer Problemlösekompetenzen .....	69
7.3  Grenzen und kritische Reflexion der Untersuchung .....	73
7.4  Forschungsperspektiven .....	77

<b>Anhang: sechs Publikationen</b> .....	87
Anhang A .....	89
<i>Reinhold Nickolaus, Stefan Behrendt, Stephan Abele</i>	
Kompetenzstrukturen bei KFZ-Mechatronikern und die Erklärungskraft des fachsystematischen KFZ-Wissens für berufsfachliche Kompetenzen .....	90
Anhang B .....	109
<i>Stephan Abele</i>	
Umgang mit Komplexität: Eine bedeutsame psychische Voraussetzung des domänenspezifischen Problemlösens? .....	110
Anhang C .....	133
<i>Stephan Abele, Raphael Ostertag, Matthias Peissner, Andreas Schuller</i>	
Eine Eye-Tracking-Studie zum diagnostischen Problemlöseprozess: Bedeutung der Informationsrepräsentation für den diagnostischen Problemlöseerfolg .....	134
Anhang D .....	159
<i>Stephan Abele</i>	
Diagnostic Problem-Solving Process in Professional Contexts: Theory and Empirical Investigation in the Context of Car Mechatronics Using Computer-Generated Log-Files .....	160
Anhang E .....	191
<i>Stephan Abele, Felix Walker und Reinhold Nickolaus</i>	
Zeitökonomische und reliable Diagnostik beruflicher Problemlösekompetenzen bei Auszubildenden zum Kfz-Mechatroniker .....	192
Anhang F .....	215
<i>Stephan Abele</i>	
Can diagnostic problem-solving competences of car mechatronics be validly assessed using a paper-pencil test? .....	216
<b>Autor</b> .....	239

# Vorwort der Herausgeber

Ein Kernziel beruflicher Bildungsmaßnahmen ist der Erwerb von Handlungskompetenzen zur Bewältigung beruflicher Anforderungen, die sich je spezifisch in einem Berufsfeld stellen. Nun konnte man in beruflichen Handlungsfeldern noch nie mit dem bloßen Einsatz von Routinen reüssieren; auf eine hoch dynamische und von digitalen Technologien durchdrungene Arbeitswelt trifft dies selbstredend noch viel weniger zu. Schon in vermeintlich einfachen, durchstrukturierten Arbeitsprozessen kann es zu Situationen kommen, in denen die Bearbeitung beruflicher Aufgaben nicht zu den erwarteten Handlungsergebnissen bzw. zum Erfolg führt. Sofern die handelnde Person keine Erklärung für diesen unerwünschten Zustand hat und sich damit nicht zufrieden geben will oder kann, muss sie nach den Ursachen suchen, um ggf. auch aus den eigenen „Fehlern“ zu lernen. Abhängig vom Komplexitätsgrad des beruflichen Tätigkeitsfeldes dürften solche Situationen, in denen etwas nicht wie erwartet bearbeitet werden kann, mehr oder weniger häufig vorkommen.

Nun gibt es allerdings berufliche Tätigkeitsfelder, zu deren genuinem Aufgabenspektrum es gehört, Ursachen für Störungen zu diagnostizieren, wie z. B. in sozialen und medizinischen Tätigkeitsfeldern oder bei Tätigkeiten in der Prozessüberwachung oder der Reparatur von Artefakten aller Art, wie etwa von Maschinen, Anlagen oder Fahrzeugen. In diesen Tätigkeitsfeldern sind Beschäftigte besonders häufig gefordert, diagnostische Probleme zu lösen, und sollten gezielt ausgebildet werden, um Ursachen für einen unerwünschten Zustand möglichst systematisch aufspüren zu können. Die für eine Ursachendiagnose erforderliche Kompetenz beschreibt Stephan Abele als „diagnostische Problemlösekompetenz“, die er in einem mehrjährigen Forschungszyklus am Beispiel des Berufsfeldes KFZ-Mechatronik beforcht. Seine theoretischen und empirischen Arbeiten liefern einen berufs- und wirtschaftspädagogischen Beitrag, der neben seiner berufspädagogisch-diagnostischen Relevanz auch in didaktischer Hinsicht bedeutsam ist, und weit über das Berufsfeld der KFZ-Mechatronik hinaus wichtige Impulse zu geben vermag.

Im Wesentlichen konzentrieren sich die Arbeiten von Stephan Abele auf die Entwicklung eines Sprachmodells und eines Erhebungsmodells für „diagnostische Problemlösekompetenz“. Die sprachlich exakte Modellierung des Konstrukts ist die unverzichtbare Voraussetzung für alle nachfolgenden Schritte der empirischen Erfassung diagnostischer Problemlösekompetenz. Forschungslücken, die sich bezüglich des Sprachmodells zeigen, bearbeitet Abele in vier empirischen Studien. Studie A bezieht sich auf die Abgrenzbarkeit diagnostischer Problemlösekompetenz von anderen beruflichen Handlungskompetenzen. Hinsichtlich der Frage nach den relevanten psychischen Komponenten, die als Basis diagnostischer Problemlösekompetenz nachweisbar sind, zielt Studie B auf die Frage, ob der Umgang mit Komplexität als bedeutsame psychische Komponente in die Modellierung einzubeziehen wäre. Mit den Studien C und D wird versucht, über zwei unterschiedliche methodische Zugänge (Arbeitsproto-

kolle und Eyetracking) eine Prozesstheorie diagnostischen Problemlösens zu entwickeln. Ein weiterer Arbeitsschwerpunkt von Abele liegt auf dem Erhebungsmodell. Hier befasst er sich in zwei weiteren Studien (E und F) mit der Frage, wie Erhebungssituationen bzw. dabei eingesetzte Items beschaffen sein müssen und in welchem Modus sie darzubieten wären, um sowohl einen spezifischen Anforderungsraum hinreichend reliabel abzubilden und darauf bezogen das Zielkonstrukt valide zu erfassen als auch unter Praktikabilitäts Gesichtspunkten zu genügen. Ein klassisches Optimierungsproblem, das unter Abwägung von Zielperspektiven ausbalanciert, aber nicht eindeutig gelöst werden kann.

Wir freuen uns sehr, über die gesammelte Publikation der Beiträge zur kumulativen Habilitation von Stephan Abele im vorliegenden Band 42 der „Pleiß-Reihe“ nun die sowohl inhaltlich als auch wissenschaftsmethodisch anregungsreiche Lektüre einem breiten Leserkreis zugänglich machen zu können.

Frankfurt und Darmstadt im August 2023

Gerhard Minnameier und Birgit Ziegler

# Berufliche Problemlösekompetenz: Modellierung und Befunde

STEPHAN ABELE

## 1 Forschungsthema und Gang der Untersuchung

### 1.1 Erfassung diagnostischer Problemlösekompetenzen als Forschungsproblem

In der vorliegenden kumulativen Habilitation geht es um die *Erfassung diagnostischer Problemlösekompetenzen in beruflichen Kontexten*. Unter einem diagnostischen Problem wird eine Situation verstanden, in der sich eine Person das Ziel gesetzt hat, die Ursache(n) für einen unerwünschten Zustand zu finden (in Anlehnung an Krems, 1994, S. 34). Der Fokus der vorliegenden Untersuchung liegt auf diagnostischen Problemen in beruflichen Kontexten. Diagnostische Probleme werden in sehr unterschiedlichen beruflichen Kontexten bearbeitet: Im medizinischen Kontext sind die Ursache(n) von Krankheiten, im Lehr-Lern-Kontext die Gründe für Lernprobleme, im Gastronomiebereich die Ursachen für unzufriedene Kund:innen, bei Bankberatungsgesprächen die Gründe für eine misslingende Kommunikation, in technischen Kontexten die Ursache(n) von Maschinenstörungen etc. zu ermitteln. Um ein diagnostisches Problem zu lösen, muss ein Individuum unterschiedliche psychische Ressourcen koordinieren und zielführend in den Problemlöseprozess einbringen. Mit dem Problemlöseprozess sind Aktivitäten von Personen angesprochen, die der Problemlösung dienen und von der Zielsetzung bis zum Auffinden der Problemursache(n) bzw. einem vorzeitigen Problemlöseabbruch reichen. Mit *diagnostische Problemlösekompetenz* werden diejenigen psychischen Grundlagen bezeichnet, die für eine erfolgreiche Gestaltung dieses Prozesses, also für die Problemlösung, benötigt werden. Der Begriff *diagnostische Problemlösekompetenz* wird hier auch im Plural verwendet. Damit wird unterstellt, dass es mehrere diagnostische Problemlösekompetenzen gibt. Das Gemeinsame dieser Problemlösekompetenzen besteht darin, dass damit die psychischen Grundlagen angesprochen werden, die zur Lösung diagnostischer Probleme erforderlich sind. Der Unterschied dieser Problemlösekompetenzen hingegen liegt darin, dass sie sich auf unterscheidbare berufliche Kontexte beziehen – z. B. den medizinischen oder den technischen Kontext.

Die Erfassung diagnostischer Problemlösekompetenzen in beruflichen Kontexten verkörpert ein wichtiges berufs- und wirtschaftspädagogisches Thema (s. ausführlicher dazu Kapitel 1.3): Diagnostische Problemlösekompetenzen repräsentieren zentrale Lehr-Lernziele der beruflichen Bildung. Ob und wie gut diese Ziele erreicht werden, lässt sich nur beurteilen, wenn diagnostische Problemlösekompetenzen (z. B. im Rahmen von Abschlussprüfungen) empirisch erfasst werden können. Darüber hinaus

müssen diagnostische Problemlösekompetenzen erfasst werden, um die Lerneffekte einschlägiger Lernumgebungen zu evaluieren. Aus einem allgemeineren Blickwinkel betrachtet ist die Erfassung diagnostischer Problemlösekompetenzen die Grundlage dafür, Lehr-Lern-Theorien zu diagnostischen Problemlösekompetenzen entwickeln und vor allem prüfen zu können (z. B. Beck, 2009).

Wie der gegenwärtige Forschungsstand verdeutlicht, gibt es im Bereich der Erfassung diagnostischer Problemlösekompetenzen noch mehrere Forschungsdesiderate (Kapitel 3.4 und Kapitel 4.3). Eine noch ungelöste Herausforderung ist z. B. die sprachliche Modellierung diagnostischer Problemlösekompetenzen, was unter anderem daran liegt, dass die diagnostische Problemlösekompetenz als psychische Grundlage nicht direkt beobachtbar ist, sondern nur über beobachtbares Problemlöseverhalten, insbesondere den beobachtbaren Problemlöseerfolg, rekonstruiert werden kann: Die diagnostische Problemlösekompetenz ist ein wirkungsvolles, aber unsichtbares (zumindest nicht direkt beobachtbares) Resultat beruflicher Lehr-Lern-Prozesse. Aus definitorischer Sicht ist ferner problematisch, dass mit diagnostischer Problemlösekompetenz ein Bündel unterschiedlicher psychischer Ressourcen sowie ein facettenreicher Problemlöseprozess angesprochen sind (Kapitel 3). Der Bezug zu beruflichen Kontexten wirft darüber hinaus die Frage auf, wie diagnostische Problemlösekompetenzen berufsauthentisch und zugleich wissenschaftlich befriedigend erfasst werden können (Kapitel 4). Die in jüngster Zeit initiierten Forschungsprogramme, wie z. B. ASCOT (Technology-based Assessment of Skills and Competences in Vocational Education and Training; Beck et al., 2016), verdeutlichen ein generelles berufspädagogisches Forschungsdefizit: Berufspädagogische Theorien sowie berufsbezogene Rahmenlehrpläne und Prüfungen werden zwar (zunehmend) kompetenzorientiert formuliert; weitgehend unklar ist jedoch bislang, wie die darin beschriebenen Kompetenzen befriedigend erfasst werden können.

Vor diesem Hintergrund wird deutlich, dass die Erfassung diagnostischer Problemlösekompetenzen einerseits ein wichtiges berufs- und wirtschaftspädagogisches Forschungsthema darstellt; andererseits bestehen aber noch dahin gehend offene Fragen, wie diese Kompetenzen wissenschaftlich befriedigend erfasst werden können. Mit diesem Forschungsproblem beschäftigt sich die vorliegende Untersuchung. Es werden einige Forschungsdesiderate im Bereich der Erfassung diagnostischer Problemlösekompetenzen bearbeitet und damit der einschlägige Forschungsstand erweitert. Dafür werden sechs aufeinander bezogene quantitativ-empirische Studien durchgeführt, die in metatheoretischer Hinsicht der Forschungslogik des Kritischen Rationalismus folgen (Popper, 2005). In diesen Studien steht die diagnostische Problemlösekompetenz von Kfz-Mechatroniker:innen im Zentrum. Bei der theoretischen Grundlegung sowie der Interpretation und Diskussion der Studienergebnisse wird aber auf Forschungsarbeiten unterschiedlicher beruflicher Kontexte zurückgegriffen. Der Schwerpunkt liegt dabei auf Studien aus dem technischen und medizinischen Kontext.

## 1.2 Systematische Einordnung des Forschungsproblems

### 1.2.1 Diagnostische Problemlösekompetenzen und berufliche Handlungskompetenz

Diagnostische Problemlösekompetenzen werden als wichtiger Bestandteil der beruflichen Handlungskompetenz aufgefasst. Mit beruflicher Handlungskompetenz wird die Gesamtheit jener psychischen Grundlagen bezeichnet, die zum erfolgreichen Handeln innerhalb eines Berufs benötigt werden. Berufliches Handeln stellt eine individuelle Aktivität dar, die ausgeführt wird, um spezifische berufliche Ziele zu erreichen (z. B. Funke, 2003, S. 18).<sup>1</sup> In erster Näherung lassen sich zwei berufliche Handlungstypen unterscheiden: Handlungen, die der Bewältigung beruflicher Aufgaben dienen, und Handlungen, die auf das Lösen beruflicher Probleme abzielen. Bei der Aufgabenbewältigung werden Anforderungen gemeistert, die kaum bis keinerlei Unsicherheiten mit sich bringen. Bei einigermaßen geübten Kfz-Mechatroniker:innen kann z. B. ein Ölwechsel an einem bekannten Fahrzeug eine Aufgabenbewältigung darstellen: Es ist klar, wie der Ölwechsel durchzuführen ist und es „drohen“ keine „Überraschungen“. Dagegen repräsentiert das berufliche Problemlösen eine Aktivität, bei der Unsicherheiten eine zentrale Rolle spielen. Beim beruflichen Problemlösen ist eine Unbekannte zu finden (Jonassen, 2000, S. 65). Beim diagnostischen Problemlösen besteht diese Unbekannte in den unbekanntem Ursachen eines unerwünschten Zustands. Hierbei ist es unerheblich, ob die Ursachen anhand von Handlungsrouninen gefunden werden oder neue Handlungsoptionen entwickelt werden müssen. Entscheidend ist, dass die Ursachen der/dem Problemlösenden nicht bekannt sind. Andere berufs- und wirtschaftspädagogische Forschungsarbeiten verwenden einen anderen Problemlösebegriff (z. B. Wuttke et al., 2015; Sembill, 1992, S. 83 f.; s. dazu auch die Ausführungen von Funke, 2003, S. 25), was in der Diskussion ausführlicher thematisiert wird (Kapitel 7.3.3). Vor dem Hintergrund dieser Ausführungen wird das diagnostische Problemlösen als eine spezifische Variante des beruflichen Problemlösens sowie des beruflichen Handelns aufgefasst. Darüber erklärt sich auch, warum diagnostische Problemlösekompetenzen als Teil der beruflichen Handlungskompetenz verstanden werden.

### 1.2.2 Diagnostischer Problemlösekompetenzen und andere berufliche Kompetenzen

Die diagnostische Problemlösekompetenz wird als eine von mehreren Kompetenzen verstanden, die innerhalb eines spezifischen Berufs<sup>2</sup> relevant sind. Berufliche Kompetenzen können erstens dadurch unterschieden werden, auf welche Anforderungsbereiche sie sich beziehen (z. B. Kfz-Service, Kfz-Reparatur etc.); zweitens können sie danach unterschieden werden, ob sie der Aufgabenbewältigung oder dem beruflichen

---

1 Im vorliegenden Kontext reicht diese grobe sprachliche Fassung des Handlungsbegriffs aus. Dass sich der Handlungsbegriff bei einer gründlichen Sprachanalyse als sehr vielschichtig entpuppt, zeigen z. B. die Ausführungen von Beck (1988), der diesen Begriff aus didaktischer Sicht und auf Basis wissenschaftstheoretischer Überlegungen differenziert rekonstruiert.

2 Gemeint ist hier die objektive Bedeutung des Begriffs *Beruf*, mit der jene Kompetenzen angesprochen sind, über die jemand verfügt bzw. verfügen sollte, der einen bestimmten Beruf hat (Beck, 2016, S. 43). Freilich ist damit nur eine von mehreren Bedeutungsvarianten des Berufsbegriffs abgedeckt.

Problemlösen dienen. Aus Sicht dieser Unterscheidungsmerkmale lassen sich innerhalb eines Berufs mehrere berufliche Kompetenzen identifizieren, die diagnostische Problemlösekompetenz ist eine davon. Bei feinerer Auflösung lassen sich diagnostische Problemlösekompetenzen innerhalb eines Berufs weiter ausdifferenzieren – z. B. diagnostische Problemlösekompetenzen im Bereich der Kfz-Elektrotechnik und der Kfz-Mechanik.

### **1.2.3 Diagnostische Problemlösekompetenzen und domänenübergreifende Problemlösekompetenzen**

Diagnostische Problemlösekompetenzen in beruflichen Kontexten werden als domänenspezifische Problemlösekompetenzen aufgefasst und begrifflich von domänenübergreifenden (auch: allgemeinen) Problemlösekompetenzen abgegrenzt. Beide Problemlösekompetenzen lassen sich hinsichtlich der Bedeutung des Fachwissens und ihres Wirkungsbereichs unterscheiden (Greiff et al., 2014, S. 162 f.). Diagnostische Problemlösekompetenzen in beruflichen Kontexten beziehen sich auf eng umrissene Anforderungen, nämlich auf berufliche Anforderungen, bei denen diagnostische Probleme zu lösen sind. Sie ermöglichen es, sehr spezifische Probleme zu lösen, setzen ein hohes Maß spezifischen Fachwissens voraus und sind nur in sehr spezifischen Kontexten nützlich. Demgegenüber spielt spezifisches Fachwissen bei allgemeinen Problemlösekompetenzen nur eine untergeordnete Rolle und es wird unterstellt, dass allgemeine Problemlösekompetenzen bei der Lösung sehr unterschiedlicher Probleme wirksam werden.

## **1.3 Bedeutung des Forschungsproblems**

### **1.3.1 Wissenschaftliche Bedeutung des beruflichen Problemlösens**

Berufliches Problemlösen wird sowohl national (z. B. Seifried et al., 2016, S. 121) als auch international (z. B. Billett, 2017, S. 52) als relevantes Forschungsthema eingeschätzt. In der wirtschaftspädagogischen Forschung in Deutschland wurde das berufliche Problemlösen in den 1990er-Jahren intensiv bearbeitet (z. B. Fürstenau, 1999; Sembill, 1992); in der jüngeren Vergangenheit geriet es im Rahmen der bereits erwähnten ASCOT-Initiative in den Forschungsfokus (z. B. Seifried et al., 2016). In der berufspädagogischen Forschung wurde das berufliche Problemlösen insbesondere von der Forschungsgruppe um Reinhold Nickolaus untersucht (z. B. Nickolaus, 2011). Insgesamt liegt aber eine überschaubare Anzahl berufs- und wirtschaftspädagogischer Forschungsarbeiten zum beruflichen Problemlösen vor. Aufgrund der skizzierten Bedeutung dieses Themas überrascht dies ebenso wie ein Blick über den Tellerrand: Das Programm der 5. Jahrestagung der GEBF (Gesellschaft für Empirische Bildungsforschung) 2017 weist trotz einer sehr großen Anzahl an Einzelbeiträgen lediglich neun Vorträge zum Thema Problemlösen aus, fünf davon zum domänenspezifischen Problemlösen.

### 1.3.2 Kontextübergreifende Bedeutung diagnostischen Problemlösens

Diagnostische Probleme werden in sehr unterschiedlichen beruflichen Kontexten bearbeitet und untersucht. Schwartz und Elstein (2008) sowie Patel et al. (1996) beschäftigen sich mit dem diagnostischen Problemlösen in medizinischen Kontexten. Katz und Anderson (1987) untersuchen das diagnostische Problemlösen bei Programmierer:innen, Walker et al. (2016a) bei Elektroniker:innen für Automatisierungstechnik, Konradt (1992) und Schaper (1994) bei Mechaniker:innen und Krems und Bachmaier (1991) bei Kfz-Mechatroniker:innen. Jonassen (2011, S.77) weist darauf hin, dass diagnostische Probleme auch bei der Kundenberatung, in der Personalentwicklung und in der Öffentlichkeitsarbeit zu lösen sind.

### 1.3.3 Curriculare Bedeutung diagnostischer Problemlösekompetenzen

Ein Blick in die Rahmenlehrpläne für berufliche Schulen zeigt, dass diagnostische Problemlösekompetenzen wichtige Lehr-Lern-Ziele der beruflichen Bildung sind. Exemplarisch seien die Ausbildungsberufe zum/zur Elektroniker:in für Automatisierungstechnik oder Betriebstechnik sowie zum/zur Kfz-Mechatroniker:in genannt (Kultusministerkonferenz, 2003b; Kultusministerkonferenz, 2003a; Kultusministerkonferenz, 2013). Es besteht wenig Zweifel, dass es sich bei der diagnostischen Problemlösekompetenz um einen relevanten Ausschnitt der Wirklichkeit und ein wichtiges berufs- und wirtschaftspädagogisches Thema handelt. Wie eine Studie von Baethge und Arends (2009) zeigt, gilt dies nicht nur für Deutschland, sondern auch für andere europäische Länder.

### 1.3.4 Unterrichtspraktische Bedeutung der Erfassung diagnostischer Problemlösekompetenzen

Die Erfassung der diagnostischen Problemlösekompetenz in beruflichen Kontexten oder, allgemeiner gesprochen, von nicht direkt beobachtbaren (unsichtbaren) psychischen Grundlagen ist eine wichtige Aufgabe der berufs- und wirtschaftspädagogischen Praxis. Im Sinne eines kognitivistischen Lernbegriffs bezieht sich Lernen auf die Veränderung interner Bedingungen, d. h. psychischer Grundlagen, die sich in spezifischem (beobachtbarem) Verhalten manifestieren (Straka & Macke, 2009, S. 5). Mit beruflichen Lehr-Lern-Arrangements wird das Ziel verfolgt, diese psychischen Grundlagen zu fördern.<sup>3</sup> Dies ist nur dann sinnvoll möglich, wenn überprüft werden kann, ob das Ziel erreicht wurde, also die psychischen Grundlagen empirisch erfasst werden können. Eine solche Erfassung erlaubt auch ein evidenzgestütztes Feedback zum individuellen Kompetenzstand und begünstigt damit die individuelle Kompetenzförderung.

---

<sup>3</sup> Dieser Forderung unterliegt eine normative Setzung, die darin besteht, dass berufliches Lehren und Lernen an einem wünschenswerten Zielzustand auszurichten sind, wobei in einem politischen und idealerweise wissenschaftlich informierten Aushandlungsprozess zu klären ist, was „wünschenswert“ bedeutet.

### **1.3.5 Akademische und wissenschaftliche Bedeutung der Erfassung diagnostischer Problemlösekompetenzen**

Auch in einer wissenschaftlichen Perspektive ist die Erfassung diagnostischer Problemlösekompetenzen bedeutsam. Erstens ist dieser Forschungsgegenstand für die Bildung von Lehrenden relevant, weil er Erkenntnisse dazu liefert, wie diagnostische Problemlösekompetenzen im Berufsschulunterricht erfasst werden können, d. h., wie eine hochwertige Lernerfolgskontrolle erfolgen kann. Zweitens ermöglichen Studien zur Erfassung diagnostischer Problemlösekompetenzen eine verständliche Beschreibung und ein tiefgründiges Verständnis eines wichtigen Fördergegenstands, was ebenfalls der akademischen Bildung von Lehrenden zugutekommen kann. Drittens müssen diagnostische Problemlösekompetenzen erfasst werden, um einschlägige Theorien zu prüfen.

### **1.4 Aufbau und Funktion des vorliegenden Texts**

Der vorliegende Text dient der theoretischen Grundlegung sowie der Zusammenfassung und Reflexion der Ergebnisse der kumulativen Habilitation. Für die theoretische Grundlegung wird im nächsten Kapitel präzisiert, was mit „Erfassung diagnostischer Problemlösekompetenzen“ gemeint ist. Dabei wird die Erfassung diagnostischer Problemlösekompetenzen als Modellierungsprozess rekonstruiert und es erfolgt eine Konzentration auf zwei zentrale Modellierungsschritte bzw. Modelle: das Sprachmodell und das Erhebungsmodell. Diese beiden Modelle werden in Kapitel 3 bzw. Kapitel 4 kriteriengeleitet und anhand systematischer Überlegungen sowie empirischer Befunde konkretisiert. Hierbei werden einige Forschungsdesiderate sichtbar, die in Kapitel 5 zu einem Forschungsprogramm gebündelt und in sechs Forschungsfragen überführt werden. Diese Forschungsfragen wurden im Rahmen von sechs empirischen Studien untersucht (Anhang A bis Anhang F), die in verschiedenen referierten Zeitschriften veröffentlicht wurden. Die Ergebnisse dieser Studien bilden den Kern der kumulativen Habilitation und werden in Kapitel 6 präsentiert. Kapitel 7 zielt darauf ab, die Ergebnisse der Einzelstudien zusammenzuführen und auf das übergeordnete Forschungsthema, die Erfassung diagnostischer Problemlösekompetenzen, zu beziehen. Dabei wird auch diskutiert, welche theoretischen Implikationen die Ergebnisse haben und welche Grenzen die vorliegende Untersuchung aufweist. Im abschließenden Ausblick werden Forschungsperspektiven skizziert, deren Bearbeitung im Anschluss an diese Untersuchung interessant sein könnte.

## **2 Erfassung diagnostischer Problemlösekompetenzen**

Die Erfassung diagnostischer Problemlösekompetenzen bezieht sich auf einen zielgerichteten Prozess. Im vorliegenden Kontext besteht das Ziel dieses Prozesses darin, sprachliche Aussagen über individuelle Ausprägungen der diagnostischen Problemlösekompetenz zu gewinnen. Es geht darum, sprachliche Aussagen darüber zu erhalten, wie die diagnostische Problemlösekompetenz einer spezifischen Person ausge-

prägt ist, was und wie viel eine Person bei einer einschlägigen Fördermaßnahme gelernt hat etc. In diesem Prozess werden die realen diagnostischen Problemlösekompetenzen in mehreren Schritten in sprachliche Aussagen über diese Kompetenz transformiert. Dieser Vorgang lässt sich als Modellierungsprozess rekonstruieren, der mehrere Modelle<sup>4</sup> umfasst und bei dem mehrere aufeinander bezogene Modellbildungen stattfinden. Dieser Modellierungsprozess wird nachfolgend beschrieben und wurde in Anlehnung an Gigerenzer (1981), Zlatkin-Troitschanskaia (2005) und Mislevy (2011) konzeptualisiert. Bei genauer Betrachtung handelt es sich um einen verwickelten und komplexen Prozess, der hier nur cursorisch, aber dennoch zweckmäßig entfaltet werden kann.

## 2.1 Modellierungsprozess

In Anlehnung an Gigerenzer (1981, S. 27) beginnt der Modellierungsprozess damit, dass ein:e Forscher:in ein Forschungsproblem identifiziert und einen damit assoziierten Forschungsgegenstand empirisch erfassen möchte. In der vorliegenden Studie stellt die Erfassung diagnostischer Problemlösekompetenzen dieses Forschungsproblem dar und „diagnostische Problemlösekompetenzen“ den Forschungsgegenstand.

Mit dem *Sprachmodell* wird möglichst präzise festgelegt, was „diagnostische Problemlösekompetenz“ bedeutet und auf welchen Wirklichkeitsbereich sie sich bezieht. In berufs- und wirtschaftspädagogischen Untersuchungen können diese begrifflichen Festlegungen sehr differenziert sein, wie einschlägige konzeptuelle Kompetenzmodelle belegen, die letztlich nichts anderes sind als Sprachmodelle (z. B. Beck, 2016). Im Rahmen dieser konzeptuellen Kompetenzmodelle wird festgelegt, dass sich die latente diagnostische Problemlösekompetenz am manifesten diagnostischen Problemlöseerfolg festmachen lässt. Der Problemlöseerfolg wird also als manifester Indikator für die latente diagnostische Problemlösekompetenz aufgefasst. Hiermit ist folgende Rekonstruktionslogik verbunden: Die individuelle Ausprägung der diagnostischen Problemlösekompetenz kann anhand des diagnostischen Problemlöseerfolgs rekonstruiert werden. Die theoretische Basis dieser Rekonstruktionslogik ist die Annahme, dass sich der sichtbare Problemlöseerfolg als Wirkung der diagnostischen Problemlösekompetenz (Ursache) darstellen lässt.<sup>5</sup> Hiermit ist aber lediglich gesagt, woran die diagnostische Problemlösekompetenz erkannt werden kann. Offen ist, welche psychischen Ressourcen damit gemeint sind. Auch das wird im Sprachmodell konkretisiert. In Anlehnung an das *evidence-centered assessment* sind im Sprachmodell die psychischen Ressourcen der diagnostischen Problemlösekompetenz sowie die darauf bezogenen externen Anforderungen begrifflich zu klären, wobei innerhalb dieses Ansatzes nicht von *Sprachmodell*, sondern von *domain modeling* gesprochen wird (Mislevy, 2011, S. 9). Das Sprachmodell hat insofern normierenden Charakter, als es die Bedeutung der diagnostischen Problemlösekompetenz sprachlich fixiert und festlegt, welche Wirklichkeitsbereiche bei der Erfassung in den Blick zu nehmen sind. Zlatkin-

4 In diesen Modellen werden spezifische Vorgänge des Modellierungsprozesses thematisch zusammengefasst. Eine präzisere Fassung des Modellbegriffs ist nicht zielführend und würde vom Thema wegführen.

5 Diese Annahme kann als theoretische Unterfütterung der definitorischen Festlegung betrachtet werden, dass sich die latente diagnostische Problemlösekompetenz am manifesten diagnostischen Problemlöseerfolg erkennen lässt.

Troitschanskaia (2005, S. 314f.) spricht in diesem Kontext von einem Theoriemodell. Da die sprachliche Modellierung diagnostischer Problemlösekompetenzen streng genommen keine Theorie darstellt, wird hier der Begriff *Sprachmodell* und nicht *Theoriemodell* verwendet. In Anlehnung an den Kritischen Rationalismus haben Theorien materiale Implikationen, d. h. sie enthalten Kausalaussagen in der Form von Ursache-Wirkungs-Beziehungen (Beck, 2009). Für die Messung der diagnostischen Problemlösekompetenz ist das Sprachmodell in ein numerisches Modell zu übersetzen (Gigenzer, 1981, S. 31). Dieser Übersetzungsvorgang umfasst mindestens zwei weitere Modelle: ein Erhebungsmodell und ein Messmodell.

Mit dem *Erhebungsmodell* wird das Sprachmodell mit der beobachtbaren Wirklichkeit verknüpft und formal (numerisch) repräsentiert. Im Erhebungsmodell ist festgelegt, welche und wie viele Items<sup>6</sup> (hier: diagnostische Probleme) zur Erfassung der diagnostischen Problemlösekompetenz verwendet werden. Zudem ist im Erhebungsmodell geregelt, in welchem Modus die diagnostischen Probleme vorgegeben und bearbeitet werden. Beispielsweise können die diagnostischen Probleme Papier- oder computerbasiert vorgegeben werden. Ein weiterer Aspekt des Erhebungsmodells bezieht sich auf die Bewertung, d. h. die Kodierung (auch: Scoring) von Probandenantworten. Die Kodierung von Probandenantworten besteht in der Zuordnung von Zahlen zu spezifischen Problemlösungsvorschlägen, wobei die Zahlenwerte die Qualität des beobachteten Proband:innenverhaltens anzeigen, d. h. den Problemlöseerfolg quantifizieren. Der Problemlöseerfolg lässt sich daran festmachen, ob diagnostische Probleme gelöst wurden oder nicht. In diesem Fall wird das Produkt des Problemlöseprozesses betrachtet. Häufig wird ergänzend berücksichtigt, wie sich Proband:innen im Problemlöseprozess verhielten – z. B., ob wichtige Problemlöseschritte vollzogen wurden. In diesem Fall wird der Problemlöseerfolg auch am Verhalten im Problemlöseprozess festgemacht.

Das *Messmodell* zielt darauf ab, die problemspezifischen Kodierungen zu einem probandenspezifischen Testwert<sup>7</sup> zusammenzufassen. Teilweise ist in diesem Kontext nicht von einem Messmodell, sondern von Skalierung die Rede (z. B. Brennan, 2001, S. 10). Eines der einfachsten Messmodelle ist die Aufsummierung von Einzelwerten einer Testperson. Neben dieser simplen Skalierungsvariante bietet der heutige Stand der Forschung viele weitere Messmodelle (z. B. Eid et al., 2010, S. 822 ff.). Da die Anwendung von Messmodellen an bestimmte Annahmen geknüpft ist, haben sich bei der Kompetenzerfassung Messmodelle etabliert, deren Voraussetzungen transparent und teilweise empirisch prüfbar sind. Letztlich basieren aber alle Messmodelle auf axiomatischen Setzungen, also auf grundlegenden Annahmen, die empirisch nicht prüfbar sind. So wird z. B. bei Messmodellen der klassischen Testtheorie in axiomatischer Hinsicht unterstellt, dass sich ein Testwert aus einem wahren Wert und einem Mess-

---

6 Mit *Item* ist eine Kombination aus Teststimulus (diagnostisches Problem, Aufgabe etc.) und Antwortformat (Multiple-Choice, offenes Antwortformat etc.) gemeint.

7 Ein Test stellt nach Lienert und Raatz (1998, S. 1) ein wissenschaftliches Routineverfahren dar, bei dem ein empirisch abgrenzbares Persönlichkeitsmerkmal (hier: diagnostische Problemlösekompetenz) mit dem Ziel erfasst wird, eine quantitative Aussage (einen Testwert) zur individuellen Ausprägung dieses Merkmals zu erhalten. Unter dieser begrifflichen Festlegung umfasst ein Test das Erhebungsmodell und das Messmodell.

fehlerwert zusammensetzt (ebd., S. 818). Im Hinblick auf Messmodelle spielt auch die theoretische Vorstellung eine Rolle, wie differenziert ein Untersuchungsgegenstand erfasst werden kann und damit auch, welches Skalenniveau die betreffenden Daten aufweisen (Beck, 2009, S. 242). Diese theoretische Vorstellung ist Teil der sprachlichen Modellierung des Forschungsgegenstands, also des Sprachmodells.<sup>8</sup>

Im *Interpretationsmodell* erfolgt die Interpretation der Testwerte, d. h. die Rückübersetzung der Testwerte bzw. numerischen Aussagen in sprachliche Aussagen. An dieser Stelle wird deutlich, dass sprachliche Aussagen über die diagnostische Problemlösekompetenz aus Testwertinterpretationen resultieren. So wird z. B. aus einem hohen Testwert interpretativ die Aussage gewonnen, dass ein Individuum über eine hohe Problemlösekompetenz verfügt. Um zu gewährleisten, dass bei diesem Übersetzungsvorgang keine wichtigen Informationen verloren gehen und keine störenden Informationen hinzukommen, muss klar sein, wie Testwerte sprachlich interpretiert werden dürfen. Dies hängt einerseits vom Sprachmodell ab, weil es die inhaltliche Bedeutung der Aussagen über die diagnostische Problemlösekompetenz festlegt. Andererseits hängt die Interpretation der Testwerte von der Qualität des gesamten Modellierungsprozesses ab, weil diese Qualität darüber entscheidet, inwiefern Testwerte die diagnostische Problemlösekompetenz unverzerrt abbilden. Gelegentlich werden Skalierungsergebnisse nicht direkt, sondern anhand statistischer Vergleichswerte (z. B. Mittelwert) oder von Normtabellen interpretiert. Dann werden zusätzlich zum Interpretationsmodell ein Datenanalysemodell und ein Normierungsmodell benötigt.

## 2.2 Konsequenzen

Was folgt nun aus diesen Ausführungen für die Erfassung diagnostischer Problemlösekompetenzen? Um diagnostische Problemlösekompetenzen empirisch zu erfassen, muss der beschriebene Modellierungsprozess realisiert werden: Es müssen das Sprachmodell, das Erhebungsmodell, das Messmodell und das Interpretationsmodell konkretisiert werden, wobei die Abhängigkeiten zwischen diesen Modellen zu beachten sind. So muss z. B. die Auswahl der Items (diagnostische Probleme) in Abhängigkeit von dem Sprachmodell erfolgen, weil dieses vorgibt, welche Items in das Bedeutungsfeld der diagnostischen Problemlösekompetenz fallen, d. h. überhaupt infrage kommen. Zudem ist zu überlegen, welchen Kriterien (theoretische Fruchtbarkeit, sprachliche Präzision, Validität etc.) der Modellierungsprozess genügen sollte. In Orientierung an diesen Kriterien kann dann der Modellierungsprozess konkretisiert werden.

Im Weiteren stehen das Sprachmodell und das Erhebungsmodell im Zentrum. Die Konzentration auf diese beiden Modelle scheint insofern angemessen, als es in einer einzelnen Habilitation kaum möglich und aus Qualitätsgründen (v. a. aus Gründen der Präzision und angemessenen Berücksichtigung des Forschungsstands) auch nicht ratsam sein dürfte, den Erkenntnisstand auf allen Modellebenen voranzutreiben.

---

8 Damit wird die Position bezogen, dass die Festlegung des Messniveaus (Skalenniveaus) per fiat, d. h. per definitorem Beschluss erfolgt. Dementsprechend ergibt sich das Messniveau daraus, ob diagnostische Problemlösekompetenzen klassifikatorisch, komparativ oder metrisch definiert werden (Stegmüller, 1989, S. 375 f.). Wie Bortz und Döring (2009, S. 70) zeigen, finden sich hierzu in der Literatur auch andere Auffassungen.

Ein konzentriertes Vorgehen scheint auch deswegen gerechtfertigt, weil trotz dieser Einschränkung bei den weiteren Ausführungen ausreichend viele Forschungsdesiderate sichtbar werden.

### 3 Konkretisierung des Sprachmodells diagnostischer Problemlösekompetenz

In diesem Kapitel wird das Sprachmodell zur diagnostischen Problemlösekompetenz konkretisiert. Zunächst wird ausgeführt, welche Methode zur Konkretisierung des Sprachmodells herangezogen wird. Dann werden Konkretisierungskriterien festgelegt. Diese Kriterien werden genutzt, um wichtige Aspekte des Sprachmodells zu bestimmen. Anschließend werden diese Sprachmodellaspekte anhand systematischer Überlegungen und empirischer Befunde elaboriert.

#### 3.1 Wie wird das Sprachmodell konkretisiert?

Die Funktion des Sprachmodells besteht darin, die Bedeutung der diagnostischen Problemlösekompetenz festzulegen. In einer „idealen“ oder zumindest (sehr) weit entwickelten Wissenschaft kann eine solche Festlegung unter Anwendung hoch formalisierter und begrifflich präzise ausgearbeiteter Theorien mit einer Nominaldefinition erfolgen. Dabei wird ein neuer Ausdruck durch einen bekannten Ausdruck eingeführt, wobei beide Ausdrücke bedeutungsidentisch sein müssen (Stegmüller, 1967, S. 335). Praktisch gesprochen bedeutet dies z. B., die diagnostische Problemlösekompetenz so zu definieren, dass das Definiendum (der neue Ausdruck, hier: diagnostische Problemlösekompetenz) in allen betreffenden Theorien gegen das Definiens (der bekannte Ausdruck) ausgetauscht werden kann, ohne dass sich die Bedeutung der Theorien verändert. Eine Nominaldefinition ist im vorliegenden Kontext aus mindestens zwei Gründen nicht möglich: Erstens ist *Kompetenz* ein Begriff der Alltagssprache, also ein vorwissenschaftlicher Begriff, der mehrdeutig ist. So wird unter Kompetenz im Alltag nicht nur eine psychische Grundlage verstanden; der Kompetenzbegriff wird auch im Sinne von Zuständigkeit gebraucht (Dudenredaktion, 2002, S. 545). Zweitens liegt im Bereich des diagnostischen Problemlösens keine Theorie vor, die hoch formalisiert und sprachlich präzise ausgearbeitet ist.

In dieser Arbeit wird das Sprachmodell, also die Bedeutung der diagnostischen Problemlösekompetenz, anhand einer Begriffsexplikation konkretisiert. Bei einer Begriffsexplikation wird ein vager Begriff (Explikandum) mit einer möglichst präzisen Bedeutung (Explikat) versehen. „Eine Begriffsexplikation besteht [...] meist aus sehr komplexen Tätigkeiten eines Wissenschaftlers; es werden darin Bedeutungsanalysen, empirische Analysen, wissenschaftliche Hypothesen und Nominaldefinitionen eine Rolle spielen“ (Stegmüller, 1967, S. 340). Dieses Zitat vergegenwärtigt, dass es sich bei Begriffsexplikationen oft um einen verwickelten und undurchsichtigen Vorgang handelt, in den, grob gesprochen, systematische Überlegungen und empirische Befunde einfließen. Letztlich, und das ist wichtig, enthält jede Begriffsexplikation aber eine kon-

ventionelle Komponente, weshalb sie nicht wahr oder falsch sein kann (Stegmüller, 1989, S. 374). Ein Sprachmodell, das mit einer Begriffsexplikation gewonnenen wird, repräsentiert in letzter Konsequenz einen konventionellen Beschluss und keine wahrheitsfähige Aussage (Horlebein, 2013, S. 5), die logisch oder empirisch prüfbar ist. Eine Begriffsexplikation kann aber mehr oder weniger adäquat sein (Poser, 2012, S. 44). So hat Carnap (1962) aus einer metatheoretischen Sicht gefordert, dass Begriffsexplikationen bestimmten wissenschaftlichen Kriterien genügen sollten (s. unten). Im Sinne vergleichbarer Forschung und des Erkenntnisfortschritts sowie einer möglichst reibungslosen Kommunikation scheint es außerdem erstrebenswert, dass sich möglichst viele Forschende auf dasselbe Verständnis von diagnostischer Problemlösekompetenz verpflichten. Eine grundlegende Annahme der weiteren Ausführungen ist, dass die Konsensfähigkeit des Sprachmodells auch davon abhängt, wie gut das Sprachmodell systematisch und empirisch begründet ist.

### 3.2 Festlegung konkretisierungsbedürftiger Sprachmodellaspekte

In diesem Abschnitt werden anhand einer Begriffsexplikation und ausgewählten Kriterien zentrale Aspekte des Sprachmodells herausgearbeitet.

#### 3.2.1 Vorüberlegungen: Begriffserläuterung, Festlegungskriterien und Ähnlichkeit

Im Vorfeld der Begriffsexplikation ist eine *Erläuterung des Explikandums* (diagnostische Problemlösekompetenz) vorzunehmen (Stegmüller, 1989, S. 374). Wie eine Google-Suche zeigt, wird der Begriff „diagnostische Problemlösekompetenz“ sowohl wissenschaftlich als auch wissenschaftlich äußerst selten verwendet (Schlagwörter: „diagnostische Problemlösekompetenz“; „diagnostic problem-solving competence“ und „diagnostic problem-solving competency“). Eine Ausnahme stellt die Publikation von Frederiksen et al. (1981) dar, in der von „clinical diagnostic problem-solving competency“ die Rede ist. Bei einer umgangssprachlichen Annäherung an diagnostische Problemlösekompetenz lassen sich drei begriffliche Bestandteile unterscheiden: Diagnose, Problem und Kompetenz. Mit *Diagnose* wird für gewöhnlich der Vorgang zur Bestimmung einer Krankheit bezeichnet (Dudenredaktion, 2002, S. 261), mit *Problem* eine schwierige Aufgabe (ebd., S. 698). *Kompetenz* wird im allgemeinen Sprachgebrauch sowohl im Sinne von Zuständigkeit als auch von Fähigkeit verwendet (ebd., S. 545). Hier ist mit diagnostischer Problemlösekompetenz – wie auch bei Frederiksen et al. (1981) – eine Fähigkeit gemeint (Kapitel 1.1). Aus einem alltagsprachlichen Gebrauch heraus könnte also vermutet werden, dass mit diagnostischer Problemlösekompetenz die Fähigkeit angesprochen ist, schwierig zu bestimmende Krankheiten zu diagnostizieren.

Um das Explikandum (diagnostische Problemlösekompetenz) zu explizieren, werden die folgenden von Carnap (1962, S. 5 ff.) vorgeschlagenen Kriterien herangezogen: Ähnlichkeit („similarity“), Fruchtbarkeit („fruitfulness“), Exaktheit („exactness“) und Einfachheit („simplicity“). Für die Wahl dieser Kriterien gibt es keine Letztbegründung; die Kriterien folgen nicht logisch aus einem Absolutkriterium. Vielmehr sind sie

aus einer pragmatischen Perspektive als zweckmäßig einzustufen (Essler, 1970, S. 58); sie erwiesen sich als wissenschaftlich brauchbar (Poser, 2012, S. 41 ff.).

Mit dem Kriterium der Ähnlichkeit wird gefordert, dass das Explikandum (diagnostische Problemlösekompetenz) und das Explikat („Resultat der Explikation“; Essler, 1970, S. 56) eine ähnliche, aber keine identische Bedeutung haben, denn im Falle einer identischen Bedeutung wäre keine Explikation nötig. Wie in Kapitel 1.1 dargestellt, wird diagnostische Problemlösekompetenz hier als psychische Grundlage aufgefasst, die benötigt wird, um die Ursache(n) eines unerwünschten Zustands zu identifizieren. Unter einem Problem wird eine Situation verstanden, bei der eine Unbekannte (hier: eine unbekannte Ursache) zu finden ist (Kapitel 1.2). Übereinstimmend mit dem zuvor beschriebenen, alltagssprachlichen Begriffsverständnis wird Kompetenz hier im Sinne von Fähigkeit verstanden. In Erweiterung dieses Verständnisses umfasst „diagnostische Problemlösekompetenz“ allerdings nicht nur schwierige Aufgaben und die Bestimmung von Krankheiten, sondern alle Anforderungen, bei denen unbekannte Ursachen für unerwünschte Zustände zu finden sind. Carnap (1962, S. 5–6) weist darauf hin, dass es gute Gründe (z. B. theoretische Fruchtbarkeit) geben kann, das Ähnlichkeitskriterium nicht zu streng auszulegen. Die hier vorgenommene Erweiterung der alltagssprachlichen Fassung von diagnostischer Problemlösekompetenz ist aus mindestens den folgenden beiden Gründen vorteilhaft: Erstens wirkt sich diese Begriffsfassung günstig auf die Operationalisierung von diagnostischer Problemlösekompetenz aus (s. dazu die Ausführungen in Kapitel 7.3.3). Zweitens ermöglicht sie es, neben Studien, die sich mit der Bestimmung von Krankheiten beschäftigen, auch Studien anderer Bereiche einzubeziehen (s. dazu Kapitel 3.3.1). Die Erweiterung trägt also zu einem theoretisch fruchtbaren Sprachmodell bei.

### 3.2.2 Anwendung des Fruchtbarkeitskriteriums

Theoretisch fruchtbar ist das Sprachmodell, wenn es an den Stand der Forschung anschließt und Erkenntnisfortschritte ermöglicht (in Anlehnung an Stegmüller, 1989, S. 375). Mit diesem Kriterium wird abgesichert, dass der Forschungsstand konsequent weiterentwickelt wird und Forschungsergebnisse aufeinander beziehbar sind. Nachfolgend werden vier Aspekte des Sprachmodells der diagnostischen Problemlösekompetenz thematisiert. Diese Aspekte wurden in Auseinandersetzung mit dem Stand der Forschung identifiziert, wobei hier nur ausgewählte Publikationen zitiert werden, die allerdings für den einschlägigen Forschungsstand als repräsentativ gelten dürfen.

Als erster Aspekt des Sprachmodells sei der *Komponentenaspekt*<sup>9</sup> genannt. Zlatkin-Troitschanskaia und Seidel (2011) bilanzierten weite Teile der nationalen und internationalen empirischen Kompetenzforschung und arbeiteten heraus, dass Kompetenzen in der berufs- und wirtschaftspädagogischen Forschung als Konstrukte anzulegen sind, die mehrere kognitive und nichtkognitive Komponenten umfassen (s. dazu auch Frederiksen et al., 1981; Sembill et al., 2013; Straka & Macke, 2009, S. 32). Dem For-

---

9 Alternativ wäre denkbar, von Facetten bzw. Aspekten zu sprechen. Weitere Forschung muss zeigen, ob sich die Komponenten- oder Facettensicht als fruchtbarer erweist.

schungsstand folgend sollte also festgelegt werden, welche psychischen Komponenten die diagnostische Problemlösekompetenz umfasst.

Ein weiterer Sprachmodellaspekt ist der *Prozessaspekt*. Üblicherweise wird unterstellt, dass sich Problemlösekompetenzen auf die Gestaltung eines Prozesses, nämlich des Problemlöseprozesses, beziehen (domänenübergreifendes Problemlösen: z. B. Greiff, 2012b, S. 87; domänenspezifisches Problemlösen: z. B. Seifried et al., 2016, S. 121 f.). Auch aus Sicht der berufs- und wirtschaftspädagogischen Lehr-Lern-Forschung ist der Prozessaspekt bedeutsam. Falls der Problemlöseprozess nicht berücksichtigt wird, beschränkt sich die Kompetenzerfassung auf das Produkt des Prozesses; der Prozess selbst und das im Prozess gezeigte Verhalten werden dagegen nicht explizit berücksichtigt. Nachteilig ist dies z. B. insofern, als ausgehend vom Prozessprodukt (Problem gelöst oder nicht gelöst) nur vage Aussagen über den zugrunde liegenden Prozess und das Verhalten im Prozess gemacht werden können. Dadurch kann verborgen bleiben, in welchem Bereich Personen relevante Lerndefizite haben und Unterstützung benötigen. Aufgrund ihres „interventionistischen Charakters“ (Beck, 1987, S. 168) ist die Berufs- und Wirtschaftspädagogik generell eher am Prozess der Leistungserstellung als an ihrem Produkt interessiert. Dementsprechend sollte das Sprachmodell zur diagnostischen Problemlösekompetenz auch Aussagen zum diagnostischen Problemlöseprozess enthalten.

Der dritte Sprachmodellaspekt, der hier berücksichtigt wird, ist der *Anforderungsaspekt*. Unter dem Anforderungsaspekt ist der externe Anforderungsbereich festzulegen, auf den sich die betreffende Kompetenz bezieht. Hier sei stellvertretend für den Forschungsstand Shavelson (2009) genannt, der auf die große Bedeutung externer Anforderungen für die Kompetenzerfassung hinweist (s. dazu z. B. auch Hartig & Klieme, 2006, S. 128 f.).

Der *Eigenständigkeitsaspekt* stellt den vierten Aspekt des Sprachmodells dar. In Anlehnung an Cronbach ist es bei der Erfassung eines psychischen Konstrukts aus theoretischer Sicht wichtig, das Konstrukt in einem nomologischen Netz von Konstrukten zu verorten (Hartig et al., 2008, S. 145 ff.). Dies erfordert, das Konstrukt auch in Bedeutungsabgrenzung zu anderen Konstrukten zu fassen. Unter dem Eigenständigkeitsaspekt ist bei der begrifflichen Klärung des Sprachmodells zu prüfen, ob die diagnostische Problemlösekompetenz als empirisch eigenständig und empirisch abgrenzbar von anderen Konstrukten aufgefasst werden kann.

### 3.2.3 Bewertung der Sprachmodellaspekte anhand weiterer Kriterien

Im Anschluss an den einschlägigen Forschungsstand sollte das Sprachmodell unter dem Komponentenaspekt, dem Prozessaspekt, dem Anforderungsaspekt und dem Eigenständigkeitsaspekt konkretisiert werden. Wie sind die vier Sprachmodellaspekte aus Sicht der Kriterien „Exaktheit“ und „Einfachheit“ zu bewerten?

„Exaktheit“ bezieht sich darauf, einen Begriff eindeutig festzulegen, d. h. ihn in „ein ganzes System wissenschaftlicher Begriffe einzuordnen“ (Stegmüller, 1989, S. 375). Dies erfordert elaborierte Theorien, die zum diagnostischen Problemlösen aber nicht verfügbar sind. Im vorliegenden Kontext wird das Kriterium in abgeschwächter Form

verwendet und es wird gefordert, dass möglichst genau und detailliert zu beschreiben ist, was unter diagnostischer Problemlösekompetenz verstanden wird. Die unter dem Kriterium „Fruchtbarkeit“ gewonnenen vier Aspekte des Sprachmodells tragen insofern zur Exaktheit bei, als sie eine differenzierte und damit sprachlich präzise Modellierung der diagnostischen Problemlösekompetenz begünstigen. Besonders deutlich wird dies am Komponentenaspekt sowie dem Anforderungs- und Prozessaspekt.

Mit dem Einfachheitskriterium wird gefordert, Begriffssysteme und Theorien möglichst einfach zu gestalten, d. h. nur so komplex anzulegen, wie nötig (in Anlehnung an Stegmüller, 1989, S. 375). Dies bedeutet z. B., auf die Einführung von Konstrukten zu verzichten, die sich nicht von bekannten und bewährten Konstrukten unterscheiden. Dieses Kriterium liefert eine weitere Begründung für den *Eigenständigkeitsaspekt*, der darauf abzielt, diagnostische Problemlösekompetenz als abgrenzbares und eigenständiges Konstrukt zu konzipieren und, falls dies *nicht* möglich sein sollte, auf dieses Konstrukt zu verzichten.

Die vier Aspekte legen das Sprachmodell der diagnostischen Problemlösekompetenz inhaltlich nicht fest. So bleibt z. B. offen, welche psychischen Komponenten die diagnostische Problemlösekompetenz konstituieren. Die Sprachmodellaspekte geben allerdings Konkretisierungsbereiche vor. Es sei ausdrücklich betont, dass weitere Aspekte denkbar sind und die Aspekte nicht logisch aus der Anwendung der Kriterien auf den Forschungsstand resultieren. Vielmehr liefern die Explikationskriterien rationale Gründe für die Wahl der Aspekte, womit Kritik möglich und (hoffentlich) ein Begriffskonsens begünstigt wird.

### 3.3 Konkretisierung der Sprachmodellaspekte

In diesem Abschnitt wird das Sprachmodell unter den genannten Sprachmodellaspekten anhand systematischer Überlegungen und empirischer Befunde konkretisiert. Der einschlägige Forschungsstand bietet zahlreiche empirische Studien. Einige dieser Studien entstanden im technischen Bereich, ein Großteil davon wurde im medizinischen Bereich angefertigt. Es ist hier nicht beabsichtigt, die Vielzahl einschlägiger empirischer Arbeiten detailliert und umfassend aufzuarbeiten, sondern auf ausgewählte und zentrale Studien einzugehen. Damit stellt sich die Frage, welche Studien zu Konkretisierung des Sprachmodells herangezogen werden.

#### 3.3.1 Einbezogene Studien

Neben Studien, die sich explizit mit diagnostischem Problemlösen beschäftigen, werden weitere verwandte Studien berücksichtigt. Wie bereits erwähnt, bezeichnet „diagnostisches Problemlösen“ Situationen, in denen nach Ursachen für einen unerwünschten Zustand gesucht wird. In *technischen* Kontexten wird in solchen Situationen auch von „Fehleranalyse“ (Nickolauset al., 2012a, S. 245 f.), „diagnostic reasoning“ (Rasmussen, 1993) oder „analytical problem solving“ (Walker et al., 2016a) gesprochen. Zudem wird im technischen Bereich häufig der Begriff *troubleshooting* verwendet (Schaafstal et al., 2000). Allerdings umfasst *troubleshooting* zwei unterscheidbare Vorgänge: die Bestimmung der Problemursache (also diagnostisches Problemlösen) sowie

die Behebung der Problemursache (Jonassen & Hung, 2006, S. 79). Dies gilt auch für „clinical reasoning“, einen Begriff, der in Artikeln des medizinischen Bereichs verwendet wird. In Anlehnung an Durning et al. (2015, S. 127) sowie Vleuten et al. (2008, S. 414) umfasst „clinical reasoning“ die folgenden Vorgänge: „diagnostic reasoning“ und „therapeutic reasoning“, wobei mit „diagnostic reasoning“ das gemeint ist, was hier „diagnostisches Problemlösen“ genannt wird.

Vor diesem Hintergrund werden nachfolgend Studien berücksichtigt, die sich mit „diagnostischem Problemlösen“, „Fehleranalyse“, „diagnostic reasoning“ sowie „analytical problem solving“ beschäftigen. Zudem werden Studien zu „troubleshooting“ und „clinical reasoning“ einbezogen. Denn: Im Anschluss an Jonassen (2011, S. 78) lässt sich vermuten, dass auch in diesen Studien der Fokus auf dem diagnostischen Problemlösen liegt, was bereits von Katz und Anderson (1987) im Programmierbereich empirisch bestätigt wurde. Für eine Inklusion von Studien zu „troubleshooting“ und „clinical reasoning“ spricht auch, dass zu diesen Themen bereits viele Studien vorliegen. Studien, die sich mit übergeordneten Konstrukten beschäftigen (clinical competence, professional competence, berufliches Problemlösen, domänenspezifisches Problemlösen etc.), werden nur am Rande berücksichtigt.

### 3.3.2 Eigenständigkeitsaspekt

Unter dem Eigenständigkeitsaspekt ist die diagnostische Problemlösekompetenz begrifflich so zu konzipieren, dass sie sich in ihrer Bedeutung von anderen relevanten Konstrukten unterscheidet. Daraus lassen sich keine empirischen Konsequenzen logisch ableiten. Wenn Studien allerdings belegen, dass sich diagnostische Problemlösekompetenzen empirisch von anderen Konstrukten unterscheiden lassen, kann darin ein überzeugender Grund – freilich kein Beweis – gesehen werden, diagnostische Problemlösekompetenzen begrifflich als eigenständiges Konstrukt zu betrachten und die Kompetenzerfassung an dieser Vorstellung auszurichten.

Aus einer begriffssystematischen Sicht wird die diagnostische Problemlösekompetenz als eine von mehreren Kompetenzen verstanden, die innerhalb eines Berufs benötigt werden (Kapitel 1.2). Auf der Bedeutungsebene unterscheiden sich berufliche Kompetenzen unter anderem durch ihren Anforderungsbezug. So beziehen sich diagnostische Problemlösekompetenzen z. B. auf die Identifikation von Ursachen für technische Störungen, wogegen sich Reparaturkompetenzen auf die Behebung dieser Störungen beziehen können. Empirisch kann dies folgendermaßen gedeutet werden: Diagnostische Problemlösekompetenzen repräsentieren einen anderen Wirklichkeitsbereich als Reparaturkompetenzen. Im Rahmen empirisch-quantitativer Forschungsansätze werden beide Wirklichkeitsbereiche typischerweise anhand von Testwerten repräsentiert. Falls sich diese Testwerte als *diskriminant valide* (empirisch unterscheidbar) erweisen, kann dies zum Anlass genommen werden, diagnostische Problemlösekompetenzen als abgrenzbare und damit eigenständige Konstrukte aufzufassen.

Auch die Homogenität von Testscores kann als empirisches Argument angeführt werden, diagnostische Problemlösekompetenzen als empirisch eigenständige Konstrukte zu betrachten. Statistisch bedeutet diese Homogenität z. B., dass Testscores der

diagnostischen Problemlösekompetenz eindimensional darstellbar sind oder unterschiedliche Testwerte der diagnostischen Problemlösekompetenz (nahezu) perfekt oder wenigstens hoch korrelieren, also *konvergent valide* sind.

### Homogenität der Testwerte und Testscores<sup>10</sup> der diagnostischen Problemlösekompetenz

Der aktuelle Forschungsstand bietet nur wenige Befunde zu der Frage, ob Testwerte der diagnostischen Problemlösekompetenz als konvergent valide aufgefasst werden können. Swanson et al. (1987) referieren in ihrem Überblicksartikel zur Erfassung klinischer Kompetenzen zwischen unterschiedlichen Testwerten der diagnostischen Problemlösekompetenz Korrelationen im Bereich von  $r = .64$  bis  $r = 1.0$ . Als einen potenziellen Grund für die am unteren Ende dieses Bereichs liegenden Korrelationen führen sie geringe Reliabilitäten einiger Testwerte an. Bei Abele (2014, S. 190) erwiesen sich in der Ausbildung zum/zur Kfz-Mechatroniker:in zwei Testwertreihen (Skalen) der diagnostischen Problemlösekompetenz als konvergent valide.

Ebenfalls im technischen Bereich fand die Studie von Walker et al. (2016b) statt. Dort zeigte sich in der Ausbildung zum/zur Elektroniker:in für Automatisierungstechnik, dass Testscores der diagnostischen Problemlösekompetenz statistisch eindimensional abbildbar sind. Die diagnostische Problemlösekompetenz von Auszubildenden der Kfz-Mechatronik erwies sich in mehreren Studien ebenfalls als eindimensional skalierbar (Abele, 2014, S. 189; Gschwendtner et al., 2009; Nickolaus et al., 2012a, S. 548).

Zusammengenommen legen die Befunde nahe, von einer Homogenität der Testscores und einer konvergenten Validität von Testwerten der diagnostischen Problemlösekompetenz auszugehen. Die Befundlage ist jedoch bisweilen eher spärlich. Interessant ist ferner, dass *diagnostische Problemlösekompetenz* in den erwähnten Studien (implizit) als metrischer Begriff betrachtet wird – dies gilt auch für viele im Weiteren zitierte Studien. Dies wird i. d. R. nicht explizit thematisiert, aber an den verwendeten statistischen Verfahren sichtbar.

### Diskriminante Validität der Testwerte der diagnostischen Problemlösekompetenz

Zu der Frage, ob sich Testwerte diagnostischer Problemlösekompetenzen von Testwerten anderer beruflicher Kompetenzen unterscheiden lassen, sind mehrere Befunde aus dem medizinischen Bereich verfügbar. Swanson et al. (1987, S. 238) bilanzierten, dass die Korrelationen zwischen diagnostischen Problemlösekompetenzen und anderen klinischen Kompetenzen in vielen Studien zwischen  $r = .1$  und  $r = .4$  lagen, also eher schwach bis mittel ausgeprägt waren. Diese Befunde können als Evidenz der diskriminanten Validität der Testwerte der diagnostischen Problemlösekompetenz betrachtet werden. Auch hier kritisierten die Autoren allerdings, dass einige Testwerte eine unbefriedigende Reliabilität aufwiesen. Ebenfalls für den Medizinbereich berichteten Wass et al. (2001, S. 329) zwischen diagnostischen Problemlösekompetenzen und

---

<sup>10</sup> Ein Testscore resultiert aus der Kodierung einer individuellen Problemlösung und nimmt z. B. den Wert null (Problem nicht gelöst) oder eins (Problem gelöst) an. Mit „Testwert“ sind dagegen die probandenspezifischen Werte gemeint, die aus der Skalierung der Testscores hervorgehen.

anderen beruflichen Kompetenzen moderate bis hohe Korrelationen ( $r \approx .5-.7$ ). Die Autoren interpretierten diese Ergebnisse einerseits im Sinne der diskriminanten Validität und andererseits als Evidenz dafür, dass diagnostische Problemlösekompetenzen und andere berufliche Kompetenzen deutlich assoziiert sind. Boulet et al. (2002) ermittelten zwischen Testwerten der diagnostischen Problemlösekompetenz und Testwerten anderer beruflicher Kompetenzen sehr hohe Korrelationen ( $r \approx .8$ ), was die Autoren letztlich dennoch dazu veranlasste, von diskriminanter Validität auszugehen. Anzumerken ist, dass in den genannten Studien manifeste Korrelationsanalysen durchgeführt wurden, der Forschungsstand aber geeignetere statistische Verfahren (mehrdimensionale latente Messmodelle) zur Prüfung der diskriminanten Validität bietet.

In der Studie von Walker et al. (2016a) wurde mithilfe einer konfirmatorischen Faktorenanalyse im Beruf des Elektrikers/der Elektrikerin für Automatisierungstechnik untersucht, ob Testwerte der diagnostischen Problemlösekompetenz von Testwerten der konstruktiven Problemlösekompetenz, also einer anderen beruflichen Kompetenz, empirisch abgrenzbar sind. Die Ergebnisse zeigten, dass die Testwerte beider beruflicher Kompetenzen sehr eng korrelierten ( $r = .77$ , latent), letztlich aber empirisch unterscheidbar waren. Im Ausbildungsberuf zum/zur Kfz-Mechatroniker:in wurde festgestellt, dass Testwerte der diagnostischen Problemlösekompetenz sehr eng mit Testwerten des Diagnosewissens korrelierten ( $r = .63$ ); auch hier erwiesen sich beide Skalen aber als empirisch trennbar (Abele, 2014, S. 194). Aus Gründen einer unbefriedigenden Modellpassung wurde in der erwähnten Publikation ein Faktorenmolell ausgeschlossen, in dem beide Testwertreihen (Diagnosewissen und Problemlösekompetenz) so hoch korrelierten ( $r = .83$ , latent), dass gewisse Zweifel an deren diskriminanten Validität angemeldet werden können (Abele, 2014, S. 194, Fußnote 13). Verstärkt werden diese Zweifel dadurch, dass sich die Erfassung des Diagnosewissens im Unterschied zur Erfassung der diagnostischen Problemlösekompetenz nicht auf die Bewältigung authentischer beruflicher Anforderungen bezog. Es ist gut vorstellbar, dass das Diagnosewissen, wenn es authentischer sowie anforderungsbezogen erfasst wird, noch enger mit der diagnostischen Problemlösekompetenz im Kfz-Bereich korrelieren könnte und empirisch möglicherweise nicht mehr von der diagnostischen Problemlösekompetenz unterscheidbar wäre.

Es spricht einiges dafür, dass Testwerte diagnostischer Problemlösekompetenzen von anderen Testwerten beruflicher Kompetenzen empirisch abgegrenzt werden können, wobei die teilweise hohen Korrelationen (v. a. im Kfz-Bereich) die diskriminante Validität auch infrage stellen und Forschungsbedarf anzeigen. Festzuhalten ist zudem, dass die Befundlage breiter sein könnte, und dies gilt besonders dann, wenn strenge forschungsmethodische Standards angelegt werden: Studien, in denen die diskriminante Validität anhand mehrdimensionaler statistischer Verfahren untersucht wurde, sind besonders selten.

### 3.3.3 Komponentenaspekt

Unter dem Komponentenaspekt sind die psychischen Komponenten der diagnostischen Problemlösekompetenz zu bestimmen. Diese Bedeutungsfestlegung lässt sich

empirisch so deuten, dass die betreffenden psychischen Komponenten für die diagnostische Problemlösekompetenz empirisch relevant sind. Um die empirische Relevanz zu untersuchen, werden häufig die interessierenden Kompetenzkomponenten und die diagnostische Problemlösekompetenz separat erfasst und anschließend die Interkorrelationen ermittelt. Deutliche Interkorrelationen werden im Sinne empirischer Relevanz interpretiert.

Nur sehr wenige Studien haben sich bislang mit der Frage beschäftigt, welche psychischen Komponenten für die diagnostische Problemlösekompetenz empirisch relevant sind. Elaboriert ist der Ansatz von Jonassen (2011), in dem die folgenden kognitiven Komponenten der diagnostischen Problemlösekompetenz unterschieden werden: Diagnosewissen, Arbeitsgedächtniskapazität und die Fähigkeit zum schlussfolgernden Denken – die Komponenten „Diagnosewissen“ und „schlussfolgernde Denkfähigkeit“ lassen sich ausdifferenzieren. Zwischen dem Diagnosewissen und der diagnostischen Problemlösekompetenz wurden mehrfach substantielle Korrelationen festgestellt (z. B. Abele, 2014, S. 194; Walker et al., 2016b; Gschwendtner et al., 2009). Wie Jonassen und Hung (2006, S. 87) ausführten, besteht auch zwischen der Arbeitsgedächtniskapazität und dem Problemlöseerfolg ein substanzieller Zusammenhang. Darüber hinaus liegen Befunde vor, die einen deutlichen Zusammenhang von fluider Intelligenz und diagnostischer Problemlösekompetenz dokumentieren (z. B. Abele, 2014, S. 197; Walker et al., 2016a). Zwischen fluider Intelligenz, Arbeitsgedächtniskapazität und der schlussfolgernden Denkfähigkeit werden gemeinhin sowohl aus theoretischen als auch aus empirischen Gründen enge Korrelationen unterstellt (Abele, 2014, S. 83 und S. 112). Es wurden keine Studien gefunden, die sich mit der Frage beschäftigen, ob auch nichtkognitive Komponenten der diagnostischen Problemlösekompetenz empirisch relevant sind. In Überblicksartikeln wird diagnostische Problemlösekompetenz i. d. R. als kognitives Konstrukt angelegt (Jonassen, 2011, S. 80 ff.; Vleuten et al., 2008, S. 413). In diesem Kontext ist die Studie von Rausch et al. (2016) interessant. Dort wurden die Komponenten „Selbstkonzept“ und „Interesse“ der beruflichen Problemlösekompetenz erfasst. Es zeigte sich, dass diese nichtkognitiven Komponenten beachtlich mit kognitiven Komponenten der beruflichen Problemlösekompetenz korrelierten.

Zusammengenommen sprechen die Befunde dafür, das Diagnosewissen, die Arbeitsgedächtniskapazität, die schlussfolgernde Denkfähigkeit und die fluide Intelligenz als kognitive Komponenten der diagnostischen Problemlösekompetenz aufzufassen. Angesichts des aktuellen Forschungsstands muss jedoch als offen gelten, ob der kognitive Komponentenbereich damit erschöpfend abgedeckt ist. Fragen ließe sich z. B., ob auch der Umgang mit Komplexität berücksichtigt werden sollte. Komplexität wird in der Problemlöseforschung als zentrales Merkmal von Problemen betrachtet (z. B. Funke, 2003, S. 34). Dörner wies verschiedentlich auf die enorme Bedeutung des Umgangs mit Komplexität beim Problemlösen hin (z. B. Dörner, 2008). Außerdem scheinen neben den erwähnten kognitiven Komponenten, auch nichtkognitive Komponenten bedeutsam zu sein, wobei die Befundlage hier besonders spärlich ist.

Aus forschungsmethodischer Sicht weist der referierte Forschungsstand zur Bedeutung einzelner Komponenten insofern Schwächen auf, als die Bedeutung der Komponenten meist mit einem Design untersucht wurde, bei dem die diagnostische Problemlösekompetenz und die Komponenten dieser Kompetenz separat erfasst werden. Anschließend wurde die Korrelation zwischen den Konstrukten ermittelt und von der Korrelationshöhe auf die empirische Relevanz der Komponenten für die Problemlösekompetenz geschlossen. Dieses Untersuchungsdesign wurde von Wuttke et al. (2015, S. 193) kritisiert: Die psychischen Komponenten wurden in den erwähnten Studien häufig mit Items erfasst, die deutlich andere Anforderungen stellen als die diagnostischen Probleme, weshalb die Bedeutung nichtkognitiver Komponenten systematisch unterschätzt wird. Renkl (2012) wies auf ein weiteres Problem dieses Designs hin: Er betonte, dass Korrelationen zwischen Konstrukten nur unter spezifischen Bedingungen kausal interpretiert werden können. So kann auf intramentaler Ebene z. B. ein Weniger an Komponente A durch ein Mehr an Komponente B ausgeglichen werden. Es ist also mit Kompensationsmechanismen zu rechnen, weshalb von Korrelationen auf Gruppenebene nicht ohne weiteres auf intramentale Zusammenhänge geschlossen werden kann.

In diesem Kontext ist wiederum die Studie von Rausch et al. (2016) interessant, weil ihr ein anderes Untersuchungsdesign zugrunde liegt und damit die erwähnten Kritikpunkte umgangen werden. In der Studie wurden unterschiedliche Komponenten (Wissensanwendung, Handlungsregulation, Selbstkonzept und Interesse) der beruflichen Problemlösekompetenz *integrativ*, d. h. bezogen auf ein spezifisches berufliches Problem und im Problemlöseprozess, erfasst. Bei dieser integrativen Erfassung wurden zwischen nichtkognitiven und kognitiven Komponenten der beruflichen Problemlösekompetenz substantielle Korrelationen beobachtet, die allerdings deutlich unterhalb einer perfekten Korrelation lagen. Diese Befunde legen die empirische Trennbarkeit der Komponenten sowie die empirische Bedeutung kognitiver und nichtkognitiver Komponenten für die berufliche Problemlösekompetenz nahe. Außerdem verdeutlichen sie das Potenzial eines Designs, bei dem die Bedeutung der Komponenten für die diagnostische Problemlösekompetenz mit einer integrativen Erfassung der Komponenten im Problemlöseprozess untersucht wird.

#### **3.3.4 Anforderungsaspekt**

Unter dem Anforderungsaspekt ist bei der begrifflichen Klärung der *diagnostischen Problemlösekompetenz* festzulegen, auf welche externen Anforderungen sich die Kompetenz bezieht. Hier können zwei Positionen unterschieden werden: In der weiten Begriffsfassung bezieht sich *diagnostische Problemlösekompetenz* auf alle diagnostischen Probleme innerhalb eines Berufs. Praktisch bedeutet dies, dass alle berufsspezifischen Anforderungen unter den Begriff *diagnostische Problemlösekompetenz* fallen, bei denen Ursachen für einen unerwünschten Zustand zu identifizieren sind. In der engen Begriffsfassung bezieht sich *diagnostische Problemlösekompetenz* auf eine Klasse von Anforderungen innerhalb eines Berufs. In diesem Fall ist festzulegen, welche Anforderungen zu einer Klasse gehören – natürlich sind auch Zwischenlösungen, also das

Zusammenlegen von Anforderungsklassen, möglich. Deutlich wird, dass die Bedeutung und die empirische Erfassung der diagnostischen Problemlösekompetenz unter dem Anforderungsaspekt deutlich variieren kann: In der weiten Begriffsfassung steht *diagnostische Problemlösekompetenz* für jene psychische Grundlage, die zur Bewältigung aller innerhalb eines Berufs auftretenden diagnostischen Probleme benötigt wird; dementsprechend sind bei der Erfassung der diagnostischen Problemlösekompetenz möglichst alle beruflichen Anforderungen zu berücksichtigen. In der engen Fassung bezieht sich diagnostische Problemlösekompetenz auf eine Anforderungsklasse; bei der Erfassung sind also nur diagnostische Probleme dieser Anforderungsklasse einzu-beziehen. Es lässt sich weder logisch ableiten noch empirisch beweisen, welche Begriffsvariante zu bevorzugen ist. Möglich ist aber, die Varianten aus der Perspektive systematischer Überlegungen und empirischer Befunde zu beleuchten. Ergänzend sei erwähnt, dass mit dem hier angesprochenen Punkt letztlich die Frage adressiert wird, wie domänenspezifisch diagnostische Problemlösekompetenzen zu fassen sind, wobei „domänenspezifisch“ synonym zu „anforderungsspezifisch“ gebraucht wird.<sup>11</sup>

Im Übersichtsbeitrag von Vleuten et al. (2008) zur Erfassung diagnostischer Problemlösekompetenzen im medizinischen Bereich werden mehrere Erhebungsverfahren diskutiert, der Anforderungsaspekt wird jedoch nicht explizit thematisiert. Durning et al. (2015) arbeiteten in ihrem Überblicksbeitrag anhand von Befunden heraus, dass diagnostisches Problemlösen anforderungsspezifisch ist (s. dazu auch Schuwirth, 2009 und Schwartz & Elstein, 2008). Aufgrund dieser Anforderungsspezifität empfehlen die Autoren, bei der Erfassung diagnostischer Problemlösekompetenzen viele diagnostische Probleme unterschiedlicher Anforderungen zu verwenden (Durning et al., 2015, S. 135). Sherbino et al. (2012) bezogen 25 diagnostische Probleme unterschiedlicher Anforderungen in ihre Studie ein. In der Publikation von Kunina-Habenicht et al. (2015) wurde die diagnostische Problemlösekompetenz in der Anforderungsklasse „shortness of breath“ erfasst. Diese Beschränkung wurde als Grenze der Studie betrachtet und es wurde angeregt, künftig diagnostische Probleme anderer Anforderungsklassen einzubeziehen (ebd., S. 1221). Insgesamt entsteht der Eindruck, dass im medizinischen Bereich eine berufsspezifische Erfassung diagnostischer Problemlösekompetenz angestrebt wird. Aus erhebungswirtschaftlicher Sicht scheint diese Strategie günstig: Sie erlaubt es, ein einzelnes Erhebungsverfahren für unterschiedliche Anforderungsklassen zu entwickeln, was den Entwicklungsaufwand deutlich reduziert. Allerdings gibt es auch Gründe, die gegen diese Strategie sprechen: Erstens können die Befunde zur Domänenspezifität des diagnostischen Problemlösens so verstanden werden, dass Kompetenzen bezogen auf eine oder wenige verwandte Anforderungsklassen erfasst werden sollten. Aus didaktischer Sicht wäre eine solche Strategie zweitens insofern vorteilhaft, als präzise festgestellt werden könnte, was im Lernprozess gelernt wurde und in welchem Bereich noch Defizite bestehen. Eine präzise Diagnostik er-

---

11 Der Domänenbegriff ist vieldeutig. Beck (2005, S. 551) meint, der Domänenbegriff werde häufig lediglich dazu genutzt, um auf den eingeschränkten Geltungsanspruch von Lehr-Lern-Theorien hinzuweisen. Hier wird mit dem Domänenbegriff in Anlehnung an Achtenhagen (2000) ein spezifischer Anforderungsbereich bezeichnet. Präzise und wissenschaftlich bewährte Kriterien, wie berufliche Anforderungen systematisch voneinander unterschieden und Anforderungsklassen definiert werden können, konnten nicht ermittelt werden.

möglicht eine anforderungsspezifische und damit gezielte Kompetenzförderung. Schließlich kann drittens aus forschungsmethodischer Sicht argumentiert werden, dass eine Homogenisierung der Anforderungen eine reliable Kompetenzerfassung begünstigt (s. dazu Kane, 2001). Bei Kunina-Habenicht et al. (2015) waren die Reliabilitäten (interne Konsistenzen) der Kompetenzmessung eher gering ( $\alpha = .48$  und  $\Omega = .63$ ). Die Studie von Sherbino et al. (2012) zeigte, dass dieses Reliabilitätsproblem nicht automatisch durch mehr diagnostische Probleme gelöst werden kann. Im Vergleich zu Kunina-Habenicht et al. (2015) setzten die Autoren 19 diagnostische Probleme mehr ein und sie ermittelten dennoch nur eine interne Konsistenz von  $\alpha = .41$ .

Walker et al. (2016b) erfassten die diagnostische Problemlösekompetenz in der Ausbildung zum Elektroniker für Automatisierungstechnik und beschränkten sich auf diagnostische Probleme aus der Anforderungsklasse „speicherprogrammierbare Steuerung“ (SPS).<sup>12</sup> In der genannten Studie suchten Auszubildende in SPS-Programmen nach Ursachen für Störungen einer technischen Anlage. Die Problemlösescores resultierten aus schriftlichen Protokollen, die von den Auszubildenden während der Erhebung angefertigt und von Kodierer:innen anhand eines Kodierleitfadens ausgewertet wurden. Die Problemlösescores erwiesen sich als eindimensional skalierbar. Zudem erbrachten die Testwerte eine akzeptable Reliabilität (*SEM-Reliabilität* = .75). Bei Auszubildenden der Kfz-Mechatronik erfassten Nickolaus et al. (2012a) die diagnostischen Problemlösekompetenzen mit elektrotechnischen diagnostischen Problemen, die sich auf zwei unterschiedliche Kfz-Systeme bezogen (Beleuchtungsanlage und Motormanagement). Elektrotechnische Kfz-Probleme zeichnen sich dadurch aus, dass die Problemursachen elektrotechnische Defekte (z. B. defekter Aktor, defekter Sensor, Kabelunterbrechung) sind. Auch hier erwiesen sich die Problemlösescores als eindimensional skalierbar, die Reliabilität der Testwerte war allerdings unbefriedigend (*EAP/PV-Reliabilität* = .55). Bei Abele et al. (2016, S. 190) waren elektrotechnische diagnostische Probleme mehrerer Kfz-Systeme ebenfalls eindimensional skalierbar.

Die Befunde aus dem Kfz-Bereich lassen sich so interpretieren, dass Problemlösescores, die auf elektrotechnischen Problemen unterschiedlicher Kfz-Systeme beruhen, statistisch eindimensional darstellbar sind. Vor diesem Hintergrund scheint es plausibel, die diagnostische Problemlösekompetenz bei Auszubildenden der Kfz-Mechatronik begrifflich anforderungsspezifisch zu fassen. Ein empirisch relevantes Beispiel für eine solche Begriffsfassung wäre die *elektrotechnische diagnostische Problemlösekompetenz*, die sich auf alle diagnostischen Probleme im Beruf des Kfz-Mechatronikers bezieht, denen elektrotechnische Ursachen zugrunde liegen.

Wie der Forschungsstand zeigt, wird unterschiedlich mit dem Anforderungsaspekt umgegangen. Im medizinischen Bereich wird eher angestrebt, berufsspezifische diagnostische Problemlösekompetenzen zu erfassen, was vor allem aus einer erhebungsökonomischen Sicht vorteilhaft ist. In den Studien aus dem technischen Bereich werden diagnostische Problemlösekompetenzen eher anforderungsspezifisch erhoben. Mit der Lösung im technischen Bereich geht eine relativ präzise sprachliche Fassung der diagnostischen Problemlösekompetenz einher. Bei der anforderungsspezifischen

---

12 Die Autoren selbst sprechen nicht von *Anforderungsklasse*.

Erfassung wurden teilweise befriedigende Reliabilitäten ermittelt, wobei die Befundlage dünn ist.

### 3.3.5 Prozessaspekt

Unter dem Prozessaspekt repräsentiert diagnostische Problemlösekompetenz eine psychische Grundlage, die zur erfolgreichen Gestaltung des diagnostischen Problemlöseprozesses benötigt wird. Zur Beschreibung des diagnostischen Problemlöseprozesses kommen prinzipiell ein normatives Prozessmodell, ein deskriptives Prozessmodell oder eine Prozesstheorie infrage. In einem *normativen Prozessmodell* wird der diagnostische Problemlöseprozess so beschrieben, wie er sein sollte. Der *Soll-Prozess* wird typischerweise von Expert:innen des diagnostischen Problemlösens festgelegt; an den Prozess wird also eine Expertennorm angelegt. In einem *deskriptiven Prozessmodell* wird der realistische diagnostische Problemlöseprozess beschrieben. Um den realen Problemlöseprozess zu rekonstruieren, wird empirisch erfasst, wie Individuen beim Lösen diagnostischer Probleme vorgehen. Hierzu werden z. B. Personen beim Problemlösen beobachtet, Laut-Denk-Studien durchgeführt oder andere Interviewtechniken angewandt. Im Unterschied zu den zuvor genannten Modellen enthält eine *Prozesstheorie* Kausalaussagen, die sich anhand von Beobachtungen falsifizieren lassen. Mit einer Prozesstheorie wird kausal erklärt, wie der diagnostische Problemlöseerfolg zustande kommt, oder anders ausgedrückt: welche Aktivitäten erfolgreiche Problemlöser:innen zwischen der Problemwahrnehmung und der Problemlösung ausführen. Im vorliegenden Kontext ist eine Prozesstheorie zu bevorzugen: Diagnostische Problemlösekompetenzen werden als psychische Grundlage verstanden, die dem Problemlöseerfolg zugrunde liegt, d. h., die ihn verursachen. Dementsprechend beziehen sich diagnostische Problemlösekompetenzen auf die erfolgreiche Gestaltung des diagnostischen Problemlöseprozesses.

Zum diagnostischen Problemlöseprozess liegt sehr viel Forschung vor, und zwar sowohl im medizinischen als auch im technischen Bereich. Hier werden lediglich jene Grundlinien des Forschungsstands nachgezeichnet, die für die weitere Untersuchung relevant sind.

Normative Prozessmodelle werden im medizinischen Bereich vor allem in Studien verwendet, die sich mit der Erfassung des diagnostischen Problemlösens beschäftigen (z. B. Norcini & McKinley, 2007, S. 240). Das Scoring bezieht sich in diesen Studien auf den Problemlöseprozess und basiert auf der Einschätzung von Expert:innen, wie ein idealer Problemlöseprozess aussehen sollte.

Die Mehrzahl der Studien im medizinischen Bereich beruht auf deskriptiven Prozessmodellen. Elstein und Kolleg:innen kamen zum Ergebnis, dass der diagnostische Problemlöseprozess im Kern einem *hypothetico-deduktiven Vorgehen* gleicht (Elstein et al., 1990): In einer Problemlösesituation wird zunächst eine Hypothese zur Problemursache entwickelt, anschließend wird der Prozess entsprechend der Hypothese strukturiert und anhand eines Hypothesentests untersucht, ob die Hypothese zutrifft. Dieser Ansatz erwies sich als sehr einflussreich (z. B. Norman, 2005), erfuhr aber unter anderem deswegen Kritik, weil die Rolle der Erfahrung (z. B. kognitive Automatisie-

rungsprozesse) darin nicht ausreichend gewürdigt wird (Norman et al., 2007). Joseph und Patel (1990, S. 44) machten darauf aufmerksam, dass Hypothesen in Abhängigkeit von diagnoserelevanten Informationen formuliert werden, wobei diese Informationen entweder in Interaktion mit der Problemumwelt gewonnen oder in Form von Diagnosewissen aus dem Langzeitgedächtnis abgerufen werden. Weiterhin gehen sie davon aus, dass die Bewertung der Hypothese (genauer: die Bewertung der Hypothese vor dem Hintergrund der Ergebnisse des Hypothesentests) ein weiterer wichtiger Bestandteil des Problemlöseprozesses sind. Wie Barrows und Feltovich (1987) unterstellten, handelt es sich bei der Informationsgewinnung sowie dem Formulieren, Testen und Bewerten von Hypothesen um wichtige Elemente des diagnostischen Problemlöseprozesses (s. dazu auch Kassirer et al., 2010; Joseph & Patel, 1990, S. 44), wobei der Prozess in der Literatur teilweise noch differenzierter dargestellt wird. In neueren Veröffentlichungen zum diagnostischen Problemlöseprozess werden meist zwei Prozessotypen unterschieden (z. B. Croskerry, 2009; Charlin et al., 2012): Der *Typ-I-Prozess* repräsentiert ein eher intuitives und schnelles Vorgehen und basiert auf der Erkennung bekannter Muster. Der *Typ-II-Prozess* entspricht einem langsamen und analytischen Vorgehen und gleicht dem oben erwähnten hypothethico-deduktiven Vorgehen.

Rouse (1983) stellte viele Studien zum diagnostischen Problemlösen zusammen und verdeutlichte, dass auch im technischen Bereich normative und deskriptive Prozessmodelle verwendet werden. Schaafstal et al. (2000) gehen davon aus, dass beim Lösen diagnostischer Probleme Informationen zum Problem zu gewinnen, Hypothesen zu Problemursachen zu formulieren, Hypothesen zu testen und schließlich Testergebnisse zu bewerten sind. Jonassen und Hung (2006) sichteten den einschlägigen Forschungsstand und übernahmen diese Struktur weitgehend. Rasmussen (1993) modellierte den diagnostischen Problemlöseprozess deutlich differenzierter. Auch Hoc und Amalberti (1995) legten den Prozess differenzierter an, griffen dabei allerdings auf das Prozessmodell von Klahr und Dunbar (1988) zurück, in dem das Formulieren, Testen und Bewerten von Hypothesen jeweils eine zentrale Rolle spielt.

Zusammenfassend ist festzuhalten, dass viele Studien zum diagnostischen Problemlöseprozess verfügbar sind und sich ausgehend von dem dargelegten Forschungsstand folgende Sub-Prozesse des Problemlöseprozesses unterscheiden lassen: Informationsrepräsentation, Hypothesenentwicklung, Hypothesentest, Hypothesenbewertung. Deutlich gemacht hat die Sichtung des Forschungsstands aber auch, dass gegenwärtig keine Prozesstheorie verfügbar ist, in welcher der diagnostische Problemlöseerfolg anhand des diagnostischen Problemlöseprozesses kausal erklärt wird.

### 3.4 Zusammenfassung und Forschungsdesiderate

Die Überlegungen und Befunde, die in diesem Kapitel thematisiert wurden, können genutzt werden, um das Sprachmodell der diagnostischen Problemlösekompetenz zu konkretisieren. Unter diagnostischen Problemlösekompetenzen wird die von anderen beruflichen Kompetenzen abgrenzbare und metrisch skalierbare Gesamtheit kognitiver und nichtkognitiver psychischer Komponenten verstanden, die in den diagnostischen Problemlöseprozess zu investieren sind, um diagnostische Probleme einer

Anforderungsklasse zu lösen. Im Hinblick auf den Anforderungsaspekt wird hier also der engen begrifflichen Fassung gefolgt.

Einige Befunde sprechen dafür, dass diagnostische Problemlösekompetenzen empirisch von anderen beruflichen Kompetenzen abgrenzbar sind. Die empirische Basis dieser Einschätzung ist aber eher schmal und auch mit Unsicherheiten verbunden. Insbesondere im Ausbildungsberuf zum/zur Kfz-Mechatroniker:in ist die Befundlage zur empirischen Eigenständigkeit nicht eindeutig. Problematisch ist ferner, dass nur wenige Studien verfügbar sind, in denen die empirische Unterscheidbarkeit anhand mehrdimensionaler Messmodelle untersucht wurde.

Es deutet sich an, dass das Interesse und das Selbstkonzept zwei empirisch relevante nichtkognitive Komponenten der diagnostischen Problemlösekompetenzen sind. Außerdem legt der Forschungsstand nahe, das Diagnosewissen, die Arbeitsgedächtniskapazität, die schlussfolgernde Denkfähigkeit und die fluide Intelligenz als wichtige kognitive Komponenten der diagnostischen Problemlösekompetenz aufzufassen, wobei weder das nichtkognitive noch das kognitive Komponentenspektrum damit erschöpfend abgedeckt sein dürfte. Aussichtsreich scheinen vor allem Studien, in denen die empirische Bedeutung der Komponenten anhand einer integrativen Erfassung der Komponenten im Problemlöseprozess untersucht wird.

Schließlich können die Informationsrepräsentation, Hypothesenentwicklung, Hypothesentest, Hypothesenbewertung als wichtige Sub-Prozesse des diagnostischen Problemlöseprozesses gelten. Eine Theorie zum diagnostischen Problemlöseprozess, die das Zustandekommen des Problemlöseerfolgs aus einer Prozessperspektive erklärt, ist gegenwärtig nicht verfügbar.

## 4 Konkretisierung des Erhebungsmodells diagnostischer Problemlösekompetenzen

In diesem Abschnitt wird das Erhebungsmodell konkretisiert. Mit dem Erhebungsmodell wird das Sprachmodell der diagnostischen Problemlösekompetenz in Zahlenwerte übersetzt. Bei der Gestaltung dieses Übersetzungsvorgangs sind mindestens die folgenden Fragen zu beantworten: Welche Items bzw. diagnostischen Probleme sollen zur Erfassung der diagnostischen Problemlösekompetenz herangezogen werden? In welchem Modus sollen die diagnostischen Probleme vorgegeben und bearbeitet werden (z. B. computerbasiert oder papierbasiert)? Wie sollten individuelle Problemlösungen kodiert, d. h., welchen individuellen Problemlösungen sollten welche Zahlen zugeordnet werden? Die Antworten auf diese Fragen hängen davon ab, welches Ziel mit der Konkretisierung des Erhebungsmodells angestrebt wird.

## 4.1 Konkretisierungsziel und Konkretisierungskriterien

### 4.1.1 Konkretisierungsziel: Valide Aussagen über diagnostische Problemlösekompetenzen

Hier ist intendiert, das Erhebungsmodell so zu konkretisieren, dass es zur validen Erfassung der diagnostischen Problemlösekompetenz bzw. validen sprachlichen Aussagen über die diagnostische Problemlösekompetenz beiträgt.<sup>13</sup> Validität wird gemeinhin als wichtigstes Ziel der Erfassung psychischer Merkmale wie z. B. diagnostischer Problemlösekompetenzen betrachtet (American Educational Research Association, American Psychological Association, national Council on Measurement in Education, 2014, S. 11).

In welchen Situationen es gerechtfertigt ist, von einer validen Erfassung einer diagnostischen Problemlösekompetenz zu sprechen, hängt davon ab, welcher Validitätsansatz verwendet wird. Im Horizont des traditionellen Validitätsansatzes sind Aussagen über die diagnostische Problemlösekompetenz dann valide, wenn sie mit einem validen Kompetenztest gewonnen werden. Von einem validen Test wird dann gesprochen, wenn der Test das Merkmal misst, das er messen soll (Moosbrugger & Kelava, 2008, S. 13). Dieser traditionelle Ansatz wird dem aktuellen Stand der Forschung allerdings nicht mehr gerecht (American Educational Research Association, American Psychological Association, national Council on Measurement in Education, 2014, S. 11–22). In der aktuellen Validitätsforschung wird ein argumentbasierter Validitätsansatz bevorzugt (Kane, 2013). Bei einem argumentbasierten Validitätsansatz wird Validität nicht als Eigenschaft eines Tests, sondern einer Testwertinterpretation verstanden. Wie die Ausführungen in Kapitel 2.1 deutlich machen, können sprachliche Aussagen über die diagnostische Problemlösekompetenz als Testwertinterpretationen verstanden werden. Der argumentbasierte Ansatz wird nachfolgend in Anlehnung an Kane (2013) skizziert.

### 4.1.2 Argumentbasierter Validitätsansatz, oder: wann darf von validen Aussagen gesprochen werden?

Im Rahmen eines argumentbasierten Validitätsansatzes liegen dann valide Aussagen über diagnostische Problemlösekompetenzen vor, wenn die Annahmen, auf denen die sprachlichen Aussagen (also die Testwertinterpretationen) beruhen, theoretisch angemessen und empirisch haltbar sind. Im vorliegenden Kontext besteht die Grundannahme darin, dass die Testwerte, die zur Erfassung der diagnostischen Problemlösekompetenz gewonnen werden, auch im Sinne dieser Kompetenz interpretiert werden dürfen. Mit Blick auf die Ausführungen zum Sprachmodell bedeutet dies, dass es gerechtfertigt sein muss, Testwerte zur diagnostischen Problemlösekompetenz als Indikatoren einer *psychischen Größe* zu interpretieren, die spezifische psychische Kom-

---

<sup>13</sup> Es sei erwähnt, dass eine valide Erfassung der diagnostischen Problemlösekompetenz valide sprachliche Aussagen über die diagnostische Problemlösekompetenz erbringt. Aufgrund dieses Zusammenhangs können beide Ausdrücke oft ohne Bedeutungsänderung gegeneinander ausgetauscht werden. Genau genommen bezieht sich eine valide Erfassung aber auf den Modellierungsprozess, aus dem die Aussagen resultieren, und valide Aussagen beziehen sich auf das Resultat dieses Modellierungsprozesses.

ponenten umfasst, sich auf einen spezifischen Prozess bezieht, metrisch skalierbar ist etc. Bei differenzierter Betrachtung beruhen Testwertinterpretationen auf vielen Einzelannahmen. Angenommen werden muss z. B., dass die Testwerte anhand geeigneter Items bzw. diagnostischer Probleme und einer geeigneten Kodierung zustande kamen. In einem Validitätsargument werden Belege zur Haltbarkeit dieser Einzelannahmen zusammengefasst. Mithilfe des Validitätsarguments kann entschieden werden, ob es gerechtfertigt ist, Testwerte auf eine spezifische Weise zu interpretieren. Das Validitätsargument bietet eine rationale Grundlage, die Validität von Testwertinterpretationen bzw. von sprachlichen Aussagen über die diagnostische Problemlösekompetenz zu beurteilen.

Validität ist keine Eigenschaft, die Aussagen zukommt oder nicht zukommt. Aussagen über die diagnostische Problemlösekompetenz können sich vielmehr graduell hinsichtlich ihrer Validität unterscheiden. Falls sich allerdings eine zentrale Annahme als nicht haltbar erweist, kann dies die gesamte Aussagenvalidität ernsthaft gefährden. In einem solchen Fall ist es auch nicht hilfreich, wenn andere Annahmen besonders gut fundiert sind. Da neue Erkenntnisse zur Revision des Validitätsarguments führen können, hat ein Validitätsurteil immer nur einen vorläufigen Charakter.

Prinzipiell sollte ein Validitätsargument alle Annahmen berücksichtigen, auf denen die Aussagen über die betreffende diagnostische Problemlösekompetenz basieren. Da solche Aussagen je nach Auflösungsgrad aber auf einer großen Anzahl an Annahmen basieren können, ist aus pragmatischen Gründen meist eine Konzentration auf ausgewählte Annahmen erforderlich. Ansonsten wären Validitätsurteile in vielerlei Hinsicht nicht möglich. Brennan (2001) empfiehlt, vor allem jene Fragen zu adressieren, die auf besonders wichtige und anfällige Annahmen des Validitätsarguments verweisen. Eine ähnliche Strategie schlägt Kane (2013) vor: Ihm zufolge sind zumindest jene Annahmen zu prüfen, die für die Validität der Aussagen besonders wichtig sind und ein besonders großes Validitätsrisiko darstellen. Falls sich diese kritischen Annahmen als haltbar erweisen, ist es gerechtfertigt, von validen Aussagen über die diagnostische Problemlösekompetenz zu sprechen.

### **4.1.3 Konkretisierungskriterien: Valide Kodierungen, Generalisierbarkeit und Übertragbarkeit**

Vor diesem Hintergrund ist bei der Konkretisierung des Erhebungsmodells mindestens darauf zu achten, dass jene Annahmen des Erhebungsmodells erfüllt sind, die besonders zentral sind und die die Aussagenvalidität besonders stark gefährden. Darin besteht im vorliegenden Kontext das übergeordnete Konkretisierungskriterium des Erhebungsmodells. Dieses Kriterium kann in Anlehnung an Kane (1992, S. 165) spezifiziert und in Unterkriterien ausdifferenziert werden: Bei der Gestaltung des Erhebungsmodells ist insbesondere darauf zu achten, dass das Erhebungsmodell erstens eine valide Kodierung individueller Problemlösungen umfasst und zweitens generalisierbare Aussagen über die diagnostische Problemlösekompetenz begünstigt. Schließlich sollte es drittens zentrale Voraussetzungen dafür schaffen, dass die Aussagen über die diagnostische Problemlösekompetenz auch auf reale berufliche Kontexte übertrag-

bar sind. Falls das Erhebungsmodell dem Kodierungskriterium, dem Generalisierbarkeitskriterium und dem Übertragbarkeitskriterium entspricht, sind wesentliche Voraussetzungen für ein belastbares Validitätsargument, d. h. für valide Aussagen über die diagnostische Problemlösekompetenz, geschaffen.

## 4.2 Konkretisierung des Erhebungsmodells unter den Validitätskriterien

In diesem Abschnitt wird anhand systematischer Überlegungen und empirischer Befunde untersucht, wie das Erhebungsmodell im Sinne der genannten Kriterien zu konkretisieren ist.

### 4.2.1 Valide Kodierung des individuellen Problemlöseerfolgs

Unter dem Kodierungskriterium ist das Erhebungsmodell so zu gestalten, dass es valide Kodierungen beinhaltet. Mit *Kodierung* ist ein Vorgang angesprochen, bei dem ein erfahrbarer Sachverhalt mit einem Zahlenwert versehen wird. *Erfahrbarer Sachverhalt* meint hier die diagnostische Problemlösekompetenz oder genauer, die Indikatoren dieser Kompetenz, also den diagnostischen Problemlöseerfolg bzw. das diesen Erfolg anzeigende diagnostische Problemlöseverhalten.<sup>14</sup> Das diagnostische Problemlöseverhalten ist der Gegenstand der Kodierung und ermöglicht die Rekonstruktion der diagnostischen Problemlösekompetenz.<sup>15</sup> Bei der Kodierung wird das Problemlöseverhalten mit einem Zahlenwert versehen und in seiner individuellen Ausprägung formal repräsentiert. Die Kodierung ist valide, wenn der Zahlenwert das repräsentiert, was er repräsentieren soll, nämlich den Problemlöseerfolg bzw. die Qualität des diagnostischen Problemlöseverhaltens eines spezifischen Individuums.

In Anlehnung an Beck (1987, S. 38 ff.) können zwei Kodierungstypen unterschieden werden: Beim ersten Typ sind Beobachter:innen ein unverzichtbarer Bestandteil der Kodierung. Dies ist z. B. der Fall, wenn das Problemlöseverhalten einzelner Personen im Berufsalltag beobachtet und von Beobachter:innen global eingeschätzt bzw. kodiert, also mit Zahlenwerten versehen wird. Beim zweiten Kodierungstyp kann prinzipiell auf Beobachter:innen verzichtet werden. Bei diesem Typ arbeiten Proband:innen z. B. an vorgegebenen diagnostischen Problemen; das dabei gezeigte Verhalten wird in objektiver Form festgehalten (schriftliche Dokumentationen, Log-File-Dateien etc.) und regelgeleitet kodiert. Auch wenn das objektivierte Probandenverhalten in der Erhebungspraxis oft von Personen kodiert wird, ist eine maschinelle Kodierung prinzipiell möglich. Mit dem zweiten Kodierungstyp ist der Anspruch verbunden, bei der Kodierung lediglich klar definierte Regeln anzuwenden und keine semantischen Transformationen durchführen zu müssen. Daran ändert auch die Tatsache nichts, dass gegenwärtig zur Erfassung diagnostischer Problemlösekompetenzen eher selten Kodiermanuale mit eindeutigen Kodierregeln verfügbar sind. Im Lichte der Ausfüh-

14 Die diagnostische Problemlösekompetenz ist als latentes Merkmal nicht direkt, sondern nur indirekt erfahrbar (z. B. Beck, 1987, S. 41). Direkt erfahrbar (d. h. beobachtbar) ist aber der diagnostische Problemlöseerfolg bzw. das individuelle Problemlöseverhalten, das diesen Erfolg anzeigt. Der Problemlöseerfolg wird hier per definitionem als Indikator der diagnostischen Problemlösekompetenz aufgefasst. Dies bedeutet, dass die diagnostische Problemlösekompetenz aus dem Problemlöseerfolg rekonstruiert werden kann und damit ein indirekt erfahrbarer Sachverhalt ist.

15 Genau betrachtet ist oft nicht das individuelle Problemlöseverhalten Gegenstand der Kodierung, sondern das in schriftlicher Form von den Problemlösern festgehaltene Problemlöseverhalten und Problemlöseergebnis.

rungen von Beck (1987, S. 181) zeigt dies einen Forschungsbedarf an und es würde zu kurz greifen, darin die prinzipielle Notwendigkeit zu sehen, Beobachter:innen in die Kodierung des Problemlöseerfolgs einzubeziehen. Wie sind die beiden Kodierungstypen aus Validitätssicht zu beurteilen?

### Beobachtungsgebundene Kodierung

Zur Erfassung der diagnostischen Problemlösekompetenz werden unterschiedliche Erhebungsverfahren eingesetzt, bei denen Beobachter:innen für die Kodierung unverzichtbar sind. Bei den *objective structured clinical examinations* bearbeiten Teilnehmende an mehreren Stationen in einer authentisch simulierten Berufsumwelt wenig komplexe diagnostische Probleme. Das dabei gezeigte Problemlöseverhalten wird typischerweise von einem/einer Beobachter:in oder mehreren Beobachter:innen global kodiert (Ilgen et al., 2012, S. 1457). Dieses Verfahren weist große Ähnlichkeit mit den praktischen Abschlussprüfungen im Kfz-Handwerk auf (Abele, 2011), bei denen Auszubildende an mehreren berufsauthentischen Prüfungsstationen spezifische diagnostische Probleme bearbeiten und das gezeigte Problemlöseverhalten von Beobachter:innen relativ global kodiert wird. Durchaus gängig sind auch Verfahren, bei denen Proband:innen im Berufsalltag und Arbeitsprozess beobachtet und das individuelle Problemlöseverhalten von Beobachter:innen kodiert wird (z. B. *mini-clinical evaluation exercises*, Norcini & McKinley, 2007, S. 246). Beobachtungsgebundene Kodierungen kommen auch beim 360° Feedback von Vorgesetzten oder dem *multisource feedback* zum Einsatz, bei dem das individuelle Problemlöseverhalten von unterschiedlichen Personen (Patient:innen, Kolleg:innen, Pflegepersonal etc.) kodiert wird (Ilgen et al., 2012, S. 1457). Im medizinischen Bereich scheint ein Trend dahin zu gehen, bei der Erfassung diagnostischer Problemlösekompetenzen verstärkt auf Beobachtungen am Arbeitsplatz und globale Kodierungen zu setzen (Tigelaar & van der Vleuten, 2014, S. 1246).

Um die Validität beobachtungsgebundener Kodierungen beurteilen zu können, ist zunächst festzuhalten, dass bei der beobachtungsgebundenen Kodierung zumindest zwei Schritte unterschieden werden können: Erstens wird ein erfahrbarer Sachverhalt erfasst<sup>16</sup> und zweitens das Erfasste mit einem Zahlenwert versehen. Sowohl (1) die Erfassung des realen Sachverhalts als auch (2) die Zuordnung eines Zahlenwerts zum Erfassten sind fehleranfällig.

1. Bei der Erfassung des Sachverhalts können mindestens zwei Fehlerarten auftreten: Erstens kann eine *verfälschte Wahrnehmung* (z. B. verursacht durch einen Vorwissensfehler, Beck, 1987, S. 185) dazu führen, dass das Problemlöseverhalten verzerrt erfasst wird. Zweitens ist denkbar, dass reale Sachverhalte erfasst werden, die nicht unter den Begriff *diagnostische Problemlösekompetenz* fallen. Mit Blick auf das im vorigen Kapitel entwickelte Sprachmodell wäre das z. B. der Fall,

<sup>16</sup> Für die nachfolgende Argumentation ist unwichtig, ob der Sachverhalt gedanklich oder sprachlich erfasst wird. Damit ist es auch nicht notwendig, auf das Basisproblem und die Frage einzugehen, wie Beobachtungsergebnisse sprachlich festgehalten werden bzw. festzuhalten sind (Stegmüller, 1989, S. 445 ff.). Mit der Verpflichtung auf das Paradigma des Kritischen Rationalismus ist ohnehin die Auffassung verbunden, dass Beobachtungen sprachlich in Basissätzen festgehalten werden, die letztlich auf Konventionen beruhen.

wenn nicht nur erfasst wird, ob ein diagnostisches Problem gelöst, sondern auch, ob bei der Problembearbeitung freundlich kommuniziert oder gelächelt wurde. Bei dieser Fehlerart bleibt der reale Bezugspunkt, d. h. die Bedeutung der Kodierung unklar. Im Extremfall können Kodierungen durchgängig und ausschließlich auf (im Lichte des Sprachmodells) irrelevanten Sachverhalten beruhen. Auch in diesem Extremfall würde die Kodierung Zahlenwerte hervorbringen, die statistisch verarbeitet werden können. Daraus resultierende Aussagen können aber niemals valide sein, d. h. die definierte diagnostische Problemlösekompetenz abbilden. Dies gilt auch dann, wenn die Zahlenwerte mit hochkomplexen und anspruchsvollen statistischen Verfahren weiterverarbeitet werden: Der erwähnte Kodierungsfehler erzeugt irreparable Schäden, die aber leicht unbemerkt bleiben.

2. Bei der Zuordnung eines Zahlenwerts zum Erfassten besteht eine Fehlerquelle erstens darin, dass bei der Zuordnung des Zahlenwerts nicht nur das berücksichtigt wird, was erfasst wurde, sondern auch irrelevante Informationen einfließen (z. B. infolge eines „Hofeffekts“, Beck, 1987, S. 185). Fehlerhafte Zuordnungen können zweitens dadurch zustande kommen, dass die Zuordnung nicht nach fachlich angemessenen Kriterien und auch nicht konsistent erfolgt. Dies führt insofern zu einer fehlerhaften Kodierung, als der Zahlenwert nicht mehr oder nur noch vage kennzeichnet, ob eine Person ein adäquates Problemlöseverhalten zeigt, d. h. in letzter Konsequenz, ob sie über diagnostische Problemlösekompetenz verfügt.

Vor diesem Hintergrund ist festzuhalten, dass beobachtungsgebundene Kodierungen aus systematischen Gründen ernstzunehmende Validitätsrisiken aufweisen: So muss prinzipiell damit gerechnet werden, dass diese Art der Kodierung durch subjektive Einflüsse verfälscht wird, und zwar sowohl bei der Erfassung des diagnostischen Problemlöseverhaltens als auch bei der Zuordnung von Zahlen zum Erfassten. Nicht auszuschließen ist ferner, dass Beobachter:innen (auch) solche Sachverhalte in den Blick nehmen, die eben nicht unter diagnostische Problemlösekompetenz fallen. Schließlich bleibt intransparent, nach welchen fachlichen Kriterien und wie konsistent das Erfasste kodiert wurde. Häufig wird argumentiert, die genannten Probleme seien zu lösen, indem Expert:innen als Kodierer:innen eingesetzt oder Beobachtungsschulungen durchgeführt werden (z. B. Tigelaar & van der Vleuten, 2014, S. 1246). Als empirischer Beleg dafür werden dann hohe Beobachtungsübereinstimmungen oder reliable Testwerte angeführt (ebd.). Darauf ist erstens zu antworten, dass sich die Erfassung diagnostischer Problemlösekompetenzen mit Beobachter:innen bislang nicht selten als wenig reliabel erwies (z. B. McGill et al., 2011). Zweitens sichern weder hohe Beobachtungsübereinstimmungen noch hohe Reliabilitäten valide Kodierungen ab: Es ist durchaus denkbar, dass Beobachtungsschulungen zu ähnlichen „Beobachtungskonzepten“ und hohen Beobachtungsübereinstimmungen führen. Ob sich die Kodierungen aber tatsächlich auf das relevante Problemlöseverhalten und damit die diagnostische Problemlösekompetenz beziehen, bleibt unklar.

Es kann aus wissenschaftlicher Sicht nicht befriedigen, wenn bei der Gestaltung des Erhebungsmodells undurchsichtig bleibt, ob wesentliche Voraussetzungen valider Aussagen erfüllt sind. Dementsprechend ist von beobachtungsgebundenen Kodierungen aus Validitätsgründen abzusehen. Bei Beck (1987) findet sich folgende prinzipielle Anmerkung zur Rolle von Beobachter:innen bei der Erfassung bzw. Messung empirischer Merkmale: „Die Unbestimmtheit und Instabilität des Menschen als ‚Meßgerät‘ steht dem Vollzug eindeutiger, wiederholbarer und vergleichbarer Tatsachenfeststellungen, die unter der Idee der Formulierung allgemeingültiger Aussagen unverzichtbar sind, im Wege. Aus diesem Grund müssen unsere [...] Erwägungen auch als *Plädoyer für die Elimination des [...] Beobachters* aus dem Vorgang der Meßwertproduktion verstanden werden“ (S. 181).

### **Kodierungen, die nicht an Beobachter:innen gebunden sind**

Zur Erfassung der diagnostischen Problemlösekompetenz wurden auch Erhebungsverfahren entwickelt, bei denen Beobachter:innen grundsätzlich verzichtbar sind. So liegen im medizinischen Bereich etwa mehrere *Multiple-Choice-Tests* zur Erfassung der diagnostischen Problemlösekompetenz vor (z. B. Durning et al., 2015, S. 135 ff.; Ilgen et al., 2012, S. 1456). In diesen Tests werden typischerweise kurze medizinische Problemfälle präsentiert und mehrere Fragen zu diesen Problemfällen gestellt. Zur Beantwortung der Fragen ist die Korrekte aus mehreren gegebenen Antworten auszuwählen. Bei *Multiple-Choice-Tests* sind die oben genannten Validitätsrisiken deshalb deutlich minimiert, weil durch die Vorgabe eines spezifischen Diagnosefalls und einer spezifischen Frage erstens klar ist, auf welchen Sachverhalt sich die Kodierung bezieht. Zweitens ist durch die Wahl einer vorgegebenen Antwort durch die Testperson das Risiko minimiert, dass das Problemlöseverhalten verzerrt erfasst wird. Drittens ist durch die Konstruktion des Tests eindeutig festgelegt, welche Antworten korrekt sind, d. h. die Zuordnung von Zahlen zum Probandenverhalten ist eindeutig und konsistent geregelt.

Ein weiteres Verfahren stellt die Administration von *Key-Feature-Problemen* dar (z. B. Vleuten et al., 2008, S. 416). *Key-Feature-Probleme* beziehen sich ebenfalls auf spezifische klinische Problemfälle, werden im Unterschied zu *Multiple-Choice-Items* aber so konstruiert, dass sie besonders kritische Teilanforderungen des diagnostischen Problemlösens abdecken. *Key-Feature-Probleme* werden üblicherweise durch die Auswahl vorgegebener Antworten oder anhand eines kurzen Texts beantwortet. Durning et al. (2015) thematisieren weitere papierbasierte Tests zur Erfassung der diagnostischen Problemlösekompetenz.

Im technischen Bereich werden zur Erfassung diagnostischer Problemlösekompetenzen unter anderem *Computersimulationen* eingesetzt, in denen Testpersonen spezifische diagnostische Probleme in einer authentisch simulierten Berufsumwelt bearbeiten (z. B. Gschwendtner et al., 2009; Nickolaus et al., 2012a; Walker et al., 2016a). Bei diesen computerbasierten Verfahren halten die Testpersonen während der Problemlösung relevante Problemlöseschritte sowie die konkrete Problemursache schriftlich fest. Die schriftlichen Dokumentationen werden unter Anwendung eines Kodier-

manuals kodiert. Für diese computerbasierten Verfahren und den Key-Feature-Ansatz gilt dasselbe wie für die erwähnten Multiple-Choice-Tests: Da Beobachter:innen eine verzichtbare Rolle spielen, sind sie nicht prinzipiell mit den oben erwähnten Validitätsrisiken verbunden. Im Weiteren werden nur noch diese drei Verfahren berücksichtigt.

#### 4.2.2 Generalisierbarkeit der Aussagen über die diagnostische Problemlösekompetenz

Unter dem Generalisierbarkeitskriterium ist das Erhebungsmodell so zu gestalten, dass anhand der ausgewählten diagnostischen Probleme Aussagen über die diagnostische Problemlösekompetenz möglich sind. Aussagen über die diagnostische Problemlösekompetenz werden interpretativ aus Testwerten gewonnen, die auf einer Auswahl diagnostischer Probleme beruhen. Aussagen über die diagnostische Problemlösekompetenz sollten allerdings nicht nur Auskunft darüber geben, wie erfolgreich eine Person eine Auswahl diagnostischer Probleme lösen kann, sondern vielmehr darüber, wie erfolgreich eine Person alle diagnostischen Probleme lösen kann, die unter diagnostische Problemlösekompetenz fallen. Eine zentrale Voraussetzung dieser Generalisierbarkeit ist, dass die ausgewählten Probleme die diagnostische Problemlösekompetenz repräsentativ abbilden. Im Lichte des hier entwickelten Sprachmodells bedeutet dies, dass die Problemauswahl spezifische Anforderungen, spezifische psychische Komponenten sowie einen spezifischen diagnostischen Problemlöseprozess abdecken sollte. Ob von einer Auswahl diagnostischer Probleme auf die diagnostische Problemlösekompetenz generalisiert werden kann, kann in Anlehnung an die Generalisierbarkeitstheorie (z. B. Brennan, 2001) unter anderem anhand der Reliabilität (genauer: der internen Konsistenz von Testwerten) beurteilt werden.<sup>17</sup> Aussagekräftig dürfte die interne Konsistenz i. d. R. aber nur dann sein, wenn die Problemauswahl auch aus Anforderungssicht als repräsentativ gelten kann.

#### Verwendung holistischer diagnostischer Probleme

Aus anforderungsbezogenen Gründen scheint der Einsatz *holistischer diagnostischer Probleme* (kurz: diagnostischer Probleme) besonders günstig für die Generalisierbarkeit. Holistisch meint hier, dass die Probleme den gesamten Prozess von der Problemwahrnehmung (Identifikation des unerwünschten Zustands) bis zur Problemlösung (Benennung der Ursache für den unerwünschten Zustand) abdecken. Holistische diagnostische Probleme dürften deshalb für die Validität günstig sein, weil mit ihnen erfasst werden kann, ob Individuen den gesamten Problemlöseprozess unter Investition der relevanten psychischen Komponenten (Ressourcen) erfolgreich gestalten können.

---

<sup>17</sup> Aus der Perspektive der Generalisierbarkeitstheorie ist die Grundlage dieser Entscheidung allerdings kompliziert. Die interne Konsistenz von Testwerten ist nicht nur eine Funktion der Problemauswahl, sondern auch davon, wie die Testwerte und die Konsistenz berechnet (statistische Modellierung) und welche Testpersonen einbezogen werden (Robitzsch et al., 2011). Erschwerend kommt hinzu, dass aus einer hohen internen Konsistenz nicht logisch zwingend geschlossen werden kann, dass die Testwerte die diagnostische Problemlösekompetenz repräsentativ abbilden. Es ist problemlos denkbar, dass die Testwerte insbesondere eine spezifische Kompetenzfacette abbilden und gerade deshalb (besonders) konsistent sind. Diese Schwierigkeit ließe sich nur umgehen, wenn die Problemauswahl aus der zufälligen Ziehung möglichst vieler diagnostischer Probleme aus einem vorab festgelegten (unendlich oder zumindest sehr großen) Problemuniversum resultiert. Da dies praktisch nur schwer einlösbar sein dürfte, ist es i. d. R. unerlässlich, auch inhaltliche bzw. anforderungsbezogene Gründe anzuführen, warum eine Problemauswahl für repräsentativ gehalten wird.

Solche Probleme gewährleisten also eine gute Repräsentation des Zielkonstrukts *diagnostische Problemlösekompetenz*.

In den oben genannten computerbasierten Verfahren aus dem technischen Bereich wurden holistische diagnostische Probleme verwendet. Bei der Analyse dieser Verfahren wurden unterschiedliche interne Konsistenzen ermittelt. Walker et al. (2016a) berechneten im Ausbildungsberuf „Elektroniker:innen für Automatisierungstechnik“ bei der Erfassung der diagnostischen Problemlösekompetenz eine interne Konsistenz von .75 (SEM-Reliabilität<sup>18</sup>). Bei Nickolaus et al. (2012b) wiesen die Testwerte bei Elektroniker:innen für Energie- und Gebäudetechnik eine Reliabilität von .67 (EAP/PV-Reliabilität<sup>19</sup>) und bei Kfz-Mechatroniker:innen eine Reliabilität von .55 (EAP/PV-Reliabilität) auf. Die anhand anforderungsbezogener Überlegungen gewonnene Einschätzung wird von den Konsistenzmaßen also nicht (vollständig) bestätigt: Obwohl holistische Probleme eingesetzt wurden, sind die internen Konsistenzen nicht durchgängig befriedigend. Dies kann mindestens die folgenden Ursachen haben: Erstens werden unterschiedliche Reliabilitätsmaße verwendet und es ist denkbar, dass bei durchgängiger Bestimmung der SEM-Reliabilität vergleichbare Konsistenzen ermittelt worden wären. Zweitens ist es möglich, dass die verwendeten Probleme unterschiedlich heterogene externe Anforderungen stellen. Auffällig ist drittens, dass in den genannten Studien nur wenige diagnostische Probleme eingesetzt wurden. Dies kann erhebliche negative Konsequenzen für die interne Konsistenz haben (z. B. Vleuten & Schuwirth, 2005, S. 311). In den Studien wurden deshalb wenige Probleme eingesetzt, weil holistische diagnostische Probleme hohe Bearbeitungszeiten verursachen.

### Verwendung von Key-Feature-Problemen und Multiple-Choice-Items

Im Hinblick auf die Bearbeitungszeit scheinen sowohl Key-Feature-Probleme als auch Multiple-Choice-Items vergleichsweise günstig zu sein. Dabei ist unter dem Generalisierungsaspekt allerdings darauf zu achten, dass die Key-Feature-Probleme bzw. Multiple-Choice-Items jeweils in ihrer Gesamtheit den kompletten diagnostischen Problemlöseprozess und die relevanten psychischen Komponenten abbilden. Aber selbst wenn die Gesamtheit aller Multiple-Choice-Items und Key-Feature-Probleme die relevanten psychischen Komponenten und den gesamten Problemlöseprozess abdeckt, besteht das prinzipielle Risiko, einen wichtigen Aspekt von diagnostischer Problemlösekompetenz zu vernachlässigen (Messick, 1994). Es ist durchaus plausibel, dass sich die diagnostische Problemlösekompetenz vor allem auch darin zeigt, ob eine Person die relevanten psychischen Komponenten koordiniert in den gesamten Problemlöseprozess einbringen kann. Dies kann weder mit einzelnen Multiple-Choice-Items noch mit Key-Feature-Problemen direkt erfasst werden.

Wenige Zweifel bestehen dahin gehend, dass *Multiple-Choice-Tests* häufig zu (sehr) hohen internen Konsistenzen führen (z. B. Durning et al., 2015). Swanson et al. (1987, S. 233) resümierten in ihrem Überblicksartikel, dass bei einer Erhebungszeit

18 Ein Konsistenzmaß, das auf Basis von Strukturgleichungsmodellen (*structural equation modeling*: SEM) gewonnen wird.

19 Ein Konsistenzmaß, das auf Basis des Raschmodells gewonnen wird.

von 2–3 Stunden und der Vorgabe von ca. 100 Multiple-Choice-Items üblicherweise interne Konsistenzen größer  $\alpha = .80$  erreicht werden.

Die von Hatala und Norman (2002) eingesetzten 15 *Key-Feature-Probleme* erbrachten bei einer Testzeit von zwei Stunden lediglich eine Reliabilität von  $\alpha = .49$  (Cronbachs Alpha). Fischer et al. (2005) erreichten bei einer Testzeit von 1,5 Stunden und 15 *Key-Feature-Problemen* eine höhere Reliabilität von  $\alpha = .65$  (Cronbachs Alpha). In den Studien wurden *Key-Feature-Probleme* aus dem Anforderungsbereich *innere Medizin* vorgegeben. Wie die Übersicht von Fischer et al. (2005) zeigt, umfasst die innere Medizin ein weites Feld und sie bezieht sich auf unterschiedliche Subsysteme des menschlichen Körpers. Es ist gut möglich, dass die interne Konsistenz höher ausgefallen wäre, wenn *Key-Feature-Probleme* eines enger umrissenen Anforderungsbereichs (einer Anforderungsklasse) eingesetzt worden wären. Als problematisch erweist sich ferner, dass bislang meist eine eher geringe Anzahl an *Key-Feature-Problemen* verwendet wurde. Dies ist auf die Bearbeitungszeit der *Key-Feature-Probleme* zurückzuführen, die in den bisherigen Studien im Vergleich zu Multiple-Choice-Items deutlich länger war. Prinzipiell besteht allerdings die Möglichkeit, *Key-Feature-Probleme* zu konstruieren, die in kürzeren Zeiten zu bearbeiten sind. In diesem Fall scheinen deutlich höhere interne Konsistenzen möglich (z. B. Vleuten & Schuwirth, 2005, S. 311).

Zusammenfassend ist festzuhalten, dass keiner der genannten Item- bzw. Problemtypen im Hinblick auf die Generalisierbarkeit restlos überzeugt. Aus anforderungsbezogenen Gründen weisen holistische diagnostische Probleme Vorteile auf. Wie insbesondere die Befunde aus dem Kfz-Bereich dokumentieren, können solche Probleme aber aufgrund langer Bearbeitungszeiten mit unbefriedigenden internen Konsistenzen einhergehen. Dies weist vermutlich darauf hin, dass die Problemauswahl die externen Anforderungen der diagnostischen Problemlösekompetenz nicht repräsentativ abbildet. Mit Multiple-Choice-Items können hohe interne Konsistenzen erzielt werden. *Key-Feature-Probleme* erbrachten bislang keine vollständig überzeugende interne Konsistenz, was vermutlich auf die relativ hohen Bearbeitungszeiten und dadurch bedingte geringe Anzahl an *Key-Feature-Problemen* bei der Testung zurückzuführen ist. *Key-Feature-Probleme* können prinzipiell allerdings so konstruiert werden, dass sie weniger Bearbeitungszeit erfordern. Sowohl *Key-Feature-Probleme* als auch Multiple-Choice-Items weisen aber ein prinzipielles Generalisierbarkeitsrisiko auf: Im Vergleich mit holistischen Problemen können sie nicht direkt abdecken, ob der gesamte Problemlöseprozess unter Investition relevanter psychischer Komponenten (Ressourcen) erfolgreich gestaltet werden kann.

### 4.2.3 Übertragbarkeit der Aussagen auf berufliche Kontexte

Unter dem Übertragbarkeitskriterium ist das Erhebungsmodell so zu konkretisieren, dass die Aussagen über die diagnostische Problemlösekompetenz auf reale berufliche Kontexte übertragbar sind. Falls einer Person in einer Erhebung eine hohe diagnostische Problemlösekompetenz bescheinigt wird, sollte sie auch im Berufsalltag viele entsprechende Probleme lösen können. Authentizität wird gemeinhin als wichtige Voraussetzung dafür gehalten, dass Aussagen über die diagnostische Problemlöse-

kompetenz auf den beruflichen Kontext übertragbar sind (Messick, 1994, S. 17 f.). Von *Authentizität* wird gesprochen, wenn zwischen den Anforderungen der Erhebung und den realen beruflichen Anforderungen eine große Ähnlichkeit besteht. Zusätzlich zu diesen Ähnlichkeitsüberlegungen werden oft statistische Analysen durchgeführt, um die Übertragbarkeit besser einschätzen zu können. Hierbei wird üblicherweise analysiert, wie gut sich anhand der Testwerte aus der Erhebung prognostizieren lässt, wie erfolgreich Probleme in realen beruflichen Kontexten gelöst werden.

### **Authentizität und Prognosekraft von Multiple-Choice-Items und Key-Feature-Problemen**

Multiple-Choice-Items aus dem medizinischen Bereich beinhalten i. d. R. kurze Fragen zu kleineren klinischen Problemfällen. Diese kurzen Fragen beziehen sich nur auf einen (sehr) eng umrissenen Aspekt eines diagnostischen Problems. Ilgen et al. (2012, S. 1456) machten deutlich, dass mit Multiple-Choice-Items prinzipiell *nicht* abgeprüft werden kann, ob Proband:innen eigenständig Ursachenhypothesen zu gesundheitlichen Beschwerden formulieren können. Dies hat mit dem geschlossenen Antwortformat zu tun. Zudem können in Multiple-Choice-Tests reale berufliche Anforderungen nur bedingt abgebildet werden. Trotz dieser Authentizitätsdefizite wird in Überblicksartikeln oft unterstellt, dass berufliche Leistungen mit Multiple-Choice-Items relativ gut prognostiziert werden können (z. B. Ilgen et al., 2012, S. 1456; Durning et al., 2015). Bei Tamblyn et al. (2007, S. 997) korrelierten Multiple-Choice-Testwerte gering bis moderat mit klinischen Kompetenzen ( $r = .10-.36$ ), bei Wenghofer et al. (2009) erwiesen sich die Ergebnisse eines schriftlichen Tests als bedeutsame Prädiktoren der späteren klinischen Kompetenz. Der letztgenannte Test bestand hauptsächlich aus Multiple-Choice-Items, enthielt aber auch Key-Feature-Probleme.

Key-Feature-Probleme werden als eine relativ authentische Erhebungsvariante eingestuft (z. B. Durning et al., 2015, S. 142). Mit Key-Feature-Problemen lassen sich prinzipiell alle Anforderungen eines Problems abdecken. So können sich Key-Feature-Probleme im Unterschied zu Multiple-Choice-Items auch auf die Formulierung von Ursachenhypothesen beziehen. Key-Feature-Probleme lassen sich sowohl mit einem geschlossenen als auch mit einem offenen Antwortformat darbieten. Eine größere Authentizität als Multiple-Choice-Items erreichen Key-Feature-Probleme auch deshalb, weil sie meist ausgehend von holistischen Problemen entwickelt werden und solche Anforderungen abbilden, die in realen beruflichen Kontexten als besonders leistungskritisch gelten. Ob Testwerte, die anhand von Key-Feature-Problemen gewonnen werden, auf reale berufliche Kontexte übertragbar sind, wurde bisher selten untersucht. Tamblyn et al. (2007) fanden heraus, dass Key-Feature-Testwerte gering bis moderat ( $r = .16-.30$ ) mit späteren klinischen Kompetenzen korrelierten (s. dazu auch Ilgen et al., 2012).

Einschränkend sei erwähnt, dass die Prognosekraft von Multiple-Choice-Items und Key-Feature-Problemen anhand beruflicher Kompetenzmaße untersucht wurde, die sich oft nicht explizit und ausschließlich auf das diagnostische Problemlösen beziehen und meist auf einer beobachtergebundenen Kodierung basieren. Angesichts dieser Einschränkungen ist fraglich, wie belastbar die betreffenden Befunde sind.

### Authentizität und Prognosekraft computersimulierter diagnostischer Probleme

Als authentisch kann die Erfassung diagnostischer Problemlösekompetenzen mit computersimulierten diagnostischen Problemen, d. h. mit den bereits mehrfach erwähnten Computersimulationen und den darin dargebotenen holistischen diagnostischen Problemen aufgefasst werden. Im Vergleich zu den anderen Erhebungsvarianten ermöglichen Computersimulationen eine besonders authentische Nachbildung realer beruflicher Anforderungen. Wichtig ist in diesem Zusammenhang auch, dass bei der Gestaltung dieser Computersimulationen eine wichtige Maxime meist darin besteht, die reale Berufsumwelt so getreu wie möglich zu simulieren (z. B. Gschwendtner et al., 2009, S. 564 ff.). Ob diese Authentizität auch zu übertragbaren Aussagen über die diagnostische Problemlösekompetenz führt, wurde mehrfach untersucht. In diesen Untersuchungen bearbeiteten Proband:innen holistische diagnostische Probleme in realen beruflichen Kontexten und in den berufsauthentischen Computersimulationen.

Walker et al. (2016a, S. 159) ermittelten im Ausbildungsberuf zum/zur Elektroniker:in für Automatisierungstechnik zwischen den Testwerten aus der realistischen und der computerbasierten Erhebung eine sehr hohe latente Korrelation ( $r \geq .85$ ) und resümierten, dass die Erfassung mit der Computersimulation (nahezu) dieselben Ergebnisse erbringt wie die Erfassung in realen beruflichen Kontexten. Wie Gschwendtner et al. (2009) im Ausbildungsberuf zum/zur Kfz-Mechatroniker:in zeigten, lassen sich mit einer Computersimulation auch im Kfz-Bereich Aussagen gewinnen, die auf reale berufliche Kontexte übertragbar sind. In der erwähnten Studie korrelierten die Testwerte beider Erhebungskontexte (real vs. computersimuliert) latent mit  $r \approx .90$ . Auch in weiteren Studien des Kfz-Bereichs erlaubten die anhand der Kfz-Computersimulation gewonnenen Testwerte eine sehr präzise Prognose des diagnostischen Problemlöseerfolgs in realen beruflichen Kontexten (Abele, 2014, S. 190; Abele et al., 2016, S. 190). Nicht übersehen werden sollte allerdings, dass der Einsatz der beschriebenen Computersimulationen auch mit Nachteilen verbunden ist: Problematisch sind unter anderem der große finanzielle und zeitliche Entwicklungsaufwand sowie die langen Erhebungszeiten, die mit dieser Kompetenzerfassung einhergehen. Erwähnenswert ist zudem die in den Studien vorgenommene latente Modellierung der Zusammenhänge: Durch die latente Modellierung wurde das Problem adressiert, dass teilweise unbefriedigende interne Konsistenzen (Reliabilitäten) vorlagen.

Zusammenfassend ist festzuhalten, dass insbesondere Aussagen, die anhand berufsauthentischer Computersimulationen und holistischer diagnostischer Probleme gewonnen werden, als generalisierbar gelten dürfen – zumindest bei einer latenten, also messfehlerbereinigten Betrachtung. Diese Aussagen basieren auf einer relativ authentischen Erhebung und führen zu Testwerten, mit denen sich reale berufliche Leistungen präzise prognostizieren lassen. Key-Feature-Probleme stellen im Vergleich mit den computerbasierten Verfahren weniger authentische, im Vergleich mit Multiple-Choice-Items allerdings authentischere Anforderungen. Zur Prognosekraft von Multiple-Choice-Items und Key-Feature-Problemen liegen wenige Studien vor. Die vorliegenden Studien sprechen dafür, dass Testergebnisse, die mit Key-Feature-Problemen und Multiple-Choice-Items ermittelt wurden, auch in realen beruflichen Kon-

texten eine gewisse Aussagekraft haben. Zu berücksichtigen ist allerdings, dass diese Befunde auf beobachtergebundenen Kodierungen beruhen, weshalb sie mit den in Kapitel 4.2.1 herausgearbeiteten Validitätsrisiken behaftet sind. Aufgrund ihrer höheren Authentizität scheinen Key-Feature-Probleme ein größeres Übertragbarkeitspotenzial aufzuweisen als Multiple-Choice-Items.

### 4.3 Zusammenfassung und Forschungsdesiderate

Das Ziel des vorliegenden Kapitels bestand darin, das Erhebungsmodell diagnostischer Problemlösekompetenzen zu konkretisieren. Bei der Konkretisierung des Erhebungsmodells ist unter anderem zu entscheiden, welche Items bzw. diagnostischen Probleme zur Erfassung der diagnostischen Problemlösekompetenz herangezogen werden sollen, in welchem Modus (z. B. papierbasiert oder computerbasiert) diese Probleme bzw. Items darzubieten sind und wie der Problemlöseerfolg zu kodieren ist. Diese Entscheidungen können nur dann rational und systematisch getroffen werden, wenn feststeht, welchen Kriterien das Erhebungsmodell genügen soll. Zu Beginn dieses Kapitels wurde im Anschluss an den einschlägigen Forschungsstand festgelegt, das Erhebungsmodell unter dem Validitätskriterium zu konkretisieren bzw. zu optimieren. Unter Anwendung des argumentbasierten Validitätsansatzes wurde herausgearbeitet, dass bei der Konkretisierung des Erhebungsmodells auf valide Kodierungen des Problemlöseerfolgs zu achten ist. Ferner sollte das Erhebungsmodell generalisierbare Aussagen über die diagnostische Problemlösekompetenz ermöglichen, die darüber hinaus auf berufliche Kontexte übertragbar sind. Falls das Erhebungsmodell die genannten Konkretisierungskriterien (Kodierungskriterium, Generalisierbarkeitskriterium und Übertragbarkeitskriterium) erfüllt, sind keine validen Aussagen garantiert, aber wichtige Voraussetzungen für valide Aussagen über die diagnostische Problemlösekompetenz geschaffen.

Wie die zurückliegenden Ausführungen verdeutlichten, ist generell von beobachtungsgebundenen Kodierungen abzusehen, um valide Kodierungen zu gewährleisten. Unter Berücksichtigung dieser Prämisse fand bei der Aufarbeitung des einschlägigen Forschungsstands eine Konzentration auf papierbasierte Multiple-Choice-Items, papierbasierte Key-Feature-Probleme und computersimulierte diagnostische Probleme statt. Bei den genannten Verfahren ist es prinzipiell möglich, auf Beobachter:innen bei der Kodierung zu verzichten. Mit diesen Verfahren ist der Möglichkeitsraum und Forschungsstand keinesfalls erschöpfend, aber in wichtigen Teilen abgedeckt (s. dazu z. B. Vleuten & Schuwirth, 2005; Ilgen et al., 2012).

Computersimulierte diagnostische Probleme können aus Sicht der *Generalisierbarkeit* als günstig gelten, da solche Probleme unmittelbar den gesamten Problemlöseprozess und die damit verbundenen psychischen Komponenten abdecken. Wie einige Studien zeigten, führte der Einsatz dieser Probleme jedoch teilweise zu Testwerten geringer interner Konsistenz (Reliabilität). Dies kann als empirisches Indiz für eine eingeschränkte Generalisierbarkeit verstanden werden. Wahrscheinlich erreichen diagnostische Probleme bezogen auf den Problemlöseprozess und die psychischen Komponenten eine sehr gute, bezogen auf die externen Anforderungen eine eher unbefriedigende Repräsentation der diagnostischen Problemlösekompetenz. Dies hat

vermutlich mit der relativ hohen Bearbeitungszeit diagnostischer Probleme zu tun. Aus verschiedenen Gründen (Feldzugang, Motivation, Ermüdung etc.) ist es aufgrund dieser Bearbeitungszeit nicht möglich, eine Problemauswahl vorzulegen, die den externen Anforderungsbereich repräsentativ abdeckt. Nachteilig an den computersimulierten diagnostischen Problemen ist ferner der hohe finanzielle und zeitliche Entwicklungsaufwand. Im Hinblick auf die *Übertragbarkeit* scheint die Kombination aus berufsauthentischen Computersimulationen und holistischen diagnostischen Problemen (sehr) vielversprechend: Wie die zitierten Studien belegen, ermöglicht diese Kombination eine authentische Erfassung diagnostischer Problemlösekompetenzen und Aussagen, die sich auf berufliche Kontexte übertragen lassen.

Die großen Stärken von Erhebungsmodellen mit *papierbasierten* (auch: schriftlichen) *Multiple-Choice-Items* bestehen in einem geringen Entwicklungsaufwand und in relativ kurzen Bearbeitungszeiten, die eine gute Abdeckung externer Problemanforderungen ermöglichen. Mit schriftlichen Multiple-Choice-Items lassen sich in vertretbaren Erhebungszeiten Testwerte hoher interner Konsistenz (Reliabilität) gewinnen. Fraglich ist allerdings, ob mit diesen Items repräsentativ erfasst werden kann, ob der (gesamte) diagnostische Problemlöseprozess unter Investition relevanter psychischer Komponenten erfolgreich gestaltet werden kann. Zweifel bestehen aber nicht nur mit Blick auf die *Generalisierbarkeit*, sondern auch hinsichtlich der *Übertragbarkeit*. Im Vergleich mit realen beruflichen Problemlöseanforderungen sind die mit Multiple-Choice-Items gestellten Anforderungen wenig authentisch. Ferner deuten die vorliegenden Befunde darauf hin, dass mit Multiple-Choice-Items gewonnene Testwerte der diagnostischen Problemlösekompetenz allenfalls moderat mit realen beruflichen Leistungen korrelieren.

Erhebungsmodelle mit schriftlichen Key-Feature-Problemen sind unter dem *Generalisierbarkeitskriterium* differenziert zu beurteilen. Einerseits ermöglichen sie es, die diagnostische Problemlösekompetenz relativ authentisch abzubilden. Andererseits erbrachten sie bislang keine Testwerte mit zufriedenstellender interner Konsistenz. Dies liegt wohl vor allem an den recht hohen Bearbeitungszeiten bisheriger Key-Feature-Probleme, die sich aber prinzipiell reduzieren ließen. Mit Key-Feature-Problemen dürfte eine authentischere Kompetenzerfassung gelingen als mit Multiple-Choice-Items. Wie die wenigen Befunde zur *Übertragbarkeit* belegen, führte dies in den bisherigen Studien allerdings nicht dazu, dass berufliche Leistungen mit Key-Feature-Testwerten besser prognostiziert werden konnten als mit Multiple-Choice-Items. Dies kann auch an der bereits erwähnten geringen internen Konsistenz (Reliabilität) der Key-Feature-Testwerte liegen. Im Vergleich mit Multiple-Choice-Items besteht der Vorteil von Key-Feature-Problemen in einer hohen Authentizität, im Vergleich mit den erwähnten Computerverfahren in einem relativ geringen finanziellen und zeitlichen Entwicklungsaufwand. Zu beachten ist auch, dass Key-Feature-Probleme bisher vergleichsweise selten eingesetzt wurden und deshalb nicht befriedigend abgeschätzt werden kann, welches diagnostische Potenzial sie haben.

Ohne Zweifel bietet der aktuelle Forschungsstand wichtige Erkenntnisse, wie das Erhebungsmodell der diagnostischen Problemlösekompetenz in Sinne der Aussagenvalidität zu gestalten ist. Einiges spricht dafür, bei der Erfassung der diagnostischen

Problemlösekompetenz insbesondere auf Erhebungsmodelle mit computersimulierten diagnostischen Problemen zu setzen. Sichtbar wurden aber auch einige Forschungsdesiderate. Als problematisch erwiesen sich unter anderem die hohen Bearbeitungszeiten sowie die teilweise geringen internen Konsistenzen (Reliabilitäten) und der hohe Entwicklungsaufwand der Kompetenzerfassung mit computersimulierten diagnostischen Problemen. Im Hinblick auf Entwicklungsaufwand, Erhebungszeit und Authentizität können Key-Feature-Probleme als Kompromiss zwischen computersimulierten diagnostischen Problemen und schriftlichen Multiple-Choice-Items betrachtet werden. Angesichts der aktuellen Befundlage ist allerdings noch zu klären, ob mit Key-Feature-Problemen tatsächlich eine valide Erfassung diagnostischer Problemlösekompetenzen möglich ist.

## 5 Forschungsprogramm und Forschungsfragen

Die zurückliegenden Ausführungen offenbarten mehrere Forschungsdesiderate im Bereich des Sprachmodells und des Erhebungsmodells. Einige dieser Desiderate wurden im Rahmen dieser Untersuchung in Forschungsfragen überführt und in sechs Einzelstudien bearbeitet.

### 5.1 Auf das Sprachmodell bezogene Forschungsfragen

Wie in Kapitel 3 herausgearbeitet, ist die aktuelle Befundlage zur empirischen Eigenständigkeit diagnostischer Problemlösekompetenzen mit Unsicherheiten verbunden. Forschungsbedarf besteht auch dahin gehend, welche psychischen Komponenten als Konstituenten diagnostischer Problemlösekompetenzen zu betrachten sind. Darüber hinaus wurde festgestellt, dass gegenwärtig keine Theorie verfügbar ist, die den Problemlöseerfolg als Wirkung des diagnostischen Problemlöseprozesses rekonstruiert. Vor diesem Hintergrund wurden in dieser kumulativen Habilitation die folgenden Forschungsfragen untersucht:

- (1) Inwiefern kann die diagnostische Problemlösekompetenz empirisch von anderen beruflichen Kompetenzen unterschieden werden? (Studie 1)
- (2) Ist der Umgang mit Komplexität eine bedeutsame psychische Komponente der diagnostischen Problemlösekompetenz? (Studie 2)
- (3) Wie kann der Prozess theoretisch modelliert werden, der dem diagnostischen Problemlöseerfolg zugrunde liegt? (Studie 3 und Studie 4)

### 5.2 Auf das Erhebungsmodell bezogene Forschungsfragen

In Anlehnung an Kapitel 4 scheint es hinsichtlich einer befriedigenden Validität (besonders) günstig, wenn das Erhebungsmodell computersimulierte diagnostische Probleme umfasst. Deutlich wurde aber auch, dass bei einem solchen Erhebungsmodell mit Testwerten geringer interner Konsistenz zu rechnen ist, was auf ein Generalisierbarkeitsproblem und ein Messfehlerproblem hinweist. Als ein Grund dieser Probleme wurden die besonders hohen Bearbeitungszeiten holistischer diagnostischer Probleme identifiziert. Als weitere Nachteile computersimulierter diagnostischer Probleme stellten

sich – wie gezeigt – die hohen Entwicklungskosten und die langen Entwicklungszeiten heraus. Schriftliche Key-Feature-Probleme weisen diesbezüglich bessere Eigenschaften auf. Offen ist allerdings, ob mit solchen Key-Feature-Problemen valide Aussagen gewonnen werden können. Ausgehend von diesen Forschungsdesideraten wurden bezüglich des Erhebungsmodells folgende Forschungsfragen untersucht:

- (5) Kann die interne Konsistenz (Reliabilität) der Erfassung diagnostischer Problemlösekompetenzen unter Wahrung des Validitätsanspruchs erhöht werden, wenn computersimulierte Key-Feature-Probleme zusätzlich zu computersimulierten diagnostischen Problemen eingesetzt werden? (Studie 5)
- (6) Können diagnostische Problemlösekompetenzen mit papierbasierten Key-Feature-Problemen valide erfasst werden? (Studie 6)

Alle der im Folgenden thematisierten Einzelstudien wurden bei Auszubildenden zum/zur Kfz-Mechatroniker:in durchgeführt und bauen auf dem skizzierten Sprachmodell auf. Die diagnostische Kfz-Problemlösekompetenz wurde jeweils anforderungsspezifisch (elektrotechnische diagnostische Kfz-Problemlösekompetenz) gefasst und es wurde unterstellt, dass diese Kompetenz aus mehreren psychischen Komponenten besteht (Diagnosewissen, schlussfolgernde Denkfähigkeit etc.) und sie sich auf einen Problemlöseprozess (Informationsrepräsentation, Hypothesenentwicklung, Hypothesentest, Hypothesenbewertung) bezieht. Zur Erfassung der diagnostischen Kfz-Problemlösekompetenz wurde durchgängig auf die bereits mehrfach erwähnte Kfz-Computersimulation und holistische diagnostische Probleme aus dem Bereich der Kfz-Elektrotechnik zurückgegriffen. Abhängig von der Fragestellung wurden ergänzend weitere psychische Merkmale erfasst und auch Key-Feature-Probleme eingesetzt.

## 6 Ergebnisse der Einzelstudien

### 6.1 Studien zum Sprachmodell

#### 6.1.1 Empirische Eigenständigkeit der diagnostischen Kfz-Problemlösekompetenz (Studie 1)

In diesem Abschnitt werden die Ergebnisse der folgenden Publikation zusammengefasst:

Nickolaus, R., Behrendt, S. & Abele, S. (2016). Kompetenzstrukturen bei Kfz-Mechatronikern und die Erklärungskraft des fachsystematischen Kfz-Wissens für berufsfachliche Kompetenzen. *Unterrichtswissenschaft, 44*, 114–130.

Die vollständige Publikation findet sich in Anhang A.

#### Thema und Forschungsfrage

In der Studie wurde untersucht, ob sich die elektrotechnische diagnostische Kfz-Problemlösekompetenz empirisch von der Kfz-Reparaturkompetenz und der Kfz-Service-

kompetenz unterscheiden lässt. Die Kfz-Reparaturkompetenz und Kfz-Servicekompetenz wurden neben der diagnostischen Problemlösekompetenz als zentrale berufsfachliche Kompetenzen im Beruf des Kfz-Mechatronikers/der Kfz-Mechatronikerin identifiziert. In der Publikation ging es auch um die Bedeutung des fachsystematischen Kfz-Wissens für die diagnostische Problemlösekompetenz.

### Forschungshypothese

Theoretisch wurde angenommen, dass die diagnostische Kfz-Problemlösekompetenz eine empirisch eigenständige Kompetenz darstellt. Empirisch wurde erwartet, dass die Testscores der diagnostischen Kfz-Problemlösekompetenz eine Dimension aufspannen, die sich statistisch von den Dimensionen *Reparaturkompetenz* und *Servicekompetenz* unterscheiden lassen. Vermutet wurde ferner ein deutlicher Einfluss des fachsystematischen Wissens auf die diagnostische Problemlösekompetenz.

### Methode

Um die Hypothesen zu prüfen, wurden Daten bei einer Stichprobe von 201 Kfz-Auszubildenden (Zahl männlich) erhoben, welche ausschließlich in Handwerksbetrieben im dritten Ausbildungsjahr angestellt waren. Zur Erfassung der diagnostischen Problemlösekompetenz wurden sechs berufstypische (holistische) diagnostische Kfz-Probleme sowie sieben computersimulierte Key-Feature-Probleme eingesetzt. Auf diese Key-Feature-Probleme wird in Kapitel 6.2.1 näher eingegangen. Die Kfz-Probleme und Key-Feature-Probleme wurden in der bereits erwähnten Kfz-Computersimulation angeboten und waren in Interaktion mit dieser zu bearbeiten. Die aus dieser Erhebung resultierenden schriftlichen Protokolle der Proband:innen wurden unter Anwendung eines Kodiermanuals kodiert. Bei der Interaktion mit der Computersimulation traten größtenteils Anforderungen auf, die weitgehend den Anforderungen der beruflichen Realität von Kfz-Mechatroniker:innen entsprachen. Die Anforderungen waren also sehr authentisch. Gschwendtner et al. (2009) zeigten, dass die mit diesem computerbasierten Verfahren gewonnenen Aussagen über die diagnostische Problemlösekompetenz auf berufliche Kontexte übertragbar sind (s. dazu auch Kapitel 4.2.3).

Die *Kfz-Reparaturkompetenz* und *Kfz-Servicekompetenz* wurden mithilfe eines videobasierten Computertests erfasst, der so konstruiert wurde, dass ebenfalls berufsauthentische Anforderungen zu bewältigen waren. Ergänzend wurden verschiedene Dimensionen des *fachsystematischen Kfz-Wissens* erhoben, wobei hier insbesondere das Wissen der folgenden Inhaltsbereiche relevant ist: Motorsteuerung, Motor und Start-Strom-Beleuchtung. Diese Inhaltsbereiche sind hier deswegen besonders relevant, weil das Wissen zur Motorsteuerung, zum Motor und zur Start-Strom-Beleuchtung eine große inhaltliche Nähe zu den vorgegebenen elektrotechnischen diagnostischen Kfz-Problemen aufweist; die Kfz-Probleme bezogen sich auf die Kfz-Subsysteme *Motorsteuerung* und *Beleuchtungsanlage*. Die curriculare Validität der eingesetzten Erhebungsverfahren wurde durch den Einbezug erfahrener Lehrkräfte, Ausbilder:innen und Prüfungsexpert:innen sowie vorliegender Tätigkeitsanalysen und einer Analyse realisierter Curricula abgesichert. Für die statistischen Analysen wurde auf Struktur-

gleichungsmodelle zurückgegriffen. Um die empirische Eigenständigkeit der diagnostischen Kfz-Problemlösekompetenz zu untersuchen, wurden konfirmatorische Faktorenanalysen (mehrdimensionale statistische Messmodelle) angewandt. Die Bedeutung des fachsystematischen Wissens für die diagnostische Problemlösekompetenz wurde mit Erklärungsmodellen untersucht.

### Ergebnisse

Bei der Erfassung der genannten Wissensdimensionen wurden EAP/PV-Reliabilitäten zwischen .79 und .87 ermittelt. Bei der diagnostischen Problemlösekompetenz lag die EAP/PV-Reliabilität bei .75, bei der Reparatur- und Servicekompetenz bei .65 und .68. Ein Vergleich der Modellpassungsgüte zeigte, dass ein dreidimensionales Faktorenmmodell besser auf die Daten passte als ein eindimensionales oder zweidimensionales Modell. Im dreidimensionalen Modell war die diagnostische Problemlösekompetenz als eigenständige Dimension der berufsfachlichen Kfz-Kompetenz und separat von den Dimensionen *Reparaturkompetenz* und *Servicekompetenz* repräsentiert. Erwähnenswert sind ferner die beachtlichen latenten Interkorrelationen zwischen der diagnostischen Problemlösekompetenz und den anderen beiden beruflichen Kompetenzen, die bei  $r = .64$  bzw.  $r = .83$  lagen. Bei Kontrolle des fachsystematischen Wissens fielen diese Interkorrelationen (deutlich) geringer aus. Darüber hinaus zeigte sich, dass das Wissen zur Motorsteuerung, zum Motor und zur Start-Strom-Beleuchtungsanlage einen beachtlichen Varianzanteil der diagnostischen Problemlösekompetenz erklärte ( $R^{22} = .56$ ). Diese spezifischen Wissensdimensionen hatten im Vergleich zu anderen fachspezifischen Wissensdimensionen einen deutlich höheren Effekt auf die diagnostische Problemlösekompetenz.

### Diskussion

Im Einklang mit den theoretischen Erwartungen sprachen die Befunde für die empirische Eigenständigkeit der diagnostischen Kfz-Problemlösekompetenz. Besonders erwähnenswert ist dies auch deswegen, weil sich die diagnostische Problemlösekompetenz empirisch von eng verwandten Kompetenzen abgrenzen ließ: Diagnostische Problemlösekompetenzen, Reparaturkompetenzen und Servicekompetenzen wurden innerhalb eines Berufs erfasst und sind im Berufsalltag vielfältig aufeinander bezogen. So ist z. B. im Anschluss an die Diagnose einer technischen Kfz-Störung normalerweise die Störung zu beheben. Es sind also Reparaturarbeiten auszuführen. Zudem bestätigten die Ergebnisse, dass das Wissen in den genannten verwandten Inhaltsbereichen einen großen und im Vergleich mit dem Wissen in anderen Inhaltsbereichen einen größeren Einfluss auf die diagnostische Problemlösekompetenz hatte. Auch dieser Befund kann als Beleg interpretiert werden, die diagnostische Problemlösekompetenz als empirisch abgrenzbar aufzufassen.

Grenzen weist die Studie insofern auf, als die Reliabilität der Erfassung der Reparatur- und Servicekompetenz nicht vollständig befriedigend war. Angesichts der latenten, d. h. messfehlerbereinigten Analyse der empirischen Eigenständigkeit der diagnostischen Problemlösekompetenz sollte dies allerdings keinen entscheidenden Einfluss

auf das Untersuchungsergebnis haben. Diskussionswürdig ist ferner, ob mit dem computerbasierten Videotest, der zur Erfassung der Reparatur- und Servicekompetenz zum Einsatz kam, tatsächlich Kompetenzen erfasst wurden oder eher Reparatur-Handlungswissen und Service-Handlungswissen. Mit dem Videotest wurde geprüft, ob jemand weiß, wie in einer spezifischen Situation zu handeln wäre. Inwieweit daraus geschlossen werden kann, ob jemand die entsprechende Handlung auch eigenständig ausführen kann, ist ungewiss. Es scheint nicht besonders plausibel, dass die diagnostische Problemlösekompetenz empirisch nicht mehr von anderen beruflichen Kompetenzen abgrenzbar ist, wenn die anderen Kompetenzen im Handlungsvollzug erfasst werden. Letztlich zeigt dieser Punkt aber weiteren Forschungsbedarf an.

### 6.1.2 Komponenten der diagnostischen Kfz-Problemlösekompetenz: Bedeutung des Umgangs mit Komplexität (Studie 2)

In diesem Abschnitt werden die Ergebnisse der folgenden Publikation zusammengefasst:

Abele, S. (2016). Umgang mit Komplexität: Eine bedeutsame psychische Voraussetzung des domänenspezifischen Problemlösens? *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 112, 37–59.

Die vollständige Publikation findet sich in Anhang B.

#### Thema und Forschungsfrage

Die Publikation beschäftigte sich mit der Bedeutung des Umgangs mit Komplexität für die diagnostische Problemlösekompetenz. Der Umgang mit Komplexität wurde als eine wichtige kognitive Komponente der diagnostischen Problemlösekompetenz identifiziert und als psychisches Personenmerkmal konzeptualisiert, das für die Verarbeitung problemlöserrelevanter Informationen benötigt wird. Je stärker dieses Merkmal ausgeprägt ist, desto erfolgreicher können *informationsintensiv* diagnostische Probleme bearbeitet werden. Informationsintensiv meint, dass die Problemlösung die Verarbeitung vieler problemspezifischer Einzelinformationen erfordert. Die Anzahl an Einzelinformationen wurde problemspezifisch theoretisch bestimmt und als Maß für die Problemkomplexität interpretiert.

#### Forschungshypothese

In Anlehnung an den Forschungsstand wurde vermutet, dass der Umgang mit Komplexität einen Einfluss auf den Problemlöseerfolg hat. Theoretisch wurde dieser Einfluss darauf zurückgeführt, dass komplexere Probleme ein höheres Ausmaß an Umgang mit Komplexität verlangen. Da dieses psychische Merkmal in einer Population (vermutlich) normalverteilt ausgeprägt ist, werden komplexe Probleme seltener gelöst als weniger komplexe Probleme. Dementsprechend wurde empirisch erwartet, dass die Komplexität eines diagnostischen Kfz-Problems einen Effekt auf die empirische Schwierigkeit eines diagnostischen Problems hat.

## Methoden

Die Hypothese wurde mit einer Stichprobe von 274 (Zahl männlich) Auszubildenden der Kfz-Mechatronik im dritten Ausbildungsjahr geprüft. Die Proband:innen bearbeiteten 13 diagnostische Kfz-Probleme<sup>20</sup>, die sich in einer Vorgängerstudie als psychometrisch brauchbar erwiesen. Um die empirische Schwierigkeit der diagnostischen Probleme zu ermitteln, wurde eine Raschanalyse durchgeführt. Die Problemkomplexität wurde theoretisch und a priori bestimmt. Der Effekt der Problemkomplexität auf die Problemschwierigkeit wurde mit einer Regressionsanalyse untersucht.

## Ergebnisse

Alle diagnostischen Probleme erwiesen sich als raschkonform. Bei der Analyse der Itemkennwerte zeigte sich, dass ein Problem zwar als sehr komplex eingestuft wurde, allerdings nur eine geringe empirische Schwierigkeit aufwies. Bei der Regressionsanalyse war kein Effekt der Problemkomplexität auf die Problemschwierigkeit feststellbar. Nach Ausschluss des auffälligen diagnostischen Problems wurde ein großer Effekt ( $\beta = .70$ ) ermittelt.

## Diskussion

Die Studie basiert auf folgender Argumentationslogik: Die diagnostische Problemlösekompetenz wurde als Ursache des diagnostischen Problemlöseerfolgs konzeptualisiert. Der Umgang mit Komplexität wurde als kognitive Komponente der diagnostischen Problemlösekompetenz aufgefasst. Daraus folgt, dass der Umgang mit Komplexität einen kausalen Einfluss auf den Problemlöseerfolg haben sollte. Da der Umgang mit Komplexität ein latentes Personenmerkmal darstellt, konnte dieser Einfluss nur indirekt untersucht werden. Es wurde analysiert, ob Probleme unterschiedlicher Komplexität mit unterschiedlichen empirischen Schwierigkeiten verbunden sind, wobei die Schwierigkeit mit steigender Komplexität zunehmen sollte. Dieser Einfluss wurde nach Ausschluss eines diagnostischen Problems beobachtet und so interpretiert, dass der Umgang mit Komplexität eine wichtige kognitive Komponente der diagnostischen Problemlösekompetenz darstellt. Durch dieses etwas verwickelte Forschungsdesign wurde es möglich, eine forschungsmethodische Schwäche vieler bisheriger Studien zu überwinden (Kapitel 3.3.3): Der Umgang mit Komplexität wurde nicht separat erfasst und anschließend korrelativ auf die diagnostische Problemlösekompetenz bezogen. Vielmehr wurde die kognitive Komponente integrativ und gemeinsam mit der diagnostischen Problemlösekompetenz erhoben und deren Bedeutung aus dem Problemlöseerfolg, also dem Indikator der diagnostischen Problemlösekompetenz, rekonstruiert.

Die Tatsache, dass ein erwartungskonformes Ergebnis erst nach Ausschluss eines auffälligen Problems zu beobachten war, wurde auf Schwierigkeiten bei der Operationalisierung der Komplexität zurückgeführt. Diese Schwierigkeiten resultierten wohl

---

<sup>20</sup> Genau genommen bearbeiteten die Proband:innen sechs diagnostische Probleme und sieben Key-Feature-Probleme.

unter anderem daraus, dass gegenwärtig keine Theorie<sup>21</sup> zum diagnostischen Problemlöseprozess vorliegt. Das Komplexitätsrating und der Problemlöseprozess stellten sich als voneinander abhängige Wirklichkeitsbereiche dar. Problematisch an der Operationalisierung war, die Anzahl an Einzelinformationen als Merkmal der externen Umwelt aufzufassen und zur Quantifizierung der Problemkomplexität heranzuziehen. Im Lichte der Argumentation von Heid (2001) kann ganz grundsätzlich und kritisch gefragt werden, wie sich Merkmale der externen Umwelt, d. h. der Komplexitätsgrad eines Problems, überhaupt objektiv rekonstruieren lassen. Vieles spricht dafür, dass die Rekonstruktion von Problemmerkmalen letztlich auf einer (oft eher impliziten) Theorie basiert, die empirisch unzutreffend sein kann. Um Merkmale wie die Problemkomplexität objektiv zu erfassen, müsste die Komplexitätsbestimmung aber auf einer elaborierten und empirisch angemessenen Theorie aufsetzen. In einer solchen Theorie wäre die Problemkomplexität z. B. als ein Interaktionsprodukt spezifischer Subjekte (deren Wahrnehmung, mentalen Verarbeitungsprozesse, Erfahrung etc.) mit spezifischen Problemen zu fassen und es wäre auch zu klären, welchen Wirklichkeitsbereich die Theorie fokussiert (z. B. objektive Anforderungen oder mentale Prozesse) und wie dieser Bereich empirisch zu erfassen ist. Damit ist ein sehr grundsätzliches und kompliziertes Thema aufgeworfen, das vom eigentlichen Thema der Habilitation wegführen würde und hier nicht weiter vertieft werden kann. Auf jeden Fall mahnt diese Kritik aber dazu, Ergebnisse aus Studien zur Bedeutung der Anforderungskomplexität für das berufliche Problemlösen und berufliche Leistungen vorsichtig zu interpretieren und sorgfältig zu reflektieren.

### 6.1.3 Diagnostischer Problemlöseprozess: Entwicklung und Prüfung einer Prozesstheorie (Studie 3)

In diesem Abschnitt werden die Ergebnisse der folgenden Publikation zusammengefasst:

Abele, S. (2017). Diagnostic Problem-Solving Process in Professional Contexts: Theory and Empirical Investigation in the Context of Car Mechatronics Using Computer-Generated Log-Files. *Vocations and Learning*, 11(1), 133–159. <https://doi.org/10.1007/s12186-017-9183-x>

Die vollständige Publikation findet sich in Anhang C.

#### Thema und Forschungsfrage

Im ersten Teil der Studie wurde im Rückgriff auf den einschlägigen Forschungsstand eine Theorie zum diagnostischen Problemlöseprozess in beruflichen Kontexten entwickelt. In dieser Theorie werden die folgenden latenten Subprozesse des diagnostischen Problemlöseprozesses unterschieden: Informationsrepräsentation, Hypothesenent-

<sup>21</sup> Wie in Kapitel 3.3.5 ausgeführt, liegen unzählige Studien zum diagnostischen Problemlöseprozess und viele normative und deskriptive Modelle zum diagnostischen Problemlöseprozess vor. Theorien, in denen der Problemlöseerfolg als Wirkung des Problemlöseprozesses rekonstruiert wird, konnten allerdings nicht gefunden werden.

wicklung, Hypothesentest und Hypothesenbewertung. Im Rahmen der Theorie wird angenommen, dass die Qualität dieser Sub-Prozesse in einem bestimmten Kausalverhältnis steht und den beobachtbaren Problemlöseerfolg determiniert. Eine theoretisch begründete Annahme der Theorie besteht darin, dass sich die Qualität des Subprozesses *Informationsrepräsentation* mit erfolgskritischem Informationsverhalten und die Qualität des Subprozesses *Hypothesentest* mit erfolgskritischem Testverhalten operationalisieren lässt.<sup>22</sup> Diese beobachtbaren Qualitätsindikatoren (auch: erfolgskritisches Problemlöseverhalten) können theoretisch bestimmt und empirisch erfasst werden. Im zweiten Teil der Studie wurde die Theorie zum diagnostischen Problemlöseprozess im Kfz-Kontext empirisch untersucht.

### Forschungshypothese

Ausgehend von der Theorie wurden folgende Hypothesen untersucht: (1) Die Qualität der Subprozesse Informationsrepräsentation und Hypothesentest hat einen Einfluss auf den diagnostischen Problemlöseerfolg. Entsprechend der oben erwähnten Operationalisierung der Subprozessqualitäten wurde empirisch erwartet, dass sowohl das erfolgskritische Informationsverhalten als auch das erfolgskritische Testverhalten einen Effekt auf den diagnostischen Problemlöseerfolg hat. (2) Die Qualität des Subprozesses Hypothesentest hat einen größeren Einfluss auf den diagnostischen Problemlöseerfolg als die Qualität von Informationsrepräsentation. Empirisch wurde erwartet, dass der Qualitätsindikator „erfolgskritisches Testverhalten“ einen größeren Effekt auf den Problemlöseerfolg hat als der Qualitätsindikator *erfolgskritisches Informationsverhalten*. (3) Der Einfluss des erfolgskritischen Informationsverhaltens auf den diagnostischen Problemlöseerfolg hängt davon ab, ob bei der Bearbeitung eines diagnostischen Problems auf Erfahrung, d. h. problemspezifisches Wissen, zurückgegriffen werden kann. Theoretisch wurde dies damit begründet, dass das erfolgskritische Informationsverhalten durch den Abruf problemrelevanten Langzeitwissens ersetzt werden kann. Empirisch wurde vermutet, dass der Effekt des erfolgskritischen Informationsverhaltens auf den diagnostischen Problemlöseerfolg durch problemspezifische Erfahrung moderiert wird. (4) Der Einfluss der Qualität von Informationsrepräsentation auf den diagnostischen Problemlöseerfolg wird durch die Qualität von Hypothesentests vermittelt. Dementsprechend wurde empirisch antizipiert, dass der Effekt des erfolgskritischen Informationsverhaltens auf den diagnostischen Problemlöseerfolg vom erfolgskritischen Testverhalten mediiert wird.

### Methode

Zur Hypothesenprüfung wurde eine Stichprobe von 339 (Zahl männlich) Auszubildenden zum/zur Kfz-Mechatroniker:in im dritten Ausbildungsjahr herangezogen. Die Hypothesen wurden auf der Basis von zwei diagnostischen Kfz-Problemen (P1 und P2) und der bereits mehrfach erwähnten Kfz-Computersimulation zur Erfassung des diagnostischen Problemlösens geprüft. Um den Problemlöseerfolg zu bestimmen, wur-

---

<sup>22</sup> Die Operationalisierung der Begriffe *erfolgskritisches Informationsverhalten* und *erfolgskritisches Testverhalten* wird weiter unten im Methodenabschnitt erläutert.

den die schriftlichen Dokumentationen analysiert, die von den Proband:innen während der Problemlösung erstellt worden waren (Interrater-Reliabilität von P1 und P2:  $\kappa \geq .95$ ). Die Qualitätsindikatoren der Subprozesse, also das erfolgskritische Informationsverhalten und das erfolgskritische Testverhalten, wurden anhand erfolgskritischer diagnostischer Hypothesen theoretisch bestimmt. Erfolgskritische diagnostische Hypothesen sind fachlich sinnvolle Annahmen zu möglichen Ursachen eines Kfz-Defekts. Sinnvoll wäre z. B. anzunehmen, eine defekte Sicherung sei für einen Abblendlichtdefekt verantwortlich; unsinnig wäre hingegen, einen leeren Tank dafür verantwortlich zu machen. Das erfolgskritische Problemlöseverhalten ergibt sich unter Anwendung domänenspezifischer Expertise aus den erfolgskritischen Hypothesen: Falls z. B. eine defekte Sicherung als Defektursache vermutet wird, wäre die Sicherung mit einem spezifischen erfolgskritischen Testverhalten (z. B. Widerstandsmessung) zu prüfen. Das erfolgskritische Informationsverhalten und das erfolgskritische Testverhalten wurden mithilfe computergenerierter Log-Dateien bestimmt. Diese Log-Dateien wurden während der Erhebung automatisch erstellt und dokumentierten das Problemlöseverhalten eines Probanden bezogen auf ein spezifisches diagnostisches Problem. Für die statistischen Analysen wurden Probit-Regressionsanalysen sowie Moderator- und Mediatoranalysen durchgeführt.

### Ergebnisse

Die Ergebnisse dokumentierten, dass das erfolgskritische Informationsverhalten und das erfolgskritische Testverhalten einen beachtlichen Einfluss auf den Problemlöseerfolg gehabt hatten (P1:  $R^2 = .56$ ; P2:  $R^2 = .63$ ). Zudem zeigte sich, dass das erfolgskritische Testverhalten (P1:  $R^2 = .52$ ; P2:  $R^2 = .61$ ) deutlich einflussreicher gewesen war als das erfolgskritische Informationsverhalten (P1:  $R^2 = .18$ ; P2:  $R^2 = .36$ ). Der erwartete Moderatoreffekt der Erfahrung war nur bei einem der beiden diagnostischen Kfz-Probleme zu beobachten. Schließlich belegten die Ergebnisse, dass der Einfluss des erfolgskritischen Informationsverhaltens auf den Problemlöseerfolg durch das erfolgskritische Testverhalten erwartungsgemäß mediiert worden war. Einige Verhaltensweisen konnten, hauptsächlich wegen geringer Varianzen, nicht zur Hypothesenprüfung verwendet werden.

### Diskussion

Die empirischen Ergebnisse standen weitgehend im Einklang mit den theoretischen Erwartungen. Der Moderatoreffekt der Erfahrung blieb allerdings bei einem der beiden diagnostischen Probleme aus. Dies wurde auf die spezifischen Anforderungscharakteristika dieses Problems zurückgeführt, die vermutlich dafür sorgten, dass keine einschlägigen Erfahrungen vorlagen. Diese ad-hoc-Erklärung wurde anhand theoretischer Überlegungen und weiterer Ergebnisse der Studie gestützt, ist in künftigen empirischen Studien aber genauer zu untersuchen.

Die Ergebnisse liefern eine empirische Rechtfertigung dafür, Informationsrepräsentation, Hypothesenentwicklung, Hypothesentest und Hypothesenbewertung als wichtige Subprozesse des diagnostischen Problemlöseprozesses und Grundlagen des diagnostischen Problemlöseerfolgs aufzufassen. In domänenübergreifender Hinsicht

wird mit der Theorie unterstellt, dass der diagnostische Problemlöseprozess in unterschiedlichen beruflichen Kontexten auf die beschriebene Weise theoretisch modelliert werden kann. In domänenspezifischer Hinsicht wird betont, dass die Inhalte des Prozesses und das erfolgskritische Problemlöseverhalten nur problemspezifisch bestimmt werden können.

Grenzen weist die Untersuchung insofern auf, als zur Entwicklung der Prozess- theorie zwar theoretische Erkenntnisse und empirische Befunde unterschiedlicher beruflicher Kontexte herangezogen wurden, die empirische Untersuchung allerdings im Kontext der Kfz-Mechatronik anhand von zwei elektrotechnischen diagnostischen Kfz-Problemen erfolgte. Die empirische Basis der Theorieprüfung ist also (sehr) schmal. Fraglich und offen ist, ob die Theorie auf andere Kfz-Probleme und andere berufliche Kontexte übertragbar ist. Dies sollte genauso in künftigen Studien untersucht werden wie der bereits erwähnte Moderatoreffekt, der nicht durchgängig beobachtet wurde.

Die Theorie zum diagnostischen Problemlöseprozess in beruflichen Kontexten wurde aus einer funktionalen Perspektive, d. h. mit Blick auf jene Anforderungen rekonstruiert, die beim diagnostischen Problemlösen zu bewältigen sind. Beim Lösen diagnostischer Probleme sind Informationen zu gewinnen, Ursachenhypothesen zu formulieren etc. Alternativ wäre denkbar gewesen, den Prozess in Anlehnung an die Erkenntnislogik von Charles Sanders Peirce formallogisch zu rekonstruieren. Diese Erkenntnislogik wurde von Minnameier (2005) unter anderem zur Analyse von Lehr-Lern-Prozessen in die Berufs- und Wirtschaftspädagogik eingebracht. Im Rahmen dieser Erkenntnislogik ließe sich die Hypothesenformulierung als logische Abduktion, die Hypothesenprüfung (zumindest teilweise) als logische Deduktion und die Bewertung der Ursachenhypothese auf der Basis von Testergebnissen als Induktion rekonstruieren. Eine formallogische Rekonstruktion des Prozesses würde im Vergleich zum gewählten Vorgehen möglicherweise zu einer präziseren und eindeutigeren Prozessbeschreibung führen. Allerdings wäre damit eine höhere Abstraktion von den konkreten Prozessanforderungen verbunden, was sich sowohl auf die wissenschaftliche als auch die praktische Rezeption der Theorie negativ auswirken könnte. Es ist hier nicht abschließend möglich, die Vor- und Nachteile beider Rekonstruktionsvarianten gegeneinander abzuwägen und zu entscheiden, welche Variante in welchem Kontext angemessener ist.

#### **6.1.4 Diagnostischer Problemlöseprozess: Bedeutung der Informationsrepräsentation (Studie 4)**

In diesem Abschnitt werden die Ergebnisse der folgenden Publikation zusammengefasst:

Abele, S., Ostertag, R., Peissner, M. & Schuler, A. (2017). Eine Eye-Tracking-Studie zum diagnostischen Problemlöseprozess: Bedeutung der Informationsrepräsentation für das Lösen diagnostischer Probleme. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 113, 86–109.

Die vollständige Publikation findet sich in Anhang D.

### Thema und Forschungsfrage

Im Beitrag ging es insbesondere um die Frage, welche Bedeutung der Subprozess *Informationsrepräsentation* des diagnostischen Problemlöseprozesses für den diagnostischen Problemlöseerfolg hat. Im Vergleich zur vorherigen Studie wurde die Bedeutung dieses Subprozesses differenzierter analysiert. Ein Nebenziel der Studie war, das Potenzial von Eye-Tracking für die Untersuchung beruflicher Problemlöseprozesse auszuloten. Ausgehend von der oben beschriebenen Prozesstheorie wurde Informationsrepräsentation als wichtiger Subprozess des diagnostischen Problemlöseprozesses und die Qualität dieses Subprozesses als Determinante des diagnostischen Problemlöseerfolgs verstanden, wobei sich Informationsrepräsentation auf die mentale Repräsentation erfolgskritischer Informationen bezog. Wie bereits in Kapitel 6.1.3 erwähnt, kann die Qualität von Informationsrepräsentation anhand von erfolgskritischem Informationsverhalten operationalisiert werden, sofern die zu repräsentierenden Informationen der Umwelt entstammen und kein einschlägiges Langzeitwissen vorliegt. Im Unterschied zur vorherigen Studie wurde hier die Qualität dieses Subprozesses allerdings nicht mithilfe computerbasierter Log-Daten, sondern mithilfe von Blickdaten operationalisiert. Diese Blickdaten beziehen sich darauf, wie lange ein:e Problemlöser:in relevante Bereiche spezifischer Informationsmaterialien fixiert.

### Forschungshypothese

Im Anschluss an den einschlägigen Forschungsstand wurde theoretisch vermutet, dass die Qualität der Informationsrepräsentation für den diagnostischen Problemlöseerfolg bedeutsam ist. Empirisch wurde erwartet, dass erfolgreiche Problemlöser:innen relevante Bereiche von Informationsmaterialien länger fixieren als weniger erfolgreiche Problemlöser:innen, und zwar deshalb, weil sie mehr Zeit zur mentalen Repräsentation relevanter Informationen verwenden. Die Fixationsdauern wurden als Qualitätsindikatoren der Informationsrepräsentation interpretiert, wobei längere Fixationsdauern eine höhere Qualität anzeigen sollten. Der Hypothese unterlag die Annahme, dass die Proband:innen die relevanten Informationen in Interaktion mit der Umwelt gewinnen mussten und nicht aus dem Langzeitgedächtnis abrufen konnten.

### Methode

An der Pilotstudie nahmen 15 Auszubildende (Zahl männlich) zum/zur Kfz-Mechatroniker:in des dritten Ausbildungsjahrs teil. Die Proband:innen bearbeiteten zwei diagnostische Kfz-Probleme in der bereits mehrfach erwähnten Kfz-Computersimulation. Der Problemlöseerfolg wurde anhand schriftlicher Protokolle ermittelt, die von den Proband:innen während der Problembearbeitung erstellt wurden. Die Fixationszeiten wurden mit Eye-Trackern als kritische totale Fixationsdauern erfasst. Diese kritischen totalen Fixationsdauern standen für die Gesamtzeit, mit der ein:e Proband:in einen kritischen Bereich von Informationsmaterialien fixierte. Die kritischen Bereiche wurden ausgehend von erfolgskritischen diagnostischen Hypothesen theoriegeleitet bestimmt.

## Ergebnisse

Die kritischen Fixationsdauern erfolgreicher und weniger erfolgreicher diagnostischer Problemlöser:innen unterschieden sich teilweise signifikant und in beachtlichem Ausmaß, teilweise waren keine Unterschiede erkennbar. Außerdem waren mitunter hohe Korrelationen zwischen den totalen kritischen Fixationsdauern und dem Problemlöseerfolg zu beobachten.

## Diskussion

Die Ergebnisse bestätigten einerseits die Erwartungen, veranlassten andererseits aber eine tiefere theoretische Auseinandersetzung und führten letztlich zu einer Ausdifferenzierung der Forschungshypothese: Die Qualität der Informationsrepräsentation erfolgreicher und weniger erfolgreicher Problemlöser:innen unterscheidet sich wohl nicht generell, sondern unter anderem abhängig von den spezifischen Anforderungen eines diagnostischen Problems.

Unterschiede sind insbesondere bei Problemen zu erwarten, die von Routinefällen<sup>23</sup> abweichen oder eine ausgeprägte mentale Modellierung der Problemsituation erfordern. Auch bei anderen diagnostischen Problemen dürfte die Informationsrepräsentation eine wesentliche Rolle spielen. Die Bedeutung der Informationsrepräsentation lässt sich mit dem verwendeten Verfahren aber nur dann empirisch belegen, wenn Unterschiede zwischen Proband:innen auftreten. Falls z. B. alle Proband:innen die nötigen Informationen aus der Problemumwelt gewinnen oder in Form von Langzeitwissen aus dem Gedächtnis abrufen können, sind keine Unterschiede (direkt) beobachtbar. Der letzte Punkt verweist darauf, dass die Bedeutung der Informationsrepräsentation nicht nur von den Anforderungen des Problems, sondern auch von den Lernerfahrungen der Problemlöser:innen abhängt. Dieser Punkt wurde in der Studie genauso wenig adressiert wie das Potenzial weiterer Indikatoren, die aus Eye-Tracking-Daten gewonnen werden können. Die empirisch angeregte und theoretisch herausgearbeitete Bedeutung der mentalen Modellierung fügt sich gut in den Forschungsstand ein: In vielen berufs- und wirtschaftspädagogischen Studien wird unterstellt, dass die Qualität der mentalen Modellierung ganz entscheidend für berufliche Leistungen ist (z. B. Nickolaus, 2014, S. 289 f.; Winther, 2010, S. 236). Wie mentale Modelle erfasst werden können, wie sie sich entwickeln und wie die mentale Modellierung gefördert werden kann, ist in beruflichen Kontexten bislang allerdings wenig erforscht.

---

23 Wie in Kapitel 1.2 erwähnt, wird in Anlehnung an Jonassen (2000, S. 65) dann von einem Problem gesprochen, wenn in einer Situation eine Unbekannte zu finden ist. Sofern abgesichert werden kann, dass die Lösung eines spezifischen diagnostischen Problems unbekannt ist, ist es dementsprechend auch dann gerechtfertigt, von einem Problem zu sprechen, wenn das Problem mit einer Handlungsroutine gelöst wurde.

## 6.2 Studien zum Erhebungsmodell

### 6.2.1 Erfassung der diagnostischen Kfz-Problemlösekompetenz mit computersimulierten Key-Feature-Problemen und computersimulierten diagnostischen Problemen (Studie 5)

In diesem Abschnitt werden die Ergebnisse der folgenden Publikation zusammengefasst:

Abele, S., Walker, F. & Nickolaus, R. (2014). Zeitökonomische und reliable Diagnostik beruflicher Problemlösekompetenzen bei Auszubildenden zum Kfz-Mechatroniker. In S. Greiff, A. Kretzschmar & D. Leutner (Hrsg.), *Problemlösen in der Pädagogischen Psychologie, Zeitschrift für Pädagogische Psychologie*, 4, 167–179.

Die vollständige Publikation findet sich in Anhang E.

#### Thema und Forschungsfrage

Die Publikation beschäftigte sich mit dem Potenzial computersimulierter Key-Feature-Probleme zur Erfassung diagnostischer Kfz-Problemlösekompetenzen. Erstens wurde die Frage untersucht, ob diagnostische Problemlösekompetenzen mit computersimulierten Key-Feature-Problemen valide erfasst werden können. Zweitens wurde untersucht, ob die ergänzende Vorgabe von computersimulierten Key-Feature-Problemen zu einer höheren Reliabilität (internen Konsistenz) führt als die alleinige Vorgabe computersimulierter (holistischer) diagnostischer Probleme. Die Kombination von computersimulierten Key-Feature-Problemen und computersimulierten diagnostischen Kfz-Problemen stellt in Aussicht, die diagnostische Kfz-Problemlösekompetenz valide, reliabel und zugleich zeitökonomisch zu erfassen. Im Beitrag wurde nicht von Key-Feature-Problemen, sondern von Teilkompetenz-Items gesprochen. Außerdem wurde anstatt des Ausdrucks *diagnostische Problemlösekompetenz* der Ausdruck *Fehlerdiagnosekompetenz* verwendet.

#### Forschungshypothesen

Aus theoretischer Sicht wurde unterstellt, dass die diagnostische Kfz-Problemlösekompetenz mit den computersimulierten Key-Feature-Problemen valide erfasst werden kann. Empirisch wurde erwartet, dass die Testwerte der computersimulierten Key-Feature-Probleme und der computersimulierten diagnostischen Probleme konvergent valide, d. h. eindimensional, skalierbar sind. Vermutet wurde ferner, dass die gemeinsame Skalierung der Testwerte zu einer höheren Reliabilität führt als die alleinige Berücksichtigung der computersimulierten diagnostischen Probleme.

#### Methode

Um die Hypothesen zu prüfen, wurden 275 (Zahl männliche) Auszubildende zum/zur Kfz-Mechatroniker:in im dritten Ausbildungsjahr sechs diagnostische Kfz-Probleme und sieben Key-Feature-Probleme vorgelegt. Sowohl die diagnostischen Probleme als auch die Key-Feature-Probleme waren in der Kfz-Computersimulation zu bearbeiten. Bei den sechs diagnostischen Problemen betrug die Erhebungszeit 125 Minuten, bei

den sieben Key-Feature-Problemen 30 Minuten. Die Key-Feature-Probleme wurden im Anschluss an empirische Befunde entwickelt und so formuliert, dass sie in relativ kurzer Zeit zu bearbeiten waren und besonders leistungskritische Anforderungen des diagnostischen Kfz-Problemlösens abdecken. Für die Prüfung der konvergenten Validität wurde mithilfe konfirmatorischer Faktorenanalysen untersucht, ob die Testwerte der Key-Feature-Probleme und der diagnostischen Probleme statistisch eindimensional skalierbar sind. Dafür wurde die Passungsgüte eines zweidimensionalen Modells, in dem die Key-Feature-Testscores und Problem-Testscores separate, aber korrelierte Faktoren repräsentierten, mit der Passungsgüte eines eindimensionalen Modells verglichen, in dem alle Scores auf einen Faktor luden. Zudem wurde die SEM-Reliabilität (Maß für die interne Konsistenz) berechnet und geprüft, ob diese Reliabilität zunimmt, wenn die Key-Feature-Probleme zusätzlich zu den diagnostischen Problemen berücksichtigt werden.

### Ergebnisse

Bei der zweidimensionalen Faktorenanalyse korrelierten beide Faktoren substantiell ( $r = .89$ ). Die Fitmaße attestierten dem eindimensionalen Modell eine ähnlich gute Passung wie dem zweidimensionalen Modell. Der inferenzstatistische Vergleich beider Modelle erbrachte eine Irrtumswahrscheinlichkeit von  $p = .04$ . Abhängig vom gewählten Signifikanzniveau ( $\alpha = .01$  oder  $\alpha = .05$ ) ergaben sich also signifikante bzw. nicht signifikante Modellunterschiede. Bei Berücksichtigung der sechs diagnostischen Kfz-Probleme wurde eine SEM-Reliabilität von .62 erreicht. Wurden zusätzlich die computersimulierten Key-Feature-Probleme einbezogen, stieg die SEM-Reliabilität auf einen Wert von .75.

### Diskussion

Im Hinblick auf die konvergente Validität war die Befundlage nicht eindeutig. Unter Anwendung des Sparsamkeitsprinzips und eines nicht allzu strengen Signifikanzkriteriums wurde letztlich die eindimensionale Lösung präferiert und konvergente Validität der computersimulierten Key-Feature-Probleme und der computersimulierten diagnostischen Probleme unterstellt. Dies bedeutet auch, dass die Testwerte der computersimulierten Key-Feature-Probleme auf berufliche Kontexte übertragbar sind: Wie frühere Studien zeigten, erbringen die computersimulierten diagnostischen Kfz-Probleme Testwerte, die sich auf (reale) berufliche Kontexte, d. h. auf reales diagnostisches Kfz-Problemlösen übertragen lassen (Kapitel 4.2.3). Damit kann die ermittelte konvergente Validität als Indiz dafür gelten, dass auch die Key-Feature-Testwerte auf berufliche Kontexte übertragbar sind.

Eindeutig sind die Befunde mit Blick auf die Reliabilität (interne Konsistenz): Der ergänzende Einsatz der computersimulierten Key-Feature-Probleme führte im Vergleich zur alleinigen Berücksichtigung der diagnostischen Kfz-Probleme zu einer beachtlichen Verbesserung der Reliabilität und das, obwohl die Erhebungszeit nur in überschaubarem Maße anstieg. Dies sprach dafür, dass die Testwerte der kombinierten Kompetenzerfassung eine höhere Generalisierbarkeit aufweisen als die Testwerte, die ausschließlich auf den diagnostischen Problemen beruhen.

Vor dem Hintergrund der in Kapitel 4.2 erwähnten Studien kann gefragt werden, ob es mit Blick auf die Studie gerechtfertigt ist, von Key-Feature-Problemen zu sprechen. Die betreffenden Items der Studie decken einen leistungskritischen Bereich authentischer diagnostischer Kfz-Probleme ab, sie knüpfen aber nicht direkt an authentische (holistische) diagnostische Probleme an und waren in relativ kurzer Zeit zu bearbeiten. Außerdem wurden einige Items in einem Multiple-Choice-Format vorgegeben. Insofern wäre es vermutlich treffender, die Items als eine Mischung aus Key-Feature-Problemen und Multiple-Choice-Items zu bezeichnen. Wenn hier dennoch von Key-Feature-Problemen gesprochen wird, geschieht dies mit der Absicht, auf den im Vergleich zu konventionellen Multiple-Choice-Items hohen Authentizitätsgrad der betreffenden Items abzuheben. Ferner soll damit darauf verwiesen werden, dass die betreffenden Items leistungskritische Anforderungen (realer) diagnostischer Kfz-Probleme abdecken.

Key-Feature-Probleme weisen einen Weg, wie berufliche Kompetenzen in akzeptablen Testzeiten valide erfasst werden können. Damit wird ein drängendes berufs- und wirtschaftspädagogisches Forschungsproblem aufgegriffen. Aus didaktischer Sicht bergen sie aber auch ein gewisses Risiko in sich: Sie könnten anregen, diagnostische Kfz-Problemlösekompetenzen mit Key-Feature-Problemen zu fördern. Auch wenn computersimulierte Key-Feature-Probleme eine reliable Diagnose von Lernbarrieren ermöglichen und somit einen gewissen didaktischen Nutzen versprechen, wurde dieser Aspekt hier nicht systematisch untersucht. Es spricht einiges dafür, bei der Förderung vor allem auf holistische diagnostische Kfz-Probleme zu setzen (van Merriënboer, 2013). Welche Folgen der Einsatz von Key-Feature-Problemen in Förderkontexten hat, ob und unter welchen Bedingungen solche Key-Feature-Probleme didaktisch fruchtbar sein können und ob sich daraus negative Folgen für berufliche Lehr-Lern-Prozesse ergeben, sind wichtige didaktische Fragen, die hier allerdings nicht vertieft werden können. Im vorliegenden Kontext steht das Potenzial von Key-Feature-Problemen für die Erfassung diagnostischer Problemlösekompetenzen im Zentrum. Besonders interessant ist in diesem Kontext, ob diagnostische Problemlösekompetenzen auch mit papierbasierten Key-Feature-Problemen valide erfasst werden können. In diesem Fall ließen sich die Erhebungszeiten, aber auch der finanzielle und zeitliche Entwicklungsaufwand enorm reduzieren.

### **6.2.2 Erfassung diagnostischer Problemlösekompetenzen mit schriftlichen Key-Feature-Problemen (Studie 6)**

In diesem Abschnitt werden die Ergebnisse der folgenden Publikation zusammengefasst:

Abele, S. (2016). Can Diagnostic Problem-Solving Competences of Car Mechatronics Be Validly Assessed Using a Paper-Pencil Test? *Journal of Technical Education (JOTED)*, 4, 190–211.

Die vollständige Publikation findet sich in Anhang F.

### Thema und Forschungsfrage

Die Publikation beschäftigte sich mit dem Potenzial von *papierbasierten* (schriftlichen) Key-Feature-Problemen zur Erfassung diagnostischer Kfz-Problemlösekompetenzen. Es wurde untersucht, ob diagnostische Problemlösekompetenzen mit authentischen schriftlichen Key-Feature-Problemen valide erfasst werden können. Wie bereits erwähnt, versprechen solche Key-Feature-Probleme im Vergleich zu computersimulierten Key-Feature-Problemen beachtliche finanzielle und zeitliche Vorteile. In der englischsprachigen Publikation wird nicht von papierbasierten Key-Feature-Problemen, sondern von „paper-based key-feature items“ gesprochen.

### Forschungshypothese

Theoretisch wurde angenommen, dass die diagnostische Problemlösekompetenz mit den authentischen papierbasierten Key-Feature-Problemen valide erfasst werden kann. Empirisch wurde unterstellt, dass die Testwerte der schriftlichen Key-Feature-Probleme statistisch eindimensional mit den Testscores der computersimulierten diagnostischen Probleme skalierbar sind (konvergente Validität). Darüber hinaus wurde eine akzeptable Reliabilität (interne Konsistenz) der papierbasierten Testwerte erwartet.

### Methode

Die Forschungshypothese wurde mit 206 (Zahl männlich) Auszubildenden zum/zur Kfz-Mechatroniker:in im dritten Ausbildungsjahr geprüft. Die Proband:innen bearbeiteten acht diagnostische Kfz-Probleme, vier in der Kfz-Computersimulation und vier unter berufsaauthentischen Bedingungen an einem Fahrzeug in einer Werkstatt. Aufgrund der sehr hohen latenten Korrelation der computerbasierten und der Werkstatt-Testwerte ( $r = .94$ ) wurde die konvergente Validität beider Skalen unterstellt und die Testwerte wurden zusammengefasst. Die zusammengefassten Testwerte wurden als Indikatoren der (realen) diagnostischen Kfz-Problemlösekompetenz interpretiert. Zusätzlich zu den diagnostischen Problemen wurden in der Studie 22 authentische schriftliche Key-Feature-Probleme entwickelt. Die Entwicklung der Key-Feature-Probleme erfolgte mit Blick auf reale berufliche Anforderungen und die Subprozesse des diagnostischen Problemlöseprozesses. Um eine möglichst hohe Generalisierbarkeit zu erzielen, wurden für jeden Subprozess schriftliche Key-Feature-Probleme entwickelt, wobei eine Gleichverteilung der Key-Feature-Probleme auf die Subprozesse angestrebt, wegen der restringierten Möglichkeiten der Papier-Bleistift-Methode aber nicht vollständig realisiert werden konnte. Aufgrund der langen Gesamterhebungszeit (350 min) wurden die schriftlichen Key-Feature-Probleme nur einer Substichprobe ( $N = 121$ ) vorgelegt. Die psychometrische Güte der schriftlichen Key-Feature-Probleme wurde anhand klassischer Itemstatistiken sowie der Passung der Testcores zum Partial-Credit-Modell (spezifisches Modell der Item-Response-Theorie) analysiert. Um die Hypothese zur konvergenten Validität zu prüfen, wurde die Passungsgüte eines eindimensionalen Partial-Credit-Modells mit der Passung eines zweidimensionalen Partial-Credit-Modells verglichen und darüber hinaus die Korrelation im zweidimensiona-

len Modell inspiziert, in dem die schriftlichen Key-Feature-Probleme und die computersimulierten diagnostischen Kfz-Probleme separate Dimensionen aufspannten.

### Ergebnisse

Zusammengenommen sprachen die statistischen Kennwerte für eine akzeptable psychometrische Güte der schriftlichen Key-Feature-Probleme. Die Testscores erreichten eine Reliabilität von  $EAP/PV = .75$  (Maß der internen Konsistenz). Bei der Prüfung der konvergenten Validität erbrachte das zweidimensionale Modell, in dem beide Dimensionen mit  $r = .76$  korrelierten, eine bessere Passung als das eindimensionale Modell. Auch beim inferenzstatistischen Vergleich beider Modellvarianten erwies sich das zweidimensionale Modell als überlegen.

### Diskussion

Im Vergleich zu bisherigen Studien (z. B. Gschwendtner et al., 2009) kann die Reliabilität der Key-Feature-Probleme als zufriedenstellend gelten (s. dazu Moosbrugger & Kelava, 2008, S. 11). Beobachtet wurde zudem, dass die schriftlichen Key-Feature-Probleme erwartungsgemäß substanziell mit der diagnostischen Kfz-Problemlösekompetenz korrelierten. Der Modellvergleich lieferte aber gewichtige Indizien gegen die konvergente Validität der schriftlichen Key-Feature-Probleme und der computersimulierten diagnostischen Probleme. Streng genommen wurde die Forschungshypothese also durch die Studienergebnisse widerlegt. Angesichts der Befunde und vor dem Hintergrund eines argumentbasierten Validitätsansatzes (s. Kapitel 4.1) würde es aber zu weit gehen, die schriftlichen Key-Feature-Probleme bzw. die darauf basierenden Testwertinterpretationen als nicht valide und zur Erfassung diagnostischer Kfz-Problemlösekompetenzen generell als ungeeignet aufzufassen. Ob die Papier-basierten Testwerte als Indikatoren der diagnostischen Problemlösekompetenz interpretiert werden können, ist vor dem Hintergrund des verfolgten Ziels und anhand eines Validitätsarguments zu beurteilen. Wie bereits in Kapitel 4.1 betont, erlaubt ein Validitätsargument kein dichotomes (valide vs. nicht valide), sondern eher ein graduelles Validitätsurteil. In einer pragmatischen Perspektive könnte die hohe Korrelation von  $r = .76$  als Konvergenzevidenz gewertet werden und den Einsatz schriftlicher Key-Feature-Probleme bei spezifischen Zielstellungen rechtfertigen. Wenn es darum geht, Testwerte mit maximaler Übertragbarkeit auf berufliche Kontexte zu gewinnen, scheinen schriftliche Key-Feature-Probleme weniger geeignet als computersimulierte diagnostische Probleme. Dagegen weisen sie mit Blick auf die interne Konsistenz (Reliabilität) und damit die Generalisierbarkeit bessere Eigenschaften auf.

Auch diese Studie erlaubt keine Aussagen zum didaktischen Potenzial von Key-Feature-Problemen. Der diagnostische Einsatz schriftlicher Key-Feature-Probleme sollte nicht als Plädoyer verstanden werden, berufliche Lehr-Lern-Prozesse zu zergliedern und unterkomplex sowie wenig authentisch anzulegen. Er stellt vielmehr einen pragmatischen Ansatz dar, die oft belastenden Testzeiten bei der Erfassung berufliche Kompetenzen unter (weitgehender) Wahrung des Validitätsanspruchs zu reduzieren.

## 7 Diskussion

Das Thema der vorliegenden Untersuchung ist die Erfassung diagnostischer Problemlösekompetenzen in beruflichen Kontexten. Wie die Ausführungen in Kapitel 1.3 zeigen, stellt die Erfassung diagnostischer Problemlösekompetenzen in beruflichen Kontexten ein wichtiges berufs- und wirtschaftspädagogisches Forschungsthema dar. In Kapitel 2 wurde die Erfassung diagnostischer Problemlösekompetenzen als Modellierungsprozess rekonstruiert, der darauf abzielt, sprachliche Aussagen über diagnostische Problemlösekompetenzen zu gewinnen. Um solche Aussagen zu erhalten, müssen mehrere Modellierungsschritte durchgeführt bzw. Modelle gebildet werden. In der vorliegenden Untersuchung standen das Sprachmodell und das Erhebungsmodell im Zentrum: Es wurde untersucht, wie beide Modelle zu realisieren bzw. zu konkretisieren sind. Das Sprachmodell und das Erhebungsmodell wurden jeweils unter Anwendung theoretisch begründeter Kriterien anhand systematischer Überlegungen und im Rückgriff auf den empirischen Forschungsstand konkretisiert. Die Konkretisierung des Sprachmodells in Kapitel 3 lieferte eine Definition diagnostischer Problemlösekompetenzen und offenbarte mehrere Forschungsdesiderate.

Die diagnostische Problemlösekompetenz wurde begrifflich als von anderen beruflichen Kompetenzen abgrenzbare und metrisch skalierbare Gesamtheit kognitiver und nichtkognitiver psychischer Komponenten gefasst, die in den diagnostischen Problemlöseprozess zu investieren ist, um diagnostische Probleme einer berufsbezogenen Anforderungsklasse zu lösen. Es wurde also eine anforderungsklassenspezifische Definition gewählt und unterstellt, dass es so viele diagnostische Problemlösekompetenzen gibt, wie Anforderungsklassen des diagnostischen Problemlösens unterschieden werden können. Als defizitär erwies sich der Forschungsstand im Bereich des Sprachmodells, unter anderem im Hinblick auf die empirische Eigenständigkeit diagnostischer Problemlösekompetenzen, die psychischen Komponenten diagnostischer Problemlösekompetenzen sowie auf den Prozess, der dem diagnostischen Problemlöseerfolg zugrunde liegt bzw. der diesen Erfolg verursacht.

Bei der Konkretisierung des Erhebungsmodells in Kapitel 4 zeigte sich unter anderem, dass computersimulierte (holistische) diagnostische Probleme eine aussichtsreiche Möglichkeit darstellen, diagnostische Problemlösekompetenzen valide zu erfassen. Key-Feature-Probleme versprechen eine reliablere (höhere interne Konsistenz) und zeitökonomischere Erfassung der diagnostischen Problemlösekompetenz als computersimulierte holistische diagnostische Probleme. Wie valide die dabei entstehenden Aussagen über die diagnostische Problemlösekompetenz sind, kann angesichts der dünnen empirischen Befundlage in diesem Bereich nicht befriedigend eingeschätzt werden.

Die erwähnten Forschungsdesiderate wurden mit der vorliegenden Untersuchung aufgegriffen und in der Ausbildung zum/zur Kfz-Mechatroniker:in im Bereich des Lösens elektrotechnischer diagnostischer Kfz-Probleme in sechs quantitativ-empirischen Studien bearbeitet, wobei der Forschungslogik des Kritischen Rationalismus gefolgt wurde. Welche Erkenntnisse brachten die sechs Studien?

## 7.1 Zusammenfassung der Erkenntnisse

### 7.1.1 Sprachmodell der elektrotechnischen diagnostischen Kfz-Problemlösekompetenz

#### Diagnostische Kfz-Problemlösekompetenz: eine empirisch eigenständige berufliche Kompetenz?

Mit der ersten Studie (Kapitel 6.1.1) zum Sprachmodell wurde die Forschungsfrage untersucht, ob die elektrotechnische diagnostische Kfz-Problemlösekompetenz empirisch von anderen beruflichen Kompetenzen unterscheidbar ist. Mit dieser Frage wurden Forschungsdesiderate aufgegriffen, die in Kapitel 3.3.2 sichtbar wurden. Dort wurde resümiert, dass die Befundlage zur empirischen Eigenständigkeit der diagnostischen Problemlösekompetenz unbefriedigend ist. Insbesondere im Kfz-Bereich veranlassten einige Befunde, an der empirischen Eigenständigkeit der diagnostischen Problemlösekompetenz zu zweifeln. Viele einschlägige Studien waren außerdem mit einem forschungsmethodischen Problem behaftet: Die empirische Eigenständigkeit der diagnostischen Problemlösekompetenz wurde meist rein korrelationsanalytisch und nicht anhand mehrdimensionaler statistischer Messmodelle untersucht.

Die Befunde der ersten Studie wurden mithilfe mehrdimensionaler Messmodelle gewonnen und legten nahe, dass die elektrotechnische diagnostische Kfz-Problemlösekompetenz im Ausbildungsberuf zum/zur Kfz-Mechatroniker:in von anderen beruflichen Kompetenzen empirisch unterschieden werden kann; gleichzeitig wurden aber auch deutliche Korrelationen zwischen den berücksichtigten beruflichen Kompetenzen ermittelt. Insgesamt scheint es vertretbar, von der empirischen Eigenständigkeit der elektrotechnischen diagnostischen Kfz-Problemlösekompetenz auszugehen.

#### Umgang mit Komplexität: eine bedeutsame psychische Komponente der diagnostischen Kfz-Problemlösekompetenz?

In der zweiten Studie (Kapitel 6.1.2) wurde der Forschungsfrage nachgegangen, ob der Umgang mit Komplexität als bedeutsame psychische Komponente der diagnostischen Kfz-Problemlösekompetenz aufgefasst werden kann. Diese Forschungsfrage ergab sich aus den Ausführungen in Kapitel 3.3.3. Dort wurde unter anderem festgestellt, dass bisherige Studien nur eine begrenzte Auswahl psychischer Komponenten der diagnostischen Problemlösekompetenz fokussiert haben. Darüber hinaus wurde ein Design als besonders aussichtsreich identifiziert, bei dem die empirische Bedeutung der psychischen Komponenten für die diagnostische Problemlösekompetenz anhand eines integrativen Untersuchungsdesigns untersucht wird. *Integrativ* heißt, dass die psychischen Komponenten während der Bearbeitung diagnostischer Probleme, also im Problemlöseprozess, erfasst wurden.

Die Befunde der zweiten Studie können als eine empirische Evidenz dafür gelten, den Umgang mit Komplexität als bedeutsame psychische Komponente der diagnostischen Problemlösekompetenz aufzufassen. Diese Befunde resultierten aus einem integrativen Untersuchungsdesign. Bei diesem Design wurden unterschiedliche komplexe Probleme vorgegeben, die empirischen Schwierigkeiten der Probleme ermittelt

und ein deutlicher Zusammenhang zwischen der Problemkomplexität und der Problemschwierigkeit angenommen. Dieser Zusammenhang wurde bestätigt und auf Basis theoretischer Überlegungen als Indiz für die Bedeutung des Umgangs mit Komplexität interpretiert. Ob dieses durchaus verwickelte Untersuchungsdesign, das in Anlehnung an Borsboom et al. (2004) entstand, zukunftsfruchtig ist, müssen weitere Studien zeigen. Als problematisch stellte sich die Operationalisierung der Problemkomplexität heraus, weil die objektive Rekonstruktion der Problemkomplexität eine elaborierte, einschlägige Operationalisierungstheorie erfordert, die aktuell allerdings nicht verfügbar ist. Auch wenn es sehr plausibel scheint, den Umgang mit Komplexität als eine wichtige Komponente der diagnostischen Problemlösekompetenz zu betrachten, bestehen diesbezüglich – unter anderem wegen der erwähnten Probleme im Bereich der Operationalisierung – noch Unsicherheiten.

### **Theorie zum diagnostischen Problemlöseprozess: der Problemlöseerfolg als Wirkung eines Prozesses**

In der dritten Studie (Kapitel 6.1.3) wurde untersucht, ob und wie der diagnostische Problemlöseerfolg als kausales Produkt (Wirkung) des diagnostischen Problemlöseprozesses rekonstruiert werden kann. Im Zentrum der Studie stand eine Theorie zum diagnostischen Problemlöseprozess, die den Problemlöseerfolg erklärt. Ausgangspunkt dieses Forschungsthemas waren die Analysen in Kapitel 3.3.5, die einerseits die Notwendigkeit einer solchen Prozesstheorie verdeutlichten und andererseits aufzeigten, dass eine solche Theorie aktuell nicht verfügbar ist. Als notwendig erscheint eine Prozesstheorie zur Erklärung des Diagnoseerfolgs an dieser Stelle besonders deshalb, weil der diagnostische Problemlöseerfolg hier per definitionem als Indikator der diagnostischen Problemlösekompetenz gilt und zudem festgestellt wurde, dass die diagnostische Problemlösekompetenz sprachlich auch unter dem Prozessaspekt zu fassen ist.

Die Befunde der dritten Studie standen weitgehend im Einklang mit der entwickelten Theorie. Dies wurde zum Anlass genommen, den diagnostischen Problemlöseerfolg als Wirkung eines Problemlöseprozesses aufzufassen, der die Sub-Prozesse *Informationsrepräsentation*, *Hypothesenentwicklung*, *Hypothesentest* und *Hypothesenbewertung* umfasst, die in einem spezifischen Kausalverhältnis stehen.

Grenzen weist die Studie insofern auf, als die Theorieprüfung lediglich auf zwei diagnostischen Kfz-Problemen basiert. Fraglich ist zudem, ob das Verfahren, mit dem das erfolgskritische Informationsverhalten und das erfolgskritische Testverhalten bestimmt, d. h. die Qualität der Subprozesse *Informationsrepräsentation* und *Hypothesentest* operationalisiert wurde, auf andere diagnostische Kfz-Probleme übertragbar ist. Darüber hinaus ist in weiteren Studien zu prüfen, ob die Theorie zum diagnostischen Problemlöseprozess auf andere berufliche Kontexte transferiert werden kann. Dies legt der domänenübergreifende Aspekt der Theorie nahe, die ausgehend vom einschlägigen theoretischen und empirischen Forschungsstand sowie unterschiedlichen beruflichen Kontexten entwickelt wurde. Eine empirische Prüfung dieses Aspekts erfolgte aber lediglich im Kfz-Kontext und auch dort nur mit den bereits genannten Einschränkungen.

kungen. Insgesamt scheint es vertretbar, die Theorie im Kfz-Kontext heranzuziehen, um das Sprachmodell der diagnostischen Problemlösekompetenz unter dem Prozessaspekt zu präzisieren.

### **Informationsrepräsentation: ein bedeutsamer Subprozess des diagnostischen Problemlöseprozesses?**

Mit der vierten Studie (Kapitel 6.1.4) wurde ebenfalls der diagnostische Problemlöseprozess untersucht. Hier ging es um die Frage, ob der Subprozess „Informationsrepräsentation“ für den diagnostischen Problemlöseprozess und den diagnostischen Problemlöseerfolg bedeutsam ist. Dieser Subprozess wurde auf Basis der oben erwähnten Prozesstheorie modelliert und mithilfe des erfolgskritischen Informationsverhaltens operationalisiert. Im Unterschied zur dritten Studie wurden aber keine Log-Daten, sondern Blickdaten (Eye-Tracking-Daten) verwendet. Mit der Studie wurde ebenfalls das in Kapitel 3.3.5 identifizierte Forschungsdesiderat aufgegriffen, dass gegenwärtig keine Prozesstheorie zur Erklärung des diagnostischen Problemlöseerfolgs vorliegt.

Die Befunde der Pilotstudie belegten teilweise, dass die Informationsrepräsentation für den diagnostischen Problemlöseerfolg bedeutsam ist; teilweise widersprachen sie aber auch der Forschungshypothese. Als mögliche Erklärung für die widersprüchlichen Befunde wurden die Anforderungen diagnostischer Probleme genannt. Diese können sehr unterschiedlich sein und dazu führen, dass die Informationsrepräsentation für den Problemlöseerfolg zwar bedeutsam, aber nicht erfolgskritisch ist. Eine weitere, anhand einer kritischen Reflexion der Befunde gewonnene Erkenntnis bestand darin, dass die Bedeutung der Informationsrepräsentation zunimmt, je intensiver eine Problemanforderung die mentale Modellierung der Problemumwelt verlangt. Dementsprechend ist die mentale Modellierung des diagnostischen Problems als wichtiger Erfolgsfaktor aufzufassen. Neben den Problemanforderungen spielen auch Lernerfahrungen eine wichtige Rolle. Falls Informationen von sehr vielen Probanden aus dem Langzeitgedächtnis abgerufen werden können, sind sie zwar für die Problemlösung relevant, die Bedeutung kann aber nicht mit statistischen Verfahren belegt werden, die interindividuelle Unterschiede fokussieren. Dies ist bei der Erfassung der diagnostischen Problemlösekompetenz zu beachten.

Eine deutliche Grenze der Studie liegt in der geringen Zahl von Proband:innen und darin, dass die Möglichkeiten von Eye-Tracking nur bedingt genutzt wurden. Neben totalen kritischen Fixationsdauern können aus Blickdaten viele weitere interessante Maße gewonnen werden. Trotz dieser Grenzen können die Studienergebnisse als empirische Belege der Bedeutung der Informationsrepräsentation für den diagnostischen Problemlöseerfolg aufgefasst werden. Der Pilotcharakter der Studie und die erwähnten Grenzen weisen aber gleichzeitig darauf hin, wie wichtig ein vorsichtiger Umgang mit den Studienergebnissen und weitere Studien sind.

## 7.1.2 Erhebungsmodell der elektrotechnischen diagnostischen Kfz-Problemlösekompetenz

### Computersimulierte Key-Feature-Probleme: eine Möglichkeit, die diagnostische Kfz-Problemlösekompetenz valide, reliabel und zeitökonomisch zu erfassen?

Die fünfte Studie (Kapitel 6.2.1) widmete sich der Frage, ob die diagnostische Kfz-Problemlösekompetenz mit computersimulierten Key-Feature-Problemen valide erfasst werden kann und ob die Kombination dieser Key-Feature-Probleme mit computersimulierten diagnostischen Kfz-Problemen in einer vertretbaren Erhebungszeit zu einer akzeptablen Reliabilität (internen Konsistenz) führt. Diese Frage wurde im Anschluss an Kapitel 4.3 entwickelt und bezieht sich auf das dort identifizierte Forschungsproblem, dass computersimulierte diagnostische Kfz-Probleme einen großen Entwicklungsaufwand verursachen, mit hohen Erhebungszeiten verbunden sind und bislang teilweise keine reliablen (konsistenten) Testwerte erbracht haben.

Die Ergebnisse der Studie waren nicht (ganz) eindeutig, wurden aber so interpretiert, dass die computersimulierten Key-Feature-Probleme valide Testwertinterpretationen, d. h. valide Aussagen über die diagnostische Kfz-Problemlösekompetenz ermöglichen. Der ergänzende Einsatz der computersimulierten Key-Feature-Probleme zu diagnostischen Kfz-Problemen führte zu einer beachtlichen Verbesserung und letztlich zufriedenstellenden Reliabilität (internen Konsistenz). Besonders relevant ist dieser Reliabilitätzuwachs deshalb, weil die sieben computersimulierten Key-Feature-Probleme erheblich weniger Erhebungszeit beanspruchten (30 Minuten) als die sechs computersimulierten Kfz-Probleme (125 Minuten).

Diskussionswürdig erscheint insbesondere die Frage, ob die auf computersimulierten Key-Feature-Problemen basierenden Testwerte und die mit computersimulierten diagnostischen Problemen gewonnenen Testwerte als konvergent valide gelten dürfen. Die Antwort auf diese Frage hängt davon ab, welche Kriterien an die Befunde angelegt und wie die Kriterien gegeneinander ausbalanciert werden: Bei Anwendung eines strengen Signifikanzkriteriums ( $\alpha = .01$ ) sprechen die Befunde eher gegen, bei Anwendung eines weniger strengen Signifikanzkriteriums ( $\alpha = .05$ ) eher für die konvergente Validität. Aus pragmatischer Sicht kann die sehr hohe Korrelation ( $r = .89$ , latent) zwischen den Testwerten beider Problemtypen zum Anlass genommen werden, konvergente Validität zu unterstellen. Aus Anforderungssicht kann angenommen werden, dass Key-Feature-Probleme das Zielkonstrukt (diagnostische Problemlösekompetenz) schlechter repräsentieren als diagnostische Probleme, was für die diskriminante Validität spricht. Im Sinne des Einfachheitskriteriums kann wiederum argumentiert werden, die Befunde rechtfertigten die Annahme der konvergenten Validität. Auf mögliche negative Konsequenzen von Key-Feature-Problemen in didaktischen Kontexten wurde in Kapitel 6.2.1 hingewiesen. Mit Blick auf diese negativen didaktischen Konsequenzen könnten die Befunde dazu führen, bei der Erfassung diagnostischer Problemlösekompetenzen auf Key-Feature-Probleme zu verzichten.

In der Studie ist die konvergente Validität der computersimulierten Key-Feature-Probleme und der computersimulierten diagnostischen Probleme gleichbedeutend mit der Übertragbarkeit der Ergebnisse auf berufliche Kontexte. In Anlehnung an bis-

herige Studien (z. B. Gschwendtner et al., 2009; Abele et al., 2016, S. 189) wurden jene Aussagen als übertragbar erachtet, die anhand der computersimulierten diagnostischen Probleme gewonnenen wurden. Dementsprechend können im Falle der konvergenten Validität auch die mit computersimulierten Key-Feature-Problemen gewonnenen Aussagen als übertragbar gelten.

Die Frage der konvergenten Validität ist für die Interpretation der Studienergebnisse entscheidend. Computersimulierte Key-Feature-Probleme ermöglichen nur dann eine zeitökonomische und reliable Erfassung der diagnostischen Problemlösekompetenz, wenn sie die diagnostische Problemlösekompetenz valide erfassen. Eine wichtige Voraussetzung dafür ist die Übertragbarkeit der Erfassungsergebnisse, also der sprachlichen Aussagen über die diagnostische Problemlösekompetenz, auf berufliche Kontexte. Falls an dieser Übertragbarkeit ernsthafte Zweifel bestehen, ist die Aussagenvalidität gefährdet. Vor diesem Hintergrund hängt nicht nur die Einschätzung der konvergenten Validität, sondern die Antwort auf die gesamte Forschungsfrage der Studie davon ab, welches der oben genannten Kriterien (Signifikanzkriterium, Einfachheitskriterium etc.) zur Beurteilung der Studienergebnisse herangezogen wird. Die Studienergebnisse erlauben also keine eindeutige Antwort darauf, ob computersimulierte Key-Feature-Probleme eine valide, reliable und zugleich zeitökonomische Erfassung der diagnostischen Problemlösekompetenz ermöglichen.

### **Papierbasierte Key-Feature-Probleme: eine Möglichkeit, die diagnostische Kfz-Problemlösekompetenz valide zu erfassen?**

In der sechsten und letzten Studie (Kapitel 6.2.2) wurde untersucht, ob die diagnostische Kfz-Problemlösekompetenz mit *papierbasierten* Key-Feature-Problemen valide erfasst werden kann. Diese Forschungsfrage ergab sich aus den Ausführungen in Kapitel 4.3 und einer pragmatischen Perspektive: Im Vergleich zu computersimulierten diagnostischen Kfz-Problemen können schriftliche Key-Feature-Probleme deutlich schneller und kostengünstiger entwickelt werden. Außerdem gehen sie mit geringeren Erhebungszeiten einher und sie haben dennoch das Potenzial, die diagnostische Problemlösekompetenz relativ authentisch zu erfassen.

In der Studie erbrachten die Testwerte der papierbasierten Key-Feature-Probleme eine zufriedenstellende Reliabilität (interne Konsistenz). Zudem legten die Ergebnisse der Studie nahe, dass diese Testwerte substanziell mit den Testwerten der computersimulierten diagnostischen Kfz-Problemen korrelierten ( $r = .76$ ), letztlich aber diskriminant valide sind. Damit sprachen die Befunde (interne Konsistenz) einerseits für die Generalisierbarkeit der Aussagen, die mit den papierbasierten Key-Feature-Problemen gewonnen wurden. Andererseits ließen sie aber an deren Übertragbarkeit auf berufliche Kontexte zweifeln.

Auch in der sechsten Studie hing die Antwort auf die Forschungsfrage davon ab, ob die Testwerte der schriftlichen Key-Feature-Probleme und der computersimulierten diagnostischen Kfz-Probleme als konvergent valide gelten dürfen. Auch hier ist letztlich entscheidend, welche der oben genannten Kriterien an die Befunde angelegt werden. Aus einer pragmatischen Sicht und im Sinne des Einfachheitskriteriums könnte

von konvergenter Validität ausgegangen werden. Allerdings ist die Befundlage hier insofern eindeutig, als sowohl die Passungsgüte der Modelle als auch der Modellvergleichstest gegen eine (vollständige) konvergente Validität sprachen. Welche Schlüsse können daraus und aus den anderen Untersuchungsergebnissen für die Erfassung diagnostischer Problemlösekompetenzen gezogen werden?

## **7.2 Implikationen der Ergebnisse für die Erfassung diagnostischer Problemlösekompetenzen**

### **7.2.1 Sprachmodell**

#### **Empirische Eigenständigkeit als Legitimationsgrundlage**

Die Befunde zur empirischen Eigenständigkeit der diagnostischen Kfz-Problemlösekompetenz können als Legitimationsgrundlage betrachtet werden, sich wissenschaftlich mit der Erfassung diagnostischer Problemlösekompetenzen zu beschäftigen. Mit Blick auf die Vielzahl an Kompetenzen, die in der einschlägigen Literatur unterschieden werden (s. dazu z. B. Zlatkin-Troitschanskaia & Seidel, 2011), und auf das Einfachheitskriterium (Kapitel 3.2.3) ist die empirische Eigenständigkeit von Kompetenzen (sehr) bedeutsam. Auch im Sinne der theoretischen Fruchtbarkeit (Kapitel 3.2.2), d. h. des Erkenntnisfortschritts, ist es nicht zweckmäßig, mehr Kompetenzen zu postulieren, zu erfassen und zu untersuchen als nötig.

#### **Theorie zum diagnostischen Problemlöseprozess**

Einen wichtigen Erkenntnisfortschritt der Untersuchung repräsentiert die Theorie zum diagnostischen Problemlöseprozess. Es spricht vieles dafür, die entwickelte Prozesstheorie im Kfz-Kontext heranzuziehen, um das Sprachmodell der diagnostischen Kfz-Problemlösekompetenz unter dem Prozessaspekt zu konkretisieren. Die Theorie wurde im Kfz-Kontext geprüft, aber auf Basis domänenübergreifender theoretischer Erkenntnisse und empirischer Befunde entwickelt und könnte von daher auch in anderen beruflichen Kontexten fruchtbar sein, in denen diagnostische Probleme zu lösen sind. Letztlich ist dies aber im Horizont der jeweils verfolgten Fragestellung und der konkreten Anforderungen der betreffenden diagnostischen Probleme zu entscheiden.

#### **Bedeutung des Sprachmodells**

Die Bedeutung des Sprachmodells für die Erfassung diagnostischer Problemlösekompetenzen ist kaum zu überschätzen. Besonders deutlich wird dies am Beispiel der Generalisierbarkeit von Aussagen über die diagnostische Problemlösekompetenz. Ob Aussagen generalisierbar sind, lässt sich nur entscheiden, wenn klar ist, worauf generalisiert werden soll. Dies wird im Sprachmodell definiert. Mit dem Sprachmodell wird festgelegt, welche psychischen Komponenten die diagnostische Problemlösekompetenz umfasst, auf welchen Prozess sich diese Kompetenz bezieht und welche externen Anforderungen damit angesprochen sind. Nur wenn diese Aspekte klar sind, kann entschieden werden, ob die Problemauswahl das Zielkonstrukt (diagnostische Problemlösekompetenz) repräsentativ abdeckt. Erkennbar wird die Bedeutung des Sprachmo-

dells auch mit Blick auf die Testwertinterpretation. Wie Testwerte der diagnostischen Problemlösekompetenz inhaltlich interpretiert werden dürfen, lässt sich nur dann sinnvoll sagen, wenn klar ist, was unter diagnostischer Problemlösekompetenz verstanden wird. Testwerte können nur das anzeigen, was im Sprachmodell angelegt wurde. Zu bedenken ist ferner, dass alle Entscheidungen, die bei der Gestaltung des Erhebungsmodells, des Messmodells und des Interpretationsmodells getroffen werden, letztlich am Sprachmodell zu relativieren sind. So hat es z. B. keinen Sinn, unabhängig vom Sprachmodell über die Auswahl von Items, den Erhebungsmodus oder die Itemkodierung nachzudenken.

Wäre es angesichts der großen Bedeutung des Sprachmodells angezeigt, bei der Erfassung diagnostischer Problemlösekompetenzen das *sprachliche Relativ* (stärker) zu berücksichtigen? In der Messtheorie hat es sich eingebürgert, zwischen einem *empirischen Relativ* und einem *numerischen Relativ* zu unterscheiden (z. B. Gigerenzer, 1981, S. 36 ff.). Mit dem empirischen Relativ sind jene Relationen angesprochen, die zwischen spezifischen Objekten der Wirklichkeit bestehen. Das empirische Relativ bezieht sich also auf die erfahrbare Wirklichkeit. Bei einer Messung wird das empirische Relativ in ein numerisches Relativ transformiert, d. h., die empirische Wirklichkeit wird in Zahlenwerte übersetzt. Bei einer adäquaten Messung repräsentiert das numerische Relativ die empirischen Relationen.

In der Perspektive des Kritischen Rationalismus ist das empirische Relativ nicht unmittelbar, sondern lediglich in Form von Beobachtungen zugänglich (Popper, 1972, S. 44). Beobachtungen bilden aber nicht einfach realistische Sachverhalte ab, sondern hängen auch von theoretischen Erwartungen, sprachlichen Festlegungen etc. ab. Popper verwendete in diesem Kontext die Scheinwerfer-Metapher (ebd., S. 47), die zum Ausdruck bringt, dass nur das beobachtet werden kann, was sich im Lichtkegel des Scheinwerfers befindet. Bei der Erfassung der diagnostischen Problemlösekompetenz gibt das Sprachmodell vor, welche Bereiche der Wirklichkeit in den Blick genommen werden oder, um im sprachlichen Bild des Lichtkegels zu bleiben: welche Wirklichkeitsbereiche im Lichtkegel stehen. Damit beeinflusst das Sprachmodell auch das empirische Relativ. Das bedeutet auch, dass das numerische Relativ nicht einfach das empirische Relativ abbildet. Vielmehr bezieht sich das numerische Relativ auch auf das Sprachmodell, also das sprachliche Relativ. Im sprachlichen Relativ werden die empirischen Beziehungen zwischen realen Objekten beschrieben. Besonders greifbar wird das bei der Frage des Skalenniveaus von Messwerten. Falls diagnostische Problemlösekompetenz im Sprachmodell als klassifikatorische Größe definiert wurde, sind bestimmte numerische Operationen nicht sinnvoll. Bei genauer Betrachtung spielen also bei der Erfassung der diagnostischen Problemlösekompetenz das empirische Relativ, das numerische Relativ *und* das sprachliche Relativ wichtige Rollen. Es ist hier nicht möglich, das Verhältnis zwischen diesen Relativen herauszuarbeiten und weiter zu präzisieren. Hier genügt es, auf die Bedeutung des sprachlichen Relativs und damit des Sprachmodells bei der Erfassung diagnostischer Problemlösekompetenzen hinzuweisen.

## 7.2.2 Erhebungsmodell

### Zentrale Gestaltungskriterien: Authentizität und Repräsentativität

Wie die zurückliegenden Ausführungen zeigen, ist bei der Gestaltung des Erhebungsmodells auf ein hohes Ausmaß an Authentizität und Repräsentativität zu achten. Sowohl die Items bzw. diagnostischen Probleme als auch die Darbietungs- und Bearbeitungsform sollten authentisch sein, d. h. möglichst genau der realen beruflichen Problemumwelt entsprechen. Das ist eine wichtige Voraussetzung dafür, bei der Erfassung der diagnostischen Problemlösekompetenz Aussagen zu gewinnen, die auf berufliche Kontexte übertragbar bzw. in beruflichen Kontexten bedeutsam sind. Wichtig ist ferner eine repräsentative Auswahl diagnostischer Probleme; ansonsten gelten die gewonnenen Aussagen nur für die Problemauswahl des Erhebungsmodells, aber sie lassen sich nicht auf die im Sprachmodell definierte (und deutlich über den Erhebungskontext hinausweisende) diagnostische Problemlösekompetenz generalisieren. Mit einem authentischen und repräsentativen Erhebungsmodell können zwei zentrale Validitätsrisiken minimiert werden: die *Konstruktunterrepräsentation* und die *konstruktirrelevante-Varianz* (Messick, 1994, S. 22). Konstruktunterrepräsentation liegt vor, wenn Testwerte nur einen (kleinen) Ausschnitt des Konstrukts (hier: diagnostische Problemlösekompetenz) abdecken. Dieses Risiko kann durch eine repräsentative Problemauswahl reduziert werden. Von konstruktirrelevanter Varianz wird gesprochen, wenn Testwertunterschiede nicht nur auf tatsächliche interindividuelle Unterschiede in der diagnostischen Problemlösekompetenz, sondern auch auf Störeinflüsse zurückzuführen sind. Dieses Validitätsrisiko kann durch eine authentische Gestaltung des Erhebungsmodells verringert werden. Für Messick (1994) stellen Konstruktunterrepräsentation und konstruktirrelevante-Varianz die Hauptvaliditätsrisiken bei der Erfassung beruflicher Kompetenzen dar.

### Das Potenzial von Key-Feature-Problemen

Eine sehr schwierige Frage ist, ob sich Key-Feature-Probleme zur Erfassung der diagnostischen Kfz-Problemlösekompetenz eignen. Im Vergleich zu (holistischen) diagnostischen Problemen sind Key-Feature-Probleme weniger authentisch und weniger repräsentativ. Mit dem Einsatz von Key-Feature-Problemen wird also prinzipiell Konstruktunterrepräsentation und konstruktirrelevante Varianz riskiert. Möglicherweise macht es aber einen Unterschied, ob Key-Feature-Probleme in einer authentischen Computersimulation oder papierbasiert vorgegeben und bearbeitet werden. Wie ein Vergleich der Ergebnisse der fünften und sechsten Studie zeigt, scheinen mit computersimulierten Key-Feature-Problemen gewonnene Aussagen besser auf berufliche Kontexte übertragbar als Aussagen, die auf papierbasierten Key-Feature-Problemen beruhen. Vielleicht liegt dies an der größeren Authentizität und besseren Repräsentativität der computersimulierten Key-Feature-Probleme. Vor diesem Hintergrund ließe sich vermuten, dass eine valide Erfassung der diagnostischen Problemlösekompetenz Items verlangt, die ein Mindestmaß an Authentizität und Repräsentativität aufweisen. Falls dieses Mindestmaß unterschritten wird, wie möglicherweise bei den papierbasierten Key-Feature-Problemen, verliert die Erfassung (deutlich) an Validität. Die

Grundlage dieser Argumentation bildet der in der fünften Studie gezogene Schluss, dass die computersimulierten Key-Feature-Probleme und die computersimulierten diagnostischen Probleme zu konvergent validen Aussagen führen. Es wurde ausführlich diskutiert, dass dieser Schluss auf einer unklaren Befundlage basiert und davon abhängt, welche Kriterien zur Beurteilung der konvergenten Validität herangezogen werden (Kapitel 7.1.2). Die Belastbarkeit der geführten Argumentation ist demzufolge offen.

### **7.2.3 Transfer der Erkenntnisse auf andere berufliche Kontexte**

Eine wichtige Frage ist, ob die im Kontext der Kfz-Ausbildung und der elektrotechnischen diagnostischen Kfz-Problemlösekompetenz gewonnenen Erkenntnisse auf andere berufliche Kontexte transferiert werden können. Einerseits kann argumentiert werden, dass die theoretischen Grundlagen der vorliegenden Untersuchung anhand prinzipieller Überlegungen und mit Studien entwickelt wurden, die aus dem medizinischen und technischen Bereich und sehr unterschiedlichen beruflichen Kontexten stammen. Dies könnte für die Transferierbarkeit der Erkenntnisse sprechen. Andererseits legt der aktuelle Forschungsstand nahe, diagnostisches Problemlösen als eine anforderungsspezifische bzw. domänenspezifische Handlung aufzufassen (z. B. Durning et al., 2015, S. 129), was den Ergebnistransfer infrage stellt. Besonders aussichtsreich scheint ein Transfer auf berufliche Kontexte, in denen ähnliche Bedingungen vorherrschen (Diagnose technischer Störungen, Diagnose elektrotechnischer Systeme etc.). Naheliegend wäre z. B. ein Transfer auf Berufe wie Elektroniker:in für Energie- und Gebäudetechnik, Elektroniker:in für Automatisierungstechnik oder Mechatroniker:in. In diesen Berufen sind Computersimulationen zur Erfassung der diagnostischen Problemlösekompetenz verfügbar (z. B. Nickolaus et al., 2012b; Walker et al., 2016a). Vertretbar dürfte jedoch auch ein weiterer Erkenntnistransfer sein, bei dem im Kontext konkreter Anforderungen und der verfolgten Fragestellung reflektiert wird, ob die hier gewonnenen Erkenntnisse eine fruchtbare Basis darstellen, die interessierende diagnostische Problemlösekompetenz zu erfassen. Letztlich werden aber weitere empirische Studien benötigt, um die Transferfrage zu klären.

### **7.2.4 Innovative Erhebungsmethoden für die Berufs- und Wirtschaftspädagogik**

Für die Berufs- und Wirtschaftspädagogik könnte diese Untersuchung auch insofern anregend sein, als teilweise innovative Erhebungsmethoden erprobt und eingesetzt wurden. Die Aufzeichnung von Blickdaten mit Eye-Trackern und von Probandenverhalten in Logfiles stellt ein objektives und reichhaltiges Datenmaterial zur Verfügung, das allerdings schwierig zu interpretieren ist. Diese Interpretationsschwierigkeiten sind wahrscheinlich nur zu bewältigen, wenn die Datenanalyse theoriebasiert erfolgt. Es besteht kein geringes Risiko, dass ein exploratives Vorgehen bei einem so reichen Datenmaterial zahlreiche signifikante ad-hoc-Befunde hervorbringt, die sich nicht reproduzieren lassen und von geringem theoretischem Wert sind. Natürlich können explorative Analysen zu inspirierenden Befunden und interessanten Forschungshypothesen führen.

## 7.3 Grenzen und kritische Reflexion der Untersuchung

### 7.3.1 Computersimulierte diagnostische Probleme als Validierungsgrundlage

Eine Grenze der fünften und sechsten Studie stellt die Tatsache dar, dass computer-simulierte diagnostische Kfz-Probleme herangezogen wurden, um zu untersuchen, ob die mit Key-Feature-Problemen gewonnen Aussagen auf berufliche Kontexte übertragbar sind. Angesichts der von Gschwendtner et al. (2009) oder Abele (2014, S. 190) referierten Befunden scheint dieses Vorgehen vertretbar. Dennoch kann nicht ausgeschlossen werden, dass die Befunde beim Einsatz realistischer diagnostischer Kfz-Probleme anders ausgefallen wären.

### 7.3.2 Konzentration auf das Sprachmodell und das Erhebungsmodell

Eine weitere Grenze der Untersuchung besteht in der Konzentration auf das Sprachmodell und das Erhebungsmodell. Zwar wurde sichtbar, dass in diesen Bereichen noch wichtige Forschungsdesiderate bestehen (Kapitel 3.4 und Kapitel 4.3), dennoch decken beide Modelle nur einen Teil des Modellierungsprozesses ab, aus dem sprachliche Aussagen über die diagnostische Problemlösekompetenz resultieren (Kapitel 2.1). Messmodelle sind für die Erfassung diagnostischer Problemlösekompetenzen ebenfalls sehr bedeutsam. Die Generalisierbarkeitstheorie weist explizit darauf hin, dass auch die Variation statistischer Modellierungen einen Einfluss auf die Varianz von Testwerten und damit auf Testwertinterpretationen haben kann (s. dazu z. B. Robitzsch et al., 2011). Dieser Punkt verweist auf eine weitere Limitation der Untersuchung.

Alle im Rahmen dieser Ausarbeitung zitierten empirischen Befunde basieren auf Messmodellen. Sprachliche Aussagen über die diagnostische Problemlösekompetenz können abhängig vom gewählten Messmodell variieren. Dieser Faktor wurde hier allerdings nicht berücksichtigt. Hier lag der Fokus auf dem Sprachmodell und dem Erhebungsmodell, der Einfluss des Messmodells wurde nicht kontrolliert. Es besteht das Risiko, dass Einschätzungen auf Befunden basieren, die auf unangemessenen Messmodellen beruhen.

### 7.3.3 Reflexion des Problemlösebegriffs

Der hier gewählte Problemlösebegriff unterscheidet sich vom Problemlösebegriff anderer berufs- und wirtschaftspädagogischer Studien, in denen in Anlehnung an Dörner nur dann von einem Problem gesprochen wird, wenn in einer spezifischen Situation ein Ziel angestrebt wird, die Mittel zur Zielerreichung aber zumindest teilweise unbekannt sind (z. B. Wuttke et al., 2015; Sembill, 1992, S. 83 f.). Ein Problem besteht demzufolge dann, wenn Individuen zur Zielerreichung eine *Lücke* in ihrem subjektiven *Handlungsplan* haben und diese schließen müssen (s. dazu auch die Problemdefinition von Funke, 2003). Im Sinne dieser Definition dürfen Anforderungen, die mit Routinehandlungen bewältigt werden können, prinzipiell nicht als Probleme bezeichnet werden: Die Bewältigung solcher Anforderungen setzt nicht zwingend voraus, dass eine Lücke im Handlungsplan zu schließen ist. Mit dieser Problemdefinition gerät ein aus didaktischer Sicht besonders interessanter Aspekt in den Blick: Problemlösen er-

weitert das individuelle Handlungsrepertoire und fördert selbstständiges Handeln. Im Hinblick auf die Erfassung von Problemlösekompetenzen bereitet dieser Problemlösebegriff allerdings Schwierigkeiten: Er verlangt, dass Individuen bei der Bewältigung der vorgegebenen Anforderungen Lücken in ihrem Handlungsplan schließen müssen. Nur dann ist es gerechtfertigt, von Problemlösen zu reden. Da Lücken im Handlungsplan abhängig von individuellen Lernerfahrungen und Voraussetzungen variieren, lässt sich nur schwer abschätzen, welche Anforderungen für ein spezifisches Individuum Probleme darstellen. Der Problemlösebegriff führt zu Schwierigkeiten bei der Operationalisierung.

In der vorliegenden Arbeit wurde in Anlehnung an Jonassen (2000, S. 65) dann von einem Problem gesprochen, wenn eine Unbekannte zu finden, d. h., eine *Wissenslücke* zu schließen ist. In diagnostischen Situationen ist dies der Fall, wenn abgesichert werden kann, dass die Ursache eines diagnostischen Problems nicht bekannt ist. Dabei spielt es keine Rolle, ob die Wissenslücke anhand einer Routinehandlung oder einer ad hoc entwickelten Handlungsoption geschlossen wurde. Der große Vorteil dieser Definition ist, dass die Operationalisierung der Problemlösekompetenz keine allzu großen Schwierigkeiten bereitet. Sofern abgesichert werden kann, dass Proband:innen eine Anforderung nicht ausschließlich aus dem Gedächtnis bewältigen können und eine Unbekannte finden müssen, ist es angemessen, von einem Problem zu sprechen. Der Nachteil dieses Problemlösebegriffs ist, dass Probleme auch mit Routinehandlungen gelöst werden können. Damit bezieht sich das Problemlösen nicht zwingend auf einen kognitiv und emotional besonders herausfordernden Vorgang. Festzuhalten ist allerdings, dass auch beim hier gewählten Problemlöseverständnis und einer entsprechenden Operationalisierung – zumindest im gewerblich-technischen Bereich – noch ausreichend viele Proband:innen *nicht* in der Lage sind, die entsprechend operationalisierten Probleme zu lösen. So konnten z. B. in der Studie von Nickolaus et al. (2012b) (sehr) viele Auszubildende zum/zur Kfz-Mechatroniker:in viele diagnostische Probleme nicht lösen, die auf die beschriebene Art operationalisiert wurden.

#### 7.3.4 Metatheoretische Reflexion

Aus Sicht des Kritischen Rationalismus spielt die Erfassung der diagnostischen Problemlösekompetenz eine wichtige Rolle. Das Zentrum des Kritischen Rationalismus bildet die empirische Prüfung von Theorien. Bei der empirischen Theorieprüfung wird, grob gesprochen, untersucht, ob empirische Prognosen, die deduktionslogisch aus einer Theorie und Randbedingungen abgeleitet wurden, durch Beobachtungen falsifiziert werden, wobei die Beobachtungen im Lichte der Theorie gewonnen werden (Popper, 1972). Die Beobachtungen werden in Basissätzen festgehalten. Die Erfassung der diagnostischen Problemlösekompetenz kann als Vorhaben aufgefasst werden, solche Basissätze zu gewinnen. Aus dieser Sicht ergeben sich mindestens zwei Diskussionspunkte.

Erstens wurde die diagnostische Problemlösekompetenz hier – wie z. B. auch in den Projekten des Forschungsprogramms ASCOT (Technology-based Assessment of Skills and Competences in Vocational Education and Training) – nicht ausgehend von

einer Theorie zur diagnostischen Problemlösekompetenz formuliert.<sup>24</sup> Die sprachliche Modellierung erfolgte vor dem Hintergrund ausgewählter Kriterien, systematischer Überlegungen und dem einschlägigen empirischen Forschungsstand (Kapitel 3). Es darf durchaus kritisch gefragt werden, inwiefern die vorliegende Arbeit überhaupt mit dem Kritischen Rationalismus vereinbar ist, d. h. zur empirischen Prüfung einer Theorie zur diagnostischen Problemlösekompetenz beitragen kann. Eine solche Theorie bezieht sich typischerweise auf ein spezifisches Forschungsproblem bzw. ein entsprechendes Forschungsziel. Die Definition der diagnostischen Problemlösekompetenz kann von diesem Forschungsziel abhängen und dementsprechend deutlich von der hier gewählten Definition abweichen. Denkbar wäre aber auch, die Ergebnisse dieser Untersuchung bei der Formulierung von Theorien einzubeziehen und zu prüfen, ob die Definition, also das Sprachmodell, übernommen werden kann. Außerdem ist zu bedenken, dass die Erfassung der Problemlösekompetenz nicht nur der Theorieprüfung dient, sondern auch z. B. diagnostische Zwecke erfüllen kann. Im diagnostischen Bereich kann die vorliegende Untersuchung jenseits der Theorieprüfung nützlich sein, sofern das hier entwickelte Sprachmodell akzeptiert wird.

Der zweite Diskussionspunkt bedarf einer ausführlicheren Besprechung. Basisätze stellen aus Sicht des Kritischen Rationalismus Konventionen, also Beschlussfassungen dar (Popper, 2005, S. 69 ff.). Dies bedeutet auch, dass die Erfassung der diagnostischen Problemlösekompetenz, d. h. die resultierenden sprachlichen Aussagen, Konventionen sind. Damit kann der Wahrheitsgehalt der sprachlichen Aussagen über die diagnostische Problemlösekompetenz weder logisch noch empirisch geprüft (i. S. von *falsifiziert*) werden. Soll die empirische Erfassung allerdings nicht einen beliebigen Vorgang darstellen und damit ein beliebiges Resultat hervorbringen, werden Kriterien benötigt, denen die empirische Erfassung genügen sollte. Der Kritische Rationalismus bietet diesbezüglich ausschließlich allgemeine Kriterien. Im Sinne der Kritisierbarkeit von Aussagen wird erstens gefordert, dass die Ergebnisse der empirischen Erfassung intersubjektiv nachprüfbar sind (ebd., S. 75). Dafür ist das Vorgehen möglichst detailliert und verständlich zu beschreiben. Zweitens sollten die Ergebnisse reproduzierbar sein (ebd.) und drittens spezifischen Regeln genügen (ebd., S. 87). Diese spezifischen Regeln scheinen bei genauer Betrachtung nur mit einem Blick in die jeweilige Objektwissenschaft konkretisiert werden zu können. So wurde hier zur Konkretisierung des Erhebungsmodells das Kriterium Validität herangezogen, das in der empirischen Sozialwissenschaft – neben weiteren Kriterien – als zentrales Gütekriterium bei der Erfassung psychischer Merkmale gilt. Ohne Zweifel lassen sich viele rationale Gründe nennen, warum die Validität ein wichtiges Gütekriterium ist. Die Bedeutung dieses Kriteriums kann aber weder logisch nachgewiesen noch falsifiziert werden. Um nicht in einen infiniten Begründungsregress zu geraten, bleibt in letzter Konsequenz nur

---

24 Dieser Feststellung unterliegt ein strenger Theoriebegriff: Im Sinne dieses Theoriebegriffs basiert die Kompetenzerfassung nur dann auf einer Theorie, wenn sie der Prüfung von sprachlichen Aussagen dient, die „deskriptive (also wertfreie) materiale Implikationen“ haben (Beck, 2006, S. 578). Im Sinne dieses Begriffs wird mit der Kompetenzerfassung entweder eine Wenn-Komponente (Ursache) oder eine Dann-Komponente (Wirkung) der sprachlichen Aussage erfasst (Beck, 2009, S. 243).

die Möglichkeit, auf den entsprechenden Konsens innerhalb der *scientific community* (wissenschaftliche Gemeinschaft) zu verweisen.

Auch Sprachmodelle bzw. Definitionen, die Bestandteil von Theorien sind, werden im Kritischen Rationalismus als Konventionen betrachtet (z. B. Beck, 2009, S. 242). Während allerdings im Bereich des Erhebungsmodells innerhalb der *scientific community* akzeptierte Kriterien (z. B. Validität) verfügbar sind, gilt das für die Konkretisierung des Sprachmodells nicht. In dieser Arbeit wurde bei der Konkretisierung des Sprachmodells auf die Methode der Explikation und die dafür oft angewandten Kriterien zurückgegriffen (Fruchtbarkeit, Exaktheit und Einfachheit). Die Explikation von Begriffen wurde in Anlehnung an Rudolf Carnap vorgenommen (Kapitel 3.2), der als ein Hauptvertreter des Logischen Empirismus gilt (Lorenz, 2004). Unter Anwendung dieser Kriterien auf den empirischen Forschungsstand wurden zentrale Aspekte des Sprachmodells herausgearbeitet. Auch in vielen anderen Arbeiten zur Kompetenz wird der Kompetenzbegriff auf Basis vorliegender empirischer Befunde definiert (z. B. Zlatkin-Troitschanskaia & Seidel, 2011). Solange dabei keine induktive Begriffsbildung stattfindet, die Definition letztlich also nicht als empirisches Geschäft, sondern als Konvention verstanden wird, scheint dieses Vorgehen mit dem Kritischen Rationalismus vereinbar. Es repräsentiert aber nur eine von prinzipiell vielen denkbaren Definitionsstrategien.

Diese Überlegungen verdeutlichen, dass Konventionen bei der Erfassung der diagnostischen Problemlösekompetenz eine wichtige Rolle spielen. Dies kann zu unterschiedlichen Definitionen und damit Operationalisierungen von diagnostischer Problemlösekompetenz und schlimmstenfalls zu widersprüchlichen Aussagen führen, was innerhalb und außerhalb der berufs- und wirtschaftspädagogischen Forschung für Irritationen sorgen und das Vertrauen in diese Forschung schwächen könnte. Um dies zu vermeiden, ist es für Wissenschaftler:innen der Berufs- und Wirtschaftspädagogik von zentraler Bedeutsamkeit, Grenzen und Unterschiede von Studien sorgfältig zu reflektieren und an der Etablierung, Reflexion und Einhaltung von Kriterien aktiv mitzuwirken. Im Sinne der theoretischen Fruchtbarkeit wäre z. B. zu fordern, bei der Klärung des Sprachmodells und des Erhebungsmodells konsequent an vorliegende theoretische und empirische Arbeiten anzuschließen. Im Hinblick auf das Sprachmodell wäre weiterhin ein hohes Maß an sprachlicher Präzision und, ganz prinzipiell, Operationalisierbarkeit einzufordern. Es kann als Limitation dieser Studie betrachtet werden, dass dieser Kriterienaspekt nicht gründlicher berücksichtigt und aufgearbeitet wurde. Eine weitere Grenze besteht darin, dass eine wissenschaftstheoretische Rekonstruktion und Reflexion des Modellierungsprozesses, aus dem die sprachlichen Aussagen über die diagnostische Problemlösekompetenz resultieren (Kapitel 2.1), nur in Ansätzen erfolgte. Möglicherweise hätte eine solche Rekonstruktion zur Wahl anderer Kriterien geführt und Konsequenzen für die Konkretisierung des Sprachmodells und des Erhebungsmodells gehabt. Darüber hinaus könnte zukünftig interessant sein, bei einer solchen wissenschaftstheoretischen Rekonstruktion aktuelle Entwicklungen der empiristischen Wissenschaftsphilosophie und den Wert anderer empiristischer For-

schungsparadigmen (z. B. verschiedene Bayesianische Ansätze) für die Kompetenzerfassung zu beleuchten.

#### 7.4 Forschungsperspektiven

In diesem letzten Abschnitt werden kursorisch einige Forschungsperspektiven aufgezeigt, die an die vorliegende Untersuchung anknüpfen und Wege weisen, wie der einschlägige Forschungsstand weiterentwickelt werden kann. Wie die vier Studien zum Sprachmodell verdeutlichten, werden weitere Studien zu der Frage benötigt, ob sich diagnostische Problemlösekompetenzen empirisch von anderen beruflichen Kompetenzen unterscheiden lassen. Bei der Untersuchung dieser Frage ist auf eine handlungsnah Operationalisierung und Erfassung der Kompetenzen zu achten. Im Hinblick auf die Operationalisierung wurden Herausforderungen bei der Erfassung von Problemmerkmalen, oder allgemeiner, beruflicher Anforderungen sichtbar. Diese Herausforderungen ergeben sich unter anderem daraus, dass berufliche Anforderungen individuell verarbeitet werden und die Erfassung von Anforderungsmerkmalen auf Annahmen beruht, wie Individuen diese Anforderungen verarbeiten und wie sie mit ihnen interagieren. Die individuelle Verarbeitung und Interaktion dürfte von der Beschaffenheit der Anforderungen, dem individuell Erlernten sowie grundsätzlichen menschlichen Eigenschaften (z. B. begrenzte Verarbeitungskapazität) abhängen. Das Zusammenspiel von Individuum und beruflichen Anforderungen ließe sich problemlos weiter ausdifferenzieren, was darauf verweist, dass belastbare Erkenntnisse in diesem Forschungsbereich nicht einfach zu gewinnen sein dürften. Gleichwohl handelt es sich hierbei um ein wichtiges berufs- und wirtschaftspädagogisches Forschungsdesiderat: Die Bearbeitung dieses Desiderats gibt Aufschluss über anforderungsbezogene und individuelle Determinanten sowie Wirkungen des Lernens in beruflichen Kontexten und liefert damit relevante Erkenntnisse für die Gestaltung beruflicher Lehr-Lern-Prozesse. Weitere Forschung ist auch im Bereich des Prozesses angezeigt, der beruflichem Problemlösen, oder allgemeiner, beruflichem Handeln zugrunde liegt. Ob und inwieweit die hier vorgelegte Theorie zum diagnostischen Problemlöseprozess auf andere berufliche Kontexte übertragbar ist, ist eine wichtige, aber bislang offene Frage. Ungeklärt ist auch, wie die mentale Modellbildung von Individuen beim beruflichen Problemlösen erfasst und gefördert werden kann und auch, wie sich die darauf bezogenen mentalen Grundlagen entwickeln. Deutlich gemacht haben die beiden Studien zum Erhebungsmodell, dass das diagnostische Potenzial von Key-Feature-Problemen nicht abschließend beurteilt werden kann, weshalb auch in diesem Bereich weitere Untersuchungen angezeigt sind. Grundsätzlichere Studien und Debatten sind dazu nötig, welche Kriterien bei der Klärung von Sprachmodellen herangezogen werden sollen und ob bzw. wie solche Kriterien aus wissenschaftstheoretischen und/oder objekttheoretischen Ansätzen gewonnen werden können. Wissenschaftstheoretische Reflexionen sind auch deswegen wichtig, weil sie die Anlage und Durchführung von Studien zur Erfassung beruflicher Kompetenzen und damit den Erkenntnisfortschritt erheblich beeinflussen können. Abschließend werden zwei Forschungsperspektiven etwas ausführlicher thematisiert.

### **7.4.1 Anwendung der Generalisierbarkeitstheorie: Wie viele diagnostische Probleme gewährleisten eine repräsentative Problemauswahl und welche Bedeutung hat die Wahl des Messmodells?**

Mindestens zwei Gründe sprechen dafür, in künftigen Studien auf die Generalisierbarkeitstheorie zurückzugreifen: Erstens könnte mithilfe dieser Theorie systematisch untersucht werden, wie viele diagnostische Probleme auszuwählen sind, um die diagnostische Problemlösekompetenz repräsentativ zu erfassen. Unter diesem Aspekt könnte in Anlehnung an die im medizinischen Bereich oft favorisierte weite Begriffsfassung von diagnostischer Problemlösekompetenz (Kapitel 3.3.4) untersucht werden, ob es möglich ist, die diagnostische Problemlösekompetenz auch dann repräsentativ zu erfassen, wenn gleichzeitig mehrere Anforderungsklassen einbezogen werden. Dies wäre auch aus testökonomischen Gründen ein interessanter Ansatzpunkt.

Zweitens könnte die Generalisierbarkeitstheorie für die Konkretisierung des Messmodells nützlich sein. Im Rahmen der Generalisierbarkeitstheorie werden Messfehler deutlich differenzierter modelliert als im Rahmen der klassischen Testtheorie (Brennan, 1998). Unterstellt wird, dass Messfehler unterschiedliche Quellen haben. So hängen Messfehler z. B. davon ab, welche Items ausgewählt und welche statistischen Modelle einbezogen werden (Robitzsch et al., 2011).<sup>25</sup> Im Rückgriff auf die Generalisierbarkeitstheorie kann mathematisch modelliert werden, welchen Einfluss Variationen in der Itemstichprobe auf die Generalisierbarkeit von Testwerten hat. Dementsprechend kann quantifiziert werden, wie viele diagnostische Probleme nötig wären, um die definierte Kompetenz mit einer hohen internen Konsistenz zu erfassen, d. h. repräsentativ abzubilden. Mit diesem Verfahren ließe sich mathematisch abschätzen, ob eine weite Begriffsfassung der diagnostischen Problemlösekompetenz zu Testwerten mit akzeptabler interner Konsistenz, d. h. Repräsentativität führt und wie viele diagnostische Probleme dafür nötig wären. Die Voraussetzung dieses Verfahrens ist allerdings, dass vorab ein Problemuniversum möglichst präzise definiert werden und die Auswahl der Probleme aus diesem Universum zufällig oder zumindest unsystematisch erfolgen müsste. Mit der Generalisierbarkeitstheorie könnte ferner systematisch untersucht werden, welcher Anteil von Testwertvarianz auf die Variation statistischer Modelle zurückzuführen ist. Dass die Auswahl statistischer Modelle einen Einfluss auf Testwertvarianzen haben kann, belegt die Studie von Robitzsch et al. (2011).

### **7.4.2 Erfassung kooperativer Problemlösekompetenzen in beruflichen Kontexten**

Eine interessante weitere Forschungsperspektive besteht darin, sich künftig mit der Erfassung, aber auch der Bedeutung kooperativer Problemlösekompetenzen in beruflichen Kontexten zu beschäftigen. Mit kooperativem Problemlösen ist ein Prozess angesprochen, bei dem mehrere Personen gemeinsam an beruflichen Problemen arbeiten. Rausch, Schley und Warwas (2015) fanden heraus, dass diese Art von Problemlösen in realen beruflichen Kontexten eine zentrale Rolle spielt. Um kooperative Problemlösekompetenzen valide zu erfassen, werden berufsauthentische Tests, d. h. vermutlich

---

<sup>25</sup> Als weitere Varianzquelle werden häufig Beobachter genannt. Aus den in Kapitel 4.2.1 genannten Gründen, wird diese Quelle hier nicht thematisiert.

computerbasierte Tests, benötigt. In PISA (*Programme for International Student Assessment*) 2015 wurden kooperative Problemlösekompetenzen computerbasiert erfasst (z. B. Greiff, 2012a). Eine interessante und leistungsstarke Computerplattform zur Erfassung und Untersuchung des kooperativen Problemlösens wird von Hao et al. (2017) vorgestellt. In diese Computerplattform können bestehende computerbasierte Tests integriert werden. Dies eröffnet die Möglichkeit, die Computersimulationen zu integrieren, die bislang im technischen Bereich zur Erfassung diagnostischer Problemlösekompetenzen verwendet wurden. Mit einer solchen Softwarelösung könnte künftig untersucht werden, wie mehrere Personen an der Lösung eines diagnostischen Problems arbeiten und welche Effekte die Kooperation hat. Dies scheint ein weiterer wichtiger Schritt in Richtung einer berufsrealen Kompetenzerfassung zu sein, die zudem vielfältige Prozessdaten zur Verfügung stellt und Mikroanalysen von Leistungserstellungs- und Lernprozessen ermöglicht.

## Literaturverzeichnis

- Abele, S. (2011). Hängt die prognostische Validität eignungsdiagnostischer Verfahren von der Operationalisierung des Ausbildungserfolgs ab? In R. Nickolaus & G. Pätzold (Hrsg.), *Lehr-Lernforschung in der gewerblich-technischen Berufsbildung. Zeitschrift für Berufs- und Wirtschaftspädagogik*. 25, 13–35 [Themenheft].
- Abele, S. (2014). *Modellierung und Entwicklung berufsfachlicher Kompetenz*. Franz Steiner.
- Abele, S., Behrendt, S., Weber, W. & Nickolaus, R. (2016). Berufsfachliche Kompetenzen von Kfz-Mechatronikern – Messverfahren, Kompetenzdimensionen und erzielte Leistungen (KOKO Kfz). In K. Beck, M. Landenberger & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung. Ergebnisse aus der BMBF-Förderinitiative ASCOT* (S. 171–203). W. Bertelsmann.
- Achtenhagen, F. (2000). Lebenslang Lernen aus Sicht des Mastery Learning. In F. Achtenhagen & W. Lempert (Hrsg.), *Lebenslanges Lernen im Beruf. Seine Grundlegung im Kindes- und Jugendalter* (S. 123–140). Leske + Budrich.
- American Educational Research Association, American Psychological Association, national Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.
- Baethge, M. & Arends, L. (2009). *Feasibility Study VET-LSA. A Comparative Analysis of Occupational Profiles and VET Programmes in 8 European Countries – International Report*. Vocational Training Research volume 8. W. Bertelsmann.
- Barrows, H. S. & Feltovich, P. J. (1987). The clinical reasoning process. *Medical Education*, 21, 86–91.
- Beck, K. (1987). *Die empirischen Grundlagen der Unterrichtsforschung. Eine kritische Analyse der deskriptiven Leistungsfähigkeit von Beobachtungsmethoden*. Hogrefe.

- Beck, K. (1988). Dimensionen des Handlungsbegriff aus didaktischer Sicht. In R. Czycholl & H. Ebner (Hrsg.), *Zur Kritik handlungsorientierter Ansätze in der Didaktik der Wirtschaftslehre* (Beiträge zur Berufs- und Wirtschaftspädagogik, Bd. 4, S. 61–75). Bibliotheks- und Informationssystem der Universität Oldenburg.
- Beck, K. (2005). Ergebnisse und Desiderate zur Lehr-Lern-Forschung in der kaufmännischen Berufsausbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 101(4), 533–556.
- Beck, K. (2006). Theorieansätze. In R. Arnold & A. Lipsmeier (Hrsg.), *Handbuch der Berufsbildung* (2., überarbeitete und aktualisierte Aufl., S. 577–585). VS Verlag für Sozialwissenschaft.
- Beck, K. (2009). Strategien empirischer Forschung zur Professionalität von Lehrpersonen. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (S. 237–247). Beltz.
- Beck, K. (2016). Individuelle Moral und Beruf - eine Integrationsaufgabe für die Ordnungsethik? In G. Minnameier (Hrsg.), *Ethik und Beruf. Interdisziplinäre Zugänge* (S. 41–54). W. Bertelsmann.
- Beck, K., Landenberger, M. & Oser, F. (Hrsg.). (2016). *Technologiebasierte Kompetenzmessung in der beruflichen Bildung. Ergebnisse aus der BMBF-Förderinitiative ASCOT*. W. Bertelsmann.
- Billett, S. (2017). Developing domains of occupational competence: workplaces and learner agency. In M. Mulder (Hrsg.), *Competence-based Vocational and Professional Education. Bridging the Worlds of Work and Education* (Technical and Vocational Education and Training: Issues, Concerns and Prospects, Bd. 23, S. 47–66). Springer International Publishing.
- Borsboom, D., Mellenbergh, G. J. & Heerden, J. (2004). The concept of validity. *Psychological Review*, 111, 1061–1071.
- Bortz, J. & Döring, N. (2009). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4., überarbeitete Aufl., Nachdruck). Springer.
- Boulet, J. R., McKinley, D. W., Norcini, J. J. & Whelan, G. P. (2002). Assessing the comparability of standardized patient and physician evaluations of clinical skills. *Advances in Health Science Education*, 7(2), 85–97.
- Brennan, R. L. (1998). (Mis)Conceptions about Generalizability Theory. *Educational Measurement: Issues and Practice*, 19(1), 5–10.
- Brennan, R. L. (2001). Some Problems, Pitfalls, and Paradoxes in Educational Measurement. *Educational Measurement: Issues and Practice*, 20(4), 6–18.
- Carnap, R. (1962). *Logical Foundations of Probability*. Chicago, IL: The University of Chicago Press.
- Charlin, B., Lubarsky, S., Millette, B., Crevier, F., Audétat, M., Charbonneau, A. et al. (2012). Clinical Reasoning Processes: Unravelling Complexity through Graphical Representation. *Medical Education*, 46(5), 454–463.
- Croskerry, P. (2009). A Universal Model of Diagnostic Reasoning. *Academic Medicine*, 84(8), 1022–1028.

- Dörner, D. (2008). Umgang mit Komplexität. In A. von Gleich & S. Gößling-Reisemann (Hrsg.), *Industrial Ecology* (S. 284–302). Vieweg+Teubner.
- Dudenredaktion. (2002). *Duden: das Bedeutungswörterbuch*: Dudenverlag.
- Durning, S. J., Rencic, J. & Schuwirth, L. (2015). Assessing Clinical Reasoning. In L. N. Pangaro & W. C. McGaghie (Hrsg.), *Handbook on Medical Student Evaluation and Assessment* (S. 127–146). Gegensatz Press.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden. Lehrbuch*. Beltz.
- 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.
- Essler, W. K. (1970). *Wissenschaftstheorie I. Definition und Reduktion*. Karl Alber.
- Fischer, M. R., Kopp, V., Holzer, M., Ruderich, F. & Jünger, J. (2005). A Modified Electronic Key Feature Examination for Undergraduate Medical Students: Validation Threats and Opportunities. *Medical Teacher*, 27(5), 1–6.
- Frederiksen, N., Ward, W. C., Case, S. M., Carlson, S. B. & Samph, T. (1981). *Development of Methods for Selection and Evaluation in Undergraduate Medical Education* (ETS Research Report Series).
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Fürstenau, B. (1999). Förderung von Problemlösefähigkeit im planspielgestützten Unterricht. *Unterrichtswissenschaft*, 27(2), 135–158.
- Gigerenzer, G. (1981). *Messung und Modellbildung in der Psychologie*. Reinhardt.
- Greiff, S. (2012a). From interactive to collaborative problem solving: Current Issues and the Programme for International Student Assessment. *Review of Psychology*, 19(2), 111–121.
- Greiff, S. (2012b). *Individualdiagnostik komplexer Problemlösefähigkeit*. Waxmann.
- Greiff, S., Kretzschmar, A. & Leutner, D. (2014). Problemlösen in der Pädagogischen Psychologie. *Zeitschrift für Pädagogische Psychologie*, 28(4), 161–166.
- Gschwendtner, T., Abele, S. & Nickolaus, R. (2009). Computersimulierte Arbeitsproben: Eine Validierungsstudie am Beispiel der Fehlerdiagnoseleistungen von Kfz-Mechatronikern. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 105, 557–578.
- Hao, J., Liu, L., von Davier, A. A., Lederer, N., Zapata-Rivera, D., Jakl, P. & Bakkenson, M. (2017). EPCAL: ETS platform for collaborative assessment and learning. *ETS Research Report Series*, 2017(1), 1–14.
- Hartig, J., Frey, A. & Jude, N. (2008). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 136–192). Springer.
- Hartig, J. & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127–143). Springer.
- Hatala, R. & Norman, G. R. (2002). Adapting the Key Feature Examination for a Clinical Clerkship. *Medical Education*, 36(2), 160–165.
- Hoc, J.-M. & Amalberti, R. (1995). Diagnosis: Some Theoretical Questions Raised by Applied Research. *Current Psychology of Cognition*, 14(1), 73–101.
- Horlebein, M. (2013). *Wissenschaftstheorie. Grundlagen und Paradigmen der Berufs- und Wirtschaftspädagogik*. Schneider.

- Ilgen, J. S., Humbert, A. J., Kuhn, G., Hansen, M. L., Norman, G. R., Eva, K. W. et al. (2012). Assessing Diagnostic Reasoning: A Consensus Statement Summarizing Theory, Practice, and Future needs. *Academic Emergency Medicine*, 19(12), 1454–1461.
- Jonassen, D. H. (2000). Toward a Design Theory of Problem Solving. *Educational Technology Research and Development*, 48(4), 63–85.
- Jonassen, D. H. (2011). *Learning to Solve Problems. A Handbook for Designing Problem-solving Learning Environments*. Routledge.
- 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.
- Kane, M. T. (1992). The Assessment of Professional Competence. *Evaluation & the Health Professions*, 15(2), 163–182.
- Kane, M. T. (2001). Current Concerns in Validity Theory. *Journal of Educational Measurement*, 38, 319–342.
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1–73.
- Kassirer, J., Wong, J. & Kopelman, R. (2010). *Learning clinical reasoning* (3. Aufl.). Lippincott Williams & Wilkins.
- Katz, I. R. & Anderson, J. R. (1987). Debugging: An Analysis of Bug-Location Strategies. *Human-Computer Interaction*, 3(4), 351–399.
- Klahr, D. & Dunbar, K. (1988). Dual Space Search during Scientific Reasoning. *Cognitive Science*, 12, 1–48.
- Konradt, U. (1992). *Analyse von Strategien bei der Störungsdiagnose in der flexiblen automatisierten Fertigung*. Universitätsverlag Brockmeyer.
- Krems, J. & Bachmaier, M. (1991). Hypothesenbildung und Strategiewahl in Abhängigkeit vom Expertisegrad. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 38, 394–410.
- Krems, J. F. (1994). *Wissensbasierte Urteilsbildung. Diagnostisches Problemlösen durch Experten und Expertensysteme*. Hans Huber.
- Kultusministerkonferenz (2003a). Rahmenlehrplan für den Ausbildungsberuf Elektroniker für Automatisierungstechnik/Elektronikerin für Automatisierungstechnik.
- Kultusministerkonferenz (2003b). Rahmenlehrplan für den Ausbildungsberuf Elektroniker für Betriebstechnik/Elektronikerin für Betriebstechnik.
- Kultusministerkonferenz (2013). Rahmenlehrplan für den Ausbildungsberuf Kraftfahrzeugmechatroniker/Kraftfahrzeugmechatronikerin.
- Kunina-Habenicht, O., Hautz, W. E., Knigge, M., Spies, C. & Ahlers, O. (2015). Assessing Clinical Reasoning (ASCLIRE): Instrument Development and Validation. *Advances in Health Science Education*, 20(5), 1205–1224.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Psychologie Verlags Union.
- Lorenz, K. (2004). Empirismus, Logischer. In J. Mittelstraß (Hrsg.), *Enzyklopädie Philosophie und Wissenschaftstheorie* (Bd. 1: A–G, S. 543–544). J. B. Metzler.

- McGill, D. A., van der Vleuten, C. P. M. & Clarke, M. J. (2011). Supervisor Assessment of Clinical and Professional Competence of Medical Trainees: a Reliability Study Using Workplace Data and Focused Analytical Literature Review. *Advances in Health Science Education, 16*(3), 405–425.
- Messick, S. (1994). The Interplay of Evidence and Consequences in the Validation of Performance Assessments. *Educational Researcher, 23*(2), 13–23.
- Minnameier, G. (2005). *Wissen und inferentielles Denken. Zur Analyse und Gestaltung von Lehr-Lern-Prozessen*. Peter Lang.
- Mislevy, R. J. (2011). *Evidence-Centered Design for Simulation-Based Assessment*. : National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Moosbrugger, H. & Kelava, A. (2008). Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 7–26). Springer.
- Nickolaus, R. (2011). Die Erfassung fachlicher Kompetenzen und ihrer Entwicklungen in der beruflichen Bildung – Forschungsstand und Perspektiven. In O. Zlatkin-Troitschanskaia (Hrsg.), *Stationen empirischer Bildungsforschung. Traditionslinien und Perspektiven* (S. 331–351). VS Verlag für Sozialwissenschaften.
- Nickolaus, R. (2014). Schwierigkeitsbestimmende Merkmale von Aufgaben und deren didaktische Relevanz. In U. Braukmann, B. Dilger & H.-H. Kremer (Hrsg.), *Wirtschaftspädagogische Handlungsfelder. Festschrift für Peter F. E. Sloane zum 60. Geburtstag* (S. 285–303). Eusl.
- Nickolaus, R., Abele, S. & Gschwendtner, T. (2012a). Valide Kompetenzabschätzung als eine notwendige Basis zur Effektbeurteilung beruflicher Bildungsmaßnahmen – Wege und Irrwege. In G. Niedermair (Hrsg.), *Kompetenzen entwickeln, messen und bewerten* (S. 537–554). Trauner.
- Nickolaus, R., Abele, S., Gschwendtner, T., Nitzschke, A. & Greiff, S. (2012b). Fachspezifische Problemlösefähigkeit in gewerblich technischen Ausbildungsberufen – Modellierung, erreichte Niveaus und relevante Einflussfaktoren. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 108*, 243–272.
- Norcini, J. J. & McKinley, D. W. (2007). Assessment Methods in Medical Education. *Teaching and Teacher Education, 23*(3), 239–250.
- Norman, G. (2005). Research in Clinical Reasoning: Past History and Current Trends. *Medical Education, 39*, 418–427.
- Norman, G., Young, M. & Brooks, L. (2007). Non-Analytical Models of Clinical Reasoning: the Role of Experience. *Medical Education, 41*, 1140–1145.
- Patel, V. L., Kaufman, D. R. & Magder, S. A. (1996). The Acquisition of Medical Expertise in Complex Dynamic Environments. In K. A. Ericsson (Hrsg.), *The Road to Excellence. The Acquisition of Expert Performance in the Arts and Sciences, Sports, and Games* (S. 127–165). Lawrence Erlbaum Associates.
- Popper, K. R. (1972). Naturgesetzte und theoretische Systeme. In H. Albert (Hrsg.), *Theorie und Realität. Ausgewählte Aufsätze zur Wissenschaftslehre der Sozialwissenschaften* (2., veränderte Aufl., S. 43–58). J. C. B. Mohr.
- Popper, K. R. (2005). *Logik der Forschung* (11. Aufl.). Tübingen: Mohr.

- Poser, H. (2012). *Wissenschaftstheorie. Eine philosophische Einführung*. Stuttgart: Reclam.
- Rasmussen, J. (1993). Diagnostic Reasoning in Action. *Systems, Man and Cybernetics, IEEE Transactions on*, 23(4), 981–992.
- Rausch, A., Schley, T. & Warwas, J. (2015). Problem Solving in Everyday Office Work – A Diary Study on Differences Between Experts and Novices. *International Journal of Lifelong Education*, 34, 448–467.
- 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.
- Renkl, A. (2012). Modellierung von Kompetenzen oder von interindividuellen Kompetenzunterschieden. Ein unterschätzter Unterschied. In J. Fleischer, D. Leutner & E. Klieme (Hrsg.), *Modellierung von Kompetenzen im Bereich der Bildung: Eine psychologische Perspektive. Psychologische Rundschau Themenheft*. 63(1), 50–53 [Themenheft]. Hogrefe.
- Robitzsch, A., Dörfler, T., Pfof, M. & Artelt, C. (2011). Die Bedeutung der Itemauswahl und der Modellwahl für die längsschnittliche Erfassung von Kompetenzen. Lesekompetenzentwicklung in der Primarstufe. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43(3), 213–227.
- Rouse, W. B. (1983). Models of Human Problem Solving: Detection, Diagnosis, and Compensation for System Failures. *Automatica*, 19(6), 613–625.
- Schaafstal, A., Schraagen, J. M. & van Berlo, M. (2000). Cognitive Task Analysis and Innovation of Training: The Case of Structured Troubleshooting. *Human Factors*, 42.
- Schaper, N. (1994). *Lernbedarfsanalysen und Trainingsgestaltung bei komplexen Diagnoseaufgaben*. Peter Lang.
- Schuwirth, L. (2009). Is Assessment of Clinical Reasoning still the Holy Grail? *Medical Education*, 43, 298–299.
- Schwartz, A. & Elstein, A. S. (2008). Clinical Reasoning in Medicine. In J. Higgs, M. A. Jones, S. Loftus & N. Christensen (Hrsg.), *Clinical Reasoning in the Health Professions* (3. Aufl., S. 223–234). Elsevier Ltd.
- Seifried, J., Rausch, A., Kögler, K., Brandt, S., Eigenmann, R., Schley, T., Siegfried, C., Eglloffstein, M., Küster, J., Wuttke, E., Sembill, D., Martens, T. & Wolf, K. D. (2016). Problemlösekompetenz angehender Industriekaufleute – Konzeption des Messinstruments und ausgewählte empirische Befunde (DomPL-IK). In K. Beck, M. Landenberger & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung. Ergebnisse aus der BMBF-Förderinitiative ASCOT* (S. 119–138). W. Bertelsmann.
- Sembill, D. (1992). *Problemlösefähigkeit, Handlungskompetenz und emotionale Befindlichkeit. Zielgrößen Forschenden Lernens*. Hogrefe.
- Sembill, D., Rausch, A. & Kögler, K. (2013). Non-cognitive facets of competence. Theoretical foundations and implications for measurement. In O. Zlatkin-Troitschanskaia & K. Beck (Hrsg.), *From Diagnostics to learning success – Proceedings in vocational education and training* (S. 199–212). Sense.

- Shavelson, R. J. (2009). Reflections on the assessment of job performance and competence. In F. Oser, U. Renold, E. G. John, E. Winther & S. Weber (Hrsg.), *VET BOOST: Towards a Theory of Professional Competencies* (S. 39–86). Sense.
- Sherbino, J., Dore, K. L., Wood, T. J., Young, M. E., Gaissmaier, W., Kreuger, S. & Norman, G. R. (2012). The relationship between response time and diagnostic accuracy. *Academic Medicine*, 87(6), 785–791.
- Stegmüller, W. (1967). Wissenschaftstheorie. In A. Diemer & I. Frenzel (Hrsg.), *Fischer-Lexikon Philosophie* (Das Fischer Lexikon, Bd. 11, S. 334–360). Fischer Taschenbuch Verlag.
- Stegmüller, W. (1989). *Hauptströmungen der Gegenwartsphilosophie. Eine kritische Einführung* (1. Bd., 7. Aufl.). Kröner.
- Straka, G. A. & Macke, G. (2009). *Neue Einsichten in Lehren, Lernen und Kompetenz*. Zugriff am 03.07.2017. Verfügbar unter <http://elib.suub.uni-bremen.de/ip/docs/00010417.pdf>
- Swanson, D. B., Norcini, J. J. & Grosso, L. J. (1987). Assessment of Clinical Competence: Written and Computer-Based Simulations. *Assessment and Evaluation in Higher Education*, 12(3), 220–246.
- Tamblyn, R., Abrahamowicz, M., Dauphinee, D., Wenghofer, E., Jacques, A., Klass, D., Smee, S., Blackmore, D., Winslade, N., Girard, N., Du Berger, R., Bartman, I., Buckridge, D. L. & Hanley, J. A. (2007). Physician Scores on a National Clinical Skills Examination as Predictors of Complaints to Medical Regulatory Authorities. *Jama*, 298(9), 993–1001.
- Tigelaar, D. E. H. & van der Vleuten, C. P. M. (2014). Assessment of Professional Competence. In S. Billett, C. Harteis & H. Gruber (Hrsg.), *International Handbook of Research in Professional and Practice-Based Learning* (S. 1237–1271). Springer.
- Vleuten, C. P. M. van der, Norman, G. & Schuwirth, L. (2008). Assessing Clinical reasoning. In J. Higgs, M. A. Jones, S. Loftus & N. Christensen (Hrsg.), *Clinical Reasoning in the Health Professions* (3. Aufl., S. 413–421). Elsevier Ltd.
- Vleuten, C. P. M. van der & Schuwirth, L. (2005). Assessing Professional Competence: From Methods to programmes. *Medical Education*, 39(3), 309–317.
- Walker, F., Link, N. & Nickolaus, R. (2016a). 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.
- Walker, F., Link, N., van Waveren, L., Hedrich, M., Geißel, B. & Nickolaus, R. (2016b). Berufsfachliche Kompetenzen von Elektronikern für Automatisierungstechnik – Kompetenzdimensionen, Messverfahren und erzielte Leistungen (KOKO EA). In K. Beck, M. Landenberger & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung. Ergebnisse aus der BMBF-Förderinitiative ASCOT* (S. 139–169). W. Bertelsmann.
- Wass, V., McGibbon, D. & van der Vleuten, C. (2001). Composite Undergraduate Clinical Examinations: How Should the Components Be Combined to Maximize Reliability? *Medical Education*, 35(4), 326–330.

- Wenghofer, E., Klass, D., Abrahamowicz, M., Dauphinee, D., Jacques, A., Smee, S., Blackmore, D., Winslade, N., Reidel, K., Bartman, I. & Tamblyn, R. (2009). Doctor Scores on National Qualifying Examinations Predict Quality of Care in Future Practice. *Medical Education*, 43(12), 1166–1173.
- Winther, E. (2010). *Kompetenzmessung in der beruflichen Bildung*. W. Bertelsmann.
- Wuttke, E., Seifried, J., Brandt, S., Rausch, A., Sembill, D., Martens, T. & Wolf, K. (2015). Modellierung und Messung domänenspezifischer Problemlösekompetenzen bei angehenden Industriekaufleuten. *Zeitschrift für Berufs und Wirtschaftspädagogik*, 111(2), 189–207.
- Zlatkin-Troitschanskaia, O. (2005). *Dynamik und Stabilität in Berufsbildungssystemen. Eine theoretische und empirische Untersuchung von Transformationsprozessen am Beispiel Bulgariens und Litauens*. Lang.
- Zlatkin-Troitschanskaia, O. & Seidel, J. (2011). Kompetenz und ihre Erfassung – das neue „Theorie-Empirie-Problem“ der empirischen Bildungsforschung? In O. Zlatkin-Troitschanskaia (Hrsg.), *Stationen Empirischer Berufsbildungsforschung* (S. 218–233). VS Verlag für Sozialwissenschaften.

## **Anhang: sechs Publikationen**



# Anhang A

## ***Publikation 1: Eigenständigkeit der diagnostischen Kfz-Problemlösekompetenz***

Nickolaus, R., Behrendt, S. & Abele, S. (2016). Kompetenzstrukturen bei Kfz-Mechatronikern und die Erklärungskraft des fachsystematischen Kfz-Wissens für berufsfachliche Kompetenzen. *Unterrichtswissenschaft*, 44, 114–130.

### **Eigener Anteil an der Studie und Begutachtungsverfahren**

Die Studienergebnisse resultierten aus zwei miteinander verbundenen Forschungsprojekten<sup>26</sup>, an deren Beantragung ich mitwirkte und in denen ich wissenschaftliche Leitungsfunktionen übernahm. In der hier behandelten Studie war ich für das Design, die Durchführung sowie die Ergebnisauswertung der Teilstudie verantwortlich, in der die diagnostische Kfz-Problemlösekompetenz erfasst wurde. Für die Publikation habe ich das Kapitel 4.2 „(Kurz-)Beschreibung der Tests und Beispielitems“ verfasst, die Daten zur diagnostischen Problemlösekompetenz aufbereitet und einen wesentlichen Beitrag zum inhaltlichen Zuschnitt, der Struktur des Artikels sowie statistischen Auswertung geleistet. Publikationen in *Unterrichtswissenschaft* durchlaufen eine zweifachblinde Begutachtung.

---

<sup>26</sup> Eines dieser Projekte ist „KOKO Kfz“, das ein Teilprojekt der Förderlinie „Technologieorientierte Kompetenzmessung in der Berufsbildung (ASCOT)“ war und vom Bundesministerium für Bildung und Forschung (BMBF) gefördert wurde. Das andere Projekt ist das seitens der Deutschen Forschungsgemeinschaft (DFG) geförderte Projekt „Multidimensionale Kompetenzstrukturen in der beruflichen Bildung“ mit dem Förderkennzeichen „NI 606/8-1“. An beiden Projekten wirkte ein größerer Kreis von Personen mit, die im Folgenden genannt seien und bei denen ich mich an dieser Stelle bedanken möchte: Prof. Dr. Tobias Gschwendtner, Dipl.-Gwl. Horst Heinzmann, Sebastian Heilig MA, Prof. Dr. Anja Sarnitz, Dipl.-Gwl. Thomas Schmidt und Thomas Trezbiatowski. Zudem möchte ich den zahlreichen Experten und Schulen für ihre Unterstützung danken.

# Kompetenzstrukturen bei KFZ-Mechatronikern und die Erklärungskraft des fachsystematischen KFZ-Wissens für berufsfachliche Kompetenzen

## *Competence Structures of Car Mechatronics and the Explanatory Power of the Professional Knowledge for Professional Abilities and Skills*

REINHOLD NICKOLAUS, STEFAN BEHRENDT, STEPHAN ABELE

In beruflichen Domänen stellt die valide Erfassung berufsfachlicher Kompetenzen eine Herausforderung dar, da das Konstrukt mehrdimensional ist und Fähigkeiten und Fertigkeiten im Gegensatz zum fachsystematischen Wissen nicht angemessen mit Papier-und-Bleistift-Tests erfasst werden können. Im Beitrag wird erstmals für die berufliche Domäne von KFZ-Mechatronikerinnen und -Mechatronikern im Rückgriff auf Simulationen und Videovignetten eine Analyse der Kompetenzstrukturen unter Einbezug der zentralen Anforderungssegmente vorgelegt. Weiter wird der Frage nachgegangen, welche prädiktive Kraft das fachsystematische Wissen für die Fehlerdiagnosekompetenz und das Handlungswissen in den Bereichen Service und Reparatur hat. Zentrale Ergebnisse: 1. Die Güte der Instrumente konnte gegenüber den vorausgegangenen Testversionen deutlich erhöht werden, in einzelnen Subdimensionen besteht jedoch noch Optimierungsbedarf. 2. Es wird eine 3-dimensionale Struktur für die Fehlerdiagnosekompetenz und das bereichsspezifische Handlungswissen dokumentiert, die bei Kontrolle des fachsystematischen Wissens schwache bis mittlere Korrelationen aufweisen. 3. Den einzelnen Dimensionen des fachsystematischen Wissens kommt bezogen auf die anderen Kompetenzdimensionen unterschiedliche prädiktive Kraft zu, abhängig davon, wie stark die Subdimensionen inhaltlich korrespondieren.

**Schlüsselwörter:** berufliche Bildung, Kompetenzstrukturen, Fachwissen, berufliche Fähigkeiten und Fertigkeiten

Valid measurement of professional competencies poses a challenge to researchers in vocational domains. This is because the construct's structure often is multidimensional and, unlike professional knowledge, skills and abilities cannot be measured adequately with paper and pencil tests. The following paper is the first to analyse the dimensional structure of professional competencies in the domain of automotive mechatronics technicians. The analysis is based on cross-sectional data that was gathered using simulations and video-based vignettes covering the main job requirements. A second question to be covered in this paper is the predictive power of professional

knowledge for the aforementioned professional competencies. The main results are: 1. Compared to preceding versions the quality of the test instruments was substantially improved. However, regarding some subtests optimization is still needed. 2. The results indicate a 3-dimensional structure of professional competencies reflecting trouble shooting, service, and maintenance and repair. After controlling professional knowledge all three show low to moderate correlations. 3. The different sub-dimensions of professional knowledge exhibit different predictive power with respect to the three competency dimensions. As could be expected, this appears largely due to the extent of overlaps in their content.

**Keywords:** vocational education, competency structures, professional knowledge, professional skills and abilities

## 1 Einleitung

In der beruflichen Bildung erweist es sich als zentrale Herausforderung, Aussagen zur Modellierung berufsfachlicher Kompetenzen bereitzustellen, da sich in den bisher einbezogenen Domänen (am Ausbildungsende) durchgängig stark ausdifferenzierte Strukturen abzeichnen (Nickolaus & Seeber, 2013) und zur reliablen und validen Abschätzung der einzelnen Subdimensionen jeweils erhebliche Testzeiten notwendig sind. Zugleich stehen bisher nur für einzelne Subdimensionen Instrumentarien zur Verfügung, weshalb Erweiterungen bzw. Neuentwicklungen von Tests zur Erfassung bisher unberücksichtigter Kompetenzfacetten notwendig sind.

Kontrovers diskutiert wird in den beruflichen Fachdidaktiken der Stellenwert fachsystematischen Wissens, das z. T. im Anschluss an Renkl (1996) ohne nähere Prüfung als träge unterstellt wird, verbunden mit der Hoffnung, die Kluft zwischen Wissen und Handeln (z. B. Mandl & Gerstenmaier, 2000) durch einen konsequenteren Handlungsbezug überwinden und berufliche Handlungskompetenz anbahnen zu können (z. B. Dörig, 2003). Zur Relevanz fachsystematischen Wissens für berufsfachliche Leistungen liegen zwar Studien vor (s. u.), allerdings ist es bisher in keinem Beruf gelungen, alle Anforderungsbereiche abzudecken, was insoweit problematisch scheint, als in den verschiedenen Anforderungsbereichen die Anforderungen erheblich variieren können (z. B. Routinetätigkeit vs. fachliches Problemlösen).

Ogleich berufliche Handlungskompetenz in der Didaktik beruflicher Bildung in der Regel umfassend definiert wird und neben berufsfachlichen soziale und personale Kompetenzen sowie kognitive Dispositionen und Bereitschaften einschließt (Sekretariat der Ständigen Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland, 2000), beschränken wir uns aus pragmatischen Gründen im Anschluss an Klieme und Leutner (2006) auf den Einbezug (erlernbarer) kognitiver Dispositionen.

In den Arbeiten zur Modellierung berufsfachlicher Kompetenzen (s. u.) dominieren gegenwärtig Arbeiten, in welchen für die berufsfachliche Kompetenz zumindest das Fachwissen und die Fähigkeit, dieses Fachwissen in (problemhaltigen) Anforder-

rungssituationen anwenden zu können, unterschieden werden. Für die Subdimensionen erfolgen domänenspezifisch weitere Ausdifferenzierungen. In diesem Beitrag gehen wir am Beispiel der KFZ-Mechatronikerin bzw. des KFZ-Mechatronikers den Fragen nach: Welche Kompetenzstruktur sich am Ausbildungsende ergibt und welche Relevanz dem fachsystematischen Wissen für die Ausprägungen der Subdimensionen berufsfachlicher Kompetenz zukommt.

## 2 Forschungsstand

### 2.1 Berufliche Anforderungen und Kompetenzstrukturen

Es liegen für verschiedene berufliche Domänen Arbeiten vor, in welchen Ergebnisse zu den Kompetenzstrukturen und der Erklärungskraft des Fachwissens für berufliche Fähigkeiten und Fertigkeiten berichtet wurden (Abele, 2014; Abele et al., 2012; Gschwendtner, Abele & Nickolaus, 2009; Nickolaus, Abele, Gschwendtner, Nitzschke & Greiff, 2012; Nickolaus, Geißel, Abele & Nitzschke, 2011; Walker, Link & Nickolaus, 2015; Winther & Achtenhagen, 2009). Das Fachwissen wurde in diesen Arbeiten curricular valide mit Papier- und Bleistift-Tests erfasst. Zur Erfassung beruflicher Fähigkeiten und Fertigkeiten kamen computersimulierte Arbeitsproben zum Einsatz. Mit den inhaltlichen Zuschnitten dieser simulativen Arbeitsproben wurden in aller Regel die Anforderungen in besonders anspruchsvollen Tätigkeitsfeldern abgebildet. So wurden z. B. bei den KFZ-Mechatronikerinnen und -Mechatronikern Diagnoseanforderungen simuliert, die in vorliegenden Tätigkeitsanalysen (z. B. Spöttl, Becker & Musekamp, 2011) als besonders anspruchsvoll ausgewiesen werden. Andere bedeutsame Anforderungssegmente, wie der Standardservice und Reparaturarbeiten, blieben in den bisherigen Studien unberücksichtigt.

Etwas breiter als bei den KFZ-Mechatronikerinnen und -Mechatronikern wurden die beruflichen Fähigkeiten und Fertigkeiten bei Industriekaufleuten (Winther & Achtenhagen, 2009) und Elektronikerinnen bzw. Elektronikern (Walker et al., 2015) erhoben. Dort wurde festgestellt, dass berufsfachliche Handlungskompetenzen (am Ausbildungsende) ein mehrdimensionales Konstrukt darstellen. Bei den Elektronikerinnen und Elektronikern für Automatisierungstechnik waren dies analytische und konstruktive Problemlösekompetenzen, die über die Fähigkeit, in steuerungstechnischen Anlagen Fehler zu diagnostizieren (analytisch), bzw. die Fähigkeit, steuerungstechnische Anlagen zu programmieren (konstruktiv), operationalisiert wurden. Gemessen an den Tätigkeitsschwerpunkten sind damit v. a. die kognitiv anspruchsvollsten Anforderungsbereiche abgedeckt (Walker et al., 2015). Ursächlich für die begrenzte Abdeckung der berufsfachlichen Fähigkeiten und Fertigkeiten ist der Sachverhalt, dass für eine reliable Erfassung der Teilkompetenzen (a) z. T. jeweils mehrstündige Testzeiten erforderlich sind,<sup>27</sup> (b) bei Anforderungen mit größeren manuellen Momenten

---

<sup>27</sup> So werden z. B. allein für die Erfassung der Fehlerdiagnosekompetenz bei KFZ-Mechatronikerinnen bzw. -Mechatronikern drei bis vier Stunden benötigt (Gschwendtner et al., 2009).

eine inhaltsvalide Testung an noch längeren Testzeiten zu scheitern droht<sup>28</sup> und (c) die Standardisierung bei inhaltsvaliden manuellen Anforderungen zusätzliche Herausforderungen mit sich bringt. Wenn hier von Fertigkeiten gesprochen wird, so sind damit in Anlehnung an Ackerman (1992) mehr oder weniger routiniert aktualisierbare Handlungspotenziale gemeint. Bezogen auf komplexere berufliche Handlungen unterstellen wir für den Ausbildungsverlauf, dass noch keine vollständige Routinebildung eintritt, was nicht ausschließt, dass Teilhandlungen (z. B. Demontagearbeiten) bereits routiniert erbracht werden können.

Folgt man Ackermans (1992) Theorie des Fertigkeitserwerbs, so sinkt die Bedeutung der kognitiven Grundfähigkeiten für die Leistungserbringung von der kognitiven über die assoziative bis zur automatisierten Phase des Fertigkeitserwerbs sukzessive ab. Theoretisch erklärt wird dies dadurch, dass das Ausmaß an bewusster Informationsverarbeitung bzw. an Aufmerksamkeit mit voranschreitendem Fertigkeitserwerb abnimmt. Vor dem Hintergrund des CLARION-Modells (Abele, 2014; Sun, 2006), in dem eine Ausdifferenzierung des Wissens in das Handlungswissen und das fachsystematische Wissen vorgenommen wird,<sup>29</sup> sind diese Routinebildungsprozesse als Hinweis deutbar, dass fachsystematisches Wissen für die Leistungserbringung weniger wichtig wird.

Rückgriffe auf das fachsystematische Wissen werden z. B. nötig, wenn das verfügbare Handlungswissen nicht hinreicht, die Anforderungen zu bewältigen. Das heißt, insbesondere im Aufbau des Handlungswissens während der Berufsausbildung sind enge Korrelationen zwischen dem Handlungswissen und dem fachsystematischen Wissen zu erwarten, die bei häufiger ausgeführten Tätigkeiten absinken sollten. Gestützt wird diese Annahme durch die bei KFZ-Mechatronikerinnen und -Mechatronikern ermittelten Befunde von Schmidt, Nickolaus und Weber (2014) im Service- und Reparaturbereich. Die dort berichteten absinkenden Korrelationen zwischen dem fachsystematischen Wissen und dem Handlungswissen verweisen zugleich auf sich im Ausbildungsverlauf vollziehende Änderungen der Kompetenzstrukturen, wozu bisher v. a. für das fachsystematische Wissen, das sich im Ausbildungsverlauf ausdifferenzieren scheint, Belege vorliegen (Gschwendtner, 2011; Schmidt et al., 2014). Weitgehend offen ist, ob sich solch ausdifferenzierte Kompetenzstrukturen auch bezogen auf das Handlungswissen und die berufsfachliche Problemlösekompetenz zeigen. Bei den KFZ-Mechatronikerinnen und -Mechatronikern ist dies plausibel, da sich bei Elektronikern selbst bei kognitiv gleichermaßen anspruchsvollen Anforderungen (analytische bzw. konstruktive Problemlösekompetenzen) Ausdifferenzierungen beobachten lassen und die Anforderungsunterschiede an KFZ-Mechatronikerinnen und -Mechatroniker bei der Fehlerdiagnose und den Servicetätigkeiten sowie den Reparaturtätigkeiten deutlich stärker ausfallen. Engere Bezüge sind für die Kompetenzen zu vermu-

---

28 Für den Wechsel eines Zahnriemens am Kraftfahrzeug sind z. B. oft mehrere Stunden erforderlich und auch für Reparaturen an den Bremsen ist mehr als eine Stunde notwendig.

29 CLARION ist eine kognitive Architektur, in der ein handlungsbezogenes und ein nicht-handlungsbezogenes Wissenssystem unterschieden wird. Wir verwenden in diesem Beitrag zur Kennzeichnung des nicht-handlungsbezogenen Wissens den Terminus fachsystematisches Wissen (s. auch Abele, 2014).

ten, die für die Bewältigung der Service- und der Reparaturaufgaben erforderlich sind, da die Tätigkeiten in diesen Segmenten inhaltliche Gemeinsamkeiten aufweisen.

## 2.2 Struktur des berufsfachlichen fachsystematischen Wissens

Relativ umfangreich ist der Kenntnisstand zur Struktur des berufsfachlichen fachsystematischen Wissens, wobei domänenübergreifend Ausdifferenzierungen entlang von Inhaltsbereichen berichtet werden (z. B. Gschwendtner, 2011; Nickolaus et al., 2011; Nickolaus & Seeber, 2013; Rosendahl & Straka, 2011; Seeber, 2008; Winther & Achtenhagen, 2009). Das gilt auch für KFZ-Mechatronikerinnen und -Mechatroniker im dritten Ausbildungsjahr, für die Gschwendtner (2011) eine 5- bzw. 6-dimensionale Modellierung vorlegte, allerdings noch völlig unbefriedigende Reliabilitäten dokumentierte. Da das 6-dimensionale Modell eine bessere Passung aufweist als das 5-dimensionale, entscheidet sich Gschwendtner für das 6-dimensionale Modell. Unter dem Gesichtspunkt der Sparsamkeit scheint es angesichts der hohen Korrelation zwischen der Motorsteuerung und dem Inhaltsbereich Start-Strom-Beleuchtung ( $r=0.87$ ) empirisch vertretbar, das 5-dimensionale Modell zu präferieren. Das von Gschwendtner verwendete Instrument wurde im Rahmen dieses Projekts einer Revision und Erweiterung unterzogen, was an anderer Stelle bereits beschrieben wurde (Schmidt et al., 2014). Zur Vorbereitung adaptiven Testens wurden in dieser neuen Testversion geschlossene Antwortformate eingesetzt. Für diese neue Testversion ergibt sich am Ende des dritten Ausbildungsjahres für das fachsystematische Wissen eine 5-dimensionale Struktur (Schmidt et al., 2014) mit den auch bei Gschwendtner (2011) ausgewiesenen Subdimensionen (Fahrwerk, Kraftübertragung, Motor, Motorsteuerung/Start-Strom-Beleuchtung<sup>30</sup>, Service). Motorsteuerung und Start-Strom-Beleuchtung, d. h. zwei elektrotechnikaffine Subdimensionen werden hier integrativ modelliert, wenngleich die latente Korrelation zwischen den beiden (potenziellen) Subdimensionen einen Wert von 0.71 (Abele, Behrendt, Weber & Nickolaus, 2015) erreicht. Das heißt, es ist offen, ob die 5- oder die 6-dimensionale Struktur besser passt. Im Zuge der Testüberarbeitung konnten die Reliabilitäten verbessert werden, wenngleich aufgrund der eingeschränkten Testzeit und des Testdesigns nicht in allen Subdimensionen befriedigende Reliabilitätswerte erreicht werden konnten (s. u.).

## 2.3 Prädiktive Kraft des fachsystematischen Wissens

Zur prädiktiven Kraft des fachsystematischen Wissens für die berufsfachlichen Problemlösekompetenzen liegen überwiegend Studien vor, in welchen das fachsystematische Wissen 1-dimensional bzw. als Generalfaktor einbezogen wurde (Gschwendtner et al., 2009; Nickolaus et al., 2011, 2012). Die Problemlösekompetenzen wurden in der Regel über Anforderungen operationalisiert, in technischen Systemen Fehler zu diagnostizieren (Kfz-Fehlerdiagnosekompetenz). In der Studie von Walker et al. (2015), in der die Einflüsse des fachsystematischen Wissens auf die berufsfachliche Problemlöse-

---

30 Die gemeinsame Modellierung der Motorsteuerung und Start-Strom-Beleuchtung, die bereits bei Gschwendtner (2011) hoch korrelierten und eine fünfdimensionale Struktur nahelegten, erweist sich bei der vergleichenden Modellprüfung der sechsdimensionalen Modellierung mit separaten Subdimensionen gemessen an den Informationskriterien als überlegen, ein signifikanter Unterschied besteht nicht (Schmidt et al., 2014).

kompetenz differenziert analysiert wurden, erwiesen sich v. a. jene Wissensbereiche als prädiktiv, die auch inhaltlich enger mit den Problemlösezuschnitten assoziiert waren (Walker et al., 2015). Das legt bezogen auf die KFZ-Mechatronikerinnen und -Mechatroniker die Vermutung nahe, dass das mechanische Wissen primär für die Reparaturkompetenzen prädiktiv wird und das elektrotechnische Wissen für die Fehlerdiagnosekompetenz in elektrotechnischen KFZ-Systemen. Im Servicebereich führt eine an der Sicherung inhaltlicher Validität ausgerichtete Operationalisierung dazu, dass Items z. T. stärkere Bezüge zu mechanischen und z. T. auch zu elektrotechnischen Inhalten aufweisen. Eine eindeutige Zuordnung zu einem mechanischen oder elektrotechnischen Anforderungsbereich scheint von daher theoretisch nicht sinnvoll. Da in unserem Test zum Service-Handlungswissen aus pragmatischen Gründen mechanische Anforderungen dominieren, ist es naheliegend, dass insbesondere das mechanische fachsystematische Wissen für das Handlungswissen prädiktiv wird.

## 2.4 Instrumente zur Erfassung berufsfachlicher Kompetenzen und deren Validität

Die Instrumentenentwicklung zur Erfassung berufsfachlicher Kompetenzen ist für den Beruf des KFZ-Mechatronikers weit fortgeschritten. Mit den verfügbaren Instrumentarien (Abele, Walker & Nickolaus, 2014; Schmidt et al., 2014) können die drei zentralen inhaltlichen Anforderungsbereiche dieses Berufs (a) Fehlerdiagnose, (b) Standardservice und (c) Reparatur so abgedeckt werden, dass neben dem einschlägigen fachsystematischen Wissen das Handlungswissen (Service, Reparatur) und eine wichtige Facette der fachspezifischen Problemlösekompetenz (Kfz-Fehlerdiagnosekompetenz) erfasst werden kann (Gschwendtner, 2011; Spöttl et al., 2011). Vor dem Hintergrund des CLARION-Modells werden mit den Instrumenten in den verschiedenen Bereichen jedoch unterschiedliche Dimensionen berufsfachlicher Kompetenz erfasst. Die Fehlerdiagnosekompetenzen (a) werden mit computersimulierten Arbeitsproben erfasst. Gschwendtner et al. (2009) zeigen, dass dieser authentische diagnostische Zugang konstruktvalide Testwerte erbringt. So lag die Korrelation zwischen der computerbasiert und an einem Kraftfahrzeug in der Werkstatt erfassten Fehlerdiagnosekompetenz bei  $r = .94$ . Als herausfordernd erwies sich bei der Fehlerdiagnose die Sicherung einer akzeptablen Reliabilität, die in den ersten Testversionen trotz ca. vier Stunden Testzeit über  $EAP/PV = .6$  nicht hinausging. Durch die Entwicklung von Teilkompetenzitems der KFZ-Fehlerdiagnose war es möglich, bei einer vertretbaren Testzeit (135 Min.) eine zufriedenstellende Reliabilität zu erreichen.<sup>31</sup> Die Teilkompetenzitems und die komplexen Problemfälle (Bearbeitungszeiten von 15–25 Min.) sind konvergent valide (ebd.).

Die Kompetenzmaße in den Bereichen des Standardservice (b) und der Reparatur (c) wurden über Videovignetten gewonnen, die zwar authentische Service- und Reparaturarbeiten wiedergeben, jedoch eher geeignet sind, das bereichsspezifische Handlungswissen abzubilden. Dieses Handlungswissen wird in Form von Handlungsplänen

---

31 Abele et al. (2014) zeigen, dass durch die Entwicklung der Teilkompetenz-Items eine deutliche EAP/PV-Reliabilitätsverbesserung von .62 auf .75 erzielt wurde.

und Handlungsbeurteilungen sowie dem situationsspezifisch notwendigen Fachwissen erhoben (Schmidt et al., 2014). Inwieweit die bei den realen Service- und Reparaturarbeiten notwendigen manuellen Fertigkeiten leistungskritisch werden, ist mit den Instrumenten nicht abschätzbar. Für die Fehlerdiagnosekompetenz zeigten Gschwendtner et al. (2009), dass die manuellen Anforderungen nicht leistungskritisch werden. Das heißt, dass in dieser Subdimension, wie bei den beiden potenziellen Subdimensionen des Handlungswissens, die kognitiven Leistungen bestimmend sind. Zu berücksichtigen ist, dass in allen Leistungsmaßen motivationale und metakognitive Momente enthalten sind, was nicht nur in Einklang mit dem CLARION-Modell steht, sondern empirisch bestätigt wurde (Nickolaus et al., 2015).

### 3 Zentrale Fragestellungen und Hypothesen

Geklärt werden soll, (1) ob sich bei den KFZ-Mechatronikerinnen und -Mechatronikern für das Handlungswissen und die Fehlerdiagnosekompetenz eine mehrdimensionale Struktur bestätigen lässt, wie das für das fachsystematische Wissen mehrfach erfolgte (Gschwendtner, 2011; Schmidt et al., 2014). Zudem wird untersucht, welche Erklärungskraft dem fachsystematischen Fachwissen für das Service- und Reparatur-Handlungswissen sowie die Fehlerdiagnosekompetenz (berufsfachliche Problemlösekompetenz) zukommt. Im Anschluss an den Forschungsstand erwarten wir, dass sich bei den KFZ-Mechatronikerinnen und -Mechatronikern unter Einbezug der Fehlerdiagnosekompetenzen und des auf Service- und den Reparaturbereich bezogenen Handlungswissens eine dreidimensionale Struktur zeigt, die durch engere Zusammenhänge zwischen dem Service- und dem Reparaturbereich als zwischen der Fehlerdiagnosekompetenz und dem Service- bzw. Reparaturbereich gekennzeichnet ist. Von Interesse ist in diesem Kontext, dass das Service- und Reparatur-Handlungswissen mit einem strukturell identischen Instrumentarium (Videovignetten) erhoben wird, was eine 1-dimensionale Skalierung dieser potenziellen Subdimensionen begünstigt. Bei Kontrolle des fachsystematischen Wissens ist zu erwarten, dass die Korrelationen zwischen den Fehlerdiagnosekompetenzen und den Service- sowie den Reparaturkompetenzen deutlich zurückgehen: Wir unterstellen, dass die in diesen Bereichen erreichten Kompetenzstände stark vom fachsystematischen Wissen abhängen und die Korrelationen zwischen den Kompetenzdimensionen z.T. auf das fachsystematische Wissen zurückzuführen sind. Geprüft werden folgende Hypothesen:

*H1:* Das 3-dimensionale Modell, das durch die Fehlerdiagnosekompetenzen sowie das Service- und Reparatur-Handlungswissen aufgespannt wird, repräsentiert die empirische Datenlage besser als ein 2-dimensionales (Fehlerdiagnose; Service und Reparatur) und ein 1-dimensionales Modell.

*H2:* Die Korrelationen zwischen dem Service- und dem Reparatur-Handlungswissen sind höher als zwischen der Fehlerdiagnosekompetenz und den beiden anderen Subdimensionen.

H3: Das fachsystematische Wissen ist ein bedeutsamer Prädiktor der Fehlerdiagnosekompetenz sowie des Service- und Reparatur-Handlungswissens.

H4: Das fachsystematische Wissen erreicht bezogen auf die Fehlerdiagnosekompetenz eine höhere prädiktive Kraft als bei dem routineträchtigeren Service- und Reparatur-Handlungswissen.

## 4 Untersuchungsanlage und Instrumente

### 4.1 Untersuchungsanlage

Den Analysen liegt ein Datensatz aus einer Querschnittsuntersuchung zugrunde, der am Ende der Ausbildung im Rahmen des Projekts KOKO Kfz<sup>32</sup> von März bis Juli 2013 erhoben wurde und insgesamt 201 fast nur männliche Probanden umfasst (ca. 98 %; Altersdurchschnitt 19.9 Jahre), die überwiegend in Handwerksbetrieben ausgebildet wurden. Die Stichprobe wurde über Berufsschulen gewonnen und war daran gebunden, dass eine Testzeit von zwölf Unterrichtsstunden verfügbar war. Zur Klärung der Forschungsfragen wurden Daten von Auszubildenden herangezogen, die alle Tests (fachsystematisches Wissen, Service- und Reparatur-Handlungswissen sowie Fehlerdiagnosekompetenz) bearbeiteten.

### 4.2 (Kurz-)Beschreibung der Tests und Beispielitems

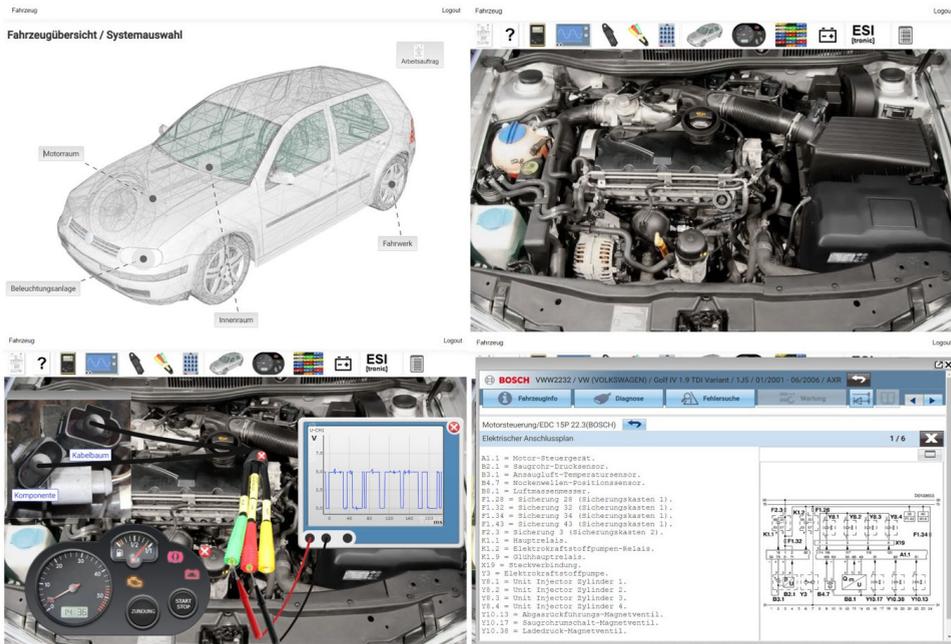
Die inhaltliche Validität der Instrumente wurde durch den Einbezug erfahrener Praktiker (Lehrkräfte aus Schulen und Betrieben, Prüfungsexperten) sowie Analysen der realisierten Curricula und vorliegende Tätigkeitsanalysen gesichert. Auch internationale Analysen zeigen, dass mit den einbezogenen Tätigkeiten wesentliche Anforderungsbereiche von KFZ-Mechatronikerinnen und -Mechatronikern abgedeckt sind (Müller & Schelten, 2009).

Zur Diagnostik der berufsfachlichen Problemlösekompetenz (*KFZ-Fehlerdiagnosekompetenz*) wurde eine Computersimulation entwickelt (Abele et al., 2012, 2014; Gschwendner et al., 2009), in der komplexe KFZ-Systeme (z. B. der Motorraum) sowie Ausschnitte der realen Berufsumwelt von KFZ-Mechatronikerinnen und -Mechatronikern (z. B. ein Oszilloskop) authentisch abgebildet und nutzbar sind (Abbildung 1). Bei der Testung wurden insgesamt 13 Fehlerdiagnose-Items vorgelegt. Die Simulation umfasst mehrere KFZ-Systeme und ein computerbasiertes Expertensystem als Standardwerkzeug von KFZ-Mechatronikerinnen und -Mechatronikern. In der Simulation sind unzählige Diagnoseoperationen<sup>33</sup> möglich, und es können verschiedene authentische KFZ-Defekte (z. B. ein defektes Abblendlicht) vorgegeben werden.

---

32 Gefördert vom Bundesministerium für Bildung und Forschung in der Förderlinie *Technologieorientierte Kompetenzmessung in der Berufsbildung* (ASCOT).

33 So sind z. B. mehrere tausend elektrotechnische Messungen möglich.



**Abbildung 1:** Screenshots aus der KFZ-Computersimulation zur Ermittlung der Fehlerdiagnosekompetenz. Oben links: Fahrzeugübersicht; oben rechts: Motorraumübersicht; unten links: Messung mit dem Oszilloskop; unten rechts: Stromlaufplan aus dem Expertensystem (Abele et al., 2014, S. 170)

Das folgende Beispiel, das sich auf einen Defekt einer Signalleitung zwischen dem Ladedruckmagnetventil (LMV) und dem Steuergerät bezieht, soll die Anforderungen der Fehlerdiagnose veranschaulichen. Bei der Diagnose dieses Defekts sind zunächst die Problembeschreibung sowie technische Spezifika des Fahrzeugs zu erfassen. Üblicherweise wird zudem mit dem computerbasierten Expertensystem der Fehlerspeicher ausgelesen. Hierdurch wird sichtbar, dass im Bereich des LMV ein Problem vorliegt, und es werden im Expertensystem verschiedene Ursachenhypothesen und Prüfstrategien vorgeschlagen. Um die Prüfungen vorzunehmen, müssen Informationen gewonnen werden, wo sich das LMV befindet und an welchen Stellen des LMV zu messen ist. Hierfür sind die in der Simulation hinterlegten Einbaulage- und Stromlaufpläne zu studieren. Im nächsten Schritt werden die vom Expertensystem vorgeschlagenen Messschritte unter Nutzung geeigneter Messgeräte vorgenommen. Anschließend werden die Messergebnisse bewertet, wobei erkennbar wird, dass die vom Expertensystem vorgeschlagenen Prüfungen nicht für eine eindeutige Ursachenbestimmung ausreichen. Zu überlegen ist nun, wie im Weiteren vorzugehen ist. Im Beispiel ist die Defektursache durch eine Widerstandsmessung an der Signalleitung des LMV eindeutig bestimmbar. Neben den komplexen, authentische Fehler abdeckenden Aufgaben wurden Teilkompetenzitems administriert, für welche z. B. aus dem Expertensystem Referenzdaten abgegriffen und im Anschluss eine Vergleichsmessung durchgeführt werden mussten (Abele et al., 2014).

Zur Erfassung des Handlungswissens in den Bereichen *Standardservice* und *Reparatur* wurde eine Computerplattform entwickelt, in der technische Informationsmaterialien verfügbar sind sowie von einem KFZ-Mechatroniker durchgeführte fachliche Handlungen mittels Videovignetten präsentiert werden (Abbildung 2). Die Durchführungsgüte dieser Handlungen wird gezielt variiert und die dargestellten Handlungssequenzen werden an geeigneten Stellen unterbrochen, um die folgenden Dimensionen zu erfassen: (a) die Qualität der individuellen Handlungspläne, (b) die Fähigkeit, die dargestellten Handlungen zu beurteilen, und (c) das für die Handlungen notwendige, situationsspezifische Fachwissen (Schmidt et al., 2014). Die Anzahl der einbezogenen Items liegt bei 54 (Service) bzw. 81 (Reparatur).

Das in Abbildung 1 gegebene Beispiel bezieht sich auf einen Ölwechsel, eine typische Service-Tätigkeit. Links ist das Bild einer Videosequenz zu sehen, die den Ölwechsel eines KFZ-Mechatronikers zeigt. Nach einer gewissen Zeit stoppt die Sequenz und es erscheint die Frage, welcher Arbeitsschritt im Anschluss an die gezeigte Sequenz zu vollziehen ist (vgl. rechte Seite der Abbildung 2). Wie aus Abbildung 2 hervorgeht, ist als nächstes der neue Dichtring zu montieren. Mit diesem Item wird das Service-Handlungswissen, d. h. die Ausprägung individueller Handlungspläne im Bereich des Ölservice erfasst.



**Abbildung 2:** Screenshots aus dem computerbasierten Test zur Erfassung des Standardservice-Handlungswissens. Links: Standbild einer Handlungssequenz zum Ölwechsel; rechts: administrierte Fragestellung

Die Darbietung der geschlossenen Items zum *fachsystematischen Wissen* (62 Items) erfolgt im Papier-und-Bleistift-Format. Ergänzend wurden partiell benötigte Stromlaufpläne und Schaubilder zu komplexen KFZ-Komponenten bereitgestellt. Beispielimitem zur Motorsteuerung: *Welche Funktion hat die Leitung zwischen Pin 44 am Steuergerät X1 und Pin 2 am Fahrgeschwindigkeitssensor B2? Zur Lösung dieser Aufgabe benötigen Sie den Stromlaufplan Nr. 1 aus dem Anhang.* Bei diesem Item müssen die Probandinnen und Probanden aus vier Antwortmöglichkeiten die korrekte auswählen, was ohne Raten nur im Rückgriff auf den Stromlaufplan Nr. 1 des Anhangs möglich ist.

### 4.3 Güte der Instrumentarien

Die Reliabilität des Instruments zur Erfassung der Fehlerdiagnosekompetenz (Testzeit 135 Min.) liegt bei  $EAP/PV = .75$ . Die Reliabilität der jeweils 2-dimensional modellier-

ten Service- und Reparaturkompetenzen<sup>34</sup> (Testzeit 180 Min.) erreicht (noch akzeptable) Werte von EAP/PV = .65–.68 und WLE = .55–.61. Für das 5-dimensionale fachsystematische Wissen (Fahrwerk, Kraftübertragung, Motor, Motorsteuerung/Start-Strom-Beleuchtung, Service) liegen die EAP/PV-Reliabilitäten in der Größenordnung von .67 bis .84, die WLE-Reliabilitäten im Bereich von .37 bis .66 (vgl. Abbildung 5). Die WLE-Reliabilität der Subdimension Kraftübertragung erweist sich als völlig unbefriedigend (.008), so dass auf deren Einbezug in den Analysen verzichtet wird. Von einer 1-dimensionalen Skalierung sehen wir trotz der partiell unbefriedigenden Reliabilitäten ab, da die Ausdifferenzierung unseres Erachtens zusätzliche Orientierung für didaktische Entscheidungen bringt.

## 5 Ergebnisse<sup>35</sup>

### 5.1 Dimensionalität der berufsfachlichen Kompetenz

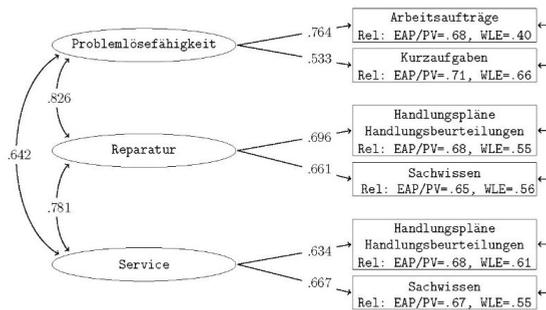
Die Prüfung von H1 erfolgte über einen Modellvergleich von drei Varianten: (a) dem postulierten 3-dimensionalen Modell mit der Fehlerdiagnosekompetenz und dem Handlungswissen in den Bereichen Service und Reparatur (Abbildung 3; Modell 3D), (b) einem 2-dimensionalen Modell mit der Fehlerdiagnosekompetenz einerseits und dem Handlungswissen im Service- und Reparaturbereich andererseits (Modell 2D) und (c) einem 1-dimensionalen Modell (Modell 1D). Die Beurteilung der Modellpassung erfolgte auf Basis folgender Fit-Grenzwerte: RMSEA  $\leq$  0.08 (Wang & Wang, 2012); CFI  $\geq$  .9 (Raykov & Marcoulides, 2006);  $X^2/df < 5$  (Hoyle, 2012). Zudem wurden der SRMR  $\leq$  .08 (Moosbrugger & Schermelleh-Engel, 2008) und die Informationskriterien AIC und BIC herangezogen. Für die Dimensionalitätsprüfung wurden die (geschachtelten) Modelle mit Hilfe des  $X^2$ -Differenzentests und der Informationskriterien AIC und BIC verglichen (Kline, 2011).

Nahezu in allen Fitmaßen ist eine Steigerung vom 1-dimensionalen über das 2-dimensionale hin zum 3-dimensionalen Modell zu verzeichnen (Tabelle 1). Das 1-dimensionale Modell erfüllt einige Grenzwerte nicht. Im Vergleich der Informationskriterien AIC und BIC lässt sich die gleiche Entwicklung, jedoch v. a. beim AIC in nicht ausreichender Größenordnung verzeichnen (ebenfalls Tabelle 1). Der  $X^2$ -Vergleich des 3-dimensionalen Modells mit dem

1- und 2-dimensionalen Modell fällt jeweils zugunsten des 3-dimensionalen Modells aus (Tabelle 2). Hervorgehoben sei nochmals, dass das Handlungswissen in den Bereichen Service und Reparatur mit strukturidentischen Videovignetten erfasst wurden und dennoch eine empirische Trennung des Handlungswissens im Service- und Reparaturbereich möglich ist.

34 Die 2-dimensionale Modellierung der Service- und der Reparaturkompetenzen erweist sich am Ende des dritten Ausbildungsjahres sowohl einer 1-dimensionalen als auch einer 3-dimensionalen Modellierung (Handlungspläne, Handlungsbeurteilungen, tätigkeitsbezogenes Fachwissen) überlegen (Schmidt et al., 2014).

35 Die statistischen Analysen erfolgten mit R (R Core Team, 2014), speziell dem Paket lavaan (Rosseel, 2012).



Standardisiertes Modell.  $N = 201$ ,  $\chi^2 = 9.400$ ,  $df = 6$ , CFI = .986, TLI = .965, RMSEA = .053, SRMR = .029, Devianz = 2892, AIC = 2922, BIC = 2972  
 Wenn nicht anders angegeben:  $p < .001$ ; \*\*;  $p < .01$ ; \*;  $p < .05$ ; n. s.: nicht signifikant.

**Abbildung 3:** Messmodell, endogene Variablen 3-dimensional

Ob diese Trennung durch Unterschiede in den Tätigkeitsanforderungen oder durch unterschiedliche Vertrautheitsgrade verursacht ist, wäre zu klären.

**Tabelle 1:** Kennwerte Messmodell endogene Variablen

Kennwert	Modell 1D	Modell 2D	Modell 3D
N	201	201	201
$\chi^2$	23.039	17.126	9.400
$df$	9	8	6
CFI	.942	.983	.986
TLI	.904	.930	.965
RMSEA	.088	.075	.053
SRMR	.049	.042	.029
Devianz	2906	2900	2892
AIC	2930	2926	2922
BIC	2970	2969	2972

**Tabelle 2:** Modellvergleich Messmodell endogene Variablen

Modelle	$\chi^2$	$df$	$p$
Modell 1D Modell 2D	5.9129	1	.01503
Modell 2D Modell 3D	7.7260	2	.02101
Modell 1D Modell 3D	13.639	3	.00344

## 5.2 Zusammenhänge der berufsfachlichen Kompetenzdimensionen

Die Korrelationen zwischen dem Handlungswissen der Bereiche Service und Reparatur fallen erwartungswidrig nicht durchgängig höher aus als jene zwischen der Diagnosekompetenz und dem jeweiligen Handlungswissen (H2). Lediglich die Korrelation zwischen der Servicekompetenz und der Diagnosekompetenz fällt erwartungskonform niedriger aus. Mit einer latenten Korrelation von  $r = .83$  zwischen der Fehlerdiagnosekompetenz und dem Handlungswissen im Reparaturbereich wird H2 falsifiziert (Abbildung 4). Die latenten Korrelationen zwischen der Fehlerdiagnosekompetenz und den beiden anderen Subdimensionen sind überraschend hoch, da die inhaltlichen Bezüge deutlich geringer sind als zwischen dem Service und der Reparatur. Begünstigt wird dies möglicherweise dadurch, dass die manuellen Momente, die im Service und der Reparatur bedeutsam sind, über die Videovignetten nicht erfasst werden und damit bei allen drei Subdimensionen kognitive Anforderungen dominieren.

## 5.3 Differenzielle prognostische Validität unterschiedlicher Dimensionen fachsystematischen Wissens

Vor diesem Hintergrund scheint von Interesse, wie sich die Bezüge zwischen den Subdimensionen verändern, wenn das fachsystematische Wissen in die Modellierung einbezogen wird. Theoretisch ist das fachsystematische Wissen ein starker Prädiktor für alle drei Subdimensionen, wobei die Pfadkoeffizienten v. a. dort relativ groß werden sollten, wo auch inhaltliche Affinitäten zwischen den Dimensionen fachsystematischen Wissens und den abhängigen Dimensionen bestehen. Wie eingangs ausgeführt, ist die Frage, ob sich, bezogen auf das fachsystematische Wissen, eine 5- oder 6-dimensionale Struktur als angemessen erweist, nicht abschließend geklärt. Aus didaktischen Erwägungen berücksichtigen wir im Weiteren die Dimensionen Motorsteuerung und Start-Strom-Beleuchtung als eigene Subdimensionen, wofür auch die relativ schwache Assoziation zwischen den beiden Konstrukten spricht ( $r = 0.71$ ). Einbezogen werden damit die Subdimensionen Fahrwerk, Motor, Motorsteuerung, Start-Strom-Beleuchtung und Service.<sup>36</sup> Wir sehen davon ab, die Dimensionen des fachsystematischen Wissens latent zu modellieren, da dies angesichts der relativ kleinen Stichprobengröße problematisch wäre. Stattdessen wählen wir zur Prüfung von H3 den Weg, die Dimensionen des fachsystematischen Wissens über die WLE-Schätzer zu modellieren, wenn gleich in einzelnen Subdimensionen unbefriedigende WLE-Reliabilitäten erreicht werden (Abbildung 5). Das in Abbildung 5 wiedergegebene Modell berücksichtigt aus Übersichtsgründen lediglich die signifikanten Pfade.<sup>37</sup> Eine Modellierung unter Einbezug der nichtsignifikanten Pfade erweist sich nicht als signifikant besser und weist im Kern die gleiche Struktur auf.<sup>38</sup>

36 Die Kraftübertragung bleibt hier, wie oben angemerkt, aufgrund unzureichender Reliabilität unberücksichtigt.

37 Die nichtsignifikanten Pfade wurden auf null gesetzt, was vertretbar erschien, da die standardisierten Koeffizienten durchgängig Werte  $< .1$  aufwiesen.

38  $X^2 = 7.0845$ ;  $df = 6$ ;  $p = .3131$

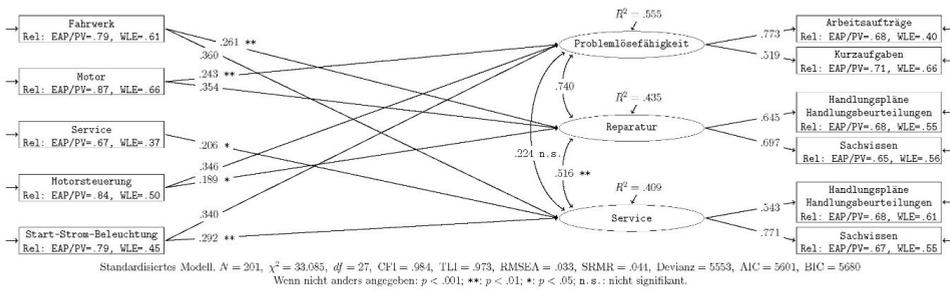


Abbildung 4: Pfadmodell Fachkompetenz (erklärungsrelevante Pfade)

Der Modellfit erweist sich als gut. Die Residualkorrelationen zwischen den drei „abhängigen“ Subdimensionen fallen erwartungskonform niedriger aus (Abbildung 4) als im vorigen Modell (Abbildung 3). Besonders deutlich fällt die Korrelation zwischen der Diagnosekompetenz und dem Service-Handlungswissen ab. Bezogen auf H3 ist von Interesse, welche prädiktive Kraft den einzelnen Wissensdimensionen für die Diagnosekompetenz und das bereichsspezifische Handlungswissen zukommt. Zu berücksichtigen sind bei den relativ gering ausfallenden Korrelationen die teils geringen WLE-Reliabilitäten der Fachwissensdimensionen. Zunächst wird deutlich, dass von den eher durch mechanisches Wissen gekennzeichneten Dimensionen des fachsystematischen Wissens (Fahrwerk, Motor) lediglich die Dimension Fahrwerk ausschließlich für die stärker durch mechanische Anforderungen gekennzeichneten Tätigkeiten im Service- und Reparaturbereich prädiktiv wird. Die Wissensdimension Motor erweist sich für die Diagnosekompetenz als bedeutsam, was plausibel scheint, da ein großer Teil der Fehlerfälle im Bereich der Motorsteuerung zu verorten ist und deren Diagnose nicht nur ein Wissen über die Ansteuerung von Motorelementen, sondern auch ein Wissen zur Funktion des Motors voraussetzt. Das über den Papier-und-Bleistift-Test erhobene Servicewissen wird lediglich für das Service-Handlungswissen prädiktiv.<sup>39</sup> Zu berücksichtigen ist in diesem Falle auch die niedrig ausfallende WLE-Reliabilität, die wohl zu einer Unterschätzung des Zusammenhangs führt (Schmidt et al., 2014). Das Wissen zur Motorsteuerung wird insbesondere für die Diagnosekompetenz prädiktiv, was plausibel ist, da ohne dieses Wissen erfolgreiche Diagnoseprozesse kaum möglich scheinen. Die Wissensdimension Start-Strom-Beleuchtung wird sowohl für die Diagnosekompetenz als auch den Service prädiktiv. Besonders stark ist dieser Einfluss erwartungskonform (aufgrund der gemeinsamen elektrotechnischen Basen) auf die Fehlerdiagnosekompetenz. Der Einfluss auf das Service-Handlungswissen ist nicht hypothesenkongruent, da in den einschlägigen Videovignetten Handlungen dominierten, die mechanische Anforderungen repräsentierten. Denkbar wäre, dass hier die nicht kontrollierten allgemeinen kognitiven Fähigkeiten bedeutsam werden. Insgesamt stützt dieses Erklärungsmodell die Grundannahme, dass v. a. jene Wissensbereiche prädiktiv werden, deren Verfügbarkeit bei der Ausführung der jeweiligen

39 Bei Einbezug der nichtsignifikanten Pfade fällt dieser Pfadkoeffizient nicht mehr signifikant aus ( $r = .17, p = .057$ ).

Tätigkeiten bedeutsam wird. Bemerkenswert sind die relativ hohen Varianzaufklärungen, die im Falle der Fehlerdiagnosekompetenz ca. 46 % und bezogen auf das bereichsspezifische Handlungswissen etwa 41 bis 44 % erreichen.

### 5.4 Prognostische Validität des fachsystematischen Wissens

Angesichts der teils unbefriedigenden WLE-Reliabilitäten prüfen wir auch ein Erklärungsmodell, in dem das fachsystematische Wissen als Generalfaktor der einzelnen Fachwissensdimensionen modelliert wird (Abbildung 5).

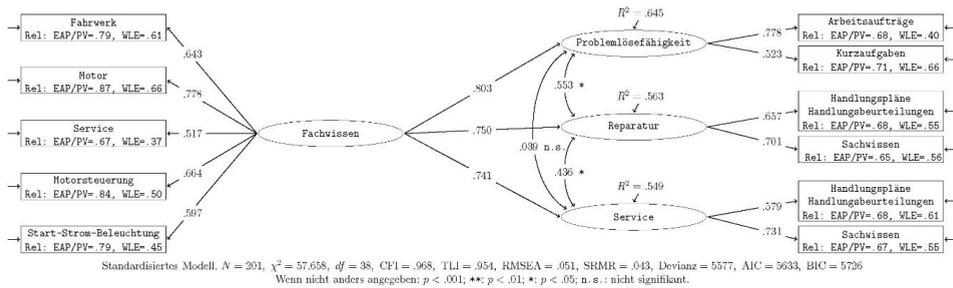


Abbildung 5: Pfadmodell Fachkompetenz (latente Modellierung Fachwissen)

Dies führt erwartungskonform bei allen endogenen Variablen zu einer deutlichen Steigerung der erklärten Varianzanteile. Überraschend hoch fallen die Erklärungsanteile (und Zuwächse) im Falle des Service und der Reparatur aus, was möglicherweise auch dem Umstand geschuldet ist, dass in diesen Fällen das Handlungswissen erfasst wird. Die von Schmidt et al. (2014) im Servicebereich und etwas schwächer auch im Reparaturbereich dokumentierte Routinebildung scheint noch nicht so weit fortgeschritten, dass das relevante Wissen nur noch implizit verfügbar ist.

## 6 Diskussion

Mit dem realisierten Untersuchungszuschnitt wurde unseres Erachtens erstmals in einem Ausbildungsberuf (nahezu) das gesamte Spektrum der berufsfachlichen Anforderungen abgedeckt. Die entwickelten Instrumentarien zur Fehlerdiagnose und zur Erfassung des Handlungswissens in den Bereichen Service und Reparatur können als innovative und weiterführende Testversionen gelten. Erwartungskonform erweisen sich die Fehlerdiagnosekompetenz, das Service- und das Reparatur-Handlungswissen als eigene Subdimensionen berufsfachlicher Kompetenz. Die Korrelation zwischen der Fehlerdiagnosekompetenz und der Reparatur fällt deutlich höher aus als zwischen der Fehlerdiagnosekompetenz und dem Service-Handlungswissen, was im Anschluss an Schmidt et al. (2014) als Effekt einer im Servicebereich relativ früh einsetzenden Routinebildung gedeutet werden kann. Die hohe Korrelation zwischen der Fehlerdiagnosekompetenz und Reparatur könnte Anlass geben, in Frage zu stellen, ob es sich

dabei tatsächlich um zwei empirisch trennbare Dimensionen handelt. Theoretisch scheint dies allerdings wenig überzeugend, da die Anforderungen in diesen Bereichen deutliche Unterschiede aufweisen.

Wünschenswert ist eine Replikation dieser Befundlage. Die bei latenter Modellierung durchgängig über 0.7 liegenden Korrelationen zwischen dem fachsystematischen Wissen und den anderen Subdimensionen evoziert die Frage, ob die aufwändigen Tests zur Fehlerdiagnosekompetenz und zum bereichsspezifischen Handlungswissen notwendig sind. Unseres Erachtens sprechen für deren Beibehaltung nicht nur die empirische Unterscheidbarkeit und die in den einzelnen Dimensionen in unterschiedlichem Grad eingelösten curricularen Ansprüche, sondern ebenso die Akzeptanz im praktischen Feld, die bei einer Reduktion auf die Erfassung des fachsystematischen Wissens nicht zu erwarten ist. Da sich auch in anderen Domänen (Industriekaufleute; Elektronikerinnen und Elektroniker) Ausdifferenzierungen auf der Anwendungsebene des Wissens dokumentieren lassen, scheinen mehrdimensionale Modellierungen in weiteren Domänen wahrscheinlich. Das hat auch Implikationen für die Prüfungspraxis, da ohne eine angemessene Berücksichtigung der Kompetenzstruktur Fehleinschätzungen der Kompetenzausprägungen riskiert werden. Das gilt v. a., wenn die Korrelationen zwischen den Subdimensionen relativ gering sind und nur ein Ausschnitt geprüft wird.

Offen ist, inwieweit im weiteren Entwicklungsprozess berufsfachlicher Kompetenzen auch Integrationsprozesse beobachtet werden können. Zu erwarten ist dies, wenn dazu von den Lehr-Lern-Arrangements oder ausgeführten Tätigkeiten entsprechende Impulse ausgehen, die eine integrative Verarbeitung unterschiedlicher Kompetenzbereiche erfordern. Bei KFZ-Mechatronikerinnen und -Mechatronikern wäre das z. B. bezogen auf die Service- und die Reparaturkompetenzen zu erwarten, die in der Praxis bezogen auf die gleichen technischen Systeme häufig sequenziell einzubringen sind (Bremsenreparatur folgt dem Bremsenservice). Die bei Kontrolle des Fachwissens und einer latenten Modellierung nahezu auf null absinkende Korrelation zwischen der Fehlerdiagnosekompetenz und dem Service-Handlungswissen lässt angesichts der unterschiedlichen kognitiven Anforderungen nicht erwarten, dass im weiteren Entwicklungsprozess eine Verschmelzung der beiden Subdimensionen beobachtet werden kann. Von Interesse wäre auch, inwieweit sich der Einfluss des fachsystematischen Wissens auf die drei abhängigen Subdimensionen über die Zeit abschwächt. Nach Schmidt et al. (2014), setzen bei KFZ-Mechatronikerinnen und -Mechatronikern bereits in der Ausbildung z. B. im Servicebereich Routinebildungsprozesse ein, die im Anschluss an die Theorie des Fertigkeitserwerbs von Ackerman (1992) eine absinkende Erklärungskraft des fachsystematischen Wissens erwarten lassen. Es ist aber auch nicht auszuschließen, dass Facharbeiter kaum noch Servicetätigkeiten ausüben und solche Aufgaben eher an Auszubildende delegiert werden.

Insgesamt dokumentieren die Ergebnisse die hohe Bedeutung des fachsystematischen Wissens für die berufsfachlichen Leistungen. Das ist auch didaktisch bedeutsam und könnte Anlass geben, dem Fachwissen in der beruflichen Didaktik wieder mehr Beachtung zu schenken, als dies gegenwärtig der Fall ist. Dafür spricht auch die ein-

gangs skizzierte Befundlage. Weitgehend bestätigt werden die inhaltlichen Affinitäten zwischen den Subdimensionen des fachsystematischen Wissens und den Kompetenzausprägungen im Anwendungsbereich. Lediglich der Pfad der Wissensdimension Start-Strom-Beleuchtung zum Service-Handlungswissen ist inhaltlich wenig plausibel. Inwieweit dafür die inkorporierten allgemeinen kognitiven Grundfähigkeiten ursächlich sind, bleibt offen, da diese nicht miterhoben werden konnten. Vorliegende Analysen mit anderen Datensätzen bei KFZ-Mechatronikerinnen und -Mechatronikern deuten darauf hin, dass die kognitiven Grundfähigkeiten bei Einbezug des fachsystematischen Wissens nur vermittelt über das Wissen indirekt wirksam werden (z. B. Abele, 2014; Nickolaus et al., 2012). Es liegen allerdings auch Erklärungsmodelle mit direkten Pfaden der kognitiven Grundfähigkeit auf berufsfachliche Kompetenzen vor (Petsch, Norwig & Nickolaus, 2015).

Eine Grenze der Untersuchung liegt in der kleinen Stichprobe, die den hohen Testaufwänden geschuldet ist. Größere Stichproben würden die Option eröffnen, auch die Subdimensionen des fachsystematischen Wissens latent zu modellieren. Entwicklungsbedarf indizieren auch die z. T. noch unbefriedigenden Reliabilitäten der Instrumente. Die großen Testzeiten, die auch bei einer schlanken Testgestaltung notwendig werden, legen bezogen auf das fachsystematische Wissen adaptives Testen nahe. Da die vorliegenden Items durchgängig geschlossen gestaltet sind und eine Stichprobe mit ca. 1.000 Probandinnen und Probanden vorliegt, ist dieser Weg naheliegend, die Reliabilität des Tests ist für adaptives Testen jedoch noch unzureichend. Für die Fehlerdiagnosekompetenz zeigten Abele et al. (2014), dass mit Teilaufgaben die gleiche Kompetenz erfasst wird, wie mit den komplexen Problemlöseaufgaben, so dass auch hier weitere Reliabilitätsverbesserungen möglich sind. Im Service- und Reparaturbereich ist neben Optimierungen der Reliabilitäten eine Validierung gemessen an den realen Leistungen wünschenswert, die manuelle Komponenten enthalten. Wir vermuten allerdings, dass im Servicebereich die manuellen Anforderungen am Ende der Ausbildung nicht mehr leistungskritisch werden.

Für die Prüfungspraxis bieten die neu entwickelten Instrumentarien neue Wege, die eine kostengünstige sowie reliable und valide Kompetenzabschätzung ermöglichen. Zumindest für die Fehlerdiagnosekompetenzen konnte auch in einer weiteren Domäne mit einem strukturidentischen Ansatz die zuvor bestehende Reliabilitäts- und Validitätsproblematik gelöst werden (Walker et al., 2015).

## Literatur

- Abele, S. (2014). *Modellierung und Entwicklung berufsfachlicher Kompetenz in der gewerblich-technischen Ausbildung*. Stuttgart: Steiner.
- Abele, S., Behrendt, S., Weber, W. & Nickolaus, R. (2016). Berufsfachliche Kompetenzen von Kfz Mechatronikern (KOKO Kfz) – Messverfahren, Kompetenzdimensionen und erzielte Leistungen. In K. Beck, M. Landenberger & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der Beruflichen Bildung* (S. 171–203). Bielefeld: Bertelsmann.

- Abele, S., Greiff, S., Gschwendtner, T., Wüstenberg, S., Nickolaus, R., Nitzschke, A. & Funke, J. (2012). Dynamische Problemlösekompetenz. Ein bedeutsamer Prädiktor von Problemlöseleistungen in technischen Anforderungskontexten? *Zeitschrift für Erziehungswissenschaft*, 15, 363–391.
- Abele, S., Walker, F. & Nickolaus, R. (2014). Zeitökonomische und reliable Diagnostik beruflicher Problemlösekompetenzen bei Auszubildenden zum Kfz-Mechatroniker. *Zeitschrift für Pädagogische Psychologie*, 28, 167–179.
- Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology*, 77, 598–613.
- Dörig, R. (2003). *Handlungsorientierter Unterricht – Ansätze, Kritik und Neuorientierung aus bildungstheoretischer, curricularer und instruktionspsychologischer Perspektive*. Stuttgart: WiKu Verlag.
- Gschwendtner, T. (2011). Die Ausbildung zum Kraftfahrzeugmechatroniker im Längsschnitt. Analysen zur Struktur von Fachkompetenz am Ende der Ausbildung und Erklärung von Fachkompetenzentwicklungen über die Ausbildungszeit. *Zeitschrift für Berufs- und Wirtschaftspädagogik, Beiheft 25*, 55–76.
- Gschwendtner, T., Abele, S. & Nickolaus, R. (2009). Computersimulierte Arbeitsproben: Eine Validierungsstudie am Beispiel der Fehlerdiagnoseleistungen von Kfz-Mechatronikern. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 105, 557–578.
- Hoyle, R. H. (2012). *Handbook of structural equation modeling*. New York, NY: Guilford Press.
- Klieme, E. & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52, 876–903.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.
- Mandl, H. & Gerstenmaier, J. (2000). *Die Kluft zwischen Wissen und Handeln*. Göttingen: Hogrefe.
- Moosbrugger, H. & Schermelleh-Engel, K. (2008). Exploratorische (EFA) und konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 305–324). Berlin: Springer.
- Müller, M. & Schelten, A. (2009). Comparativ international analysis of occupational tasks and qualification requirements for the labor market and assessment tasks at the end of VET in participating countries. *Vocational Training Research*, 8, 33–47.
- Nickolaus, R., Abele, S., Gschwendtner, T., Nitzschke, A. & Greiff, S. (2012). Fachspezifische Problemlösefähigkeit in gewerblich-technischen Ausbildungsberufen. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 108, 243–272.
- Nickolaus, R., Geißel, B., Abele, S. & Nitzschke, A. (2011). Fachkompetenzmodellierung und Fachkompetenzentwicklung bei Elektronikern für Energie- und Gebäudetechnik im Verlauf der Ausbildung – Ausgewählte Ergebnisse einer Längsschnittstudie. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 25. Beiheft, 77–94.
- Nickolaus, R., Maier, A., Nitzschke, A., Schnitzler, A., Velten, S. & Dietzen, A. (2015). Zur Relevanz mathematischer Kompetenzen für die Entwicklung berufsfachlicher Kompetenzen bei Auszubildenden der Mechatronik und Fachinformatik. *Unterrichtswissenschaft*, 3, 263–281.

- Nickolaus, R. & Seeber, S. (2013). Berufliche Kompetenzen: Modellierungen und diagnostische Verfahren. In A. Frey, U. Lissmann & B. Schwarz (Hrsg.), *Handbuch Berufspädagogische Diagnostik* (S. 166–195). Weinheim: Beltz.
- Petsch, C., Norwig, K. & Nickolaus, R. (2015). Berufsfachliche Kompetenzen in der Grundstufe Bautechnik. In A. Rausch, J. Warwas, J. Seifried & E. Wuttke (Hrsg.), *Konzepte und Ergebnisse ausgewählter Forschungsfelder der beruflichen Bildung* (S. 59–88). Baltmannsweiler: Schneider.
- R Core Team. (2014). *R: A language and environment for statistical computing*. Wien: R Foundation for Statistical Computing. Zugriff am 10.2.2016 unter <http://www.R-project.org/>
- Raykov, T. & Marcoulides, G. A. (2006). *A first course in structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Renkl, A. (1996). Träges Wissen. Wenn erlerntes nicht genutzt wird. *Psychologische Rundschau*, 47, 78–92.
- Rosendahl, J. & Straka, G. (2011). Kompetenzmodellierungen zur wirtschaftlichen Fachkompetenz angehender Bankkaufleute. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 107, 190–217.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. Zugriff am 10.2.2016 unter <http://www.jstatsoft.org/v48/i02/>
- Schmidt, T., Nickolaus, R. & Weber, W. (2014). Modellierung und Entwicklung des fachsystematischen und handlungsbezogenen Fachwissens von Kfz-Mechatronikern. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 110, 549–574.
- Seeber, S. (2008). Ansätze zur Modellierung beruflicher Fachkompetenz in kaufmännischen Ausbildungsberufen. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 104, 74–97.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (Hrsg.). (2000). *Handreichungen für die Erarbeitung von Rahmenlehrplänen der Kultusministerkonferenz (KMK) für den berufsbezogenen Unterricht in der Berufsschule und ihre Abstimmung mit Ausbildungsordnungen des Bundes für anerkannte Ausbildungsberufe*.
- Spöttl, G., Becker, M. & Musekamp, F. (2011). Anforderungen an KFZ-Mechatroniker und Implikationen für die Kompetenzerfassung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 25. Beiheft, 37–53.
- Sun, R. (2006). The CLARION cognitive architecture: Extending cognitive modeling to social simulation. In R. Sun (Ed.), *Cognition and multi-agent interaction* (pp. 79–102). New York, NY: Cambridge University Press.
- Walker, F., Link, N. & Nickolaus, R. (2015). Berufsfachliche Kompetenzstrukturen bei Elektronikern für Automatisierungstechnik am Ende der Berufsausbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 111, 222–241.
- Wang, J. & Wang, X. (2012). *Structural equation modeling. Applications using Mplus*. Chichester: Wiley; Higher Education Press.
- Winther, E. & Achtenhagen, F. (2009). Skalen und Stufen kaufmännischer Kompetenz. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 105, 521–556.

# Anhang B

## ***Publikation 2: Bedeutung des Umgangs mit Komplexität***

Abele, S. (2016). Umgang mit Komplexität: Eine bedeutsame psychische Voraussetzung des domänenspezifischen Problemlösens? *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 112, 37–59.

### **Eigener Anteil an der Publikation und Begutachtungsverfahren**

Auch diese Publikation resultierte aus einer Teilstudie des Projekts „KOKO Kfz“ der Förderlinie „Technologieorientierte Kompetenzmessung in der Berufsbildung (ASCOT)“ des Bundesministeriums für Bildung und Forschung (BMBF)<sup>40</sup>, an dessen Beantragung ich mitwirkte. In dieser Teilstudie hatte ich wissenschaftliche Leitungsfunktionen. Außerdem war ich für die Entwicklung der diagnostischen Probleme, das Studiendesign, die Durchführung sowie die Auswertung verantwortlich. Die Publikation habe ich konzipiert und verfasst, die einschlägigen statistischen Auswertungen stammen ebenfalls von mir. Publikationen in *Zeitschrift für Berufs- und Wirtschaftspädagogik* durchlaufen ein dreifach-blindes Begutachtungsverfahren.

---

40 An dem Teilprojekt wirkte ebenfalls ein größerer Kreis von Personen mit, die im Folgenden genannt seien und bei denen ich mich an dieser Stelle ebenfalls bedanke: Dipl.-Gwl. Horst Heinzmann, Sebastian Heilig MA, Dipl.-Gwl. Thomas Schmidt und Thomas Trezbiatowski. Zudem möchte ich den zahlreichen Experten und Schulen für ihre Unterstützung danken.

# Umgang mit Komplexität

## *Eine bedeutsame psychische Voraussetzung des domänenspezifischen Problemlösens?*

STEPHAN ABELE

### **Kurzfassung**

In der Studie wurde untersucht, ob der Umgang mit Komplexität für das domänenspezifische Problemlösen bedeutsam ist. Der Umgang mit Komplexität wurde als interindividuell variierende psychische Ressource konzeptualisiert, die eine (sehr) große inhaltliche Überlappung mit der Arbeitsgedächtniskapazität aufweist. Theoretisch wurde vermutet, dass bei domänenspezifischen Problemen unterschiedlicher Komplexität ein unterschiedliches Ausmaß an Umgang mit Komplexität erforderlich ist und diese Probleme deshalb von unterschiedlich vielen Personen gelöst werden, d. h. unterschiedlich schwierig sind. Empirisch wurde erwartet, dass die Problemkomplexität die Problemschwierigkeit beeinflusst. Untersucht wurde diese Hypothese regressionsanalytisch mit Auszubildenden der Kfz-Mechatronik ( $n = 275$ ) und 13 Kfz-Fehlerdiagnose-Problemen, deren Komplexität a priori kodiert und deren Schwierigkeit Rasch-basiert bestimmt wurde. Der erwartete Effekt war erst nach Ausschluss eines Problems festzustellen. Neben diesem Widerspruch wird im Beitrag auch diskutiert, warum die in Anlehnung an den Forschungsstand vorgenommene und mit vielen domänenspezifischen Kompetenzmodellen assoziierte Konzeptualisierung des Umgangs mit Komplexität für die fachdidaktische Forschung problematisch ist.

### **Abstract**

The aim of the study was to investigate the role of coping with complexity in domain-specific problem-solving. According to the state of the art coping with complexity was considered a mental resource, which is limited, varies among individuals and can largely be understood as working memory capacity. Theoretically, it was expected that problems of varying levels of complexity demand different “quantities” of coping with complexity. Thus, they are solved by a varying number of people, which means that their difficulty varies. Empirically, it was assumed that the complexity of problems influences the problems’ difficulty. This hypothesis was examined by means of a sample of car mechatronics ( $n = 275$ ) and 13 problems (troubleshooting scenarios). The complexity of the problems was defined a priori and their difficulties were determined applying the Rasch model. The results showed that there is an influence of the problems’ complexity on their difficulty. However, one problem must be excluded from the statistical analysis. Finally, this unexpected finding is discussed. Furthermore, the discussion addresses the conceptualisation of coping with complexity as working memory capacity, which can be problematic in didactical contexts.

Die aktuelle Flüchtlingskrise und Großprojekte wie der Flughafenbau in Berlin vergegenwärtigen auf zugespitzte Weise, dass wir im Alltag oft schwierige Probleme zu lösen haben. Mit welcher Qualität diese Probleme gelöst werden, hängt entscheidend vom Problemlöseniveau einzelner Individuen (Politiker, Ökonomen, Ingenieure, Facharbeiter, Handwerker etc.) ab. Hier kommt institutionalisierten Bildungsprozessen eine wichtige Aufgabe zu: Sie sollten Lernende dazu befähigen, Probleme des privaten und beruflichen Alltags hochwertig zu lösen. Solche Problemlösungen setzen meist domänenspezifisches Lernen voraus (z. B. Feltovich/Prietula/Ericsson 2006, S. 47 ff.), weshalb in Bildungskontexten die Förderung des *domänenspezifischen Problemlösens* zentral ist.

In der domänenspezifischen Lehr-Lernforschung, insbesondere der fachdidaktischen Forschung, wird oft unterstellt, dass Komplexität eine wichtige Ursache dafür ist, warum domänenspezifische Anforderungen schwierig sind (z. B. Schumann/Eberle 2011; Kauertz u. a. 2010). Dementsprechend wird angenommen, dass der *Umgang mit Komplexität* eine wichtige Voraussetzung domänenspezifischer Leistungen ist. Es liegen nur wenige Studien vor, die diese Annahme im Bereich des domänenspezifischen Problemlösens untersuchen. Ein generelles Defizit des Forschungsstands besteht darin, dass der Umgang mit Komplexität kaum systematisch theoretisch konzeptualisiert wird. Meist wird er als bedeutsam erachtet, was genau darunter zu verstehen ist und welche psychischen Ressourcen in Lehr-Lernprozessen zu fördern sind, damit der Umgang mit Komplexität gelingt, bleibt weitgehend unklar. Im Hinblick auf die Förderung des domänenspezifischen Problemlösens ist dies ein schwer wiegendes Defizit.

Dieses Defizit wird in der vorliegenden Studie bearbeitet: Im theoretischen Teil wird geklärt, was unter dem Umgang mit Komplexität im Kontext des domänenspezifischen Problemlösens verstanden wird und wie er theoretisch konzeptualisiert werden kann. Im empirischen Teil wird untersucht, ob der auf spezifische Weise konzeptualisierte Umgang mit Komplexität für das domänenspezifische Problemlösen bedeutsam ist.

## **1 Domänenspezifisches Problemlösen und der Umgang mit Komplexität**

Ein domänenspezifisches Problem liegt vor, wenn ein Individuum in einer spezifischen Umwelt ein Ziel verfolgt und (noch) nicht weiß, wie es dieses Ziel erreichen kann (Mayer/Wittrock 2006, S. 288). Domänenspezifisches Problemlösen lässt sich als zielorientierte, auf die Umwelt bezogene Aktivität eines Individuums verstehen. Hierbei interagiert das Individuum unter Investition seiner psychischen Ressourcen idealerweise solange mit der Umwelt, bis der erwünschte Zielzustand erreicht, d. h. das Problem gelöst ist. Im Unterschied zu domänenübergreifendem (oft auch: allgemeinem) Problemlösen erfordert domänenspezifisches Problemlösen kontextspezifische Lernerfahrungen und daraus resultierendes Vorwissen (Greiff/Kretzschmar/Leutner

2014, S. 162), dessen Wirksamkeit auf einen bestimmten Umweltausschnitt, d. h. eine Domäne<sup>41</sup> beschränkt ist.

Die psychischen Ressourcen, die dem domänenspezifischen Problemlösen zugrunde liegen, können mit dem Begriff „domänenspezifische Problemlösekompetenz“ zusammengefasst werden. Mit Zlatkin-Troitschanskaia/Seidel (2011) können Kompetenzen als komplexe Personmerkmale aufgefasst werden, die sich ausdifferenzieren lassen. Hier wird unterstellt, dass der *Umgang mit Komplexität* eine wichtige Facette der domänenspezifischen Problemlösekompetenz ist. Zudem wird angenommen, dass diese psychische Ressource interindividuell variiert und einen entscheidenden Einfluss darauf hat, ob domänenspezifische Probleme unterschiedlicher Komplexität gelöst werden können, wobei Folgendes zu betonen ist: Der Umgang mit Komplexität ist nur eine von mehreren Facetten der domänenspezifischen Problemlösekompetenz oder anders ausgedrückt: nur eine von mehreren psychischen Ressourcen, die den Problemlöseerfolg beeinflussen.

## 2 Bedeutung des Umgangs mit Komplexität

### 2.1 Untersuchungsansatz

Ob ein Individuum über domänenspezifische Problemlösekompetenzen verfügt, wird üblicherweise daran festgemacht, ob es domänenspezifische Probleme lösen kann. Als psychisches Merkmal ist die domänenspezifische Problemlösekompetenz nicht direkt beobachtbar, sie kann nur aus den beim domänenspezifischen Problemlösen gezeigten Leistungen erschlossen werden (s. dazu auch Beck 1987, S. 165 ff.). Die gezeigten Leistungen resultieren aus dem Zusammenwirken unterschiedlicher psychischer Ressourcen und hier wird angenommen, dass der Umgang mit Komplexität dabei eine wichtige Ressource ist. In einer Validitätsperspektive bedeutet dies, dass die Problemlöseleistungen zu einem gewissen Teil auch als Ausdruck der individuellen Ausprägung des (als psychische Ressource konzeptualisierten) Umgangs mit Komplexität interpretiert werden können, sofern die Leistungen auf Problemen unterschiedlicher Komplexität beruhen. Falls diese Validitätsannahme trägt, bedeutet dies, dass der Umgang mit Komplexität eine bedeutsame psychische Voraussetzung des domänenspezifischen Problemlösens ist.

Empirisch werden solche Validitätsannahmen typischerweise korrelationsanalytisch untersucht: Die domänenspezifische Problemlösekompetenz sowie der Umgang mit Komplexität werden separat erfasst, anschließend wird die Korrelation zwischen beiden Konstrukten berechnet und anhand der Korrelationshöhe beurteilt, ob es empirisch gerechtfertigt ist, den Umgang mit Komplexität als Facette der domänenspezifischen Problemlösekompetenz zu betrachten (Hartig/Frey/Jude 2008, S. 145 ff.). Dieser

---

41 Der Domänenbegriff ist vieldeutig. Hier wird damit in Anlehnung an Achtenhagen (2000) ein Realitätsbereich bezeichnet, dessen Grenzen entlang fachlicher (z. B. mathematischer) oder beruflicher (z. B. Kfz-mechatronischer) Anforderungen (wohl oft nicht eindeutig) verlaufen. Ob diese fachliche „Grenzziehung“ gerechtfertigt ist oder ob der Domänenbegriff letztlich „nur“ eingeschränkte Geltungsansprüche von Lehr-Lern-Theorien zum Ausdruck bringt, wie etwa Beck (2005) vermutet (S. 551), kann hier nicht vertieft werden.

Validierungsansatz wird von Borsboom/Mellenbergh/van Heerden (2004) aus verschiedenen Gründen scharf kritisiert. Ihr Hauptargument besteht darin, dass die ermittelte Korrelation nichts darüber aussagen muss, ob zwischen beiden Konstrukten ein Kausalverhältnis besteht (s. dazu auch Renkl 2012). Aber genau darin, so die Autoren, bestehe der Kern einer Validitätsannahme. Stattdessen schlagen die Autoren einen Validierungsansatz vor, bei dem geprüft wird, ob aus einer von außen induzierten Variation des interessierenden psychischen Merkmals theoriekonforme Wirkungen resultieren, was eine präzise Modellierung des psychischen Merkmals und seiner Funktion impliziert.

Der hier verwendete Validierungsansatz von Embretson (1998) steht im Einklang mit dem Ansatz von Borsboom/Mellenbergh/van Heerden (2004). Im Zentrum dieses Ansatzes steht eine Theorie, die, etwas vereinfacht ausgedrückt, Aussagen dazu enthält, welche psychische Ressource (Prozess, Strategie etc.) mit welchen Items erfasst werden soll, welche Itemmerkmale die Ressource „aktivieren“ und inwieweit eine systematische Variation der Itemmerkmale (und damit Manipulation der psychischen Ressource) die Itemschwierigkeit beeinflusst. Die Itemschwierigkeit wird üblicherweise mit Stichprobendaten unter Anwendung des Rasch-Modells geschätzt (z. B. Daniel/Embretson 2010; Hartig/Frey 2012). Falls der Einfluss des Itemmerkmals auf die Itemschwierigkeit empirisch feststellbar ist, kann dies als (starkes) Argument dafür gelten, Testwerte im Sinne der zu erfassenden Ressource zu interpretieren (s. dazu auch Hartig/Frey/Jude 2008, S. 154). Auf den vorliegenden Kontext übertragen bedeutet dies: Hier wird unterstellt, dass der Umgang mit Komplexität eine interindividuell variierende psychische „Größe“ darstellt und eine spezifische Ausprägung dieser „Größe“ das Lösen von Problemen spezifischer Komplexität erlaubt. Bei Annahme einer Normalverteilung dieser „Größe“ in einer Population werden, eine Rasch-Modellierbarkeit unterstellt, domänenspezifische Probleme mit zunehmender Komplexität von weniger Personen gelöst und damit schwieriger (s. dazu auch Kauertz 2008, S. 47 ff.). Falls diese Annahme gilt, d. h. ein Effekt der Problemkomplexität auf die Problemschwierigkeit zu beobachten ist, kann der Umgang mit Komplexität als wichtige psychische Voraussetzung des domänenspezifischen Problemlösens betrachtet werden.

## 2.2 Präzisierung des Komplexitätsbegriffs

Darüber, was unter einem komplexen Problem zu verstehen und wie Komplexität zu operationalisieren ist, liegen sehr unterschiedliche Vorstellungen vor (z. B. Campbell 1988; Minnameier 2005, S. 57 ff.; Quesada/Kintsch/Gomez 2005). Im Bereich der domänenübergreifenden Problemlöseforschung fungiert Komplexität einmal als Oberbegriff für einen bestimmten (besonders anspruchsvollen) Problemtyp und einmal als Eigenschaft eines spezifischen Problems (Funke 2006, S. 379). Hier ist mit Komplexität eine spezifische Eigenschaft domänenspezifischer Probleme gemeint.

In der Literatur wird Komplexität häufig eher im Sinne von Problemschwierigkeit definiert (ebd.; Klieme u. a. 2001, S. 187 f.). Komplexe Probleme sind dann meist schwierige Probleme und solche, die sich auf schwierige (anspruchsvolle) kognitive

Prozesse beziehen (z. B. Winther 2010, S. 235). Diese Definition ist im vorliegenden Kontext wenig hilfreich: Um die empirische Bedeutung des Umgangs mit Komplexität zu bestimmen, wird der Zusammenhang zwischen der Problemkomplexität und der Problemschwierigkeit untersucht. Bei einer synonymen Verwendung von Problemschwierigkeit und Problemkomplexität stehen beide Begriffe für denselben Wirklichkeitsbereich, weshalb eine Zusammenhangsanalyse keinen Sinn hat.

In der wohl gängigsten, wenngleich nicht unumstrittenen (z. B. Minnameier 2005, S. 57 ff.) Definition stellt Komplexität ein *Maß für die Anzahl an auf spezifische Weise vernetzten Elementen* dar (z. B. Dörner 2008, S. 284 ff.; Funke 2003, S. 128; Kauertz 2008, S. 38). Im Bereich des Problemlösens wird diese Komplexitätsdefinition oft auf die folgenden beiden Bereiche angewandt: den psychischen Problemlöseprozess und die objektive Problemanforderung. Damit sind zu unterscheidende Wirklichkeitsbereiche angesprochen. Der psychische Problemlöseprozess bezieht sich auf jene psychischen Aktivitäten, die zwischen der Zielsetzung und (idealiter) der Problemlösung liegen. Der mentale, also nicht direkt beobachtbare Problemlöseprozess kann aus unterschiedlich vielen Elementen (z. B. Teilschritten oder kognitiven Vorgängen: schlussfolgern, analysieren etc.) bestehen. Die objektive Problemanforderung bezieht sich dagegen auf die inhaltliche Struktur der Problemumwelt und die daraus resultierende Anzahl an Elementen, die zur Problemlösung zu verarbeiten sind (z. B. in der Problemstellung enthaltene oder bei der Problemlösung zu berücksichtigende Informationseinheiten). Einmal wird Komplexität also als Eigenschaft des Problemlöseprozesses aufgefasst und einmal als Eigenschaft der objektiven Problemanforderung. Vor diesem Hintergrund kann sich der Umgang mit Komplexität einmal darauf beziehen, wie gut ein Individuum mehr oder weniger komplexe Problemlöseprozesse beherrscht und einmal darauf, wie erfolgreich es unterschiedlich komplexe Problemanforderungen bearbeitet. Im Weiteren werden zu beiden Auffassungen empirische Befunde referiert. Studien, die eine von der vorliegenden Publikation abweichende Komplexitätsdefinition vornehmen (z. B. Winther 2010; Bernholt/Parchmann/Commons 2009; Musekamp/Schlömer/Mehrafza 2015), werden aus Gründen der Vergleichbarkeit nicht berücksichtigt.

### **2.3 Umgang mit Komplexität als Beherrschung komplexer Problemlöseprozesse**

Einige Studien konzeptualisieren den Umgang mit Komplexität als Beherrschung komplexer Problemlöseprozesse. Bei diesem Ansatz wird der Komplexitätsgrad typischerweise mit der Anzahl an Problemlöseschritten (z. B. Petsch/Norwig/Nickolaus 2015) oder der Anzahl an Teilzielen (z. B. Daniel/Embretson 2010) bestimmt. Die Schwierigkeit hierbei besteht darin, dass der psychische Problemlöseprozess nicht beobachtbar ist. Deshalb sollte, wie Embretson/Daniel (2008) überzeugend darstellen, die Prozesskomplexität auf Basis eines theoretischen Modells zum Problemlöseprozess operationalisiert werden. Bei genauer Betrachtung reicht ein solches Modell allerdings nicht aus. Zusätzlich zu einem Prozessmodell wird eine Theorie dazu benötigt, welche Problemanforderungen welche Prozesskomplexitäten induzieren: Aus den ob-

jektiven Anforderungen eines Problems kann nicht ohne weiteres auf die Komplexität des Problemlöseprozesses geschlossen werden. Die Problemanforderung und der Problemlöseprozess stellen unterscheidbare Wirklichkeitsbereiche dar, zwischen denen kein deterministisches Verhältnis besteht. Wie Befunde der Expertiseforschung zeigen, ist bei identischen Anforderungen abhängig von domänenspezifischen Lernerfahrungen mit unterschiedlichen Problemlöseprozessen zu rechnen (Chi/Glaser 1982). So wenden Personen mit viel Lernerfahrung häufig Rückwärts- und solche mit wenig Lernerfahrung häufig Vorwärtsstrategien an (Krems 1994, S. 52), wobei Rückwärtsstrategien meist (deutlich) weniger komplexe Problemlöseprozesse implizieren als Vorwärtsstrategien. Auch Kauertz (2008) belegt anhand empirischer Befunde, dass bei einer erfolgreichen Problembearbeitung mit unterschiedlichen Lösungswegen, d. h. unterschiedlich komplexen Problemlöseprozessen zu rechnen ist (S. 24ff.). Um also fundiert zu untersuchen, ob die Komplexität von Problemlöseprozessen einen Einfluss auf die Problemschwierigkeit hat, wird eine sehr anspruchsvolle Theorie benötigt. Eine solche Theorie sollte Aussagen dazu enthalten, welche Problemanforderungen unter welchen Bedingungen welche Prozesskomplexitäten induzieren. Davon sind wir in vielen Bereichen des domänenspezifischen Problemlösens weit entfernt. Auch die durchaus beachtlichen Studien, die einen Einfluss der Problemkomplexität auf die Problemschwierigkeiten nachweisen (Petsch/Norwig/Nickolaus 2015; Daniel/Embretson 2010; Embretson/Daniel 2008), ändern nichts an dieser Einschätzung: Sie gründen ihre Operationalisierung im erwähnten Sinn nicht auf ein systematisch ausgearbeitetes Theoriemodell und geben künftigen Studien allenfalls eine grobe Orientierung. Eine Konzeptualisierung des Umgangs mit Prozesskomplexität als Beherrschung komplexer Problemlöseprozesse scheidet angesichts des gegenwärtigen Forschungsstands also aus, sofern eine theoretisch gut fundierte Untersuchung angestrebt wird.

#### **2.4 Umgang mit Komplexität als Beherrschung komplexer Problemanforderungen**

Ein anderer Ansatz besteht darin, den Umgang mit Komplexität als *Beherrschung einer komplexen Problemanforderung* zu konzeptualisieren. Hier bezieht sich die Komplexität auf die *objektiven* Anforderungen eines Problems. Es wird unterstellt, dass die Anforderungskomplexität unabhängig vom Bearbeitungsprozess und anhand inhaltlicher Überlegungen bestimmt werden kann (Kauertz 2008, S. 29). Hierzu wird typischerweise von der Musterlösung her überlegt, wie viele Informationseinheiten bei einer erfolgreichen Bearbeitung zwingend berücksichtigt bzw. verarbeitet werden müssen und in welcher Relation diese Einheiten stehen (z. B. Neumann u. a. 2007, S. 106). Sehr dominant ist diese Art der Komplexitätsbestimmung im Bereich der naturwissenschaftlichen Didaktikforschung, in der allerdings nicht die Komplexität von Problemanforderungen, sondern von Aufgaben zur Kompetenzmessung untersucht wird. In sehr unterschiedlichen naturwissenschaftlichen Fächern zeigte sich, dass die Anforderungskomplexität die Aufgabenschwierigkeit substantiell beeinflusst (Physik: Kauertz 2008; Schoppmeier 2013; Chemie: Walpuski/Ropohl 2011; Biologie: Wellnitz u. a. 2012), wobei nicht selten Varianzaufklärungen von mehr als 25 % beobachtet wurden.

Teilweise konnte dieser Einfluss allerdings nicht repliziert werden (Viering/Fischer/Neumann 2010). Replikationsschwierigkeit traten auch insofern auf, als sich die Aufgabenschwierigkeiten nicht erwartungskonform auf die verschiedenen Komplexitätsstufen verteilten (Leutner u. a. 2008, S. 177; Schoppmeier 2013, S. 67), weshalb die Anzahl an Komplexitätsstufen, d. h. der Differenzierungsgrad des Komplexitätsmaßes mehrfach modifiziert wurde.

Petsch/Norwig/Nickolaus (2015) greifen den erwähnten Ansatz auf und wenden ihn auf das domänenspezifische Problemlösen von Auszubildenden im Baubereich (Zimmerer, Maurer und Stuckateure) an. Sie stellten fest, dass die Anforderungskomplexität auch bei regressionsanalytischer Kontrolle weiterer Problemmerkmale einen erheblichen Einfluss auf die Problemschwierigkeit hat ( $\beta = .48$ ). Von zentraler Bedeutung ist hier die Studie von Nickolaus u. a. (2012) zum domänenspezifischen Problemlösen bei Auszubildenden der Kfz-Mechatronik, weil sie den Ausgangspunkt dieser Untersuchung darstellt. Dort korrelierten die Komplexität der Problemanforderung und die Problemschwierigkeit manifest mit  $r = .90$ . Beide Studien zum domänenspezifischen Problemlösen basieren auf post-hoc Analysen, d. h. die Komplexität wurde nicht im Vorfeld der empirischen Untersuchung bestimmt. Zudem enthalten sie keine theoretische Modellierung der psychischen Ursachen des Einflusses von Komplexität auf Schwierigkeit.

## 2.5 Konzeptualisierung des Umgangs mit Komplexität als Arbeitsgedächtniskapazität

Die zitierten Befunde legen nahe, den Umgang mit Komplexität als bedeutsame psychische Ressource des domänenspezifischen Problemlösens bzw. zur Bewältigung komplexer Problemanforderungen aufzufassen, wobei in den meisten Studien nicht thematisiert wird, welche psychische Ressource damit gemeint ist. Angesichts der gewählten Operationalisierung der Komplexität (Anzahl an zu verarbeitenden Informationseinheiten) drängt es sich allerdings auf, dass der Arbeitsgedächtniskapazität (zumindest implizit) eine wichtige Bedeutung beigemessen wird. Diese Feststellung lässt sich mit der *cognitive load theory* fundieren (s. dazu auch Kauertz 2008, S. 45), die zwar zur Gestaltung von Lehr-Lern-Prozessen entwickelt wurde (Pass/Renkl/Sweller 2003), sich aber auch auf das domänenspezifische Problemlösen anwenden lässt (s. dazu auch Bley/Wiethe-Körprich/Weber 2015).

Im Licht der bisherigen Ausführungen und der *cognitive load theory* erscheint der Umgang mit Komplexität als psychische Ressource, die der Informationsverarbeitung dient und eine sehr große inhaltliche Überlappung mit der Arbeitsgedächtniskapazität aufweist (Sweller 2011). Je mehr Informationen bei der Problemlösung simultan verarbeitet werden müssen, desto größer ist die Belastung des Arbeitsgedächtnisses, d. h. die *intrinsic load* der Problemanforderung. Die *intrinsic load* entspricht dem Teil der kognitiven Gesamtbelastung beim Problemlösen (*cognitive load*), der durch die Problemanforderung verursacht wird, und lässt sich anhand der Anforderungskomplexität bestimmen, also der Anzahl an Informationseinheiten, die zur Bewältigung der Problemanforderung simultan zu verarbeiten sind (ebd., S. 58). Die Kapazität des Ar-

beitsgedächtnisses ist prinzipiell begrenzt, variiert interindividuell (z. B. Stern/Grabner/Schumacher 2007, S. 75 ff.) und dürfte bei der gewählten Operationalisierung die (Haupt)Ursache für den Einfluss der Problemkomplexität auf die Problemschwierigkeit sein.

Zusammenfassend ist festzuhalten, dass der Umgang mit Komplexität hier in Anlehnung an bisherige Studien als Arbeitsgedächtniskapazität konzeptualisiert wird und die empirischen Befunde insgesamt dafür sprechen, dass ein so konzeptualisierter Umgang mit Komplexität für das domänenspezifische Problemlösen bedeutsam ist, wobei der Forschungsstand folgende Defizite aufweist: Erstens erwies es sich bislang teilweise als schwierig, Befunde zum Einfluss der Komplexität auf die Schwierigkeit zu replizieren. Zweitens untersuchen die wenigen verfügbaren Studien den Einfluss post-hoc, was auch Leutner u. a. (2008, S. 179) monieren. Drittens liegen zum domänenspezifischen Problemlösen keine Studien vor, die den Effekt von Anforderungskomplexität auf Schwierigkeit theoriebasiert, d. h. mit einer klaren Vorstellung zu den psychischen Ursachen dieses Effekts untersuchen.

### 3 Forschungsfrage und Hypothese

Vor diesem Hintergrund wurde im vorliegenden Beitrag untersucht, ob der Umgang mit Komplexität eine bedeutsame psychische Voraussetzung des domänenspezifischen Problemlösens ist. Theoretisch wurde unterstellt, dass der als Arbeitsgedächtniskapazität konzeptualisierte Umgang mit Komplexität für das domänenspezifische Problemlösen bedeutsam ist, und zwar deshalb, weil er eine prinzipiell begrenzte und auch interindividuell variierende psychische Ressource darstellt, die bestimmt, wie erfolgreich unterschiedlich komplexe Problemanforderungen gelöst werden können bzw. wie schwierig domänenspezifische Probleme sind. Empirisch wurde im Anschluss an die von Nickolaus u. a. (2012) ermittelte Korrelation von  $r = .90$  ein sehr großer Effekt der Komplexität auf die Problemschwierigkeit erwartet (H1). Geprüft wurde ferner, ob sich der Befund von ebd. (2012) replizieren lässt, d. h. ob sich die hier ermittelte Korrelation signifikant von  $r = .90$  unterschied (H2). Beide Hypothesen wurden im Bereich des Problemlösens in der Kfz-Domäne untersucht.

## 4 Methode<sup>42</sup>

### 4.1 Stichprobe

Bei der Hypothesenprüfung wurde auf eine Stichprobe von 274 überwiegend männlichen (98,1%) Auszubildenden der Kfz-Mechatronik des dritten Ausbildungsjahrs zurückgegriffen, die im Durchschnitt 19,8 Jahre ( $SD = 2,4$ ) alt waren und meist einen Hauptschul- (50,5%) bzw. Realschulabschluss (44,8%) und in seltenen Fällen eine

---

42 Die folgenden Analysen wurden durch die Forschungsinitiative ASCOT ([www.ascot-vet.net](http://www.ascot-vet.net)) des BMBF ermöglicht.

Hochschulzugangsberechtigung (4,7%) besaßen. 80,3% der Stichprobe hatten die deutsche, 8,9% die türkische und 4,2% die italienische Staatsbürgerschaft. Die Daten wurden in 15 Klassen und an 13 Berufsschulen, 11 davon lagen in Baden-Württemberg, erhoben. Alle Auszubildenden absolvierten den berufspraktischen Teil ihrer dualen Ausbildung in Handwerksbetrieben.

## 4.2 Diagnostische Maße

### 4.2.1 Umgang mit Komplexität als Teil der domänenspezifischen Problemlösekompetenz

Im Einklang mit den bisherigen Ausführungen wurde in der Untersuchung die domänenspezifische Problemlösekompetenz erfasst und unterstellt, dass damit implizit zu einem beachtlichen Teil auch der Umgang mit Komplexität diagnostiziert wurde. Zur Erfassung der Problemlösekompetenz wurden den Probanden Probleme im Bereich der elektrotechnischen Kfz-Fehlerdiagnose vorgelegt. Die Kfz-Fehlerdiagnose stellt einen wichtigen Tätigkeitsbereich von Kfz-Mechatronikern dar (Rauner u. a. 2002). Bei den vorgelegten Problemen bestand das Ziel darin, die Ursache für einen technischen Kfz-Defekt zu finden. Um die Defektursache zu identifizieren, mussten ausgehend von einer Problembeschreibung (z. B. das Fahrzeug lässt sich nicht starten) typischerweise defektrelevante Fahrzeugkomponenten lokalisiert (z. B. ein spezifisches Kabel) und im Rückgriff auf Stromlaufpläne mehrere elektrotechnische Messungen an spezifischen Messstellen zur Funktionsprüfung (z. B. eine Durchgangsprüfung des Kabels per Widerstandsmessung) systematisch vorgenommen werden.

Dargeboten wurden die Probleme in einer Kfz-Computersimulation, in der Ausschnitte der realen Berufsumwelt von Kfz-Mechatronikern authentisch abgebildet sind und mit der berufsalltagsgetreu interagiert werden kann (ausführlichere Informationen zur Simulation finden sich bei Gschwendtner/Abele/Nickolaus 2009). Die Auszubildenden wurden aufgefordert, ihre Diagnoseschritte handschriftlich zu notieren und die Defektursache eindeutig zu benennen.

In einer Validierungsstudie zeigte sich, dass die Kompetenzmessung mit der Computersimulation zu (nahezu) identischen Ergebnissen führt wie eine Kompetenzmessung in der Realität, d. h. in einer Kfz-Werkstatt und an einem Fahrzeug. Die Kompetenzmessung mit der Computersimulation erwies sich also als ökologisch valide und zudem als ausreichend reliabel (SEM-Reliabilität als Maß der internen Konsistenz: .75; Abele/Walker/Nickolaus 2014).

Insgesamt bearbeiteten die Probanden 13 berufsauthentische Probleme, wovon 6 sehr zeitintensiv und 7 weniger zeitintensiv waren. In der Studie von ebd. (2014) stellten sich beide Problemtypen als konvergent valide und eindimensional skalierbar heraus: Die 13 Problemlösemaße können als Indikatoren eines latenten Merkmals, d. h. der Problemlösekompetenz (und damit zum Teil auch des Umgangs mit Komplexität) verstanden werden, weshalb im Weiteren nicht mehr zwischen ihnen unterschieden wird.

Das Scoring der acht dichotom und der fünf dreistufig kodierten Probleme erfolgte auf Basis der handschriftlichen Dokumentationen der Probanden und eines Kodierleitfadens, der von drei Experten der Kfz-Fehlerdiagnose ausgearbeitet wurde (zwei

Ausbildungs-/Werkstattmeister, ein wissenschaftlicher Mitarbeiter). Den höchsten Score erhielten Probanden, die die Defektursache eindeutig nannten; Teilscores wurden vergeben, wenn bestimmte diagnoserelevante und eindeutig identifizierbare Teilergebnisse dokumentiert wurden. Das Scoring wurde von zwei unabhängigen Kodieren vorgenommen. In sehr seltenen Fällen abweichender Kodierungen wurde unter Anwendung fachlicher Kriterien gemeinsam entschieden, wie zu kodieren ist.

#### 4.2.2 Komplexität der Problemanforderung

Bei Wahrung der ökologischen Validität ist es unmöglich, Kfz-Fehlerdiagnose-Probleme beliebiger Komplexität zu konstruieren. Vielmehr müssen berufsauthentische Probleme ausgewählt und anschließend deren Komplexität bestimmt werden. Die Anforderungskomplexität der 13 hier ausgewählten Probleme wurde im Anschluss an Nickolaus u. a. (2012) a priori bestimmt. Für das Komplexitätsrating wurde für jedes Problem die Anzahl an Informationseinheiten ermittelt, die zur eindeutigen Identifikation der Defektursache, d. h. der Problemlösung zwingend zu berücksichtigen und zu verarbeiten sind. Als zwingend nötig wurden bspw. Informationen zur Einbaulage einer Komponente bzw. zu spezifischen Messstellen oder diagnoserelevante Messergebnisse erachtet. Die Komplexität eines Problems nahm also z. B. mit zunehmender Anzahl relevanter Messergebnisse zu. Hier wurde die Kodierung nur von einer Person vorgenommen. Die im Ergebnisteil berichtete Effekthöhe der Komplexität legt nahe, dass eine (zumindest akzeptable) objektive, reliable und valide Kodierung gelang (Rost 2004, S. 389 ff.).

#### 4.3 Untersuchungsdesign und Durchführung

Die Komplexität der Problemanforderung wurde, wie bereits erwähnt, im Vorfeld der Datenerhebung ermittelt, die domänenspezifische Problemlösekompetenz in Computerräumen an Berufsschulen. Bei der Erfassung der Problemlösekompetenz bzw. des Umgangs mit Komplexität wurden die sechs zeitintensiven Probleme aus Zeitgründen im *youden square design* (Frey/Hartig/Rupp 2009) dargeboten, wobei jeder Proband fünf Probleme bearbeitete. Die sieben weniger zeitintensiven Probleme wurden von allen Probanden bearbeitet und mit einem *latin square design* (ebd.) dargeboten. Beide Designs erlauben die Kontrolle von Positions-, d. h. auch Ermüdungseffekten, was angesichts der Testzeit von 135 min nötig war.

Aus Standardisierungsgründen erfolgte die Einführung in die Computersimulation mit einem 10-minütigen Video, das jeder Proband mit Kopfhörer an seinem Monitor anschaute. Anschließend bearbeiteten die Probanden ein Übungsblatt, das sich auf zentrale Funktionen der Simulation bezog (z. B. auf die Aktivierung von Systemkomponenten) und von einem der beiden Testleiter ausgewertet wurde. In seltenen Fällen falsch bearbeiteter Aufgaben wurden die betreffenden Simulationsfunktionen nochmals im Klassenverbund anhand der an die Wand projizierten Simulation erläutert. Insofern kann unterstellt werden, dass letztlich keine entscheidenden nutzungsbedingten Leistungsdifferenzen mehr vorlagen. Die Einführung dauerte ca. 40 min, woraus eine Gesamterhebungszeit von ca. 175 min resultierte.

#### 4.4 Hypothesenprüfung

Für die Hypothesenprüfung wurde in einem ersten Schritt die Problemschwierigkeit unter Anwendung des Partial-Credit-Modells (eindimensionales polytomes Rasch-Modell) und der Maximum-Likelihood-Methode geschätzt (Bond/Fox 2007, S. 123 ff.), wofür auf die Statistiksoftware ConQuest (Wu u. a. 2007) zurückgegriffen wurde. Aus dieser Schwierigkeitsschätzung resultierten Logitwerte, bei denen ein metrisches Skalenniveau unterstellt werden kann (Greiff 2012, S. 140). Im Vorfeld der Schätzung wurde überprüft, ob die anhand der Probleme (Items) erfassten Daten mit dem gewählten Modell vereinbar sind. Hierzu wurden Item-Fit-Maße, und zwar sowohl die ungewichteten (Outfit) als auch die gewichteten (Infit) Mean-Square-Werte (*MNSQ*) inspiziert (Bond/Fox 2007, S. 137 ff.). Als Grenzwerte galten im Anschluss an Wilson (2005, S. 129) ein *MNSQ*  $< 0,75$  und  $> 1,33$  und die zugehörigen *t*-Werte, die ab einem Wert  $> |2|$  eine signifikante Abweichung zum verwendeten Modell anzeigen. Da bei größeren Stichproben häufig mit signifikanten Abweichungen zu rechnen ist, wurde, wie von Wilson vorgeschlagen, ein Item nur ausgeschlossen, wenn sowohl der *t*-Wert als auch die *MNSQ*-Werte auffällig waren. Zu beachten ist ferner, dass Werte von  $t < -2$  einen Overfit, also eine „zu gute“ Modelpassung anzeigen (Rost 2004, S. 374) und als unproblematisch erachtet werden können. Bei Abele/Walker/Nickolaus (2014) erwiesen sich die Probleme als eindimensional skalierbar und es waren zwischen den Problemen keine Residualkorrelationen zu beobachten, womit die Daten zwei wichtige Modell-Annahmen zu erfüllen scheinen (Rost 2004, S. 345): Eindimensionalität und lokale stochastische Unabhängigkeit. Die Trennschärfen der Probleme stellten sich in der genannten Studie ebenso als mindestens akzeptabel heraus und werden hier nicht weiter thematisiert.

In einem nächsten Schritt wurde mit der Statistiksoftware SPSS 22 (IBM 2013) mithilfe einer linearen Regressionsanalyse untersucht, welchen Effekt die Problemkomplexität auf die Problemschwierigkeit hat (H1), wobei bei beiden Datenreihen ein metrisches Skalenniveau unterstellt wurde. In dieser Regressionsanalyse fungierten die Probleme bzw. deren „Stufen“ als Fälle. Für jedes der acht dichotomen Items wurde jeweils ein Schwierigkeitsparameter geschätzt, für die fünf dreistufigen Items jeweils zwei, woraus eine Fallzahl von 18 resultierte. Bei einer Fallzahl von 18 und einer Teststärke von 80 % ( $\alpha = ,05$  und  $\beta = ,20$ , Eid/Gollwitzer/Schmitt 2010, S. 241) sollte das standardisierte Regressionsgewicht  $\beta \geq ,52$  bzw. der Effekt der Komplexität auf die Schwierigkeit  $f^2 \geq ,37$  ( $R^2_{adj} \geq ,27$ ) betragen, was mit G\*Power 3.1.7 (Faul u. a. 2007) berechnet wurde und angesichts der von Nickolaus u. a. (2012) ermittelten Korrelation von  $r = ,90$  realistisch schien. Bei einer einfachen linearen Regression entspricht  $\beta$  der Korrelation zwischen dem Prädiktor und dem Kriterium. Ob, wie mit H2 angenommen, die erwartete Korrelationshöhe repliziert werden kann, wurde mithilfe des von Eid/Gollwitzer/Schmitt (2010, S. 548) dargestellten Verfahrens durch einen Vergleich der von Nickolaus u. a. (2012) und hier ermittelten Korrelation inferenzstatistisch ( $\alpha = ,05$ ) abgesichert.

## 5 Ergebnisse

Im Ergebnisteil werden zunächst die Kennwerte zur Modellpassung der Probleme sowie deren Schwierigkeit und Komplexität referiert. Im Anschluss daran werden die Ergebnisse der Prüfung von H1 und H2 besprochen.

### 5.1 Modellpassung der Probleme sowie deren Schwierigkeit und Komplexität

Tabelle 1 zeigt die *MNSQ*- und *t*-Werte der Probleme. Bei P1.2 überschritt der *t*-Wert des Outfit den Grenzwert von  $|2|$ . Da der zugehörige *MNSQ*-Wert nicht auffällig war, wurde dieses Problem *nicht* ausgeschlossen. Bei P11 nahm sowohl der *t*-Wert des Outfit als auch des Infit einen negativen Wert an, der jenseits der Grenze lag, aber mit einem akzeptablen *MNSQ*-Wert verbunden und auch insofern unproblematisch war, als er einen Overfit anzeigte (Rost 2004, S. 374). Die restlichen Werte waren unauffällig. Bei den weiteren Analysen wurden alle Probleme berücksichtigt.

**Tabelle 1:** Kennwerte zur Modellpassung der Probleme (Items), deren Schwierigkeit sowie Komplexität

Probleme	Infit		Outfit		Schwierigkeit	Komplexität
	<i>MNSQ</i>	<i>t</i>	<i>MNSQ</i>	<i>t</i>		
P1.1	1,01	0,30	1,05	0,60	0,32	1
P1.2	1,02	0,20	1,25	2,50	1,43	3
P2	0,96	-0,50	0,89	-1,20	1,33	6
P3.1	1,01	0,10	1,00	0,00	0,30	7
P3.2	1,01	0,10	0,85	-1,60	2,03	10
P4.1	0,97	-0,30	0,96	-0,40	0,60	3
P4.2	1,00	0,00	0,98	-0,10	1,00	7
P5	0,98	-0,20	0,95	-0,50	-1,54	15
P6.1	1,09	1,50	1,13	1,30	0,52	6
P6.2	0,97	-0,10	0,97	-0,30	1,75	12
P7	1,07	0,60	1,11	1,30	-1,85	1
P8	0,96	-0,70	0,94	-0,70	0,29	1
P9	1,02	0,40	1,00	0,00	0,62	3
P10	0,98	-0,50	1,00	0,10	0,31	3
P11	0,87	-2,40	0,80	-2,40	0,71	4
P12	0,94	-1,20	0,91	-1,00	-0,41	3
P13.1	1,00	0,10	1,00	0,10	-0,92	1
P13.2	1,07	1,00	1,07	0,80	0,48	3

*Anmerkungen.* P: dargebotenes Problem, dichotom (z. B. P2) oder dreistufig (z. B. P1.1 und P1.2) kodiert; *MNSQ*: weighted (Infit) bzw. unweighted (Outfit) Mean-Square-Wert; *t*: standardisierter *t*-Wert des *MNSQ*

In der Tabelle sind zudem die Schwierigkeiten der Probleme als Logitwerte sowie die Ergebnisse der Komplexitätskodierung dargestellt. Hier fiel auf, dass P5 das höchste Komplexitätsrating, aber nur eine geringe Schwierigkeit aufwies. Die anderen Schwierigkeiten und Komplexitätsmaße waren unauffällig.

## 5.2 Effekt der Problemkomplexität auf die Problemschwierigkeit

Bei der regressionsanalytischen Berechnung des Effekts der Problemkomplexität auf die Problemschwierigkeit ergab sich ein nicht signifikantes  $\beta$  von ,19 ( $t = 0,79$ ,  $df = 16$ ;  $p = ,44$ ). Nach Ausschluss von P5, das im vorigen Abschnitt auffällig wurde, ergab sich ein  $\beta$  von ,70 ( $t = 3,83$ ,  $df = 15$ ;  $p = ,00$ ) bei  $R^2_{adj} = ,46$  bzw.  $f^2 = ,98$ , also einem sehr starken Effekt (Eid/Gollwitzer/Schmitt 2010, S. 583). Die ermittelte Korrelation ( $r = ,70$ ) unterschied sich nicht signifikant von der erwarteten Korrelation von  $r = ,90$  ( $z = 1,45$ ;  $p = ,07$ ), wobei die Teststärke mit 47% (sehr) gering ausfiel, d. h. der Test aufgrund der kleinen Fallzahl mit einem relativ großen  $\beta$ -Fehler behaftet war.

# 6 Diskussion

## 6.1 Bedeutung des Umgangs mit Komplexität

### 6.1.1 Ergebnisse der Untersuchung

In der vorliegenden Studie wurde untersucht, ob der Umgang mit Komplexität eine bedeutsame psychische Voraussetzung des domänenspezifischen Problemlösens ist. Der Umgang mit Komplexität wurde im Anschluss an den einschlägigen Forschungsstand als psychische Ressource und Arbeitsgedächtniskapazität konzeptualisiert, die interindividuell variiert und bei der Bearbeitung unterschiedlich komplexer Probleme wirksam wird. Theoretisch wurde vermutet, dass die individuelle Ausprägung des Umgangs mit Komplexität einen Einfluss darauf hat, ob Probleme unterschiedlicher Komplexität gelöst werden können bzw. unterschiedlich schwierig sind. Dementsprechend wurde empirisch erwartet, dass die Problemkomplexität die statistisch definierte Problemschwierigkeit beeinflusst. Bei Berücksichtigung aller domänenspezifischen Probleme ließ sich dieser Effekt regressionsanalytisch nicht feststellen. Erst nach Ausschluss von P5 war der erwartete Effekt in der erwarteten Höhe (H1) zu beobachten und wurde der Befund von Nikolaus u. a. (2012) repliziert (H2), wobei dieser Replikationsbefund aufgrund seiner geringen Testpower (sehr) vorsichtig zu interpretieren ist. Warum war der erwartungskonforme Effekt erst nach Ausschluss von P5 zu beobachten?

### 6.1.2 Konfundierung von Komplexitätsrating und Problemlöseprozess als Störfaktor

Wahrscheinlich hat dies damit zu tun, dass das Rating der Problemkomplexität und der Problemlöseprozess konfundiert sind. Bei genauer Betrachtung fällt auf, dass das hier bestimmte Komplexitätsmaß (Anzahl an vernetzten Informationseinheiten) davon abhängt, wie bei der Problemlösung vorgegangen wird: Beim Komplexitätsrating

wurde durchgängig (implizit) unterstellt, dass die Problemlösung schrittweise von der Komponente (z. B. Leuchtmittel) zum „Versorger“ (z. B. Fahrzeugbatterie) erfolgt. Dementsprechend sollte die Anzahl an zu verarbeitenden Informationen zunehmen, je „näher“ das Problem am „Versorger“ lokalisiert ist. Bei P5 (einem Lichtdefekt infolge einer defekten Sicherung) gingen viele Auszubildende allerdings wohl anders vor: Sie haben relativ früh die Funktion der Sicherung geprüft und damit bei der Problemlösung deutlich weniger Informationseinheiten verarbeitet als angenommen. Ein solches Vorgehen wird von erfahrenen Ausbildern häufig empfohlen.<sup>43</sup> Das Komplexitätsrating von P5 ist also vermutlich nicht valide und es ist gut vorstellbar, dass dieses Validitätsproblem auch in anderen Kontexten auftritt.

Es lassen sich mindestens zwei Gründe nennen, warum die Konfundierung von Komplexitätsrating und Problemlöseprozess in den bisherigen Studien keine Schwierigkeiten bereitete: (1) Es ist denkbar, dass aufgrund spezifischer Charakteristika domänenspezifischer Probleme nur *ein* Lösungsweg möglich ist und damit alle Probanden zur Problemlösung zwingend eine bestimmte Anzahl an Informationen verarbeiten mussten. (2) Es ist vorstellbar, dass alle Probanden aus irgendeinem Grund (z. B. aufgrund in der Ausbildung erworbener Strategien) denselben Lösungsprozess durchlaufen. Anzunehmen ist, dass die Homogenität von Lösungsprozessen auch davon abhängt, wie intensiv bei der Problemlösung auf eigenständige Lernerfahrungen zurückgegriffen werden kann. Nicht selten führen individuelle Lernprozesse z. B. dazu, dass effizientere, d. h. „kürzere“ Wege zum Ziel entdeckt werden (Feltovich/Prietula/Ericsson 2006, S. 56, et passim; Krems 1994, S. 79 ff.), wobei sich ab einem bestimmten Punkt die Frage stellt, ob ein Problem gelöst oder eine Aufgabe bearbeitet wird (Funke 2003, S. 25).

Aus dem erwartungswidrigen Befund ist also Folgendes zu lernen: Ein valides Rating der Komplexität der Problemanforderung, d. h. der Anzahl an bei der Problemlösung zu verarbeitenden Informationseinheiten kann ein empirisch belastbares Modell zum Problemlöseprozess voraussetzen, worüber wir aktuell in den meisten Domänen nicht verfügen. Liegt ein solches Modell nicht vor, ist zu befürchten, dass die Untersuchung des Effekts von Komplexität auf Problemschwierigkeit durch einen Störfaktor verzerrt wird: die Konfundierung von Komplexitätsrating und Problemlöseprozess. Wie die vorliegende Studie und P5 zeigen, kann dieser Störfaktor die Ergebnisse ebenso erheblich verzerren, wie der im nächsten Abschnitt betrachtete Aspekt.

### 6.1.3 Moderatorwirkung des domänenspezifischen Lernens

Die Aufarbeitung des Forschungsstands resultierte unter Anwendung der *cognitive load theory* in einem als Arbeitsgedächtniskapazität konzeptualisierten Umgang mit Komplexität. Es wurde unterstellt, dass die Problemschwierigkeit von der *intrinsic*

---

43 Aufgrund meiner eigenen Erfahrungen im Bereich der Kfz-Fehlerdiagnose und Gesprächen mit Kfz-Ausbildern war mir bewusst, dass bei P5 mit dem beschriebenen Effekt zu rechnen ist: Eine Sicherungsprüfung ist häufig ein effizientes Mittel der elektrotechnischen Fehlerdiagnose. Da es in der vorliegenden Studie aber auch um die Replikation der Befunde von Nickolaus et al. (2012) ging, wurde durchgängig das dort vorgenommene Komplexitätsrating umgesetzt. Zu bedenken ist, dass die Annahme, bei elektrotechnischen Fehlerdiagnosen wird häufig frühzeitig die Sicherungsfunktion geprüft, zwar plausibel ist, aus empirischer Sicht aber trotzdem falsch sein kann.

*load*, also von der kognitiven Belastung abhängt, die Probleme unterschiedlicher Anforderungskomplexität erzeugen. Allerdings hängt die *intrinsic load* nicht nur von der Anforderungskomplexität, sondern auch von domänenspezifischem Lernen ab. Domänenspezifisches Lernen kann die *intrinsic load* mindestens in zweifacher Hinsicht reduzieren: (1) Die *intrinsic load* nimmt durch den Erwerb und die Anwendung von Wissensschemata ab. In solchen im Langzeitgedächtnis gespeicherten Schemata liegen Informationseinheiten in organisierter Form vor (Eysenck 2006, S. 275). Ein aus dem Langzeitgedächtnis abgerufenes Schema entspricht im Arbeitsgedächtnis nur *einer* Informationseinheit, weshalb die Aktivierung von Schemata die kognitive Belastung beim Problemlösen deutlich minimieren kann (Sweller 2011, S. 59). (2) Die *intrinsic load* wird durch den Erwerb und die Anwendung von Problemlösestrategien reduziert. Solche Strategien ermöglichen eine sukzessive und damit kognitiv entlastende Informationsverarbeitung. Beide Punkte sprechen dafür, dass der Einfluss der Problemkomplexität auf die Problemschwierigkeit mit zunehmender domänenspezifischer Lernerfahrung abnehmen kann, weil die von der Problemkomplexität verursachte *intrinsic load* zusehends geringer wird. Daraus folgt, dass ein als Arbeitsgedächtniskapazität konzeptualisierter Umgang mit Komplexität für domänenspezifisches Problemlösen v. a. bei eher unbekanntenen Anforderungen bedeutsam ist, bei welchen domänenspezifisches Lernen (noch) keine (allzu) große Rolle spielt.

Diese Argumentation wird durch empirische Befunde gestützt: Es zeigte sich, dass der Einfluss der Komplexität auf die Schwierigkeit davon abhängt, ob lösungsrelevante Fachinformationen in der Aufgabenstellung enthalten sind. Bei gegebenen Fachinformationen ist der Einfluss der Komplexität höher als wenn die Fachinformationen aus dem Langzeitgedächtnis abgerufen werden müssen (Walpuski/Ropohl 2011). Durch die Vorgabe von Fachinformationen verlieren das domänenspezifische Wissen und damit das domänenspezifische Lernen für Problemlöseleistungen an Bedeutung, wogegen die Arbeitsgedächtniskapazität leistungskritischer wird. Kapazitätsnachteile können dann nicht mehr durch umfangreiches Wissen kompensiert werden.

Bei Schoppmeier (2013) erklärt die Anforderungskomplexität in einer Leistungskurstichprobe im Fach Physik 8 % weniger Varianz als in einer Grundkurstichprobe (S. 65). Unter der plausiblen Annahme, dass die Schülerinnen und Schüler des Leistungskurses über ein umfangreicheres domänenspezifisches Wissen verfügen, passt auch dieser Befund gut zur hier geführten Argumentation: Es ist zu unterstellen, dass die Bedeutung des Umgangs mit Komplexität durch das domänenspezifische Lernen moderiert wird, wobei die Bedeutung mit voranschreitendem domänenspezifischem Lernen abnehmen dürfte.

#### 6.1.4 Zusammenfassendes Fazit

Welche Bedeutung kommt nun dem Umgang mit Komplexität für das domänenspezifische Problemlösen zu? Es ist anzunehmen, dass der Umgang mit Komplexität eine wichtige Bedeutung für das domänenspezifische Problemlösen hat. Allerdings lässt sich dies empirisch nur nachweisen, wenn ein valides Rating der Komplexität der Problemanforderung gelingt, was aufgrund der Konfundierung von Komplexitätsrating

und Problemlöseprozess eine (sehr) anspruchsvolle Aufgabe darstellt. Zu beachten ist ferner, dass das domänenspezifische Lernen die Bedeutung des Umgangs mit Komplexität moderiert. Möglicherweise sind die erwähnte Konfundierung und der Modatoreffekt Gründe dafür, weshalb Befunde zum Einfluss der Komplexität auf die Schwierigkeit teilweise nicht repliziert werden konnten.

## 6.2 Implikationen für die Modellierung und Förderung domänenspezifischer Kompetenzen

### 6.2.1 Modellierung domänenspezifischer (Problemlöse-)Kompetenzen: Was bedeutet der Effekt der Problemkomplexität auf die Problemschwierigkeit?

Es wurde herausgearbeitet, dass der Einfluss der Komplexität auf die Schwierigkeit (wohl) v. a. dann feststellbar ist, wenn domänenspezifisches Lernen nicht (allzu) bedeutsam ist. Von daher ist fraglich, ob die im Bereich der naturwissenschaftlichen Fachdidaktik vorgelegten Kompetenzmodelle (z. B. Kauertz u. a. 2010; Wellnitz u. a. 2012), in denen die Anforderungskomplexität eine wichtige Anforderungsdimension darstellt, nicht ihr eigentliches Ziel – zumindest teilweise – verfehlen. Mit diesen Modellen wird (meist implizit) behauptet, dass der Umgang mit Komplexität eine wichtige Facette der domänenspezifischen Kompetenz ist. Dementsprechend werden bei der Konstruktion von Kompetenztests gezielt Aufgaben unterschiedlicher Anforderungskomplexität entwickelt und die erzielte Aufgabenqualität u. a. danach beurteilt, wie präzise die Aufgabenschwierigkeit anhand der Anforderungskomplexität prognostiziert werden kann. Im Sinne der bisherigen Argumentation ist diese Prognosekraft ein Indiz für die Bedeutung des Umgangs mit Komplexität, also v. a. der Arbeitsgedächtniskapazität. Die Arbeitsgedächtniskapazität steht in sehr engem Verhältnis zur fluiden Intelligenz (Stern/Grabner/Schumacher 2007, S. 75 f.), ist zu einem erheblichen Teil vererbt bzw. frühkindlich geprägt (Cattell 1987) und nicht durch domänenspezifisches Lernen veränderbar (Fries/Souvignier 2009, S. 410). Die erwähnten Kompetenzmodelle „provizieren“ die Erfassung der Arbeitsgedächtniskapazität und gefährden ihr eigentliches Ziel: die Bilanzierung domänenspezifischer Lernprozesse.

Walpuski/Ropohl (2011) interpretieren den Effekt von Komplexität auf Schwierigkeit anders. Sie argumentieren, ein hoher Effekt der Komplexität auf die Schwierigkeit belege, dass Kompetenzen im eigentlichen Sinn erfasst werden. Der hohe Effekt, der v. a. bei Vorgabe von Fachinformationen mit der Aufgabenstellung zu beobachten ist, zeige an, wie erfolgreich mit Fachinformationen umgegangen werden kann (Walpuski/Ropohl 2011, S. 85). Eine tiefergehende theoretische Begründung dieser Interpretation fehlt. Mit dieser Untersuchung schlage ich auf Basis der *cognitive load theory* eine andere Interpretation vor: Der hohe Effekt zeigt – zumindest tendenziell – an, dass die Arbeitsgedächtniskapazität eine wichtige und das domänenspezifische Lernen eine weniger bedeutsame Rolle spielt. In künftigen Studien zum domänenspezifischen Problemlösen sollte überlegt werden, wie ein nachgewiesener Effekt der Komplexität auf die Schwierigkeit zu interpretieren ist bzw. welche psychischen Ressourcen diesen Effekt verursachen.

### 6.2.2 Welche Bedeutung hat der Umgang mit Komplexität für die Förderung domänenspezifischer Problemlösekompetenzen?

Ein als Arbeitsgedächtniskapazität konzeptualisierte Umgang mit Komplexität ist für das domänenspezifische Problemlösen prinzipiell bedeutsam, lässt sich durch pädagogische Interventionen aber nicht fördern (Fries/Souvignier 2009, S. 410). Dennoch spielt er bei Interventionen eine wichtige Rolle: Erstens ist bei der Förderung domänenspezifischer Problemlösekompetenzen auf individuelle Kapazitätsgrenzen zu achten und gegebenenfalls mit Komplexitätsreduktion zu reagieren. Zweitens sollten Problemlösestrategien vermittelt werden, die eine sukzessive Problemlösung ermöglichen und dadurch kognitiv entlasten, d.h. die *intrinsic load* einer Problemanforderung minimieren. Petsch/Norwig/Nickolaus (2014) entwickelten einen Interventionsansatz zur Förderung allgemeiner und domänenspezifischer Problemlösestrategien, der sich in mehreren Interventionsstudien als wirksam erwies (s. dazu auch Zinn u. a. 2015).

### 6.3 Grenzen der Untersuchung

Die Ergebnisse und Schlussfolgerungen dieser Studie basieren auf einer spezifischen Komplexitätsdefinition und einer spezifischen psychologischen Konzeptualisierung des Umgangs mit Komplexität, sind nur vor diesem Hintergrund sinnvoll zu verstehen und weisen insofern Grenzen auf. Alternative theoretische Entscheidungen sind denkbar und hätten möglicherweise zu einer anderen Studienanlage und anderen Ergebnissen geführt. Wenn der Umgang mit Komplexität hier als Arbeitsgedächtniskapazität konzeptualisiert wurde, so geschah dies in Anlehnung an den gegenwärtigen Forschungsstand und mit dem Ziel, dieses relativ unbekanntes Konstrukt in einem ersten Zugriff theoretisch besser zu fassen. Dörner (2008) weist darauf hin, dass beim Umgang mit Komplexität weitere psychische Merkmale (z. B. Emotionen wie Angst, Stress etc.) bedeutsam sind. Wahrscheinlich würde aber auch die Berücksichtigung dieser Merkmale nichts daran ändern, dass er für das domänenspezifische Problemlösen bedeutsam ist.

Es liegen unterschiedliche Modelle zum Arbeitsgedächtnis und dazu vor, was unter der Arbeitsgedächtniskapazität zu verstehen ist (Baddeley/Eysenck/Anderson 2009, S. 60 ff.). Gemeinsam ist diesen Modellen, dass sie interindividuell variierende Kapazitätsgrenzen annehmen. Die Komplexität einiger hier einbezogener Probleme (z. B. P6.2) dürfte diese Grenzen prinzipiell überschreiten, was deutlich macht, dass der Effekt der Komplexität auf die Schwierigkeit (wohl) nicht alleine mit der Arbeitsgedächtniskapazität erklärt werden kann. Hier wird angenommen, dass der Effekt hauptsächlich kapazitätsbedingt auftritt.

Kritisch kann gefragt werden, ob es zulässig ist, ein komplexes Konstrukt wie die domänenspezifische Problemlösekompetenz analytisch in Einzelfacetten zu unterteilen und die Bedeutung einer Einzelfacetten (Umgang mit Komplexität) unabhängig von den anderen Facetten zu untersuchen. Auch wenn das Ganze (domänenspezifische Problemlösekompetenz) möglicherweise mehr ist als die Summe der wie auch immer beim Problemlösen zusammenwirkenden Einzelfacetten, halte ich diese Forschungsstrategie u. a. aus einer Förderperspektive für hilfreich: Sie ermöglicht es, die Bedeu-

tung der Einzelfacetten für das domänenspezifische Problemlösen zu untersuchen und sie theoretisch präzise zu beschreiben, was wiederum zu einer effektiven Förderung der domänenspezifischen Problemlösekompetenz beitragen dürfte. Es sei betont, dass die Unterteilung der Problemlösekompetenz in Einzelfacetten nicht wie oftmals üblich dazu führte, den Umgang mit Komplexität separat zu erfassen. Vielmehr wurde er integrativ, d. h. als Teil der domänenspezifischen Problemlösekompetenz diagnostiziert. Dies steht im Einklang mit Wuttke u. a. (2015, S. 193) und entgeht von daher auch der dort aufgeführten Kritik an einer isolierten Erfassung von Facetten der Problemlösekompetenz.

Bei der Rezeption der Ergebnisse und Aussagen dieser Studie ist zu beachten, dass das Komplexitätsrating vermutlich eng mit anderen Anforderungsmerkmalen assoziiert ist: Die Anforderungskomplexität scheint nicht unabhängig von anderen Problemmerkmalen variierbar (Nickolaus u. a. 2012), weshalb der Effekt der Komplexität auf die Schwierigkeit neben der Bedeutung des Umgangs mit Komplexität auch die Bedeutung anderer psychischer Merkmale für das Problemlösen spiegelt. Petsch/Norwig/Nickolaus (2015) zeigen, dass die Komplexität auch bei regressionsanalytischer Kontrolle weiterer Merkmale bedeutsam ist, allerdings nicht im hier festgestellten Ausmaß. Bedenkenswert ist auch, ob aus es den regressionsanalytisch, anhand von Gruppenstatistiken gewonnenen Effekten auf Wirkungen auf Individualebene geschlossen werden darf (z. B. Renkl 2012). Alternativ wären hier Experimentalstudien denkbar.

Einschränkend sei zudem erwähnt, dass die Kritik an den Kompetenzmodellen aus der naturwissenschaftlichen Fachdidaktik auf dem Moderatoreffekt des domänenspezifischen Lernens beruht, der unter Anwendung der cognitive load theory herausgearbeitet, mit einigen empirischen Befunden belegt, allerdings nicht systematisch empirisch geprüft wurde. Außerdem basiert die Untersuchung auf Problemen aus dem Bereich der Kfz-Fehlerdiagnose und es ist offen, inwieweit die Erkenntnisse generalisiert werden können.

Aus fachdidaktischer Sicht erweist sich eine Konzeptualisierung des Umgangs mit Komplexität als Arbeitsgedächtniskapazität als wenig befriedigend. Denn dadurch gerät ein psychisches Merkmal in den Fokus, das zwar für domänenspezifisches Lernen (sehr) relevant ist, durch pädagogische Interventionen jedoch nicht verändert werden kann. Angesichts der prinzipiell begrenzten individuellen Arbeitsgedächtniskapazität und der oft enormen Komplexität domänenspezifischer Probleme ist davon auszugehen, dass deren Lösung zu einem erheblichen Teil durch domänenspezifisches Lernen ermöglicht wird.

#### **6.4 Ausblick auf eine vielversprechende fachdidaktische Forschungsperspektive: Umgang mit Komplexität als Resultat domänenspezifischen Lernens**

Abschließend wird in Anlehnung an Embretson/Daniel (2008) eine alternative Konzeptualisierung des Umgangs mit Komplexität vorgeschlagen, und zwar eine Konzeptualisierung als Beherrschung komplexer domänenspezifischer Problemlöseprozesse.

Dieser Ansatz wurde bereits im Theorieteil diskutiert und angesichts des (noch) defizitären Forschungsstands aus rein pragmatischen Gründen nicht weiter verfolgt.

Der Problemlöseprozess bezieht sich auf jene psychischen Aktivitäten, die dem Problemlösen zugrunde liegen und prinzipiell meist auf einen enorm großen „Aktivitätsraum“. In Anlehnung an die ACT-R Theorie (*adaptive control of thought – rational*, Anderson 2007) hängt die erfolgreiche Gestaltung domänenspezifischer Problemlöseprozesse maßgeblich davon ab, ob dieser „Raum“ zielführend „durchschritten“ werden kann, d. h. Problemlösestrategien verfügbar sind. Problemlösestrategien implizieren ein schrittweises Vorgehen (Edelmann/Wittmann 2012, S. 185), strukturieren den Problemlöseprozess und ermöglichen eine zieldienliche Informationsgewinnung, eine stimmige mentale Problemrepräsentation, die Formulierung sowie Prüfung von Hypothesen etc. Es ist sowohl theoretisch als auch empirisch gut fundiert, dass domänenspezifische Problemlösestrategien insbesondere im Rückgriff auf domänenspezifisches Wissen entwickelt werden (z. B. Anderson 2007, S. 135 ff.).

Vor diesem Hintergrund lässt sich der Umgang mit Komplexität als Anwendung erworbener Problemlösestrategien zur erfolgreichen Gestaltung komplexer Problemlöseprozesse konzeptualisiert. Hier stehen domänenspezifische Strategien und domänenspezifisches Wissen, also domänenspezifisches Lernen im Fokus, wobei auch nicht-kognitive Problemlösefacetten zu berücksichtigen sind.

Für die Umsetzung dieser Konzeptualisierung werden empirisch belastbare Modelle zum domänenspezifischen Problemlöseprozess benötigt, die Aussagen dazu enthalten, unter welchen Bedingungen (Problemanforderungen, Lernerfahrungen etc.) Prozesse welcher Komplexität zu erwarten sind. Die Entwicklung solcher Modelle ist nicht zuletzt deswegen ein wichtiges fachdidaktisches Forschungsdesiderat, weil sie ein präzises Verständnis (komplexer) domänenspezifischer Problemlöseprozesse und damit assoziierter psychischer Ressourcen impliziert, was für die Förderung domänenspezifischer Problemlösekompetenzen zentral ist.

## Literaturverzeichnis

- Abele, S./Walker, F./Nickolaus, R. (2014): Zeitökonomische und reliable Diagnostik beruflicher Problemlösekompetenzen bei Auszubildenden zum Kfz-Mechatroniker. In: Zeitschrift für Pädagogische Psychologie 28, H. 4, S. 167–179.
- Achtenhagen, F. (2000): Lebenslanges Lernen aus Sicht des Mastery Learning. In: Achtenhagen, F./Lempert, W. (Hrsg.): Lebenslanges Lernen im Beruf. Seine Grundlegung im Kindes- und Jugendalter. Opladen: Leske+Budrich, S. 123–140.
- Anderson, J. R. (2007): How can the human mind occur in the physical universe? New York: Oxford University Press.
- Baddeley, A. D./Eysenck, M. W./Anderson, M. C. (2009): Memory. Hove: Psychology Press.
- Beck, K. (1987): Die empirischen Grundlagen der Unterrichtsforschung. Eine kritische Analyse der deskriptiven Leistungsfähigkeit von Beobachtungsmethoden. Göttingen: Hogrefe.

- Beck, K. (2005): Ergebnisse und Desiderate zur Lehr-Lern-Forschung in der kaufmännischen Berufsausbildung. In: *Zeitschrift für Berufs- und Wirtschaftspädagogik*, H. 4, S. 533–556.
- Bernholt, S./Parchmann, I./Commons, M. L. (2009): Kompetenzmodellierung zwischen Forschung und Unterrichtspraxis. In: *Zeitschrift für Didaktik der Naturwissenschaften* 15, S. 219–244.
- Bley, S./Wiethe-Körprich, M./Weber, S. (2015): Formen kognitiver Belastung bei der Bewältigung technologiebasierter authentischer Testaufgaben – eine Validierungsstudie zur Abbildung von beruflicher Kompetenz. In: *Zeitschrift für Berufs- und Wirtschaftspädagogik* 111, H. 2, S. 268–294.
- Bond, T. G./Fox, C. M. (2007): *Applying the Rasch Model: Fundamental measurement in the Human Sciences*. Mahwah, NJ: Erlbaum.
- Borsboom, D./Mellenbergh, G. J./van Heerden, J. (2004): The concept of validity. In: *Psychological Review* 111, S. 1061–1071.
- Campbell, D. J. (1988): Task complexity: a review and analysis. In: *Academy of Management Review* 13, H. 1, S. 40–52.
- Cattell, R. B. (1987): *Intelligence. Its structure, growth and action*. Amsterdam: North-Holland.
- Chi, M. T. H./Glaser, R. (1982): Expertise in Problem Solving. In: Sternberg, R. J. (Hrsg.): *Advances in the psychology of human intelligence*. Hillsdale, NJ: Erlbaum, S. 7–70.
- Daniel, R./Embretson, S. E. (2010): Designing cognitive complexity in mathematical problem-solving items. In: *Applied Psychological Measurement* 34, H. 5, S. 348–364.
- Dörner, D. (2008): *Umgang mit Komplexität*. In: *Industrial Ecology*: Springer, S. 284–302.
- Edelmann, W./Wittmann, S. (2012): *Lernpsychologie*. Weinheim: Beltz.
- Eid, M./Gollwitzer, M./Schmitt, M. (2010): *Statistik und Forschungsmethoden*. Lehrbuch. Weinheim: Beltz.
- Embretson, S. E. (1998): A Cognitive Design System Approach to Generating Valid Tests: Application to Abstract Reasoning. In: *Psychological Methods* 3, H. 3, S. 380–396.
- Embretson, S. E./Daniel, R. (2008): Understandings and quantifying cognitive complexity level in mathematical problem solving items. In: *Psychology Science Quarterly* 50, H. 3, S. 328–344.
- Eysenck, M. W. (2006): *Fundamentals of Cognition*. Hove: Psychology Press.
- Faul, F./Erdfelder, E./Lang, A.-G./Buchner, A. (2007): G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. In: *Behavior Research Methods*, H. 39, S. 175–191.
- Feltovich, P. J./Prietula, M. J./Ericsson, K. A. (2006): Studies of expertise from psychological perspectives. In: Ericsson, K. A./Charness, N./Feltovich, P. J./Hoffman, R. R. (Hrsg.): *The Cambridge handbook of expertise and expert performance*. Cambridge: Cambridge University Press, S. 41–67.
- 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. In: *Educational Measurement: Issues and Practice* 28, H. 3, S. 39–53.

- Fries, S./Souvignier, E. (2009): Training. In: Wild, E./Möller, J. (Hrsg.): Pädagogische Psychologie. Heidelberg: Springer, S. 406–428.
- Funke, J. (2003): Problemlösendes Denken. Stuttgart: Kohlhammer.
- Funke, J. (2006): Komplexes Problemlösen. In: Funke, J. (Hrsg.): Denken und Problemlösen. Göttingen: Hogrefe, S. 375–445.
- Greiff, S. (2012): Individualdiagnostik komplexer Problemlösefähigkeit. Münster: Waxmann.
- Greiff, S./Kretschmar, A./Leutner, D. (2014): Problemlösen in der Pädagogischen Psychologie. In: Zeitschrift für Pädagogische Psychologie 28, H. 4, S. 161–166.
- Gschwendtner, T./Abele, S./Nickolaus, R. (2009): Computersimulierte Arbeitsproben: Eine Validierungsstudie am Beispiel der Fehlerdiagnoseleistungen von Kfz-Mechatronikern. In: Zeitschrift für Berufs- und Wirtschaftspädagogik 105, S. 557–578.
- Hartig, J./Frey, A. (2012): Konstruktvalidierung und Skalenbeschreibung in der Kompetenzdiagnostik durch die Vorhersage von Aufgabenschwierigkeiten. In: Psychologische Rundschau 63, H. 1, S. 43–49.
- Hartig, J./Frey, A./Jude, N. (2008): Validität. In: Moosbrugger, H./Kelava, A. (Hrsg.): Testtheorie und Fragebogenkonstruktion. Berlin: Springer, S. 136–192.
- IBM (2013): SPSS. Chicago.
- Kauertz, A. (2008): Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben. Berlin: Logos.
- Kauertz, A./Fischer, H. E./Mayer, J./Sumfleth, E./Walpuski, M. (2010): Standardbezogene Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe I. In: Zeitschrift für Didaktik der Naturwissenschaften 16, S. 135–153.
- Klieme, E./Funke, J./Leutner, D./Reimann, P./Wirth, J. (2001): Problemlösen als fächerübergreifende Kompetenz. Konzeption und erste Resultate aus einer Schulleistungsstudie. In: Zeitschrift für Pädagogik 47, S. 179–200.
- Krems, J. F. (1994): Wissensbasierte Urteilsbildung. Diagnostisches Problemlösen durch Experten und Expertensysteme. Bern: Hans Huber.
- Leutner, D./Fischer, H. E./Kauertz, A./Schabram, N./Fleischer, J. (2008): Instruktionspsychologische und fachdidaktische Aspekte der Qualität von Lernaufgaben und Testaufgaben im Physikunterricht. In: Thonhauser, J. (Hrsg.): Aufgaben als Katalysatoren von Lernprozessen. Eine zentrale Komponente organisierten Lehrens und Lernens aus der Sicht von Lernforschung, allgemeiner Didaktik und Fachdidaktik. Münster: Waxmann, S. 169–181.
- Mayer, R. E./Wittrock, M. C. (2006): Problem Solving. In: Alexander, P. A./Winne, P. H. (Hrsg.): Handbook of educational psychology. Mahwah, NJ: Lawrence Erlbaum, S. 287–303.
- Minnameier, G. (2005): Wissen und inferentielles Denken. Frankfurt am Main: Lang.
- Musekamp, F./Schlömer, B./Mehrafza, M. (2015): Fachliche Anforderungen an Ingenieure in der Technischen Mechanik – eine empirische Analyse von Aufgabenmerkmalen. In: Musekamp, F./Spöttl, G. (Hrsg.): Kompetenz im Studium und in der Arbeitswelt. Nationale und internationale Ansätze zur Erfassung von Ingenieurkompetenzen. Frankfurt am Main: Lang, S. 181–204.

- Neumann, K./Kauertz, A./Lau, A./Notarp, H./Fischer, H. E. (2007): Die Modellierung physikalischer Kompetenzen und ihrer Entwicklung. In: *Zeitschrift für Didaktik der Naturwissenschaften* 13, S. 101–121.
- Nickolaus, R./Abele, S./Gschwendtner, T./Nitzschke, A./Greiff, S. (2012): Fachspezifische Problemlösefähigkeit in gewerblich technischen Ausbildungsberufen – Modellierung, erreichte Niveaus und relevante Einflussfaktoren. In: *Zeitschrift für Berufs- und Wirtschaftspädagogik* 108, S. 243–272.
- Pass, F./Renkl, A./Sweller, J. (2003): Cognitive load theory and instructional design: Recent developments. In: *Educational Psychologist* 38, H. 1, S. 1–4.
- Petsch, C./Norwig, K./Nickolaus, R. (Hrsg.) (2014): Kompetenzförderung leistungsschwächerer Jugendlicher in der beruflichen Bildung – Förderansätze und ihre Effekte.
- Petsch, C./Norwig, K./Nickolaus, R. (2015): Berufsfachliche Kompetenzen in der Grundstufe Bautechnik – Strukturen, erreichte Niveaus und relevante Einflussfaktoren. In: Rausch, A./Warwas, J./Seifried, J./Wuttke, E. (Hrsg.): *Konzepte und Ergebnisse ausgewählter Forschungsfelder der beruflichen Bildung – Festschrift für Detlef Sembill*. Baltmannsweiler: Schneider Verlag Hohengehren GmbH, S. 59–88.
- Quesada, J./Kintsch, W./Gomez, E. (2005): Complex problem-solving: a field in search of a definition? In: *Theoretical Issues in Ergonomics Science* 6, H. 1, S. 5–33.
- Rauner, F./Hitz, H./Spöttl, G./Becker, M. (2002): Aufgabenanalysen für die Neuordnung der Berufe im Kfz-Sektor. <http://www.itb.uni-bremen.de/uploads/media/Kfz-Neuordnung-Bericht4.pdf>. 15.05.2014.
- Renkl, A. (2012): Modellierung von Kompetenzen oder von interindividuellen Kompetenzunterschieden. Ein unterschätzter Unterschied. In: Fleischer, J./Leutner, D./Klieme, E. (Hrsg.): *Modellierung von Kompetenzen im Bereich der Bildung: Eine psychologische Perspektive*, S. 50–53.
- Rost, J. (2004): *Lehrbuch Testtheorie – Testkonstruktion*. Bern: Huber.
- Schoppmeier, F. (2013): *Physikkompetenz in der gymnasialen Oberstufe*. Berlin: Logos.
- Schumann, S./Eberle, F. (2011): Bedeutung und Verwendung schwierigkeitsbestimmender Aufgabenmerkmale für die Erfassung ökonomischer und beruflicher Kompetenzen. In: Fasshauer, J./Fürstenau, B./Wuttke, E. (Hrsg.): *Grundlagenforschung zum Dualen System und Kompetenzentwicklung in der Lehrerbildung*. Opladen: Barbara Budrich, S. 77–89.
- Stern, E./Grabner, R./Schumacher, R. (2007): *Lehr-Lern-Forschung und Neurowissenschaften – Erwartungen, Befunde, Forschungsperspektiven*. Berlin: BMBF.
- Sweller, J. (2011): Cognitive Load Theory. In: *Psychology of Learning and Motivation* 55, S. 37–74. Viering, T./Fischer, H. E./Neumann, K. (Hrsg.) (2010): Die Entwicklung physikalischer Kompetenz in der Sekundarstufe I. Projekt Physikalische Kompetenz.
- Walpuski, M./Ropohl, M. (2011): Einfluss des Testaufgabendesigns auf Schülerleistungen in Kompetenztests. In: *Naturwissenschaften im Unterricht Chemie* 22, 124/125, S. 82–86.

- Wellnitz, N./Fischer, H. E./Kauertz, A./Mayer, J./Neumann, I./Pant, H. A./Sumfleth, E./Walpuski, M. (2012): Evaluation der Bildungsstandards – eine fächerübergreifende Testkonzeption für den Kompetenzbereich Erkenntnisgewinnung. In: Zeitschrift für Didaktik der Naturwissenschaften 18, S. 261–291.
- Wilson, M. (2005): Constructing Measures. An item response modeling approach. Mahwah, NJ: Lawrence Erlbaum Associates.
- Winther, E. (2010): Kompetenzmessung in der beruflichen Bildung. Bielefeld: Bertelsmann.
- Wu, M. L./Adams, R. J./Wilson, M. R./Haldane, S. A. (2007): ACER ConQuest Version 2.0: Generalised item response modelling software. Camberwell, VIC: ACER Press.
- Wuttke, E./Seifried, J./Brandt, S./Rausch, A./Sembill, D./Martens, T./Wolf, D. K. (2015): Modellierung und Messung domänenspezifischer Problemlösekompetenz bei angehenden Industriekaufleuten. In: Zeitschrift für Berufs- und Wirtschaftspädagogik 111, H. 2, S. 189–207.
- Zinn, B./Wyrwal, M./Sari, D./Louis, A. (2015): Förderung von Auszubildenden im Berufsfeld Metalltechnik. In: Zeitschrift für Berufs- und Wirtschaftspädagogik 111, H. 1, S. 56–78.
- Zlatkin-Troitschanskaia, O./Seidel, J. (2011): Kompetenz und ihre Erfassung – das neue „Theorie-Empirie-Problem“ der empirischen Bildungsforschung? In: Zlatkin-Troitschanskaia, O. (Hrsg.): Stationen Empirischer Berufsbildungsforschung. Wiesbaden: VS Verlag, S. 218–233.

# Anhang C

## ***Publikation 3: Bedeutung der Informationsrepräsentation***

Abele, S., Ostertag, R., Peissner, M. & Schuler, A. (2017). Eine Eye-Tracking-Studie zum diagnostischen Problemlöseprozess: Bedeutung der Informationsrepräsentation für das Lösen diagnostischer Probleme. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 113, 86–109.

### ***Eigener Anteil an der Publikation und Begutachtungsverfahren***

Die Publikation resultierte aus einer eigenfinanzierten Pilotstudie, die ich initiierte, plante und durchführte. Die Publikation habe ich angelegt und geschrieben. Bei der Datenerhebung, Datenauswertung und Ergebnisinterpretation hat mich R. Ostertag unterstützt; das Fraunhofer-Institut für Arbeitswirtschaft und Organisation in Stuttgart (Dr. M. Peissner, A. Schuler) stellte die Eye-Tracking-Technik, übernahm den technischen Support und brachte sich in die Anlage der Studie sowie die Datenauswertung ein. Publikationen in der *Zeitschrift für Berufs- und Wirtschaftspädagogik* durchlaufen ein dreifach-blindes Begutachtungsverfahren.

# Eine Eye-Tracking-Studie zum diagnostischen Problemlöseprozess: Bedeutung der Informationsrepräsentation für den diagnostischen Problemlöseerfolg<sup>44</sup>

*An eye tracking study on the problem-solving process in professional contexts: relevance of “representing information” for the diagnostic problem-solving success*

STEPHAN ABELE, RAPHAEL OSTERTAG, MATTHIAS PEISSNER, ANDREAS SCHULLER

## Kurzfassung

Die Ziele des Beitrags waren, das Potential von Eye-Tracking für empirische Studien im Bereich der beruflichen Bildung auszuloten und die Bedeutung der Informationsrepräsentation, eines wichtigen Subprozesses des diagnostischen Problemlöseprozesses, für den diagnostischen Problemlöseerfolg zu untersuchen. Die Qualität der Informationsrepräsentation wurde anhand totaler kritischer Fixationsdauern operationalisiert. Es wurde vermutet, dass die Informationsrepräsentation für den diagnostischen Problemlöseerfolg bedeutsam ist und erfolgreiche Problemlöser bei der Informationsrepräsentation eine höhere Qualität erreichen als weniger erfolgreiche Problemlöser. Die Ergebnisse der Studie bestätigten diese Vermutung. Deutlich wurde aber auch, dass sich die Qualität der Informationsrepräsentation guter und schwacher Problemlöser nicht generell unterscheidet. Vielmehr dürfte der Qualitätsunterschied von den spezifischen Anforderungen eines diagnostischen Problems abhängen. Als wichtig erwies sich, in welchem Ausmaß ein diagnostisches Problem eine mentale Modellierung der Problemsituation erfordert und ob die Informationsrepräsentation von Routinefällen abweicht.

**Schlagnote:** Domänen-spezifisches Problemlösen; Logfile-Daten; Fixationsdauer; computerbasierte Diagnostik; Kfz-Mechatroniker

## Abstract

The present study aimed at investigating the potential of eye-tracking for research in vocational education and the relevance of “representing information” for the diagnostic problem-solving success. “Representing information” was conceptualized as a sub-pro-

---

44 Den anonymen Gutachtern danken wir herzlich für die wertvollen Hinweise zu einer früheren Version.

cess of the diagnostic problem-solving process and its quality was measured using total critical fixation durations. It was hypothesized that “representing information” is relevant to the problem-solving success and that successful problem solvers achieve a higher quality in “representing information” than poor problem solvers. Taken together, the results of the study supported this hypothesis. The results, however, also suggested that successful and poor problem solvers do not generally differ in terms of the quality of “representing information”; rather, it is to suppose that quality differences depend on specific requirements of a diagnostic problem. It appeared that quality differences in “representing information” especially are connected to diagnostic problems involving non-routine information collection and requiring a certain degree of mental modeling.

**Keywords:** domain-specific problem solving; logfile data; fixation duration; computer-based assessment; automotive technicians; car mechatronics

## 1 Notwendigkeit und Ziele der Untersuchung

Diagnostisches Problemlösen ist ein wichtiger Lernbereich der beruflichen Bildung und in vielen Berufen bedeutsam (Jonassen 2011, S.77). Mit diagnostischem Problemlösen sind Situationen angesprochen, in denen Ursachen für einen unerwünschten Zustand (fehlerhafte Kostenkalkulation, fehlerhaftes Computerprogramm, technischer Defekt etc.) zu bestimmen sind (Schaafstal/Schraagen/Van Berlo 2000, S.75). Um dieses Ziel zu erreichen, sind mehrere Aktivitäten über einen gewissen Zeitraum zielführend zu koordinieren: Es geht darum, den *diagnostischen Problemlöseprozess* erfolgreich zu gestalten.

Aus didaktischer Sicht sind wissenschaftlich gesicherte Erkenntnisse zum Problemlöseprozess außerordentlich wichtig. Im Sinne eines kognitivistischen Lernbegriffs zielt Lernen auf die Veränderung mentaler Prozesse ab (Beck 1987, S. 166 ff.). In der gegenwärtig oft dominanten Kompetenzforschung wurden in den letzten Jahren zweifellos beachtliche Fortschritte erzielt (z. B. Beck/Landenberger/Oser 2016; Fleischer/Leutner/Klieme 2012), das Interesse einschlägiger Forschungsarbeiten galt allerdings meist Prozessresultaten und nicht den Prozessen, die diesen Resultaten vorausgingen (Renkl 2012). Teilweise wurden in diesen Arbeiten aber auch Kompetenzniveau-modelle genutzt, um Aussagen zu Prozessen zu treffen (z. B. Abele 2016b; Nickolaus u. a. 2012; Petsch/Norwig/Nickolaus 2015; Winther/ Klotz 2014, S. 23). Hierbei wurden durchaus wichtige und plausible Erkenntnisse gewonnen. Problematisch an diesen Studien ist allerdings (s. dazu Abele 2016b), dass der Prozess meist nicht theoretisch modelliert wurde, eine Replikation von Befunden nur eingeschränkt gelang und mit hoch inferenten Schlüssen gearbeitet werden musste, weil keine Prozessdaten zur Verfügung standen. Wir denken, dass für gesicherte Erkenntnisse zum Problemlöseprozess letztlich Prozessdaten benötigt werden. Die Erfassung von Prozessdaten ist jedoch aufwändig.

Vor diesem Hintergrund werden mit der vorliegenden Pilotstudie die folgenden Ziele verfolgt: Erstens wird der diagnostische Problemlöseprozess in der Kfz-Domäne untersucht. Dabei geht es um die Bedeutung der Informationsrepräsentation, also eines wichtigen Subprozesses des diagnostischen Problemlöseprozesses, für den Problemlöseerfolg. Zweitens wird ausgelotet, welches Potential mithilfe von Eye-Tracking erfasste Prozessdaten für die Analyse beruflicher Problemlöseprozesse haben. In einigen sozial- und wirtschaftswissenschaftlichen Bereichen stellt Eye-Tracking einen festen Bestandteil des Methodenrepertoires dar; in der beruflichen Bildung spielt diese Erhebungsmethode (bislang) keine bedeutsame Rolle.

## 2 Bedeutung der Informationsrepräsentation für den diagnostischen Problemlöseerfolg

### 2.1 Theoretische Modellierung

#### 2.1.1 Was wird unter Informationsrepräsentation verstanden?

In der vorliegenden Untersuchung wird die *Informationsrepräsentation* als Subprozess des diagnostischen Problemlöseprozesses konzeptualisiert. Mit diesem Subprozess sind von außen nicht direkt beobachtbare mentale Aktivitäten angesprochen. Diese Aktivitäten zielen darauf ab, Informationen mental verfügbar zu machen, die für das Lösen eines diagnostischen Problems relevant, d. h. erfolgskritisch sind. Solche Informationen haben im Problemlöseprozess unterschiedliche Funktionen: Sie können dazu dienen, Hypothesen zu Ursachen des unerwünschten Zustands zu formulieren oder dazu, diese Ursachenhypothesen zu prüfen oder zu bewerten. Falls z. B. die Ursache für ein defektes Kfz-Abblendlicht zu finden ist, könnten Informationen hilfreich sein, welche Sicherung(en) das Abblendlicht „absichern“, wo sich diese Sicherung(en) im Fahrzeug befinden oder über welche Kabelstränge das Abblendlicht mit Spannung versorgt wird. Unabhängig von ihrer Funktion, können diese Informationen entweder in Form von Wissen aus dem Langzeitgedächtnis abgerufen werden oder sie entstammen der Problemumwelt. Um erfolgskritische Informationen aus der Problemumwelt zu gewinnen, muss mit der Umwelt interagiert und erfolgskritisches Informationsverhalten ausgeführt werden.

#### 2.1.2 Woran lässt sich die Qualität der Informationsrepräsentation erkennen?

*Erfolgskritisches Informationsverhalten* dient der mentalen Repräsentation erfolgskritischer Informationen, ist auf die Problemumwelt gerichtet und direkt beobachtbar. Beim Lösen diagnostischer Probleme im Kfz-Bereich kann sich ein solches Verhalten z. B. darauf beziehen, einen Einbaulageplan oder einen Stromlaufplan zur Informationsgewinnung zu verwenden.

Das erfolgskritische Informationsverhalten kann anhand des folgenden Dreischritts bestimmt werden: Zuerst werden die erfolgskritischen diagnostischen Hypothesen zu einem spezifischen Problem ermittelt; ausgehend von diesen kritischen

Hypothesen wird dann überlegt, welche Informationen zur Formulierung, Prüfung und Bewertung dieser Hypothesen vorliegen müssen; daraus werden schließlich erfolgskritische Informationsverhaltensweisen abgeleitet. Falls z. B. untersucht wird, ob eine defekte Sicherung für ein defektes Abblendlicht verantwortlich ist (erfolgskritische diagnostische Hypothese), muss bekannt sein, wo sich die betreffende Sicherung befindet und wie sie geprüft werden kann (erfolgskritische diagnostische Informationen). Um diese Informationen mental zu repräsentieren, kann es notwendig sein, einen Sicherungseinbaulageplan heranzuziehen (erfolgskritisches Informationsverhalten). Das theoretische Zentrum dieses Dreischritts bilden erfolgskritische diagnostische Hypothesen.

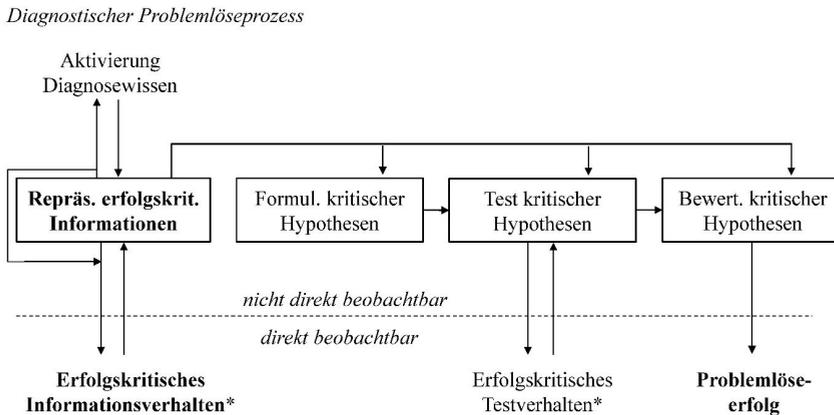
*Erfolgskritische diagnostische Hypothesen* geben sinnvolle Ursachen eines unerwünschten Zustands (diagnostischen Problems) an. Sinnvoll ist eine Ursache dann, wenn sie den unerwünschten Zustand tatsächlich erklären kann, d. h. aus fachlicher Sicht und mit Blick auf ein spezifisches Problem angemessen ist. Sinnvoll wäre es z. B., eine defekte Sicherung für einen Abblendlichtdefekt verantwortlich zu machen; unsinnig wäre es hingegen, einen defekten Drehzahlsensor oder einen „leeren“ Tank als Problemursache zu vermuten. Zu einem diagnostischen Problem lassen sich meist mehrere erfolgskritische diagnostische Hypothesen formulieren. So können neben der defekten Sicherung, ein defektes Leuchtmittel, verschiedene Kabelunterbrechungen etc. den Abblendlichtdefekt verursachen. Wie oben erwähnt, lässt sich aus den erfolgskritischen diagnostischen Hypothesen unter Anwendung fachlicher Expertise das erfolgskritische Informationsverhalten für ein spezifisches Problem bestimmen.

Das erfolgskritische Informationsverhalten kann als *Qualitätsindikator* der Informationsrepräsentation aufgefasst werden. Der mentale Subprozess „Informationsrepräsentation“ löst einerseits ein spezifisches Informationsverhalten aus, andererseits beeinflusst dieses Verhalten die Qualität der Informationsrepräsentation. So kann z. B. beim Lösen eines diagnostischen Problems ein spezifisches Informationsverhalten erforderlich sein – z. B. ein Verhalten, das den Einbaulageplan von Sicherungen verfügbar macht. Anhand des Einbaulageplans wiederum können erfolgskritische Informationen mental repräsentiert werden, was für die Qualität der Informationsrepräsentation entscheidend ist.

### **2.1.3 Rolle der Informationsrepräsentation im diagnostischen Problemlöseprozess**

Wie aus der in Abbildung 1 dargestellten Theorie zum diagnostischen Problemlöseprozess hervorgeht, hat die Informationsrepräsentation einen Einfluss auf den diagnostischen Problemlöseerfolg. In der Abbildung wird die Informationsrepräsentation (Grafik: Repräs. erfolgskrit. Informationen) als Subprozess des mentalen diagnostischen Problemlöseprozesses aufgefasst. Der diagnostische Problemlöseprozess determiniert den Problemlöseerfolg und umfasst neben der Informationsrepräsentation drei weitere Subprozesse. Die Informationsrepräsentation wirkt kausal auf diese anderen Subprozesse und damit in letzter Konsequenz auf den Problemlöseerfolg. Die Qualität der Informationsrepräsentation lässt sich am erfolgskritischen Informationsverhalten er-

kennen, das von diesem Subprozess ausgelöst wird und ihn wiederum beeinflusst, was die Pfeile zwischen „Informationsrepräsentation“ und „erfolgskritisches Informationsverhalten“ anzeigen. Im Sinne der Theorie ist anzunehmen, dass das erfolgskritische Informationsverhalten als Qualitätsindikator der Informationsrepräsentation mit dem Problemlöseerfolg korreliert.



\* Erfolgskritisches Verhalten folgt aus kritischen diagnostischen Hypothesen, die problemspezifisch von Domänenexperten zu bestimmen sind.

**Abbildung 1:** Bedeutung der Informationsrepräsentation für den diagnostischen Problemlöseerfolg (Abele 2016a)

Die Theorie verbindet domänenübergreifende und domänenspezifische Aspekte. In domänenübergreifender Hinsicht wird angenommen, dass sich diagnostische Problemlöseprozesse in unterschiedlichen beruflichen Kontexten theoretisch ähnlich „fassen“ lassen; in domänenspezifischer Hinsicht wird unterstellt, dass die konkreten mentalen Inhalte der Subprozesse und das erfolgskritische Informationsverhalten nur für eine spezifische Domäne bzw. ein spezifisches diagnostisches Problem konkretisiert werden können. Hier erfolgt die Konkretisierung anhand diagnostischer Probleme aus dem Kfz-Bereich. Weitere Aspekte der Theorie zum diagnostischen Problemlöseprozess werden bei Abele 2016a beschrieben.

## 2.2 Empirische Fundierung

### 2.2.1 Informationsrepräsentation: Bedeutsam für den diagnostischen Problemlöseerfolg?

Die empirische Forschung in der beruflichen Bildung bietet nur wenig gesicherte Erkenntnisse zur Bedeutung der Informationsrepräsentation für den diagnostischen Problemlöseerfolg. Nickolaus (2014) berichtet, dass der Problemlöseerfolg in beruflichen Kontexten davon abhängt, ob im Problemlöseprozess eigenständig Informationen erschlossen und geeignete mentale Modelle zur Problemsituation gebildet werden

können. Die in dem Überblicksartikel zusammengefassten Erkenntnisse zum Problemlöseprozess basieren auf Niveaumodellen, also einem eingangs bereits erwähnten Ansatz, der mit forschungsmethodischen Schwierigkeiten verbunden ist (s. dazu Abele 2016b). Im vorliegenden Kontext ist besonders bedeutsam, dass die aus diesem Ansatz resultierenden Erkenntnisse ohne Prozessdaten gewonnen wurden.

Im Bereich der Expertiseforschung wurden einige Studien zum diagnostischen Problemlösen und zur Bedeutung der Informationsrepräsentation durchgeführt, die auf Prozessdaten basieren und in denen sich die Qualität der Informationsrepräsentation als bedeutsam für den Problemlöseerfolg erwies. So liegen Befunde vor, die dokumentieren, dass Experten mehr Informationsverhaltensweisen zeigen als Novizen (Nendaz u. a. 2005, S. 419). Wie die von Elstein/Shulman/Sprafka (1990, S. 11) erwähnten Ergebnisse nahelegen, besteht beim diagnostischen Problemlösen zwischen der Gründlichkeit bzw. dem Umfang des Informationsverhaltens und der Qualität der Informationsrepräsentation allerdings kein Zusammenhang. Diagnostische Problemlöseexperten scheinen sich vielmehr durch ein selektives Informationsverhalten auszuzeichnen (ebd.). Für den diagnostischen Problemlöseerfolg ist also nicht entscheidend, möglichst viele, sondern die kritischen Informationen zu sammeln, d. h. erfolgskritisches Informationsverhalten auszuführen. Dementsprechend zeigten Groves/O'Rourke/Alexander (2003, S. 310), dass Experten beim Lösen diagnostischer Probleme mehr *erfolgskritische* Informationsverhaltensweisen zeigen als Novizen. Dieser ebenfalls aus dem Medizinbereich stammende Befund wird von Morris/Rouse (1985, S. 504) für das Lösen diagnostischer Probleme in technischen Kontexten bestätigt. Diese Befunde lassen sich im Sinne der oben skizzierten Theorie so interpretieren, dass die kritische Informationsrepräsentation für den Problemlöseerfolg bedeutsam ist.

Auch der Forschungsstand der Expertiseforschung ist jedoch mit Unsicherheiten verbunden. In den einschlägigen Studien wurden die Informationen oft unter kontrollierten Bedingungen vorgegeben. Die Probanden bearbeiteten also keine authentischen diagnostischen Probleme, was sowohl die Aussagekraft (Barrows/Feltovich 1987, S. 90) als auch die Generalisierbarkeit (Schwartz/Elstein 2009, S. 250) der Befunde in Frage stellt. Schließlich erlauben die referierten Studien keine Aussagen dazu, ob anhand der oben skizzierten Theorie tatsächlich erfolgskritisches Informationsverhalten identifiziert werden kann, d. h. Verhalten, das die Qualität der Informationsrepräsentation anzeigt.

### **2.2.2 Erfolgskritisches Informationsverhalten: ein Qualitätsindikator der Informationsrepräsentation?**

Bei Abele (2016a) hat sich das mit der Theorie bestimmte erfolgskritische Informationsverhalten als aussagekräftiger Qualitätsindikator der Informationsrepräsentation erwiesen. In der Studie zum diagnostischen Problemlösen im Kfz-Bereich wurden zwei unterschiedlich schwierige elektrotechnische Kfz-Probleme eingesetzt und anhand der skizzierten Theorie problemspezifisch mehrere erfolgskritische Verhaltensweisen bestimmt. Das einfache Problem konnte durch Abarbeiten einer Anleitung gelöst werden, die in einem computerbasierten Expertensystem hinterlegt war. Ein

solches Expertensystem stellt ein Standardwerkzeug von Kfz-Mechatronikern dar. Um das schwierige Problem zu lösen, musste eine Problemlösestrategie eigenständig entwickelt werden. Für beide Probleme wurden zunächst die erfolgskritischen diagnostischen Hypothesen bestimmt, anschließend wurde daraus das erfolgskritische Informationsverhalten abgeleitet. Bei beiden Problemen zeigte sich, dass das *Aufrufen eines relevanten Einbaulageplans* mit dem Problemlöseerfolg korrelierte ( $r \approx .20$ ). Im Kontext der dort verwendeten Probleme ist dieser Befund insofern sinnvoll, als die Problemlösung mit elektrotechnischen Messungen an einer spezifischen Kfz-Komponente verbunden und damit das Auffinden der Komponente ein wichtiger Problemlöseschritt war. Gezeigt hat sich dort auch, dass die Bedeutung dieses Verhaltens durch Erfahrung moderiert wurde, d. h. das Aufrufen des Einbaulageplans war nur dann bedeutsam, wenn nicht bereits einschlägige Vorerfahrungen vorlagen, d. h. die Einbaulage nicht bereits bekannt war. Falls bereits Wissen zur Einbaulage vorlag, wurde das beobachtbare Informationsverhalten (wohl) durch eine mentale Aktivität (Wissensabruf) ersetzt. Beim schwierigen Problem hatte v. a. das *Aufrufen eines relevanten Stromlaufplans* einen bedeutsamen Einfluss auf den Problemlöseerfolg ( $\beta = .52$ ). Dieser Einfluss resultierte wahrscheinlich daraus, dass beim schwierigen Problem eigenständig eine zielführende Problemlösestrategie zu entwickeln war, was eine mentale Modellierung der Problemlösesituation voraussetzte. Dafür waren relevante Ausschnitte des betreffenden Kfz-Systems mental zu repräsentiert, wofür auf den Stromlaufplan zurückgegriffen wurde, der wichtige Systeminformationen enthält.

In der erwähnten Studie wurde das erfolgskritische Verhalten anhand von computergenerierten Logfile-Dateien erfasst und berücksichtigt, ob z. B. der Stromlaufplan aufgerufen wurde oder nicht. Dies stellt insofern ein relativ grobes Vorgehen dar, als das Aufrufen von Informationsmaterialien wenig darüber Aussagen muss, ob auch die erfolgskritischen Informationen gewonnen wurden.

### 2.2.3 Kritische Fixationsdauer: eine geeignete Variante, erfolgskritisches Informationsverhalten zu operationalisieren?

Die kritische Fixationsdauer stellt eine weitere Variante dar, das erfolgskritische Informationsverhalten zu operationalisieren. Die kritische Fixationsdauer dokumentiert z. B., wie lange ein kritischer Bereich von Informationsmaterialien fokussiert wurde. In der Eye-Tracking-Forschung hat sich die Annahme als fruchtbar erwiesen, die Fixationsdauer als validen Indikator dafür zu betrachten, welche Informationen mit welcher Intensität verarbeitet werden (van Gog/Jarodzka 2013, S. 144). Gemeinhin wird unterstellt (Kok/Jarodzka 2016, S. 31), dass die Fixation einerseits von mentalen Zielen (Top-down-Steuerung) und andererseits von den fokussierten Stimuli abhängt (Bottom-up-Steuerung). Diese Annahme steht im Einklang mit der oben skizzierten Theorie, die davon ausgeht, dass der Subprozess „Informationsrepräsentation“ einerseits das Informationsverhalten auslöst und andererseits von den gewonnenen Informationen beeinflusst wird.

Wie vorliegende Befunde belegen, kann die kritische Fixationsdauer einen relativ präzisen Indikator für die Qualität der Informationsrepräsentation darstellen. In einer

Untersuchung zum diagnostischen Problemlösen in einem elektrischen Schaltkreis fanden van Gog/Paas/van Merriënboer (2005) heraus, dass erfolgreiche Problemlöser häufiger die kritischen Informationsbereiche des Schaltkreises fokussierten als weniger erfolgreiche. Unsere Recherchen erbrachten keine weiteren Eye-Tracking-Studien zum diagnostischen Problemlösen, die im vorliegenden Kontext relevant sind. Studien aus anderen Bereichen liefern jedoch einige hier relevante Hinweise. Wie Charness u. a. (2001) zeigten, werden kritische Bereiche einer Figurenkonstellation auf einem Schachbrett von Schachexperten länger fokussiert als von Novizen, wenn über den bestmöglichen nächsten Schachzug nachgedacht wird. In der Studie von Jarodzka u. a. (2010) fokussierten Experten bei einer Klassifikationsaufgabe im Biologiebereich kritische Bereiche des zu klassifizierenden Objekts länger als Novizen. Schließlich fanden Tsai u. a. (2012) für das Lösen naturwissenschaftlicher Probleme heraus, dass Experten kritische Informationsbereiche relativ lange fokussierten. Die Fixationsdauer stellte sich also sowohl domänenübergreifend (s. dazu auch van Gog/Scheiter 2010, S. 97) als auch mit Blick auf das Problemlösen als Qualitätsindikator der Informationsrepräsentation heraus.

### 3 Forschungshypothese

Vor diesem Hintergrund nahmen wir in der Pilotstudie an, dass die Qualität der Informationsrepräsentation den diagnostischen Problemlöseerfolg beeinflusst. Die Qualität der Informationsrepräsentation wurde anhand von kritischen Fixationsdauern, d. h. einer spezifischen Variante des erfolgskritischen Informationsverhaltens operationalisiert. Empirisch wurde erwartet, dass erfolgreiche Probanden beim Lösen diagnostischer Kfz-Probleme erfolgskritische Informationsbereiche länger fokussieren als weniger erfolgreiche Probanden. Der Untersuchung lag die Annahme zugrunde, dass die einbezogenen erfolgskritischen Informationen nicht bereits in Form von Langzeitwissen vorlagen, sondern aus der Problemumwelt interaktiv zu gewinnen waren.

## 4 Methode

### 4.1 Stichprobe und Design

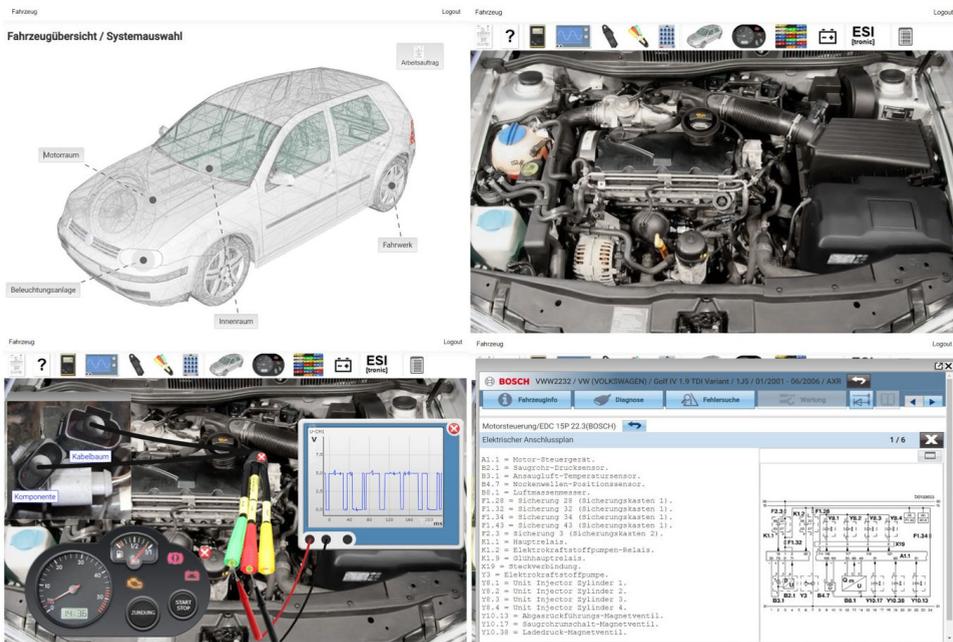
Die Stichprobe der Pilotstudie umfasste 15 Kfz-Mechatroniker des dritten Ausbildungsjahrs. Die Auszubildenden stammten aus vier Klassen zweier baden-württembergischer Berufsschulen, waren ausschließlich männlich und im Durchschnitt 21,7 Jahre alt (Range: 18–37). Auf Basis des Fachlehrerurteils wurden aus jeder Klasse zwei sehr schwache und zwei sehr starke diagnostische Problemlöser gezogen. Eine testbasierte Auswahl von „Novizen“ und „Experten“ war aus organisatorischen Gründen nicht möglich. Wie die Ergebnisse zeigen, war dies für die Studie nicht problematisch. Bei einer Erhebung führten zeitliche Verzögerungen dazu, dass nur drei Probanden einbezogen werden konnten. Jeder Proband bearbeitete nacheinander zwei typische dia-

gnostische Kfz-Probleme in einer authentischen Kfz-Computersimulation: ein Problem lag im Bereich des Drehzahlgebers (DZG) und eins im Bereich des Ladedruckmagnetventils (LMV).

## 4.2 Materialien und Apparate

### 4.2.1 Kfz-Computersimulation und diagnostische Probleme

Die diagnostischen Probleme waren in einer Kfz-Computersimulation zu bearbeiten. Diese Simulation wird in Anlehnung an Abele/Walker/Nickolaus (2014) kurz beschrieben. In der Computersimulation sind relevante Ausschnitte ausgewählter Kfz-Systeme sowie berufstypische Arbeitsmittel von Kfz-Mechatronikern verfügbar. Bei der Problembearbeitung können die Probanden diese Arbeitsmittel weitgehend analog zur Arbeit in einer Kfz-Werkstatt nutzen. Abbildung 2 zeigt Screenshots der Computersimulation.



**Abbildung 2:** Screenshots aus der Kfz-Computersimulation (oben links: Fahrzeugübersicht; oben rechts: Motorraumübersicht; unten links: Signalmessung mit dem Oszilloskop; unten rechts: Stromlaufplan aus dem computerbasierten Expertensystem)

Auf der Einstiegsseite bzw. der Fahrzeugübersicht (in der Abbildung oben links dargestellt) können die Problembeschreibung sowie verschiedene Fahrzeugsysteme per Mausklick aktiviert werden. Im vorliegenden Beitrag ist nur das System *Motormanagement* relevant, da sich beide diagnostische Probleme darauf beziehen. Zur Problemlösung können akustische Prüfungen (z. B. Prüfung des Motorgeräuschs), Sichtprü-

fungen (z. B. Prüfung, ob eine Sicherung „durchgebrannt“ ist) und elektrotechnische Messungen (z. B. eine Widerstandsmessung) durchgeführt werden. Außerdem wurden lösungsrelevante und nicht lösungsrelevante Pfade eines computerbasierten Expertensystems repräsentiert, das ein Standardwerkzeug von Kfz-Mechatronikern darstellt. In diesem Expertensystem sind Informationsmaterialien, wie z. B. der Einbaulageplan von Kfz-Komponenten und Stromlaufpläne, verfügbar. Zudem stellt das System für manche Problemstellungen eine Anleitung zur Verfügung, wie bei der Problemlösung vorzugehen ist. Die Interaktion mit der Simulation sei beispielhaft anhand einer elektrotechnischen Messung veranschaulicht.

In Abbildung 2 ist rechts oben der Motorraum zu sehen, der zur Problemlösung im Motormanagement benötigt wird und in dem über die obere Menüleiste verschiedene Arbeitsmittel verfügbar sind (z. B. der Sicherungskasten des Fahrzeuginnenraums). Um Messungen durchzuführen, ist das entsprechende Symbol des benötigten Messinstruments (Multimeter, Oszilloskop oder Strommesszange) auszuwählen. In der Abbildung unten links wurde das Oszilloskop ausgewählt und das Signal des Ladedruckmagnetventils (LMV) gemessen. Dafür mussten – wie beim „realen“ diagnostischen Problemlösen – die entsprechende Steckverbindung geöffnet, das Adapterkabel angeschlossen und, zum Starten des Fahrzeugs, die Instrumentenanzeige aktiviert werden.

Im Motormanagement sind außer den 10 Elementen der Menüleiste 13 Kfz-Komponenten und mehrere Hundert Messstellen verfügbar sowie mehrere Tausend distinkte Messungen möglich. Die untere rechte Ecke der Abbildung zeigt einen Stromlaufplan, der über das Expertensystem aufgerufen wurde. Gschwendtner, Abele und Nickolaus (2009) belegten, dass das diagnostische Problemlösen mit der Simulation valide erfasst werden kann. Außerdem geben sie weitere Informationen zur Computersimulation.

Das Drehzahlgeber-Problem lässt sich prinzipiell durch das Abarbeiten der im computerbasierten Expertensystem verfügbaren Problemlöseanleitung lösen. Umso überraschender war es, dass es sich in einer Vorgängerstudie mit einer Lösungsquote von 15,2% als äußerst schwierig herausstellte (Abele/Walker/Nickolaus 2014, S. 174). Das Problem im Bereich des Ladedruckmagnetventils (LMV) erfordert eine eigenständige Strategieentwicklung und erwies sich in derselben Studie erwartungsgemäß als sehr schwierig.

#### **4.2.2 Eye-Tracking-Technik**

Zur Erfassung der kritischen Fixationsdauern setzten wir das System „TobiiX120“ mit einer zeitlichen Auflösung von 60 Hz ein, das unterhalb des Computerbildschirms angebracht war. Für die Ermittlung der Fixationsdauern wurde die Software „Tobii Studio 3.3.2“ verwendet.

### **4.3 Durchführung**

Die Studie umfasste vier Erhebungstermine: an drei Terminen waren vier Probanden und an einem drei Probanden einbezogen. Zu Beginn der Datenerhebung erhielten

die Probanden eine Einführung in die Kfz-Computersimulation. Aus Standardisierungsgründen erfolgte die Einführung mit einem 10-minütigen Video, das jeder Proband mit Kopfhörer an seinem Monitor anschaute. Anschließend bearbeitete jeder Proband ein Übungsblatt, das sich auf zentrale Funktionen der Simulation bezog (z. B. auf die Aktivierung von Arbeitsmitteln). Nach 18 Minuten wurden den Probanden die Lösungen zu diesem Übungsblatt präsentiert und in seltenen Fällen falsch bearbeiteter Aufgaben die betreffenden Simulationsfunktionen nochmals im Klassenverbund anhand der an die Wand projizierten Simulation erläutert. Insofern ist zu unterstellen, dass keine entscheidenden nutzungsbedingten Differenzen zwischen den Probanden vorlagen. Die Einführung dauerte ca. 40 min. Anschließend bearbeitete *ein* Proband die beiden diagnostischen Problem an einem Computer, der mit dem Eye-Tracker ausgestattet war. Die anderen Probanden bearbeiteten andere diagnostische Probleme im selben Computerraum. Vor der Erfassung der Blickdaten wurde der Proband in die Eye-Tracking-Erhebung eingeführt und es wurde ihm erklärt, worauf er zu achten hat (z. B. dass er seinen Kopf nicht aus dem Aufnahmebereich nehmen soll). Danach wurde der Eye-Tracker auf die individuellen Besonderheiten der Testperson kalibriert, indem die Testperson einen Punkt fixierte, der sich auf dem Monitor bewegte. Nach der erfolgreichen Kalibrierung bearbeitete der Proband die beiden diagnostischen Probleme eigenständig, wobei seine Blickbewegung aufgezeichnet wurde. Für das erste Problem (Drehzahlgeber) standen 20 Minuten, für das zweite Problem (Ladedruckmagnetventil) 25 Minuten zur Verfügung. Der Proband hatte die Möglichkeit, das zweite Problem vorzeitig aufzurufen. Spätestens nach 45 Minuten erfolgte ein Wechsel: Der Proband des Eye-Tracking-Rechners wechselte an einen anderen Computerplatz und bearbeitete dort weitere diagnostische Probleme, ein anderer Proband wechselte an den Eye-Tracking-Platz usw. Die Probandenreihenfolge wurde insofern permutiert, als am Eye-Tracking-Platz abwechselnd leistungsstarke und leistungsschwache Probanden saßen. Zur Dokumentation erhielten die Auszubildenden einen vorstrukturierten Bogen, auf dem sie relevante Problemlöseschritte und die konkrete Problemursache notieren sollten.

## 4.4 Maße

### 4.4.1 Problemlöseerfolg

Der Problemlöseerfolg wurde anhand der schriftlichen Dokumentationen der Probanden ermittelt. Das Scoring erfolgte auf Basis eines Kodiermanuals, das von drei Kfz-Experten (zwei Ausbildungs-/Werkstattmeister, ein wissenschaftlicher Mitarbeiter) ausgearbeitet wurde und sich in einer Vorgängerstudie bewährte (Abele/Walker/Nickolaus 2014). Falls die korrekte Problemursache genannt und mit geeigneten Evidenzen belegt wurde, wurde der höchste Score von „2“ vergeben. Falls bestimmte diagnoserelevante und leicht identifizierbare Teilschritte eindeutig dokumentiert wurden (z. B. Dokumentation einer spezifischen Messung und Angabe des korrekten Messwerts), wurde mit „1“ kodiert.

#### 4.4.2 Kritische totale Fixationsdauer

Außerdem wurde die totale kritische Fixationsdauer erfasst. Die totale kritische Fixationsdauer bezeichnet die Gesamtzeit, mit der ein Proband einen kritischen Bereich eines Informationsdokuments anschaute. Um diese Fixationsdauer zu erfassen, waren kritische Bereiche, d. h. *areas of interest* (AOIs) zu definieren. Dazu wurden zunächst für jedes Problem die kritischen diagnostischen Hypothesen ermittelt, daraus problemspezifisch die erfolgskritischen Informationen abgeleitet und anschließend die AOIs bestimmt. Letztlich wurden aber nicht alle, sondern im Rückgriff auf die Studie von Abele (2016a) nur ausgewählte AOIs berücksichtigt: Bei einem Problem, das mit dem DZG-Problem vergleichbar ist, erwies sich dort der Einbaulageplan und bei einem mit dem LMV-Problem vergleichbaren Problem der Einbaulageplan sowie der Stromlaufplan als besonders relevant.

Die drei AOIs des Drehzahlgeber-Problems sind in Abbildung 3 dargestellt. Sie beziehen sich auf den Einbaulageplan, der zum Auffinden des Steckanschlusses des Drehzahlgebers nötig war. Die linke Seite des Einbaulageplans zeigt eine Liste mit verschiedenen Kfz-Komponenten und zugehörige Abkürzungen, auf der rechten Seite wird die Einbaulage der Komponenten grafisch angezeigt. Um die Einbaulage des Steckanschlusses des Drehzahlgebers zu finden, ist idealtypisch zunächst der Steckanschluss und seine Abkürzung in der Liste zu fixieren (AOI 1), dann muss die relevante Abkürzung in der Grafik gefunden werden (AOI 2) und schließlich muss die Einbaulage in der Grafik fokussiert werden (AOI 3).

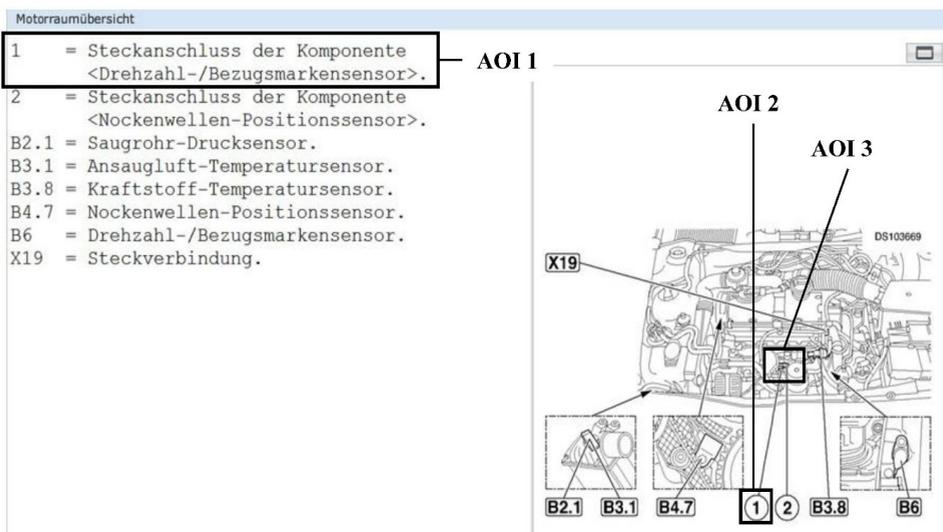
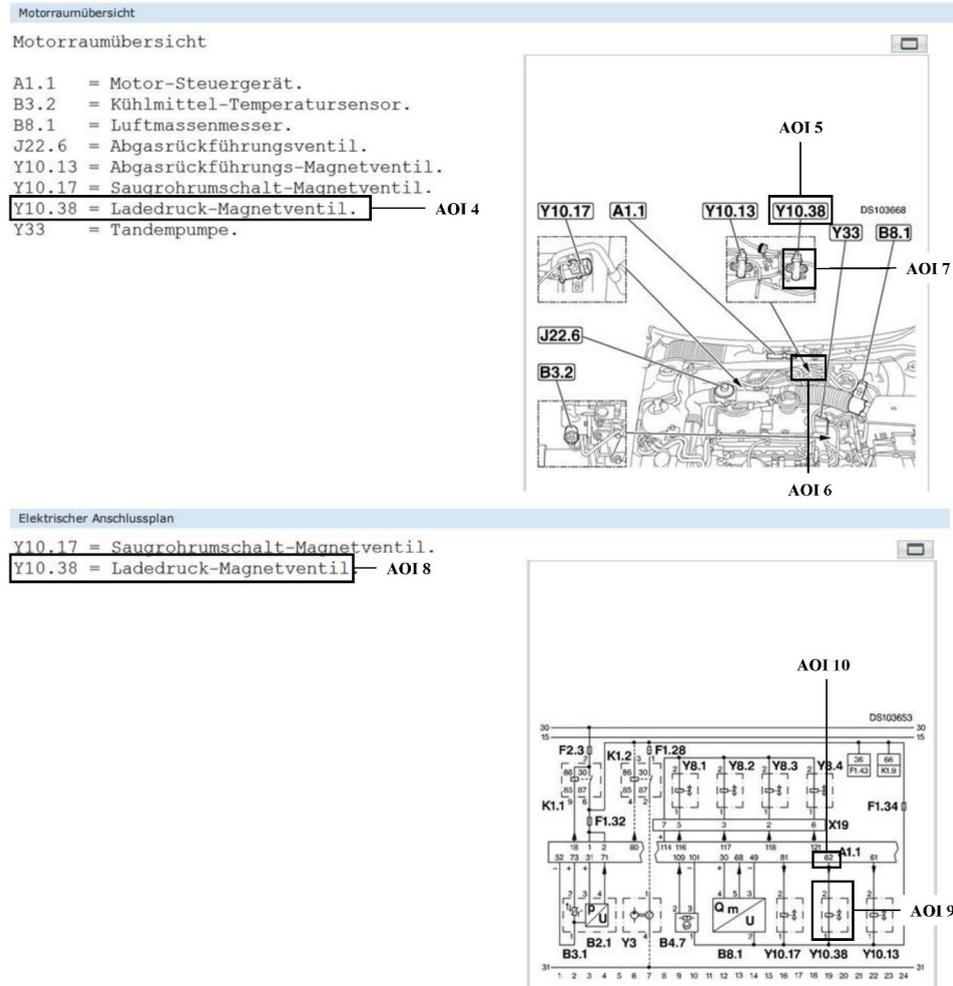


Abbildung 3: Areas of interest des Einbaulageplans des Drehzahlgebers

Abbildung 4 zeigt die AOIs des LMV-Problems. Dem oben abgebildeten Plan konnte die Einbaulage des Ladedruckmagnetventils in mehreren Schritten entnommen werden: Zunächst war der links dargestellten Liste die Abkürzung des LMV zu entnehmen

(AOI 4), dann die Abkürzung in der Grafik zu fokussieren (AOI 5), daran anschließend die Einbaulage in der Grafik zu fixieren (AOI 6) und schließlich die Einbaulage der Komponente der Grafik zu erfassen (AOI 7).

Im unten dargestellten Stromlaufplan galt es zunächst ebenfalls, die Abkürzung der Komponente zu fokussieren (AOI 8). Anschließend konnte die Komponente im Stromlaufplan fokussiert (AOI 9) und letztlich Pin 62 fixiert werden (AOI 10), der die Verbindung mit dem Motorsteuergerät anzeigt und für die Problemlösung relevant war.



**Abbildung 4:** Area of interests des Einbaulageplans (oben) und des Stromlaufplans (unten) des Ladedruck-magnetventils

## 4.5 Datenanalyse

Zur Prüfung, ob sich die kritischen Fixationsdauern erfolgreicher und weniger erfolgreicher Problemlöser unterschieden, wurde aus den folgenden Gründen der Mann-Whitney-Test, also ein non-parametrischer Test eingesetzt (Bortz/Lienert 2008, S. 59 f.): Erstens basiert die Studie auf einer kleinen Stichprobe. Zweitens sind Fixationsdauern meist nicht normalverteilt (Pannasch 2003, S. 6). Um die Bedeutung der Gruppenunterschiede zu beurteilen, wurde der exakte Signifikanztest verwendet (Field 2013, S. 547) und die Effektstärke  $r$  berechnet (Field 2013, S. 550). Für die Interpretation der Effektstärke wurde auf die Konvention von Cohen zurückgegriffen, der zufolge ein  $r$  von 0,1 einen kleinen, ein  $r$  von 0,3 einen mittleren und ein  $r$  von 0,5 einen großen Effekt anzeigt (ebd., S. 550).

## 5 Ergebnisse

### 5.1 Ausgewählte deskriptive Befunde

Deskriptive Befunde werden zu beiden diagnostischen Problemen dargestellt. Es erfolgt aber jeweils eine Beschränkung auf die totale kritische Fixationsdauer, die für die weiteren Analysen besonders bedeutsam ist.

In Tabelle 1 sind die Häufigkeiten der totalen kritischen Fixationsdauern der Probanden für jeden DZG-Problemlösescore und AOI 3 abgetragen. Erwartungsgemäß waren die fünf Probanden, die die Einbaulage des DZG-Steckanschlusses in der Grafik nicht fixierten, auch nicht erfolgreich. Auffällig ist jedoch auch, dass ein Proband AOI 3 für fünf Sekunden, also relativ lange fokussierte und dennoch erfolglos blieb. Ergänzend sei erwähnt, dass der Einbaulageplan beim DZG-Problem von 12 der 15 Probanden und der Stromlaufplan von nur drei Probanden aufgerufen wurde.

**Tabelle 1:** Häufigkeiten der totalen kritischen Fixationsdauern von AOI 3 (Einbaulage des DZG-Steckanschlusses in Grafik) je Problemlösescore

Score	Totale kritische Fixationsdauer											Gesamt
	0	0,27	1,51	1,55	2,00	4,70	5,00	6,05	6,17	6,72	7,46	
0	5	0	1	1	0	0	1	0	0	0	0	8
1	0	0	0	0	0	1	0	1	0	1	0	3
2	0	1	0	0	1	0	0	0	1	0	1	4

In Tabelle 2 sind die Häufigkeiten der Fixationsdauern von AOI 10 des LMV-Problems dargestellt. Wie zu sehen ist, weist ein erfolgloser Proband eine sehr lange Fixationsdauer von 10,23 Sekunden auf. Dieser Proband wies auch bei den anderen AOIs des Ladedruckmagnetventils eine weit überdurchschnittliche totale Fixationsdauer auf. Ferner war auffallend, dass der Proband unter Berücksichtigung aller bearbeiteter

Problemstellungen (neben den beiden hier relevanten Problemen wurden vier weitere bearbeitet) überdurchschnittlich gut abschnitt. Aufgrund dieser widersprüchlichen Befunde wurde dieser Proband von den weiteren Analysen des Ladedruckmagnetventils ausgeschlossen. Erwähnenswert ist zudem, dass der Einbaulageplan des Ladedruckmagnetventils von 14 Probanden und der Stromlaufplan von 10 Probanden aufgerufen wurden.

**Tabelle 2:** Häufigkeiten der totalen kritischen Fixationsdauern von AOI 10 (Pin 62 des LMV an Motorsteuergerät) je Problemlösescore

Totale kritische Fixationsdauer											
Score	0	0,08	0,13	0,17	0,58	1,05	5,05	6,21	10,23	11,94	Gesamt
0	2	1	1	0	0	0	1	0	1	0	6
1	4	0	0	0	0	1	0	1	0	0	6
2	0	0	0	1	1	0	0	0	0	1	3

## 5.2 Hypothesentest: Weisen erfolgreiche Probanden tatsächlich eine höhere totale kritische Fixationsdauer auf als weniger erfolgreiche Probanden?

Um die Forschungshypothese zu testen, wurden die Probanden problemspezifisch in zwei Gruppen eingeteilt. Bei der Gruppeneinteilung wurden zwei Varianten gewählt: Bei der ersten Variante wurden die erfolglosen Problemlöser (Score: 0) den erfolgreichen und teilweise erfolgreichen (Score: 1 und 2) gegenübergestellt, bei der zweiten Variante wurden die erfolglosen und teilweise erfolgreichen (Score: 0 und 1) mit der Gruppe der vollständig erfolgreichen Problemlöser (Score: 2) verglichen.

Tabelle 3 zeigt die Ergebnisse beider Varianten für die drei AOI des Drehzahlgeber-Problems. Bei der ersten Variante unterschieden sich die Fixationsdauern aller drei AOI deutlich. Dies dokumentierten die Signifikanzwerte, aber v. a. die Effektstärken, die bei AOI 3 mit  $r = -.70$  besonders hoch ausfiel. Bei der zweiten Gruppenbildung fielen die Effektstärken geringer und die  $p$ -Werte deutlich höher aus. Auch hier waren die Unterschiede bei AOI 3 am höchsten.

Am Rande sei Folgendes erwähnt: Das Aufrufen des Einbaulageplans (0: nicht aufgerufen und 1: aufgerufen) und der Problemlöseerfolg (Score: 0, 1, 2) korrelierten mit  $r = .43$  ( $p = .055$ ), die totale kritische Fixationsdauer von AOI 3 und der Problemlöseerfolg merklich höher, und zwar mit  $r = .53$  ( $p = .021$ ).

**Tabelle 3:** Gruppenunterschiede: Totale kritische Fixationsdauer beim **DZG**-Einbaulageplan

	Gruppe 1: Score 0 (n = 8)	Gruppe 2: Score 1 oder 2 (n = 7)	
	Median	Median	Mann-Whitney Test
AOI 1	0,13	1,60	$U = 14,0; z = -1,65; p = ,057$ $r = -0,43$
AOI 2	0,00	0,25	$U = 13,5; z = -1,73; p = ,046$ $r = -0,47$
AOI 3	0,00	6,05	$U = 5,0; z = -2,71; p = ,002$ $r = -0,70$
	Gruppe 1: Score 0 und 1 (n = 11)	Gruppe 2: Score 2 (n = 4)	
	Median	Median	Mann-Whitney Test
AOI 1	0,48	1,42	$U = 20,0; z = -0,27; p = ,420$ $r = -0,07$
AOI 2	0,13	0,64	$U = 12,5; z = -1,28; p = ,115$ $r = -0,33$
AOI 3	1,51	4,09	$U = 11,0; z = -1,46; p = ,085$ $r = -0,38$

*Anmerkung.* AOI 1: Kurzbezeichnung für den Steckanschluss des Drehzahlgebers in der Liste, AOI 2: Kurzbezeichnung für den Steckanschluss des Drehzahlgebers in der Grafik, AOI 3: Einbaulage des Steckanschlusses des Drehzahlgebers in der Grafik; r: Effektstärke.

In Tabelle 4 finden sich die Ergebnisse zu den Gruppenunterschieden beim Einbaulageplan des Ladedruckmagnetventils. Bei beiden Gruppierungsvarianten waren hier zwischen den erfolgreichen und weniger erfolgreichen Probanden keine bedeutsamen Unterschiede zu beobachten.

**Tabelle 4:** Gruppenunterschiede: Totale kritische Fixationsdauern beim **LMV**-Einbaulageplan

	Gruppe 1: Score 0 (n = 6)	Gruppe 2: Score 1 oder 2 (n = 9)	
	Median	Median	Mann-Whitney Test
AOI 4	1,32	0,78	$U = 21,0; z = -0,71; p = ,255$ $r = -0,18$
AOI 5	0,00	0,00	$U = 25,0; z = -0,27; p = ,445$ $r = -0,07$
AOI 6	0,29	0,32	$U = 27,0; z = 0,00; p = ,527$ $r = 0,00$
AOI 7	0,43	1,07	$U = 22,0; z = -0,59; p = ,302$ $r = -0,15$

(Fortsetzung Tabelle 4)

	Gruppe 1: Score 0 und 1 (n = 12)	Gruppe 2: Score 2 (n = 3)	
	Median	Median	Mann-Whitney Test
AOI 4	0,78	1,37	$U = 11,0; z = -1,01; p = ,176$ $r = -0,26$
AOI 5	0,00	0,00	$U = 15,0; z = -0,49; p = ,343$ $r = -0,13$
AOI 6	0,45	0,00	$U = 15,0; z = -0,46; p = ,369$ $r = -0,12$
AOI 7	0,64	1,27	$U = 17,0; z = -0,15; p = ,475$ $r = -0,04$

Anmerkung. AOI 4: Abkürzung für das Ladedruckmagnetventil in der Liste, AOI 5: Abkürzung für das Ladedruckmagnetventil in der Grafik, AOI 6: Einbaulage des Ladedruckmagnetventil in der Grafik 1, AOI 7: Einbaulage des Ladedruckmagnetventil in der Grafik 2; r: Effektstärke.

Tabelle 5 enthält die Ergebnisse zu den AOI des LMV-Stromlaufplans. Bei AOI 9 waren hier bei beiden Gruppierungsvarianten schwache bis mittlere Unterschiede festzustellen, die allerdings nicht signifikant waren. Was AOI 10 und die zweite Gruppierungsvariante anbelangt, war ein (fast) signifikanter deutlicher Gruppenunterschied zu beobachten. Bei Einschluss des oben erwähnten, ausgeschlossenen Probanden lag die Effektstärke bei diesem AOI immer noch bei  $r = .38$  und der Signifikanzwert bei  $p = .084$ .

**Tabelle 5:** Gruppenunterschiede: Totale kritische Fixationsdauern beim LMV-Stromlaufplan

	Gruppe 1: Score 0 (n = 5)	Gruppe 2: Score 1 oder 2 (n = 9)	
	Median	Median	Mann-Whitney Test
AOI 8	0,68	0,37	$U = 22,0; z = -0,07; p = ,502$ $r = -0,02$
AOI 9	0,32	3,13	$U = 16,0; z = -0,89; p = ,213$ $r = -0,24$
AOI 10	0,08	0,17	$U = 19,0; z = -0,49; p = ,335$ $r = 0,13$
	Gruppe 1: Score 0 und 1 (n = 12)	Gruppe 2: Score 2 (n = 3)	
	Median	Median	Mann-Whitney Test
AOI 8	0,68	0,37	$U = 16,0; z = -0,08; p = ,500$ $r = -0,02$
AOI 9	0,32	3,13	$U = 10,0; z = -1,04; p = ,176$ $r = -0,28$
AOI 10	0,00	0,58	$U = 6,0; z = -1,70; p = 0,60$ $r = -0,45$

Anmerkung. AOI 8: Abkürzung für das Ladedruckmagnetventil in der Liste, AOI 9: Ladedruckmagnetventil im Stromlaufplan, AOI 10: Pin des Ladedruckmagnetventils am Motorsteuergerät; r: Effektstärke.

## 6 Diskussion

### 6.1 Zusammenfassung und theoretische Vertiefung der Befunde

Die Ziele dieser Untersuchung bestanden darin, das Potential von Eye-Tracking für die Gewinnung empirischer Erkenntnisse im Bereich der beruflichen Bildung auszuloten und die Bedeutung der Informationsrepräsentation für den diagnostischen Problemlöseerfolg zu untersuchen. Die Informationsrepräsentation wurde als Subprozess des diagnostischen Problemlöseprozesses konzeptualisiert und die Qualität dieses Subprozesses anhand spezifischen Informationsverhaltens, nämlich der totalen kritischen Fixationsdauer, operationalisiert. Es wurde vermutet, dass erfolgreiche Problemlöser bei der Informationsrepräsentation eine höhere Qualität erreichen als weniger erfolgreiche Problemlöser, d. h. erfolgreiche Probanden fokussieren kritische Informationsbereiche länger als weniger erfolgreiche Probanden. In der Untersuchung wurden zwischen erfolgreichen und erfolglosen Problemlösern teilweise keine bedeutsamen Unterschiede bezüglich der kritischen Fixationsdauer gefunden, teilweise waren die Unterschiede (sehr) bedeutsam. Beobachtet wurde zudem, dass die kritischen Fixationsdauern deutlich höher mit dem Problemlöseerfolg korrelieren können als die Tatsache, ob bestimmte Informationsmaterialien aufgerufen werden oder nicht. Kritische Fixationsdauern können demnach ein präziserer Indikator für die Qualität der Informationsrepräsentation sein als binäre Verhaltensdaten, wie sie z. B. computergenerierten Logfile-Dateien zu entnehmen sind (s. z. B. Abele 2016a). Beim Drehzahlgeber-Problem waren v. a. dann bedeutsame Gruppenunterschiede festzustellen, wenn die Fixationsdauern der erfolglosen Problemlöser mit den Fixationsdauern der Problemlöser verglichen wurden, die zumindest einen Score von „1“ erreichten. Dies ist insofern plausibel, als der Score von „1“ dann vergeben wurde, wenn die Probanden den Steckanschluss des Drehzahlgebers gefunden und dies mit eindeutigen Evidenzen belegten. Die Fixationsdauern des Drehzahlgebers bezogen sich allesamt auf die Einbaulage des DZG-Steckanschlusses. Bei der anderen Gruppierungsvariante fielen die Unterschiede weniger hoch aus, bei zwei AOI waren aber auch dann noch mittlere bis starke Effekte zu beobachten, die bei einer größeren Stichprobe vermutlich auch statistisch signifikant ausgefallen wären. Beim Problem des Ladedruckmagnetventils waren beim Einbaulageplan hingegen allenfalls (sehr) schwache Gruppenunterschiede festzustellen. Wie lässt sich dieser Unterschied zwischen den beiden diagnostischen Problemen erklären?

Die bedeutsamen Gruppenunterschiede bei den Fixationsdauern des DZG-Einbaulageplans, weisen sehr wahrscheinlich auf eine problemspezifische Besonderheit hin. Um das Drehzahlgeber-Problem zu lösen, ist nicht der Drehzahlgeber selbst, sondern der in einigem Abstand befindliche Steckanschluss dieser Komponente zu lokalisieren (Abbildung 3). Dies stellt insofern einen Sonderfall dar, als Steckanschluss und Komponente bei vielen anderen diagnostischen Kfz-Problemen eine Einheit bilden. Viele Probanden wandten wohl diese Erfahrung an und suchten nach dem Drehzahlgeber und *nicht* nach dem Steckanschluss des Drehzahlgebers. Dafür spricht auch das hier nicht weiter behandelte Ergebnis, dass die meisten Probanden für eine relativ

lange Zeit im Einbaulageplan den Drehzahlgeber selbst (B6) fixierten. Wie die Befunde von Besnard/Bastien-Toniazzo (1999) belegen, kann Erfahrung in diagnostischen Situationen ein Nachteil sein, die vom Regelfall abweichen. Dies würde auch erklären, warum sich das Drehzahlgeber-Problem bislang als überraschend schwierig erwies (Abele/Walker/Nickolaus 2014). Um das Ladedruckmagnetventil (LMV) zu lokalisieren, kann dagegen dem gewohnten Verhaltensmuster gefolgt werden, weshalb beim LMV-Problem vermutlich keine (deutlichen) Gruppenunterschiede beim Fixieren des Einbaulageplans festgestellt wurden.

Bedeutsame Gruppenunterschiede wurden beim LMV-Problem allerdings bei den kritischen Fixationsdauern gefunden, die sich auf den Stromlaufplan beziehen. Diese Unterschiede waren v. a. dann deutlich, wenn die erfolglosen bzw. teilweise erfolgreichen (Score: 0 und 1) mit den erfolgreichen (Score: 2) Problemlösern kontrastiert wurden. Dies ist deshalb plausibel, weil erst für den Score von „2“ eine eigenständige Strategie entwickelt werden musste und der Stromlaufplan relevant war; der Score von „1“ ließ sich hingegen ohne Stromlaufplan durch Abarbeiten der Anleitung des computerbasierten Expertensystems erreichen. Wie lässt sich diese Relevanz der kritischen Fixationsdauern beim Stromlaufplan theoretisch erklären?

Die Gruppenunterschiede beim LMV-Stromlaufplan deuten wohl darauf hin, dass zur Lösung des LMV-Problems ein mentales Modell zur Problemlösesituation entwickelt werden musste. In diesem mentalen Situationsmodell waren vermutlich Informationen zum Ladedruckmagnetventil (z. B. Funktion und Pinbelegungen) und dessen Verbindung zu anderen Komponenten (z. B. dem Motorsteuergerät) repräsentiert. Dieses Modell diente wohl der Formulierung erfolgskritischer diagnostischer Hypothesen und hielt Informationen dazu bereit, wie die Hypothesen getestet und die Testergebnisse bewertet werden können. Ein solches Modell war sehr wahrscheinlich die Voraussetzung dafür, eigenständig eine zielführende Strategie zu entwickeln (s. dazu z. B. Kluwe/Haider 1999). Das LMV-Problem ließ sich im Unterschied zum Drehzahlgeber-Problem nicht lösen, indem einfach eine Anleitung des computerbasierten Expertensystems abgearbeitet wurde. Die mentale Modellierung stellte sich auch in anderen Studien zum diagnostischen Problemlösen als zentral heraus (Perez 1991, S. 122 ff.).

Die Befunde weisen ferner darauf hin, dass zwischen der totalen kritischen Fixationsdauer und der Qualität der Informationsrepräsentation wohl kein linearer Zusammenhang besteht. Einige Probanden hatten relativ hohe kritische Fixationsdauern, waren aber nicht erfolgreich. Vermutlich zeigen kritische Fixationsdauern ab einem bestimmten Punkt nicht mehr die Qualität der Informationsrepräsentation, sondern eher Schwierigkeiten bei der Verarbeitung des Stimulusmaterials, also Schwierigkeiten bei der Informationsrepräsentation an (s. dazu z. B. Pannasch 2003, S. 5 f.). Möglicherweise wäre es für die Erfassung der Informationsrepräsentation gewinnbringend, Fixationsdauern zu kategorisieren: In die erste Kategorie könnten jene Fälle gefasst werden, die keine oder eine zu geringe Fixationsdauer aufweisen, in die zweite Kategorie jene, die eine zu hohe Fixationsdauer aufweisen, und schließlich in die dritte Kategorie jene, deren Fixationsdauern zwischen den beiden anderen Kategorien liegen. Die Grenzen dieser Kategorien wären, wenn möglich, anhand bisheriger Befunde

oder exploratorisch zu bestimmen. Die Qualität der Informationsrepräsentation sollte entlang dieser Kategorien steigen. Aufgrund der Stichprobengröße konnte dieses Verfahren hier nicht erprobt werden. Möglich ist auch, dass zwischen der kritischen Fixationsdauer und der Qualität der Informationsrepräsentation ein invertierter u-förmiger Zusammenhang besteht (s. dazu z. B. Greiff u. a. 2016).

Zusammenfassend sprechen die Befunde dafür, dass die Informationsrepräsentation erfolgreicher Problemlöser eine höhere Qualität aufweist als die Informationsrepräsentation weniger erfolgreicher Problemlöser. Die Informationsrepräsentation ist also für den diagnostischen Problemlöseerfolg bedeutsam. Angesichts der Ergebnisse ist aber nicht davon auszugehen, dass sich die Qualität der Informationsrepräsentation guter und schwacher Problemlöser generell unterscheidet. Vielmehr dürfte die Gültigkeit dieser Aussage von spezifischen Anforderungen (Notwendigkeit der mentalen Modellierung und Abweichung von Routinefällen) eines diagnostischen Problems abhängen.

## 6.2 Implikationen der Studie

In forschungsmethodischer Hinsicht hat sich gezeigt, dass mit Eye-Tracking aufschlussreiche Prozessdaten gewonnen werden können, die gegenüber computergenerierten Logfile-Daten Vorteile bieten. Zwar sind die Anschaffungskosten geeigneter Eye-Tracker und der nötigen Auswertungssoftware relativ hoch, wie die vorliegende Studie belegt, kann diesbezüglich aber auch erfolgreich mit anderen Organisationen oder Einrichtungen kooperiert werden. Falls Blickdaten an beruflichen Schulen erfasst werden, sollten mobile Eye-Tracker mit einer relativ großen Kopfbewegungstoleranz eingesetzt werden: Es war zu beobachten, dass es für Auszubildende teilweise schwierig war, den Kopf innerhalb eines vorgegebenen Bereichs zu bewegen.

Was die Gestaltung computerbasierter Expertensysteme anbelangt, legen die Befunde nahe, der Gestaltung von Informationsmaterialien die nötige Aufmerksamkeit zu schenken. Wie an der überraschend geringen Lösungsquote des eigentlich einfachen Drehzahlgeber-Problems erkennbar wurde, kann die Gestaltung von Informationsmaterialien erhebliche praktische Konsequenzen haben. Die Lösungsquote beim Drehzahlgeber-Problem würde wahrscheinlich beachtlich steigen, wenn der Drehzahlgeber und der einschlägige Steckanschluss auf dem Einbaulageplan in einer gemeinsamen Kategorie abgebildet würden. Vielleicht könnte auch ein einleitender Satz in dieser Kategorie auf die Besonderheit hinweisen, dass im Einbaulageplan einerseits der Drehzahlgeber und andererseits der DZG-Steckanschluss dargestellt sind. Damit würde diesem Sonderfall explizit Rechnung getragen.

In einer didaktischen Perspektive regt die Studie an, bei der Förderung des diagnostischen Problemlösens der Informationsrepräsentation die nötige Beachtung zu schenken. Es geht allerdings nicht darum, oberflächliches Informationsverhalten anzutrainieren, sondern darum, mentale Grundlagen zur Repräsentation jener Informationen zu schaffen, die für den diagnostischen Problemlöseprozess und Problemlöseerfolg relevant sind: Es sollte gelernt werden, kritisches Informationsverhalten zu initiieren, auszuführen und zu nutzen, um relevante, also für den Problemlöseerfolg

erfolgskritische Informationen mental zu repräsentieren und gegebenenfalls ein mentales Modell bilden zu können. Im Anschluss an van Merriënboer (2013) kommt es wohl darauf an, die Informationsrepräsentation nicht isoliert, sondern mit „vollständigen“ diagnostischen Problemen zu fördern. Insgesamt scheint die Befundlage aber dahingehend noch nicht eindeutig, ob Subprozesse auch isoliert oder immer im Gesamtprozess gefördert werden sollten.

### 6.3 Grenzen und Ausblick

Die Ergebnisse der Studie basieren auf einer kleinen Stichprobe und zwei diagnostischen Problemen, weshalb unsicher ist, welche Bedeutung die Informationsrepräsentation generell beim diagnostischen Problemlösen hat und von welchen Anforderungsbedingungen diese Bedeutung abhängt. Zu bedenken ist allerdings, dass hier trotz einer kleinen Stichprobe teilweise signifikante und große Effekte gefunden wurden und sich diese Effekte theoretisch gut plausibilisieren lassen. Grenzen weist diese Pilotstudie auch insofern auf, als lediglich die totale Fixationsdauer berücksichtigt wurde. Eye-Tracking erlaubt jedoch die Ableitung vieler weiterer interessanter Indikatoren (Holmqvist u. a. 2011). Außerdem sei darauf hingewiesen, dass aufgrund des mehrfachen Hypothesentests mit derselben Stichprobe das Risiko der Alpha-Fehlerkumulierung besteht.

Im Bereich der Prozessdatenerfassung könnten in künftigen Studien Eye-Tracking-Brillen genutzt werden, um Blickdaten im „realen“ Arbeitsprozess zu erfassen. In zukünftigen Experimentalstudien könnte zudem systematisch der Frage nachgegangen werden, wie computerbasierte Informationssysteme zu gestalten sind, damit Nutzer ideal beim beruflichen Problemlösen unterstützt werden. Hier empfiehlt sich eine enge Kooperation zwischen Wissenschaftlern aus dem Bereich „Mensch-Maschine-Interaktion“ und der beruflichen Bildung. Aus einer didaktischen Sicht könnten Untersuchungen interessant sein, inwieweit die Informationsrepräsentation durch die Konfrontation mit Sonderfällen verbessert werden kann. Gewinnbringend könnten außerdem Eye-Tracking-Studien zum Thema sein, wie Stromlaufpläne gelesen und mentale Modelle gebildet werden. Daran könnten Interventionsstudien zur Förderung der mentalen Modellbildung und des diagnostischen Problemlösens anschließen. Der große Vorteil von Prozessdaten besteht darin, sich hypothesenprüfend an das „herantasten“ zu können, was Auszubildenden und Facharbeitern im Problemlöseprozess besondere Schwierigkeiten bereitet und deshalb besonderer Unterstützung im Lehr-Lern-Prozess bedarf.

## Literatur

- Abele, S. (2016a): Theory of the Diagnostic Problem-solving Process in Professional Contexts and its Evaluation Using Computer-generated Log-file Data, Manuskript zur Veröffentlichung eingereicht.
- Abele, S. (2016b): Umgang mit Komplexität. In: Zeitschrift für Berufs- und Wirtschaftspädagogik 112, H. 1, S. 37–59.

- Abele, S./Walker, F./Nickolaus, R. (2014): Zeitökonomische und reliable Diagnostik beruflicher Problemlösekompetenzen bei Auszubildenden zum Kfz-Mechatroniker. In: *Zeitschrift für Pädagogische Psychologie* 28, H. 4, S. 167–179.
- Barrows, H. S./Feltovich, P. J. (1987): The clinical reasoning process. In: *Medical Education* 21, S. 86–91.
- Beck, K. (1987): Die empirischen Grundlagen der Unterrichtsforschung. Eine kritische Analyse der deskriptiven Leistungsfähigkeit von Beobachtungsmethoden. Göttingen: Hogrefe.
- Beck, K./Landenberger, M./Oser, F. (Hrsg.) (2016): Technologiebasierte Kompetenzmessung in der beruflichen Bildung. Ergebnisse aus der BMBF-Förderinitiative ASCOT. Bielefeld: Bertelsmann.
- Besnard, D./Bastien-Toniazzo, M. (1999): Expert error in trouble-shooting: an exploratory study in electronics. In: *International Journal of Human-Computer Studies* 50, S. 391–405.
- Bortz, J./Lienert, G. A. (2008): *Kurzgefasste Statistik für die klinische Forschung: Leitfaden für die verteilungsfreie Analyse kleiner Stichproben*: Springer-Verlag.
- Charness, N./Reingold, E. M./Pomplun, M./Stampe, D. M. (2001): The perceptual aspect of skilled performance in chess: Evidence from eye movements. In: *Memory & cognition* 29, H. 8, S. 1146–1152.
- Elstein, A. S./Shulman, L. S./Sprafka, S. A. (1990): Medical Problem Solving. A Ten-Year Retrospective. In: *Evaluation & the Health Professions* 13, H. 1, S. 5–36.
- Field, A. (2013): *Discovering statistics using IBM SPSS statistics*. London, Thousand Oaks, New Delhi, Singapore: Sage Publications.
- Fleischer, J./Leutner, D./Klieme, E. (Hrsg.) (2012): *Modellierung von Kompetenzen im Bereich der Bildung: Eine psychologische Perspektive*. Göttingen: Hogrefe.
- Greiff, S./Niepel, C./Scherer, R./Martin, R. (2016): Understanding students' performance in a computer-based assessment of complex problem-solving. An analysis of behavioral data from computer generated log files 61, S. 36–46.
- Groves, M./O'Rourke, P./Alexander, H. (2003): The clinical reasoning characteristics of diagnostic experts. In: *Medical Teacher* 25, H. 3, S. 308–313.
- Gschwendtner, T./Abele, S./Nickolaus, R. (2009): Computersimulierte Arbeitsproben: Eine Validierungsstudie am Beispiel der Fehlerdiagnoseleistungen von Kfz-Mechatronikern. In: *Zeitschrift für Berufs- und Wirtschaftspädagogik* 105, S. 557–578.
- Holmqvist, K./Nyström, M./Andersson, R./Dewhurst, R./Jarodzka, H./van de Weijer, J. (2011): *Eye tracking: A comprehensive guide to methods and measures*: OUP Oxford.
- Jarodzka, H./Scheiter, K./Gerjets, P./van Gog, T. (2010): In the eyes of the beholder: How experts and novices interpret dynamic stimuli. In: *Learning and Instruction* 20, H. 2, S. 146–154.
- Jonassen, D. H. (2011): *Learning to solve problems: A handbook for designing problem-solving learning environments*. New York, NY: Routledge.
- Kluwe, R. H./Haider, H. (1999): Modelle zur internen Repräsentation komplexer technischer Systeme. In: *Sprache und Kognition* 9, H. 4, S. 173–192.

- Kok, E. M./Jarodzka, H. (2016): Before your very eyes: The value and limitations of eye tracking in medical education. In: *Medical Education*.
- Morris, N. M./Rouse, W. B. (1985): Review and evaluation of empirical research in troubleshooting. In: *Human Factors* 27, S. 503–530.
- Nendaz, M. R./Gut, A. M./Perrier, A./Louis-Simonet, M./Reuille, O./Junod, A. F./Vu, N. V. (2005): Common strategies in clinical data collection displayed by experienced clinician-teachers in internal medicine. In: *Medical Teacher* 27, H. 5, S. 415–421.
- Nickolaus, R. (2014): Schwierigkeitsbestimmende Merkmale von Aufgaben und deren didaktische Relevanz. In: Braukmann, U./Dilger, B./Kremer H.-H. (Hrsg.): *Wirtschaftspädagogische Handlungsfelder. Festschrift für Peter F. E. Sloane zum 60. Geburtstag*. Detmold: Eusl, S. 285–303.
- Nickolaus, R./Abele, S./Gschwendtner, T./Nitzschke, A./Greiff, S. (2012): Fachspezifische Problemlösefähigkeit in gewerblich technischen Ausbildungsberufen – Modellierung, erreichte Niveaus und relevante Einflussfaktoren. In: *Zeitschrift für Berufs- und Wirtschaftspädagogik* 108, S. 243–272.
- Pannasch, S. (2003): Ereignisbezogene Veränderungen der visuellen Fixationsdauer. Dissertation. Technische Universität Dresden.
- Perez, R. S. (1991): A view from troubleshooting. In: Smith, M. U. (Hrsg.): *Toward a Unified Theory of Problem Solving*. Hillsdale, NJ: Lawrence Erlbaum Associates, S. 115–153.
- Petsch, C./Norwig, K./Nickolaus, R. (2015): Berufsfachliche Kompetenzen in der Grundstufe Bautechnik – Strukturen, erreichte Niveaus und relevante Einflussfaktoren. In: Rausch, A./Warwas, J./Seifried, J./Wuttke, E. (Hrsg.): *Konzepte und Ergebnisse ausgewählter Forschungsfelder der beruflichen Bildung – Festschrift für Detlef Sembill*. Baltmannsweiler: Schneider Verlag Hohengehren, S. 59–88.
- Renkl, A. (2012): Modellierung von Kompetenzen oder von interindividuellen Kompetenzunterschieden. Ein unterschätzter Unterschied. In: Fleischer, J./Leutner, D./Klieme, E. (Hrsg.): *Modellierung von Kompetenzen im Bereich der Bildung: Eine psychologische Perspektive*. Göttingen: Hogrefe, S. 50–53.
- Schaafstal, A./Schraagen, J. M./van Berlo, M. (2000): Cognitive task analysis and innovation of training: the case of structured troubleshooting. In: *Human Factors* 42.
- Schwartz, A./Elstein, A. S. (2009): Clinical problem solving and diagnostic decision making: a selective review of the cognitive research literature. In: Knottnerus, J. A./Buntinx, F. (Hrsg.): *The evidence base of clinical diagnosis: Wiley-Blackwell*, S. 237–255.
- Tsai, M.-J./Hou, H.-T./Lai, M.-L./Liu, W.-Y./Yang, F.-Y. (2012): Visual attention for solving multiple-choice science problem: An eye-tracking analysis. In: *Computers & Education* 58, H. 1, S. 375–385.
- van Gog, T./Jarodzka, H. (2013): Eye Tracking as a Tool to Study and Enhance Cognitive and Metacognitive Processes in Computer-Based Learning Environments. In: Azevedo, R./Alevin, V. (Hrsg.): *International Handbook of Metacognition and Learning Technologies*. New York, NY: Springer New York, S. 143–156.

- van Gog, T./Paas, F./van Merriënboer, J. J. G. (2005): Uncovering expertise-related differences in troubleshooting performance: combining eye movement and concurrent verbal protocol data. In: *Applied Cognitive Psychology* 19, H. 2, S. 205–221.
- van Gog, T./Scheiter, K. (2010): Eye tracking as a tool to study and enhance multimedia learning. In: *Learning and Instruction* 20, H. 2, S. 95–99.
- van Merriënboer, J. (2013): Perspectives on problem solving and instruction. In: *Computers & Education*, H. 64, S. 153–160.
- Winther, E./Klotz, V. K. (2014): Spezifika der beruflichen Kompetenzdiagnostik–Inhalte und Methodologie. In: *Zeitschrift für Erziehungswissenschaft* 17, H. 1, S. 9–32.



# Anhang D

## ***Publikation 4: Theorie zum diagnostischen Problemlöseprozess***

Abele, S. (2017). Diagnostic Problem-Solving Process in Professional Contexts: Theory and Empirical Investigation in the Context of Car Mechatronics Using Computer-Generated Log-Files. *Vocations and Learning* (2018) 11:133–159. DOI 10.1007/s12186-017-9183-x

### **Eigener Anteil an der Publikation und Begutachtungsverfahren**

Die Publikation resultierte aus meinem laufenden DFG-Projekt mit dem Förderkennzeichen Ab 441 1–1, in dem ich auf Daten aus dem Projekt „Mathematisch-naturwissenschaftliche Kompetenzen in der beruflichen Erstausbildung“ (ManKobE) zurückgreife.<sup>45</sup> Für die Beantragung, die Anlage, Durchführung und Auswertung des Projekts bin ich ebenso allein verantwortlich wie für die gesamte Publikation. Publikationen in *Vocations and Learning* durchlaufen ein dreifach-blindes Begutachtungsverfahren.

---

45 Das Projekt wurde vom Leibniz Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN) unter anderem in Kooperation mit der Abteilung für Berufs-, Wirtschafts- und Technikpädagogik der Universität Stuttgart durchgeführt. Den Professoren Dr. O. Köller sowie Dr. R. Nickolaus gilt mein Dank, dass ich auf die betreffenden Daten zurückgreifen durfte.

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

STEPHAN ABELE

## 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.

**Keywords:** Diagnostic problem-solving process . Troubleshooting . Clinical reasoning . Log-file analysis . Car mechatronics . Automotive technicians

## 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 Feltoovich (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).<sup>46</sup>

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 Feltoovich 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.

---

<sup>46</sup> 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.

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 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,<sup>47</sup> 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 subprocess. 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

---

47 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).

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 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 be-

come 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; Fel'tovich 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 Prechtel (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, 77 ff.) 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) <b>Representing information</b>	To mentally represent information relevant to the diagnostic problem	Retrieving and using external information material
(2) <b>Generating diagnostic hypotheses</b>	To generate diagnostic hypotheses providing potential but untested problem solutions	Usually none <sup>a</sup>
(3) <b>Testing diagnostic hypotheses</b>	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) <b>Evaluating diagnostic hypotheses</b>	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)

<sup>a</sup> 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).

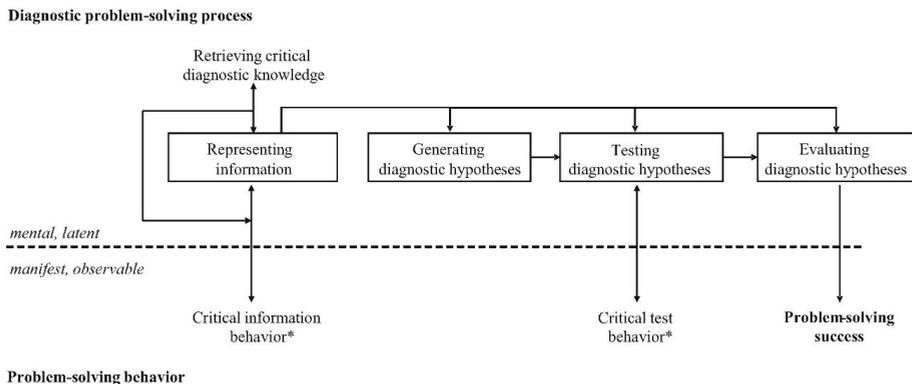
## **Aim and Hypotheses of the Empirical Study**

The aim of the empirical study is to investigate the theory of the diagnostic problem-solving process. Abbildung 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).



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

**Figure 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 oper-

ationalization 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.

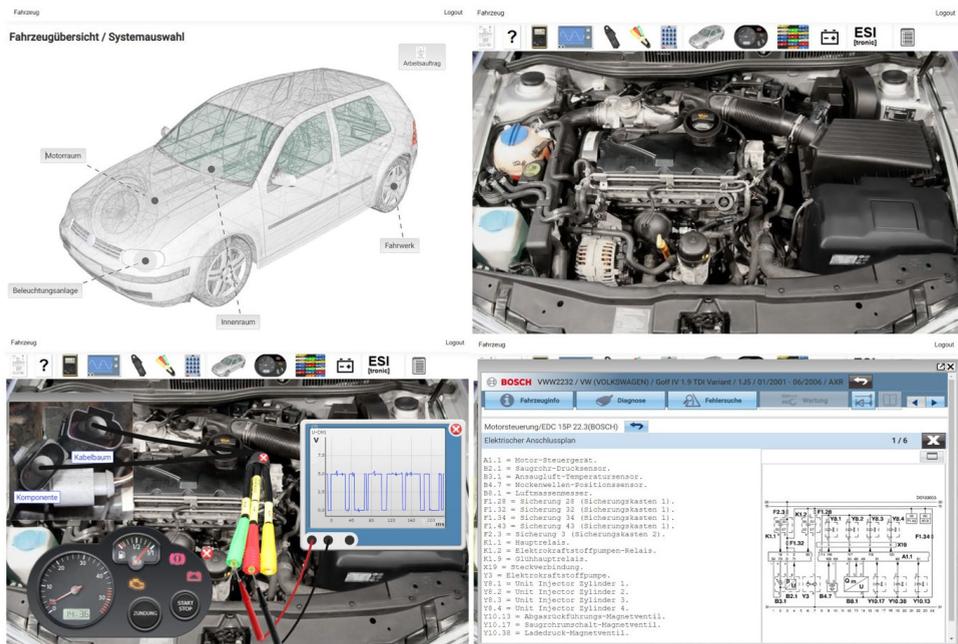
Please note that RH2 and RH3 are not logically derivable from, but in line with, the theory. Regarding these empirical hypotheses, the empirical results do not have logical implications for the theory but they provide evidence to evaluate its empirical plausibility.

## 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 (Abbildung 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 "Belectronic engine management" is relevant. In this system, 17 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.



**Figure 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 “Beleuchtung 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)

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 ( $N = 67$ ):  $\kappa = .95$  (P1) and  $.97$  (P2).

To determine the *critical behavior*, the problems were analyzed and the critical diagnostic hypotheses were identified (see first line of Tabelle 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 (Abbildung 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 Tabelle 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 (Tabelle 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%).

**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 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="00" motorState="100" timeStamp="+00:02:44" id="8">ignition on: true</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:02:45" id="9">shortCut open: DiagnoseSoftware</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:02:45" id="10">ESITRONIC screen: 00_01</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:02:47" id="11">ESITRONIC screen: 00_02</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:02:56" id="12">ESITRONIC screen: g4_00_01</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:02:58" id="13">ESITRONIC screen: g4_03_01_01</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:03:03" id="14">ESITRONIC screen: g4_03_01_13_01</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:03:10" id="15">ESITRONIC screen: g4_03_01_13_m01</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:03:13" id="16">ESITRONIC screen: g4_03_01_13_m01_m08_m02_04_01</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:03:18" id="17">ESITRONIC Fehlerspeicher lesen:</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:03:22" id="18">ESITRONIC screen: g4_03_01_13_m01_m08_m02_04_02</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:03:57" id="19">ESITRONIC screen: g4_44A3_01</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:04:12" id="20">button close: null</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:04:21" id="21">module open: Motorabdeckung</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:04:23" id="22">module zoom in: Kraftstofftemperatursensor</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:04:25" id="23">module open: Kraftstofftemperatursensor</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:04:27" id="24">module close: Kraftstofftemperatursensor</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:04:29" id="25">module open: Kraftstofftemperatursensor</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:04:34" id="26">shortCut open: Multimeter</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:04:42" id="27">Messspitze aufgesetzt: red/01_04-03-01</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:04:45" id="28">Messspitze aufgesetzt: black/01_04-03-02</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:04:45" id="29">Multimeter measure: 01_04-03-01/01-04-03-02 (5 Udc)</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:05:20" id="30">button close: Multimeter</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:05:22" id="31">shortCut open: DiagnoseSoftware</activity>
  <activity lightState="00" motorState="100" timeStamp="+00:05:27" id="32">ESITRONIC screen: g4_44A3_02</activity>
  <activity lightState="00" motorState="100" 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="+00:05:56" id="37">Messspitze aufgesetzt: red/01_04-03-04</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>
  <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>
    
```

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

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 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. After-

wards, 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)			
				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 ( $\beta$ ) 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  $\chi^2$  value, ratio of the  $\chi^2$  value and degrees of freedom  $\leq 3$ , root mean square error of approximation (RMSEA)  $\leq .08$ , weighted root mean square residual (WRMR)  $\leq 1$  and comparative fit index (CFI)  $\geq .95$  (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 BT3<sup>o</sup>, BT4<sup>o</sup> and BT5<sup>o</sup> (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 BT3<sup>o</sup> 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, Tabelle 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 (Tabelle 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 ( $\Phi = .89$ , Tabelle 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 R2 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		
	<i>b</i>	<i>SE b</i>	$\beta$	<i>b</i>	<i>SE b</i>	$\beta$	<i>b</i>	<i>SE b</i>	$\beta$
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	
<i>R</i> <sup>2</sup>	.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		
	<i>b</i>	<i>SE b</i>	$\beta$	<i>b</i>	<i>SE b</i>	$\beta$	<i>b</i>	<i>SE b</i>	$\beta$
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	
<i>R</i> <sup>2</sup>	.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 *R2* 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		
	<i>b</i>	<i>SE b</i>	$\beta$	<i>b</i>	<i>SE b</i>	$\beta$
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	
<i>R</i> <sup>2</sup>	.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 Tabelle 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.

Tabelle 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  $R$  (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.<sup>48</sup> 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.

**Table 7:** Moderation effect of experience problem analyzing Problem 2

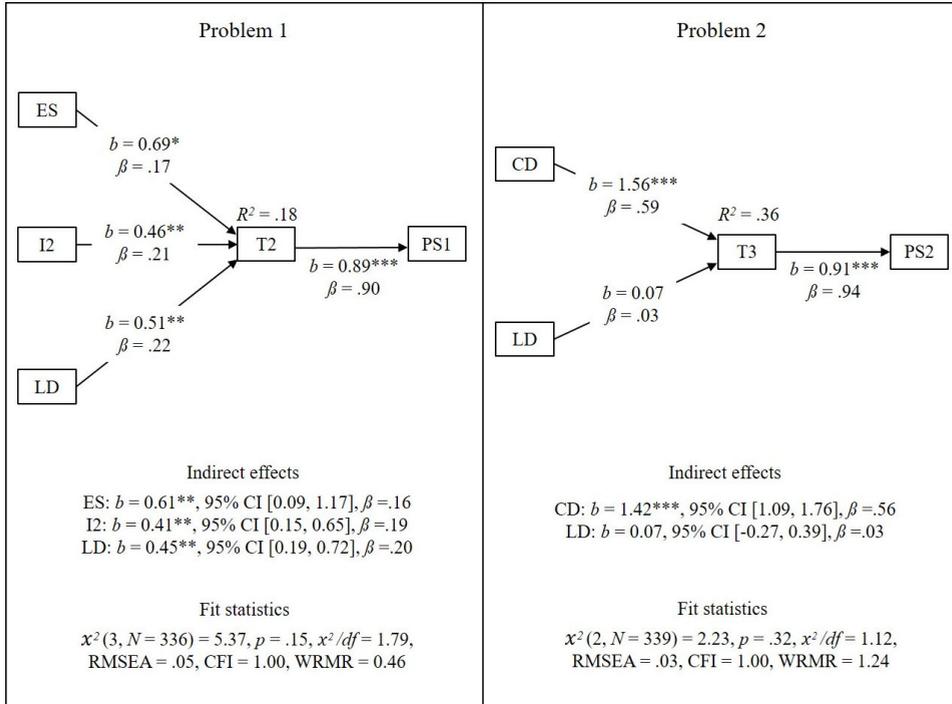
Critical information behavior	Model 1: Inexperienced group			Model 2: Experienced group		
	<i>b</i>	<i>SE b</i>	$\beta$	<i>b</i>	<i>SE b</i>	$\beta$
I1	-0.26	0.38	-.09	-.64	0.35	-.23
I2	-0.21	0.35	-.08	-.25	0.29	-.10
LD	0.52	0.45	.19	0.06	0.45	.02
CD	1.61***	0.40	.57	1.42***	0.34	.56
Constant	2.02***	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 \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$

The right part of Abbildung 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

<sup>48</sup> As mentioned before, I3 was not considered due to a perfect correlation with I2.

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.



**Figure 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 Bcorrect“ 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 Bmay 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 Merriënboer (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 Merriënboer 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.

**Acknowledgements** This research was funded by the German Research Foundation (DFG: Ab 441/1–1) and the Leibniz Institute for Science and Mathematics Education at Kiel University (IPN). I am grateful to Reinhold Nickolaus and Olaf Köller for the permission to use data from the project "Mathematics and Science Competencies in Vocational Education and Training" (ManKobE: <http://www.ipn.uni-kiel.de/de/forschung/projekte/mankobe>). Additionally, I would like to thank Didem Atik, Julian Etzel, Christoph Lindner and Jan Retelsdorf for their excellent support in collecting and transferring the data.

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

- 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.
- 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 Merriënboer, 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. *IEEE Transactions on Systems, Man and Cybernetics*, 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.



# Anhang E

## ***Publikation 5: Erfassung der diagnostischen Kfz-Problemlösekompetenz mit computersimulierten Key-Feature-Problemen***

Abele, S., Walker, F. & Nickolaus, R. (2014). Zeitökonomische und reliable Diagnostik beruflicher Problemlösekompetenzen bei Auszubildenden zum Kfz-Mechatroniker. In S. Greiff, A. Kretzschmar & D. Leutner (Hrsg.), *Problemlösen in der Pädagogischen Psychologie*, Zeitschrift für Pädagogische Psychologie, 4, 167–179.

### **Eigener Anteil an der Publikation und Begutachtungsverfahren**

Die Publikation resultierte aus einer Teilstudie des Projekts „KOKO Kfz“ der Förderlinie „Technologieorientierte Kompetenzmessung in der Berufsbildung (ASCOT)“ des Bundesministeriums für Bildung und Forschung (BMBF)<sup>49</sup>, an dessen Beantragung ich mitwirkte. In dieser Teilstudie hatte ich wissenschaftliche Leitungsfunktionen und war ich für die Entwicklung der computerbasierten Key-Feature-Probleme, das Studiendesign, die Durchführung sowie die Datenauswertung verantwortlich. Die Publikation habe ich konzipiert und verfasst, die einschlägigen statistischen Auswertungen stammen ebenfalls von mir. Des Weiteren leisteten Prof. Dr. R. Nickolaus sowie Juniorprofessor Dr. F. Walker wichtige Beiträge zur Publikation. Publikationen in der *Zeitschrift für Pädagogische Psychologie* durchlaufen ein dreifach-blindes Begutachtungsverfahren.

---

49 An dem Teilprojekt wirkte ein größerer Kreis von Personen mit, die im Folgenden genannt seien und bei denen ich mich bedanken möchte: Prof. Dr. Tobias Gschwendtner, Dipl.-Gwl. Horst Heinzmann, Sebastian Heilig MA, Prof. Dr. Anja Sarnitz, Dipl.-Gwl. Thomas Schmidt und Thomas Trezbiatowski. Zudem möchte ich den zahlreichen Experten und Schulen für ihre Unterstützung danken.

# Zeitökonomische und reliable Diagnostik beruflicher Problemlösekompetenzen bei Auszubildenden zum Kfz-Mechatroniker

STEPHAN ABELE, FELIX WALKER UND REINHOLD NICKOLAUS

## Zusammenfassung

Im vorliegenden Beitrag wird eine bislang ungelöste Herausforderung bei der Erfassung beruflicher Problemlösekompetenzen behandelt: Aufgrund langer Bearbeitungszeiten ökologisch valider beruflicher Problemlöse-Szenarien und begrenzter Testzeiten können zur Diagnostik nur wenige Szenarien vorgegeben werden. Das damit verbundene Reliabilitätsproblem wird am Beispiel von Kfz-Mechatronikern des 3. Ausbildungsjahrs ( $N = 275$ ) bearbeitet, deren Problemlösekompetenz sich insbesondere bei der Kfz-Fehlerdiagnose zeigt. Vorgestellt wird ein diagnostischer Ansatz, bei dem Kfz-Fehlerdiagnose-Szenarien und Teilkompetenz(TK)-Items eingesetzt werden, die relativ schnell zu bearbeiten sind und leistungskritische Anforderungen der Kfz-Fehlerdiagnose abdecken. Die Ergebnisse zeigen, dass die TK-Items und die Fehlerdiagnose-Szenarien eindimensional modellierbar sind und der ergänzende Einsatz der TK-Items zeitökonomisch einen substantiellen Reliabilitätszuwachs bringt. Abschließend wird diskutiert, ob der Ansatz auf andere Domänen übertragbar ist und pädagogische Interventionen bereichern kann.

**Schlüsselwörter:** berufliche Problemlösekompetenz, Fehlerdiagnose, Kfz-Mechatroniker, Reliabilität, Computergestützte Diagnostik

## Time-Saving and Reliable Diagnostics in Measuring Professional Problem-Solving Competence in the Domain of Car Mechatronics

### Abstract

The aim of this article is to overcome the challenge often associated with the measurement of professional problem-solving competence: Ecologically valid problem-solving scenarios require a lot of testing time. However, since this time is usually limited, only a few scenarios can be selected for competence measurement, leading to unreliable results. This reliability problem is addressed in the domain of car mechatronics, where problem-solving competence can be measured with troubleshooting scenarios. The article presents an innovative diagnostic approach, which uses such scenarios *and* some additional items covering certain critical troubleshooting requirements in comparably short testing time. The analyses are based on a sample of car mechatronics ( $N = 275$ , 3rd year of training) and suggest that the scenario measurements and the additional

items reflect one empirical dimension. Furthermore, the implementation of the additional items brought substantial gains in reliability within acceptable test times. Finally, whether the approach is transferable to other domains and could enhance educational processes will be discussed.

**Keywords:** professional problem-solving competence, troubleshooting, car mechatronics, reliability, computer-based measurement

## 1 Problemstellung und Problembearbeitung<sup>50</sup>

Ein wichtiges Ziel von Bildungssystemen besteht darin, erfolgreiche Problemlöser auszubilden. Diskussionen um ein internationales Large-Scale-Assessment in der beruflichen Bildung (VET-LSA) vergegenwärtigen, dass die Qualität beruflicher Bildungssysteme besonders daran festzumachen ist, wie erfolgreich Auszubildende konkrete *fachliche Probleme* lösen (Baethge et al., 2006). Eine im Hinblick auf ein VET-LSA sehr bedeutsame und bislang ungelöste Herausforderung ist, solche beruflichen Problemlösekompetenzen *reliabel* zu erfassen (Nickolaus & Seeber, 2013). Dies ist hauptsächlich darauf zurückzuführen, dass bei der Diagnostik aus motivationalen und organisatorischen Gründen nur wenige ökologisch valide Problemlöse-Szenarien vorgegeben werden können, da deren Bearbeitung lange dauert.

Im vorliegenden Beitrag wird ein Ansatz vorgestellt, mit dem das angesprochene *Reliabilitätsproblem* zeitökonomisch überwunden werden soll. Dieser Ansatz wird für die berufliche Ausbildung entwickelt und bezieht sich auf die Erfassung beruflicher Problemlösekompetenzen bei Kfz-Mechatronikern<sup>51</sup>.

## 2 Problemlösen in der Domäne Kfz-Mechatronik

In internationalen (Müller & Schelten, 2009, S. 36 f.) und nationalen Studien (Spöttl, Becker & Musekamp, 2011) stellte sich die Fehlerdiagnose als besonders relevanter Ausschnitt der Domäne Kfz-Mechatronik heraus, was v. a. auf die besonders problemhaltigen Anforderungen dieser Tätigkeit zurückzuführen ist (Rauner, Hitz, Spöttl & Becker, 2002, S. 76). Mit der Fokussierung der Kfz-Fehlerdiagnose wird auch der Anschluss an die umfassenden Forschungsarbeiten zur Fehlerdiagnose sichergestellt (international: z. B. Kontogiannis & Moustakis, 2002; Ross & Orr, 2009; national: z. B. Rowold, Hochholdingner & Schaper, 2008; Sonntag & Schaper, 1997). Die Kfz-Fehlerdiagnose lässt sich in Anlehnung an Spöttl et al. (2011) als eine Tätigkeit verstehen, die auf die Ursachenbestimmung für ein fehlerhaftes Kfz-System in Abhängigkeit von auftre-

---

50 Die Autoren danken den anonymen Gutachtern dieser Zeitschrift für die hilfreichen Kommentare zu einer früheren Beitragsfassung.

51 Um den Lesefluss nicht zu stören, verwenden wir im Text die maskulinen Wortformen stellvertretend für beide Geschlechter.

tenden Fehlersymptomen abzielt (s. auch Schaafstal, Schraagen & van Berlo, 2000, S.75). Erst wenn die Fehlerursache bekannt ist, kann das defekte Kfz-System wiederhergestellt werden. Da bei der Fehlerbehebung andere und häufig weniger problemhaltige Anforderungen (z. B. manuelle und relativ anspruchslose Anforderungen wie der Tausch einfacher Komponenten; Rauner et al., 2002, S.17) auftreten als bei der Fehlerdiagnose, wird die Fehlerbehebung hier *nicht* einbezogen. Im vorliegenden Beitrag geht es ausschließlich um Fehlerdiagnosen in *elektrotechnischen Kfz-Systemen*, genauer, in der *Beleuchtungsanlage* und im *elektronischen Motormanagement*, die als besonders praxisrelevant und anspruchsvoll betrachtet werden können (ebd., S. 85 f.).

Das Problem besteht in einer Kfz-Fehlerdiagnosesituation in der Unkenntnis der Ursache für einen Systemdefekt und die Problemlösung in der eindeutigen Rückführung beobachtbarer Symptome auf die Fehlerursache. Fehlerdiagnosen gelten in der Problemlöseforschung als bedeutsamer Problemlösetyp (Funke, 2006; Leutner, Fleischer, Wirth, Greiff & Funke, 2012). Wie für Problemlösen charakteristisch (z. B. Fischer, Greiff & Funke, 2012), wird auch bei der Kfz-Fehlerdiagnose eine unbefriedigende Situation (fehlerhaftes Kfz-System) in einen befriedigenden Zielzustand (Kenntnis der Fehlerursache) überführt.

Bei Kfz-Fehlerdiagnosen gehen die für die Problemlösung notwendigen Informationen nur teilweise aus der Problembeschreibung hervor. Die wichtigsten lösungsrelevanten Informationen müssen durch gezielte Eingriffe in das defekte elektrotechnische System gewonnen werden, das i. d. R. aus zahlreichen vernetzten Elementen besteht. Insofern wird die *Kfz-Fehlerdiagnose in elektrotechnischen Kfz-Systemen* als *komplexes Problemlösen in der Domäne Kfz-Mechatronik* aufgefasst (s. Greiff, Kretzschmar & Leutner, 2014).

## 3 Fehlerdiagnose in elektro-technischen Kfz-Systemen

### 3.1 Anforderungen

Zur strukturierten Anforderungsbeschreibung bei der Kfz-Fehlerdiagnose werden in Anlehnung an die allgemeine Problemlöseforschung vier Fehlerdiagnosephasen unterschieden (OECD, 2013, S. 126): (1) *exploring and understanding*: In dieser Phase besteht das Ziel darin, aktiv lösungsrelevante Informationen zur Fehlerdiagnosesituation zu sammeln. Hierzu sind die Problembeschreibung sowie technische Kfz-Spezifika zu erfassen und oft weitere technische Informationsquellen (z. B. ein computerbasiertes Expertensystem) zu nutzen. (2) *representing and formulating*: Daran anknüpfend ist im Zusammenspiel mit spezifischem Vorwissen ein kohärentes mentales Modell zur Fehlerdiagnosesituation zu entwickeln, das die Formulierung von Hypothesen zur Fehlerursache erlaubt.

(3) *planning and executing*: Im Anschluss daran werden eine Gesamtstrategie entwickelt, Teilziele gebildet und die Prüfstrategie umgesetzt. Zur Prüfung werden vorwiegend elektrotechnische Messungen durchgeführt. (4) *monitoring and reflecting*:

Schließlich wird bewertet, ob die Strategie zielführend war und ob die Ursachenhypothese zutrifft oder ob eine neue Hypothese zu formulieren ist.

Bei der folgenden exemplarischen Anforderungsbeschreibung wird auf ein Fehlerdiagnose-Szenario des *elektronischen Motormanagements* zurückgegriffen. Bei diesem Szenario liegt die Fehlerursache in einer unterbrochenen Signalleitung zwischen dem Ladedruckmagnetventil (LMV) und dem Motorsteuergerät. *Exploring and understanding*: Um Informationen zur Fehlerdiagnosesituation zu sammeln, sind die Problembeschreibung, in der die Fehlersymptome gegeben sind, sowie technische Spezifika des Fahrzeugs zu erfassen und die Fehlersymptome nachzuvollziehen. In der vorliegenden Situation wird üblicherweise zudem mit dem computerbasierten Expertensystem der Fehlerpeicher ausgelesen. Dabei wird sichtbar, dass im Bereich des LMV ein Problem vorliegt, und es werden im Expertensystem verschiedene Ursachenhypothesen und Prüfstrategien vorgeschlagen. Um die Prüfungen vorzunehmen, müssen überdies Informationen gewonnen werden, wo sich das LMV befindet und an welchen Pins des LMV zu messen ist. Hierfür sind Einbaulage- und Stromlaufpläne zu studieren. *Representing and formulating*: Auf Basis der gewonnenen Informationen wird im Rückgriff auf Vorwissen ein situationsspezifisches mentales Modell entwickelt, das die vorgeschlagene Fehlerursache und prüfrelevante Informationen (z. B. Prüfstellen und relevante Messgeräte) enthält. *Planning and executing*: Im Weiteren werden auf Basis dieses mentalen Modells die vom Expertensystem vorgeschlagenen Messschritte unter Nutzung eines Multimeters oder eines Oszilloskops durchgeführt. *Monitoring and reflecting*: Anschließend werden die Prüfergebnisse bewertet, wobei erkennbar wird, dass die durchgeführten Prüfungen hier nicht für eine eindeutige Ursachenbestimmung ausreichen. *Representing and formulating*: Vielmehr ist das mentale Modell anhand der ermittelten Ergebnisse und eines Stromlaufplans zu modifizieren und eine neue Ursachenhypothese zu formulieren, d. h. es ist zu überlegen, welche Komponente den Defekt verursachen könnte. *Planning and executing*: Im Anschluss daran wird eine Strategie zur Hypothesenprüfung entwickelt und die Hypothese geprüft. Im vorliegenden Fall ist die Fehlerursache durch eine Widerstandsmessung an der Signalleitung des LMV eindeutig bestimmbar. *Monitoring and reflecting*: Der an der relevanten Stelle zu messende Widerstand ist unendlich groß, was eine unterbrochene Signalleitung anzeigt und die beobachtbaren Fehlersymptome erklärt. Das Beispiel verdeutlicht, dass das Szenario, wie auch die anderen in die vorliegende Untersuchung einbezogenen Szenarien (s. Abschnitt 7.2 und vgl. Tab. 1), alle oben beschriebenen Anforderungen umfasst, allerdings werden diese oft iterativ und nicht immer vollständig in der oben idealtypisch dargestellten Reihenfolge durchlaufen.

### 3.2 Leistungskritische Teilanforderungen

Einige der genannten Anforderungen sind für die weitere Argumentation besonders zentral. Zu beachten ist, dass Anforderungen, wie sie für Fehlerdiagnosen in mechanischen Systemen typisch sind (z. B. Druckmessung, Undichtigkeit ermitteln etc., Schaper, 1994, S. 197), hier nicht betrachtet werden.

Nickolaus, Abele, Gschwendtner, Nitzschke und Greiff (2012, S. 255) arbeiteten auf Basis eines empirisch ermittelten Niveaumodells besonders leistungskritische Teilkompetenzen und damit assoziierte Teilanforderungen der Kfz-Fehlerdiagnose in elektrotechnischen Kfz-Systemen heraus. Diese Teilkompetenzen lassen sich mit Teilkompetenz-Items erfassen (s. Abschnitt 5.2) und beziehen sich auf folgende Anforderungen: (1) Nutzung des computerbasierten Expertensystems zur Informationsbeschaffung, Lesen von Stromlaufplänen und Plänen zur Einbaulage von Systemkomponenten sowie Erfassung relevanter technischer Kfz-Spezifika (z. B. Komponententyp, Verlauf des elektrischen Energieflusses, Vernetzung von Systemkomponenten). Solche Anforderungen gehören zum *exploring and understanding* und werden im Weiteren als *Informationsgewinnung* bezeichnet. (2) Entwicklung einer Strategie zur Funktionsprüfung elektrotechnischer Komponenten in Abhängigkeit von den gewonnenen Informationen (*Entwicklung einer Prüfstrategie*). (3) Korrekter Einsatz elektrotechnischer Messgeräte, Anwendung der gewonnenen Informationen und Einbezug anspruchsvoller Komponenten in die Prüfung (z. B. Motorsteuergerät) (*Umsetzung der Prüfstrategie*). Die beiden letztgenannten Anforderungen beziehen sich auf *planning* (Entwicklung einer Prüfstrategie) und *executing* (Umsetzung der Prüfstrategie).

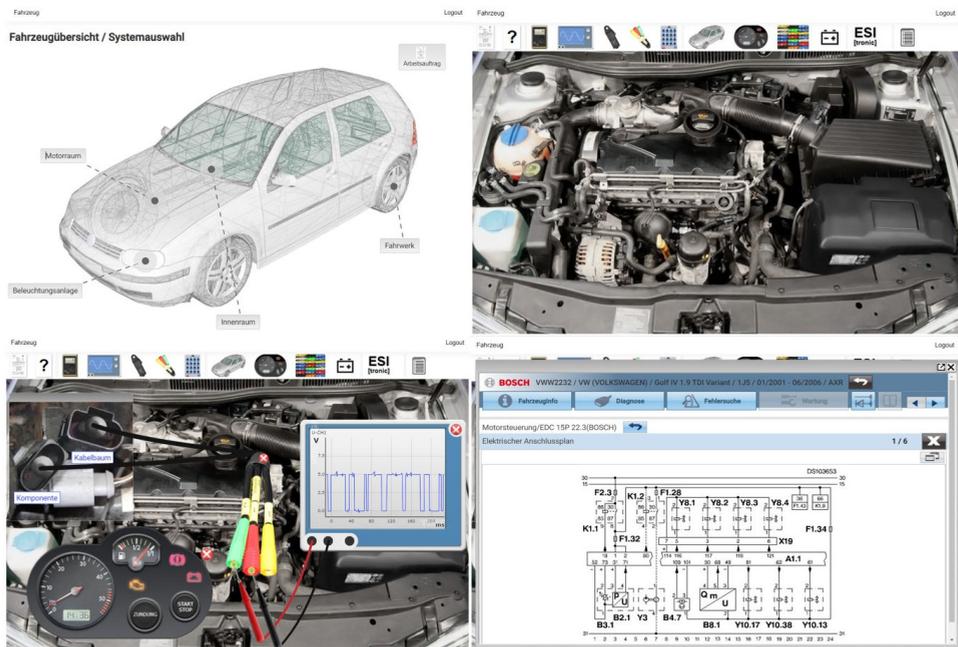
### 3.3 Kfz-Fehlerdiagnosekompetenz

Um die Anforderungen der Kfz-Fehlerdiagnose zu bewältigen, werden Fehlerdiagnosekompetenzen benötigt. Unter der Kfz-Fehlerdiagnosekompetenz in elektrotechnischen Kfz-Systemen verstehen wir die psychische Voraussetzung dafür, dass eine Person die Ursache für ein fehlerhaftes elektrotechnisches Kfz-System eindeutig identifizieren kann. In Anlehnung an Weinert (2002, S. 27) stellt Wissen eine zentrale Dimension der Kfz-Fehlerdiagnosekompetenz dar, eine Hypothese, die empirisch gut belegt ist (Abele, 2014, Kap. 3 und 4). Es ist anzunehmen, dass sowohl *Fehlerdiagnose-Sachwissen* (z. B. Wissen, wie elektrotechnische Komponenten in einem Stromlaufplan dargestellt werden) als auch *Fehlerdiagnose-Handlungswissen* (z. B. Wissen, wie die Funktion einer elektrotechnischen Komponente zu prüfen ist) relevant sind (Süß, 1996, S. 64 ff.). Die Kfz-Fehlerdiagnosekompetenz wird als differentielles Konstrukt aufgefasst, dessen Vermittlung curricular im 2. und 3. Ausbildungsjahr vorgesehen ist.

## 4 Computergestützte Erfassung der Kfz-Fehlerdiagnosekompetenz

Zur Erfassung der Kfz-Fehlerdiagnosekompetenz wurde auf Fehlerdiagnose-Szenarien und eine Computersimulation zurückgegriffen, in der die Berufsumwelt von Kfz-Mechatronikern ausschnitthaft und realitätsgetreu abgebildet ist und mit der authentisch interagiert werden kann.

## 4.1 Kfz-Computersimulation



**Abbildung 1:** Screenshots aus der Kfz-Computersimulation (oben links: Fahrzeugübersicht; oben rechts: Motorraumübersicht; unten links: Signalmessung mit dem Oszilloskop; unten rechts: Stromlaufplan aus dem computerbasierten Expertensystem)

Die linke obere Ecke von Abbildung 1 zeigt die Einstiegsseite der Kfz-Computersimulation.

In dieser Ansicht können die Problembeschreibung sowie verschiedene Fahrzeugsysteme per Mausklick aktiviert werden; im vorliegenden Beitrag sind nur die Systeme *Motormanagement* und *Beleuchtung* relevant. Zur Fehlerdiagnose können systemübergreifend Sichtprüfungen (z. B. zur Prüfung, ob das Abblendlicht funktioniert) und elektrotechnische Messungen durchgeführt werden. In der Computersimulation sind alle relevanten Ausschnitte der betreffenden Kfz-Systeme sowie Arbeitsmittel eines Kfz-Mechatronikers abgebildet, die für die Bearbeitung der hier einbezogenen Fehlerdiagnose-Szenarien benötigt werden. Damit können die Probanden analog zur Fehlerdiagnose in einer Kfz-Werkstatt vorgehen. Die Interaktion mit der Simulation sei beispielhaft anhand einer elektrotechnischen Messung und des Motormanagements veranschaulicht.

In Abbildung 1 ist rechts oben der Motorraum zu sehen, der zur Fehlerdiagnose im Motormanagement benötigt wird und in dem über die obere Menüleiste verschiedene Arbeitsmittel verfügbar sind (z. B. die Problembeschreibung oder der Sicherungskasten des Fahrzeuginnenraums). Um Messungen durchzuführen, ist das entsprechende Symbol des benötigten Messinstruments (Multimeter, Oszilloskop oder

Strommesszange) auszuwählen. In der Abbildung unten links wurde das Oszilloskop ausgewählt und das Signal des Ladedruckmagnetventils gemessen. Dafür mussten – wie bei der «realen» Kfz-Fehlerdiagnose – die entsprechende Steckverbindung geöffnet, das Adapterkabel angeschlossen und, zum Starten des Fahrzeugs, die Instrumentenanzeige aktiviert werden. Allein im *Motormanagement* sind außer den 10 Elementen der Menüleiste 13 Kfz-Komponenten und mehrere Hundert Messstellen verfügbar sowie mehrere Tausend distinkte Messungen möglich. Zusätzlich zu den Kfz-Systemen wurden lösungsrelevante und nicht lösungsrelevante Pfade eines computerbasierten Expertensystems repräsentiert, das ein Standardwerkzeug der Fehlerdiagnose in Kfz-Werkstätten darstellt. Die untere rechte Ecke der Abbildung zeigt einen Stromlaufplan, der über das Expertensystem aufgerufen wurde. Ausführlichere Informationen zur Computersimulation geben Gschwendtner, Abele und Nickolaus (2009).

## 4.2 Befunde zur Reliabilität

Die Reliabilität, d. h. interne Konsistenz der computergestützten Messung der Kfz-Fehlerdiagnosekompetenz wurde in zwei Studien untersucht: Abele et al. (2012) ermittelten eine EAP/PV-Reliabilität von .55, Nickolaus, Geißel, Abele und Nitzschke (2011) erzielten eine nur geringfügig höhere EAP/PV-Reliabilität von .62. Ein wichtiger Grund für die unbefriedigende interne Konsistenz liegt darin, dass die Bearbeitung eines ökologisch validen Fehlerdiagnose-Szenarios sehr zeitintensiv ist. Gschwendtner et al. (2009) legten aufgrund von Experteneinschätzungen eine Bearbeitungszeit von 30 min je Szenario fest (s. dazu auch Hochholdinger, Schaper & Sonntag, 2008, S. 153). Damit konnten trotz Testzeiten von vier Unterrichtsstunden und einer Instruktionszeit von ca. 40 min nur vier Szenarien eingesetzt werden. Das beschriebene Problem scheint keine Besonderheit der Domäne Kfz-Mechatronik zu sein: Auch in anderen Domänen gelang bislang entweder ebenfalls keine hinreichend reliable Messung der Fehlerdiagnosekompetenz (Kester, Kirschner & Merrienboer, 2006), oder es wurden nur wenige Szenarien eingesetzt und das damit einhergehende Reliabilitätsproblem nicht thematisiert (z. B. Hochholdinger & Schaper, 2013; Schaper, Hochholdinger & Sonntag, 2004; Schaafstal et al., 2000).

## 4.3 Befunde zur Validität

Zur *faktoriellen Validität* (Hartig, Frey & Jude, 2008, S. 154) von Fehlerdiagnose-Szenarios des Motormanagements und der Beleuchtungsanlage liegen ebenfalls Befunde vor. Die Befunde legen nahe, dass sich die Szenarios systemübergreifend durch ein einfaktorielles Modell ( $\lambda = .52\text{--}.78$ , Passung des Einfaktormodells:  $\chi^2/df = 1.2$ ,  $RMSEA = .05$ ,  $CFI = .96$ , Abele, 2014, S. 189) darstellen lassen. Theoretisch lässt sich dies auf die gemeinsamen physikalischen Funktions- und Wirkungsprinzipien zurückführen, die elektrotechnischen Kfz-Systemen zugrunde liegen und ähnliche Anforderungen bei der Fehlerdiagnose implizieren (s. Abschnitt 3.1). Im Anschluss an diese Erkenntnislage wird die Kfz-Fehlerdiagnosekompetenz in elektrotechnischen Kfz-Systemen als *relativ homogenes Konstrukt* aufgefasst.

Angesichts der geringen Reliabilität der computergestützten Kompetenzmessung wurde die *Konstruktvalidität*, wie von Hartig et al. (2008) empfohlen, messfehlerbereinigt untersucht. Gschwendtner et al. (2009) zeigten anhand eines Within-subjects-Design und zweier Stichproben ( $n_1 = 134$  und  $n_2 = 123$ ), dass mit der Computersimulation und in realen Arbeitskontexten, d. h. in der Werkstatt an Fahrzeugen erfasste Fehlerdiagnosekompetenzmaße *latent* sehr hoch korrelieren ( $r_1 = .94$  und  $r_2 = .94$ ). Wie Nickolaus, Abele und Gschwendtner (2012) verdeutlichen, lassen sich diese mit vier «realen» und vier «simulierten» Fehlerdiagnose-Szenarien erhobenen Maße als Indikatoren *eines* latenten Faktors interpretieren (Passung des Einfaktormodells:  $\chi^2 = 13.46$ ;  $df = 14$ ;  $\chi^2/df = .96$ ,  $RMSEA = .00$ ,  $CFI = 1.00$ ). Bei Abele (2014, S. 189 f.) korrelierten mit der Simulation und in realen Arbeitskontexten erhobene Kompetenzmaße *latent* ebenfalls sehr hoch ( $r = .90$ , manifest:  $r = .62$ ) und erwiesen sich ebenfalls als konvergent valide und das, obwohl im Unterschied zur ersten Studie größtenteils andere Szenarien eingesetzt wurden. Diese *messfehlerbereinigten* Ergebnisse legen nahe, dass mit der Computersimulation die in realen Arbeitskontexten benötigte Fehlerdiagnosekompetenz in elektrotechnischen Kfz-Systemen weitgehend valide erfasst werden kann.

## 5 Ansatz zur zeitökonomischen Überwindung des Reliabilitätsproblems

In den zitierten Validierungsstudien konnte das oben erwähnte Reliabilitätsproblem durch latente Modellierungen umgangen werden. Bei einer Individualdiagnostik und in Large-Scale-Assessments, in denen auf Basis von *vier* Szenarien geschätzte Kompetenzwerte zu ungenau sind (Walter & Rost, 2011, S. 134ff.), kann ein solcher Ansatz nicht weiterhelfen.

### 5.1 Diskussion möglicher Lösungsansätze

In solchen Fällen wäre eine Reliabilitätssteigerung durch *adaptives Testen* prinzipiell möglich. Allerdings sind die beim adaptiven Testen geforderten Stichprobengrößen (Walter, 2011, S. 396) und Itemumfänge (Lang, 2011, S. 417) bei der computergestützten Erfassung der Kfz-Fehlerdiagnosekompetenz – zumindest gegenwärtig – nicht realisierbar.

In Anlehnung an den *MicroDYN-Ansatz* aus der domänenübergreifenden komplexen Problemlöseforschung (Greiff, 2012, S. 77 ff.) wäre denkbar, weniger komplexe Fehlerdiagnose-Szenarien als bisher zu entwickeln, was kürzere Bearbeitungszeiten impliziert. Mit diesem Ansatz kann das Reliabilitätsproblem zwar abgemildert, aber nicht gelöst werden: Die Bearbeitungszeiten lassen sich nur begrenzt reduzieren (s. dazu Abschnitt 7.2), sofern die Szenarien ökologisch valide sein sollen, was in der beruflichen Bildung eine zentrale Forderung darstellt. Ökologische Validität bedeutet hier, dass sich die Szenarien auf authentische elektrotechnische Kfz-Systeme beziehen müssen, deren Komplexität hoch und unveränderlich, d. h. durch die Systeme vorgege-

ben ist. Da es also kaum gelingen dürfte, geeignete Szenarien mit kurzen Bearbeitungszeiten zu entwickeln, scheidet eine *simple Testverlängerung*, d. h. eine Erhöhung der Szenarienzahl aus.

## 5.2 Computergestützte Erfassung leistungskritischer Teilkompetenzen der Kfz-Fehlerdiagnose

Vor diesem Hintergrund wählten wir hier den Weg, zusätzlich zu den Fehlerdiagnose-Szenarien *Teilkompetenz (TK)-Items* vorzugeben, die sich in relativ kurzen Zeiten bearbeiten lassen und sich auf leistungskritische Teilanforderungen der Kfz-Fehlerdiagnose beziehen. Wir gehen davon aus, dass wir mit solchen *leistungskritischen TK-Items* die interindividuellen Kompetenzunterschiede der Kfz-Fehlerdiagnose genauso gut erfassen können wie mit den Fehlerdiagnose-Szenarien. Falls diese Annahme trägt, ist aus einer empirischen Sicht die eindimensionale Modellierung der TK-Items und Szenarien gerechtfertigt. Bei Berücksichtigung der TK-Items und der Szenarien sollte die Reliabilität der Diagnostik, so die Annahme, deutlich höher sein als bei exklusiver Berücksichtigung der Szenarien. Es sei betont, dass unser Ziel *nicht* darin besteht, mit den TK-Items alle Anforderungen der Kfz-Fehlerdiagnose abzudecken, sondern darin, Kompetenzunterschiede reliabel zu erfassen.

In einer früheren Studie wurden die TK-Items mit einem schriftlichen Test dargeboten. Dabei zeigte sich, dass die Scores der TK-Items und der Fehlerdiagnose-Szenarien zwar beachtlich korrelieren ( $r = .63$ , latent), letztlich aber wohl empirisch trennbare Kompetenzdimensionen abbilden (Abele, 2014, S. 194). Theoretisch führten wir dies u. a. darauf zurück, dass mit der Computersimulation im Unterschied zum schriftlichen Test auch implizites Wissen erfasst wird (s. dazu auch die Befunde von Renkl, Gruber, Mandl & Hinkofer, 1994). Wir nehmen an, dass eine eindimensionale Modellierung empirisch gerechtfertigt ist, wenn auch die TK-Items computergestützt, d. h. in Interaktion mit der Kfz-Computersimulation zu bearbeiten sind.

Insgesamt wurden 7 TK-Items entwickelt. Um eine möglichst große strukturelle Anforderungsschnittmenge zwischen den TK-Items und den Szenarien zu gewährleisten, decken fünf TK-Items gleichzeitig die Anforderungen *Informationsgewinnung* (IG), die *Entwicklung einer Prüfstrategie* (EP) und die *Umsetzung dieser Prüfstrategie* (UP) ab, zwei Items beschränken sich auf IG (s. Tab. 1). Die IG-Items wurden entwickelt, um die Differenzierungsqualität im unteren Kompetenzbereich zu erhöhen. Im angesprochenen Niveaumodell erreichten 17% der Probanden nicht einmal das erste Niveau (Nickolaus, Abele, Gschwendtner, Nitzschke & Greiff, 2012, S. 252). Um eine strenge Dimensionalitätsprüfung zu gewährleisten, beziehen sich die TK-Items auf andere Systemkomponenten als die einbezogenen Fehlerdiagnose-Szenarien (vgl. Tab. 1). Damit lassen sich Lerneffekte und Autokorrelationen, d. h. Korrelationen zwischen den Szenarien und TK-Items infolge identischer Anforderungen ausschließen.

Abschließend sei das Item TK5 erläutert, dessen Aufgabenstellung folgendermaßen lautet: *Bitte prüfen Sie die Signalleitung des Nockenwellen-Positionssensors (NWP), die vom Stecker des Sensors zum Motorsteuergerät führt.* Um TK5 zu bearbeiten, sind unter Nutzung des computerbasierten Expertensystems Einbaulage- und Stromlaufpläne

zu studieren, relevante Prüfstellen zu identifizieren und im Rückgriff auf spezifisches Vorwissen die Sensorfunktion und seine Interaktion mit dem Motorsteuergerät zu erfassen (IG). Daran anknüpfend ist für die Funktionsprüfung der Signalleitung eine geeignete Prüfstrategie zu entwickeln (EP). Abschließend ist diese Strategie umzusetzen, wobei das Motorsteuergerät einzubeziehen und das geeignete Messgerät auszuwählen sowie korrekt an den richtigen Stellen anzulegen ist, wofür angesichts vieler und nahe beieinander liegender Steckverbindungen und Messstellen mehrere erfolgskritische Entscheidungen zu treffen sind (UP).

## 6 Fragestellung und Hypothesen

Die entscheidende Frage ist nun, ob mit dem entwickelten Ansatz tatsächlich eine Reliabilitätsverbesserung bei der Erfassung der Kfz-Fehlerdiagnosekompetenz erreicht wird. Wie bereits erwähnt, nahmen wir an, dass (1) die TK-Items und die Fehlerdiagnose-Szenarien eindimensional skalierbar sind (*Hypothese zur Eindimensionalität*) und dass (2) die eindimensionale Modellierung der TK-Items und der Szenarien zeitökonomisch zu einer deutlich höheren Reliabilität führt als die alleinige Berücksichtigung der Szenarien (*Hypothese zum Reliabilitätswachst*).

## 7 Methode

### 7.1 Stichprobe<sup>52</sup>

Zur Hypothesenprüfung wurde eine Stichprobe von 275 Kfz-Mechatronikern herangezogen, die Ihre Ausbildung 2010 in baden-württembergischen Handwerksbetrieben aufnahmen und sich zum Erhebungszeitpunkt im 3. Ausbildungsjahr befanden. Für die Rekrutierung wurden landesweit Berufsschulen kontaktiert. Letztlich nahmen 13 Berufsschulen und 16 Berufsschulklassen aus ganz Baden-Württemberg an der Studie teil; die Teilnahmequote in den Klassen lag insgesamt bei 82.7%.

Das Durchschnittsalter der Stichprobe betrug 19.73 Jahre ( $SD = 1.81$ ). Ein Großteil der Auszubildenden war männlich (97.4%, Population: 97.6%, Bundesinstitut für Berufsbildung, 2010) und hatte die deutsche Staatsbürgerschaft (85.4%, Population: 88.2%; türkische: 8.6%; italienische: 3.3%; andere: 2.6%). Ungefähr die Hälfte wies einen Hauptschulabschluss auf (49.7%, Population: 49.4%), 46.3% hatten einen Realschulabschluss (Population: 43.9%) und nur wenige eine Hochschulzugangsberechtigung (4.0%, Population: 4.5%).

---

52 Die folgenden Analysen wurden durch die Forschungsinitiative ASCOT ([www.ascot-vet.net](http://www.ascot-vet.net)) des BMBF ermöglicht.

## 7.2 Diagnostische Maße

**Tabelle 1:** Erläuterungen und deskriptive Befunde zu den Fehlerdiagnose-Szenarien und den Teilkompetenz-Items

Code	Elektrotechnisches Kfz-System/ System-komponente	Anforderungen	Score	<i>P</i>	<i>r</i> <sub>it</sub>	<i>n</i>
<i>Kfz-Fehlerdiagnose-Szenarien</i>						
FDS1	Elektronisches Motormanagement/ Drehzahlgeber	Identifikation der Fehler- ursache	0	50.4	.66	230
			1	34.4		
			2	15.2		
FDS2	Elektronisches Motormanagement/ Kraftstofftemperatursensor	Identifikation der Fehler- ursache	0	74.9	.61	227
			1	25.1		
FDS3	Elektronisches Motormanagement/ Ladedruckmagnetventil	Identifikation der Fehler- ursache	0	55.0	.66	202
			1	36.6		
			2	8.4		
FDS4	Elektronisches Motormanagement/ Kühlmittel-Glühkerzen-Relais	Identifikation der Fehler- ursache	0	58.6	.71	198
			1	26.3		
			2	15.2		
FDS5	Beleuchtungsanlage/ Abblendlicht I	Identifikation der Fehler- ursache	0	21.5	.42	232
			1	78.5		
FDS6	Beleuchtungsanlage/ Abblendlicht II	Identifikation der Fehler- ursache	0	59.0	.55	200
			1	31.5		
			2	9.5		
<i>Teilkompetenz-Items der Kfz-Fehlerdiagnose</i>						
TK1	Elektronisches Motormanagement/ Motorsteuergerät (MSG)	IG: Lokalisierung MSG mit Einbaulageplan	0	17.0	.38	258
			1	83.0		
TK2	Elektronisches Motormanagement/ Motorsteuergerät (MSG)	IG: Pinbestimmung MSG mit Stromlaufplan	0	55.5	.57	256
			1	44.5		
TK3	Elektronisches Motormanagement/ Luftmassenmesser (LMM)	IG: Lokalisierung & Pin- bestimmung LMM EP: Prüfstrategie Versor- gungsspannung UP: Messung der Versor- gungsspannung	0	62.4	.48	258
			1	37.6		

(Fortsetzung Tabelle 1)

Code	Elektrotechnisches Kfz-System/ System-komponente	Anforderungen	Score	P	$r_{it}$	n
TK4	Elektronisches Motormanagement/ Abgasrückführungs-Magnetventil (AGR)	IG: Lokalisierung & Pin- bestimmung AGR	0	55.9	.58	254
		EP: Prüfstrategie Tastver- hältnis	1	44.1		
		UP: Messung des Tastver- hältnis				
TK5	Elektronisches Motormanagement/ Nockenwellenpositionssensor (NWP)	IG: Lokalisierung & Pin- bestimmung NWP	0	63.7	.68	245
		EP: Prüfstrategie Signal- leitungsfunktion	1	36.6		
		UP: Widerstandsmes- sung an Signalleitung				
TK6	Elektronisches Motormanagement/ Kühlmittel-Glühkerzenrelais (KGR)	IG: Lokalisierung & Pin- bestimmung KGR	0	40.4	.60	245
		EP: Prüfstrategie Spulen- funktion	1	59.6		
		UP: Widerstandsmes- sung an Spule				
TK7	Beleuchtungsanlage/ Standlicht (SL)	IG: Lokalisierung & Pin- bestimmung SL	0	24.0	.65	254
		EP: Prüfstrategie SL-Funk- tion	1	41.3		
		UP: Spannungsmessung an SL	2	34.7		

*Anmerkungen:* FDS: Fehlerdiagnose-Szenario, TK: Teilkompetenz-Item; IG: Informationsgewinnung, EP: Entwicklung einer Prüfstrategie, UP: Umsetzung der Prüfstrategie; P: scorespezifischer Probandenanteil;  $r_{it}$ : Trennschärfen, für die Fehlerdiagnose-Szenarien und die Teilkompetenz-Items jeweils separat berechnet; n: Stichprobengröße.

Zur Kompetenzmessung wurden sechs Fehlerdiagnose-Szenarien eingesetzt, die in Interaktion mit der beschriebenen Kfz-Computersimulation zu bearbeiten waren. Wie aus Tabelle 1 hervorgeht, beziehen sich vier dieser Szenarien auf das Motormanagement und zwei auf die Beleuchtungsanlage. Die Anforderungen dieser Szenarien wurden im Abschnitt 3.1 beispielhaft herausgearbeitet, wobei auf FDS3 zurückgegriffen wurde. Im Vergleich zu früheren Studien, in denen die Bearbeitungszeit je Szenario 30 min betrug, wurde die Bearbeitungszeit hier auf Basis von Experteneinschätzungen und Pilotierungen etwas reduziert (FDS2, FDS3 und FDS6: 25 min; FDS4: 20 min; FDS1 und FDS5: 15 min). Es wurde allerdings deutlich, dass dieser Zeiteinsparung bei Wahrung der ökologischen Validität klare Grenzen gesetzt sind. Während der Erhebung waren die Auszubildenden aufgefordert, ihre Fehlerdiagnoseschritte präzise handschriftlich zu notieren und die Fehlerursache konkret zu benennen. Das zwei- bzw. dreistufige manuelle Scoring (s. Tab. 1) erfolgte auf Basis eines Leitfadens, der

von drei Experten der Kfz-Fehlerdiagnose (zwei Ausbildungs-/Werkstattmeister, ein wissenschaftlicher Mitarbeiter) ausgearbeitet wurde. Den höchsten Score erhielten Probanden, die die korrekte Fehlerursache eindeutig nannten, Zwischenscores wurden vergeben, wenn bestimmte diagnoserelevante und leicht identifizierbare Teilschritte (z. B. Dokumentation einer wichtigen Messung und Angabe des korrekten Messwerts) eindeutig dokumentiert wurden.

Zur Erfassung der Teilkompetenzen wurden die in Abschnitt 5.2 erwähnten sieben Items eingesetzt (s. Tab. 1), die sich auf die Informationsgewinnung, die Entwicklung einer Prüfstrategie und die Umsetzung dieser Strategie beziehen, ebenfalls in Interaktion mit der Computersimulation und *zusammen* in 30 min zu bearbeiten waren. Bei sechs dieser Items wurde ein zwei-, bei einem (TK7) ein dreistufiges Scoring vorgenommen. Auch hier erfolgten die Dokumentation handschriftlich und das Scoring leitfadengestützt und auch hier wurde abhängig von eindeutig bestimmbareren Diagnoseschritten kodiert. Für TK5, d. h. die Funktionsprüfung der Signalleitung des NWP (s. Abschnitt 5.2) bedeutet dies z. B., dass der höchste Score nur vergeben wurde, wenn der korrekte Messwert (ein Wert, der nur an der betreffenden Systemstelle zu messen ist) genannt wurde. Zufällig richtige Antworten können wir aufgrund der sehr hohen Komplexität der Simulation (weitgehend) ausschließen (Goldhammer, Kröhne, Keßel, Senkbeil & Ihme, 2014).

### 7.3 Erhebungsdesign und Erhebungsdurchführung

Wie in Vorgängerstudien (z. B. Abele et al., 2012), schien eine gesamte Testzeit von *vier Schulstunden* angemessen. Dies erwies sich für die Berufsschulen, in deren Computerräumen die Testung stattfand, als (noch) akzeptabel. Um alle sechs Szenarien einsetzen zu können und Positionseffekte zu kontrollieren, wurde ein *youden square design* (Frey, Hartig & Rupp, 2009) verwendet: Jeder Proband bearbeitete fünf der sechs Szenarien, die vollständig permutiert wurden. Es kamen sechs verschiedene Testhefte zum Einsatz. Im Anschluss an die Erfassung der Kfz-Fehlerdiagnosekompetenz wurden die TK-Items dargeboten. Um Positionseffekte zu kontrollieren, erhoben wir hier mit einem *latin square design* (ebd.): Jeder Proband bearbeitete alle TK-Items, wobei die einzelnen Cluster vollständig permutiert wurden. Die gesamte Bearbeitungszeit für die fünf Szenarien und die sieben TK-Items betrug ca. 135 min.

Aus Standardisierungsgründen erfolgte die Einführung in die Computersimulation mit einem 10-minütigen Video, das jeder Proband mit Kopfhörer an seinem Monitor anschaute. Anschließend mussten die Probanden ein Übungsblatt bearbeiten, das sich auf zentrale Funktionen der Simulation bezog (z. B. auf die Aktivierung von Systemkomponenten) und von einem der beiden Testleiter ausgewertet wurde. In seltenen Fällen falsch bearbeiteter Aufgaben wurden die betreffenden Simulationsfunktionen nochmals im Klassenverbund anhand der an die Wand projizierten Simulation erläutert. Insofern kann unterstellt werden, dass letztlich keine entscheidenden *nutzungsbedingten* Leistungsdifferenzen mehr vorlagen. Die Einführung dauerte ca. 40 min, woraus eine Gesamterhebungszeit von ca. 175 min resultierte.

## 7.4 Hypothesenprüfung

Um die Hypothese zur Eindimensionalität (H1) zu prüfen, wurden konfirmatorische Faktorenanalysen für kategoriale Indikatoren durchgeführt, bei denen Modellparameter auf Basis der empirischen polychorischen bzw., bei zwei dichotomen Indikatoren, tetrachorischen Korrelationen geschätzt werden (Wang & Wang, 2012, S. 68). Bei der Spezifikation der Modelle wurde die Faktorvarianz jeweils auf 1 fixiert. Zur Schätzung der Modellparameter wurde das WLSMV-Verfahren verwendet (Eid, Gollwitzer & Schmitt, 2010, S. 874) und auf die Software Mplus (Muthén & Muthén, 2008) zurückgegriffen.

Vor der Hypothesenprüfung wurden für die Szenarien und die TK-Items separat der Probandenanteil je Scorestufe ( $P$ ) sowie die Trennschärfen ( $rit$ ) bestimmt. Dann wurde für die Szenarien und TK-Items jeweils eine konfirmatorische Faktorenanalyse durchgeführt und separat für beide Modelle (M1 und M2) die faktorielle Validität der Indikatoren beurteilt. Anschließend fand eine vergleichende Modellanalyse statt: Zuerst galt es, ein zweifaktorielles Modell (M3) zu analysieren, in dem die Szenarien Indikatoren eines Faktors und die TK-Items Indikatoren des anderen Faktors waren und die Faktorkorrelation frei geschätzt wurde. Anschließend wurde unter Einbezug der Szenarien *und* der TK-Items ein Einfaktormodell analysiert (M4). Dieses Modell ist ein Spezialfall eines Zweifaktormodells mit einer auf  $r = 1.0$  fixierten Faktorkorrelation. Von daher stellen M3 und M4 verschachtelte Modelle dar, weshalb der Passungsunterschied beider Modelle inferenzstatistisch geprüft werden kann (Eid et al., 2010, S. 882). Falls H1 zutrifft, sollte zwischen beiden Modellen kein bedeutsamer Passungsunterschied feststellbar sein und M4 eine ähnlich gute Modellpassung aufweisen wie M3. Die Beurteilung der Modellpassung erfolgte auf Basis der folgenden bei Moosbrugger und Schermelleh-Engel (2008, S. 319) dargestellten Fit-Werte:  $\chi^2/df \leq 3$ ,  $RMSEA \leq .08$ ,  $CFI \geq .95$ . Zudem wurde der WRMR herangezogen, der kleiner 1 sein sollte (Wang & Wang, 2012, S. 70). Ob zwischen M3 und M4 bedeutsame Unterschiede bestehen, wurde mit einem  $\chi^2$ -Differenztest inferenzstatistisch abgesichert, wobei ein Signifikanzniveau von  $\alpha = .01$  herangezogen wurde. Hierbei kamen Satorra-Bentler-korrigierte  $\chi^2$ -Werte (Eid et al., 2010, S. 883) und die in Mplus verfügbare Option «DIFF-Test» zum Einsatz. Falls zwischen M3 und M4 kein bedeutsamer Passungsunterschied feststellbar ist, ist ein wichtiger diagnostischer Fortschritt erzielt. Da es angesichts der hierarchischen Datenstruktur bei der Parameterschätzung zu einer Unterschätzung der Standardfehler kommen kann (Nagy, 2005, S. 114), werden bei allen Schätzungen mit der Option «TYPE = COMPLEX» und der Klassenzugehörigkeit als Clustermerkmal die Standardfehler korrigiert.

**Tabelle 2:** Fit-Werte der verschiedenen Modelle und Ergebnis des Modellvergleichs

Modelle	<i>n</i>	$\chi^2$	<i>df</i>	$\chi^2/df$	RMSEA (KI)	CFI	WRMR
M1	271	25.02	9	2.78	.08 (.04–.12)	.89	.94
M2	272	21.83	14	1.56	.05 (.00–.08)	.94	.79
M3	275	79.22	64	1.24	.03 (.00–.05)	.95	.90
M4	275	81.25	65	1.25	.03 (.00–.05)	.94	.91

$\chi^2$ -Differenztest: M3 und M4:  $\Delta\chi^2 (df=1) = 4.221, p = .04$

Anmerkungen: KI: 90 %-Konfidenzintervall; 1da beim Modellvergleich Satorra-Bentler-korrigierte  $\chi^2$ -Werte verwendet wurden, ergibt sich  $\Delta\chi^2$  nicht direkt aus den angegebenen  $\chi^2$ -Werten von M3 und M4.

Um H2 zu prüfen, wurde zuerst jeweils die Reliabilität für M1 (Szenarien) und M2 (TK-Items) bestimmt. Dann wurde die Reliabilität für M4 (Szenarien und TK-Items) berechnet, die im Sinne von H2 deutlich höher ausfallen sollte als die von M1. Die bisherigen Befunde legen nahe, dass es sich bei der Kfz-Fehlerdiagnosekompetenz um ein relativ homogenes Konstrukt handelt (s. Abschnitt 4.3), weshalb wir zur Reliabilitätsbestimmung Konsistenzanalysen durchführten. Da wir nur über eine begrenzte Indikatoranzahl verfügen, ein unterrichtsrelevantes, d. h. zeitveränderliches Konstrukt untersuchen und aus organisatorischen Gründen keine zweite Testung durchführen konnten, schlossen wir andere Methoden der Reliabilitätsbestimmung (Schermelleh-Engel & Werner, 2008) aus. Obwohl zusätzliche Items nicht zwangsläufig zu besseren Reliabilitäten führen (Yousfi, 2005), entschieden wir uns für den vorgestellten Ansatz: Er versprach eine praktikable Lösung des Reliabilitätsproblems und ermöglichte eine empirische Prüfung der Reliabilitätsannahme. Die Reliabilitäten (SEM-Reliabilität) berechneten wir mit der für nicht-lineare Strukturgleichungsmodelle entwickelten Formel von Green und Yang (2009, Formel 21) und, da gegenwärtig kein einschlägiges R-Paket verfügbar ist, eigenständig mit der Statistik-Software R (R Core Team, 2014).<sup>53</sup>

## 8 Ergebnisse

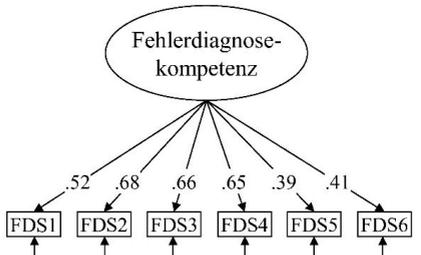
### 8.1 Deskriptive Befunde

Wie Tabelle 1 zu entnehmen ist, wiesen die Szenarien (FDS) gute Trennschärfen auf. Zudem waren weder Decken- noch Bodeneffekte zu beobachten: Das leichteste Szenario (FDS5) wurde von 78.5 % der Probanden gelöst, das schwierigste von 8.4 % (FDS3). Im unteren Tabellenbereich sind die unabhängig von den Szenarien berechneten Kennwerte der TK-Items zu sehen. Auch hier waren durchweg gute Trennschärfen und keine Decken- und Bodeneffekte zu beobachten. Damit erwiesen sich sowohl die Szenarien als auch die TK-Items als hinreichend trennscharf.

53 Stefan Behrendt gilt unser herzlicher Dank für die Unterstützung bei der Reliabilitätsberechnung.

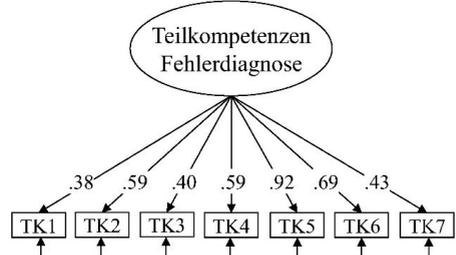
### 8.2 Prüfung der Hypothese zur Eindimensionalität

a) M1: Einfaktormodell Fehlerdiagnose-Szenarien



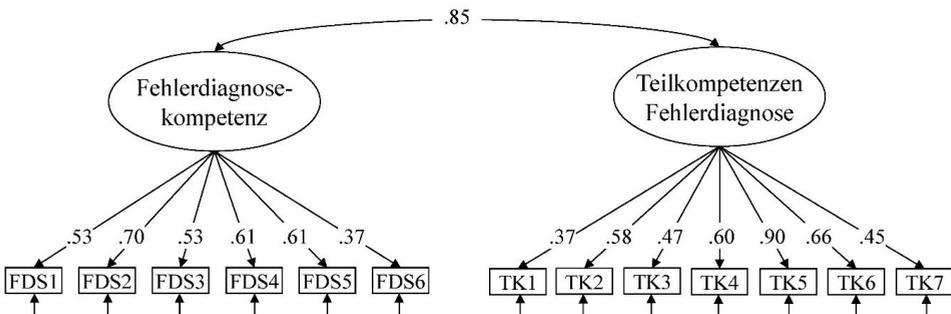
$\chi^2(df = 8, n = 271) = 18.23; \chi^2/df = 2.29, CFI = .90, RMSEA = .07, WRMR = .80; p(\lambda) < .001$

b) M2: Einfaktormodell Teilkompetenz-Items



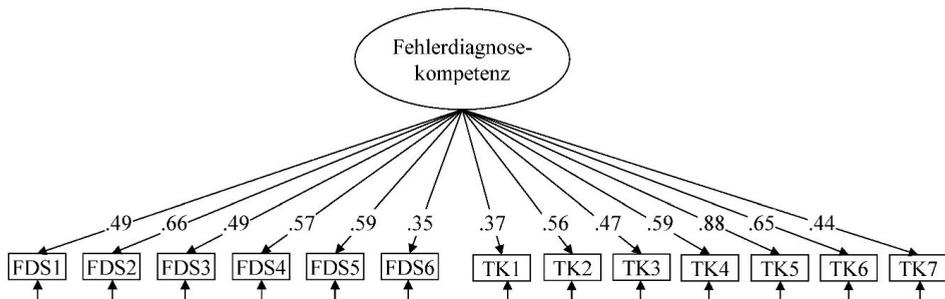
$\chi^2(df = 13, n = 272) = 19.77; \chi^2/df = 1.52, CFI = .97, RMSEA = .04, WRMR = .74; p(\lambda) < .001$

c) M3: Zweifaktormodell mit frei geschätzter Faktorkorrelation



$\chi^2(df = 45, n = 275) = 59.03; \chi^2/df = 1.31, CFI = .97, RMSEA = .03, WRMR = .82; p(\lambda, r) < .001$

d) M4: Eindimensionales Modell



$\chi^2(df = 45, n = 275) = 62.30; \chi^2/df = 1.38, CFI = .96, RMSEA = .04, WRMR = .85; p(\lambda) < .001$

**Abbildung 2:** Ergebnisse der konfirmatorischen Faktorenanalysen zur Prüfung der Eindimensionalitätshypothese (angegeben sind jeweils standardisierte Modellparameter)

Im oberen Teil von Abbildung 2 finden sich die Ergebnisse für M1 und M2, deren Ladungen mindestens befriedigende ( $\lambda \geq .32$ ) und durchweg signifikante ( $p < .001$ ) Werte erreichten. Mit Blick auf M1 fällt auf, dass die Ladungen von FDS5 und FDS6 etwas geringer ausfielen als die der anderen Szenarien. Ähnliches galt für die Indikatoren TK1 und TK7 bei M2. Die Anpassungsgüte von M2 war durchweg zufriedenstellend und bei M1 abgesehen vom *CFI* ebenfalls akzeptabel (s. Tab. 2). Da in der Literatur mitunter auch ein geringerer *CFI*-Grenzwert angesetzt wird (.90, Nagy, 2005, S. 114), wurde auch bei M1 insgesamt von einer annehmbaren Passung ausgegangen.

Mit Blick auf das Zweifaktormodell M3 (s. Abbildung 2 Mitte und Tabelle 2) waren durchweg mindestens akzeptable Ladungshöhen und gute Fit-Werte festzustellen. Die frei geschätzte Faktorkorrelation fiel sehr hoch aus.

Auch das eindimensionale M4 näherte die empirische Datenlage gut an. Sichtbar wurde hier aber auch, dass die Ladungen merklich variierten, wobei diese Variation unabhängig davon zu sein schien, ob Szenarien oder TK-Items betrachtet wurden. Wie ein Modellvergleich zeigte (s. Tabelle 2), waren die  $\chi^2$ -Differenzen nicht signifikant ( $p = .04$ ), wobei die Signifikanzgrenze nur knapp überschritten wurde. Auch weitere Befunde sprachen für das sparsamere Modell M4: Die Fit-Werte beider Modelle unterschieden sich nur unwesentlich. Die Prüfung von H1 abschließend sei erwähnt, dass die Ladungen korrespondierender Indikatoren – abgesehen von FDS3 und FDS5 – modellübergreifend relativ stabil waren.

### 8.3 Prüfung der Hypothese zum Reliabilitätszuwachs

Im Hinblick auf die Reliabilität zeigte sich Folgendes: Bei M1 und M2 lagen die SEM-Reliabilitäten bei .62 bzw. .65. Erwartungskonform stieg die Reliabilität bei M4, d. h. bei einer simultanen Berücksichtigung der Szenarien und der TK-Items deutlich an, und zwar auf einen Wert von .75.

## 9 Diskussion

Das Ziel des vorliegenden Beitrags besteht darin, eine bislang ungelöste Herausforderung bei der Erfassung beruflicher Problemlösekompetenzen zu überwinden: Aufgrund langer Bearbeitungszeiten ökologisch valider beruflicher Problemlöse-Szenarien und begrenzter Testzeiten können bei der Diagnostik nur wenige Szenarien eingesetzt werden. Es wurde ein innovativer Ansatz vorgestellt, mit dem komplexe Problemlösekompetenzen in der Domäne Kfz-Mechatronik, d. h. Fehlerdiagnosekompetenzen in elektrotechnischen Kfz-Systemen mit Fehlerdiagnose-Szenarien *und* mit relativ schnell zu bearbeitenden leistungskritischen TK-Items diagnostiziert werden können. Angenommen wurde, dass die TK-Items und die Szenarien eindimensional modellierbar sind und zusammen zu einer reliableren Kompetenzdiagnostik führen als die Szenarien allein. Beide Hypothesen wurden von den Ergebnissen bestätigt: Die TK-Items und die Szenarien erwiesen sich als eindimensional modellierbar und mit

den Szenarien und TK-Items wurde zeitökonomisch eine deutlich höhere Reliabilität erreicht als mit den Szenarien allein.

Kritisch könnte eingewandt werden, dass die merklich unterhalb von  $r=1$  liegende Faktorkorrelation in M3 und der nahe an der Signifikanzgrenze liegende Modellvergleichstest von M3 und M4 die Eindimensionalitätsannahme in Frage stellen. Diesem Einwand ist in Folgestudien nachzugehen. Angesichts der akzeptablen und nur unwesentlich von M3 verschiedenen Passungsgüte von M4 spricht aktuell wohl mehr dafür, die Annahme zur Eindimensionalität beizubehalten.

Der entwickelte Ansatz weist einen Weg, wie die noch nicht ganz zufriedenstellende Reliabilität ohne Testzeiterhöhung verbessert werden kann: Die Szenarienzahl könnte weiter reduziert und die Zahl der TK-Items erhöht werden. Angesichts der Ergebnisse wäre es sogar legitimierbar, komplett auf Szenarien zu verzichten und dadurch die Testzeit bei gleichzeitiger Erhöhung der Reliabilität zu minimieren, wobei insbesondere zwei Punkte zu beachten sind: Erstens kann diese Art der Reliabilitäts-erhöhung Auswirkungen auf die Konstruktvalidität des Tests haben (Rost, 1996, S. 397). Zweitens nimmt die Testreliabilität nur zu, wenn die zusätzlichen TK-Items bestimmte Bedingungen erfüllen (z. B. unkorrelierte Fehlervarianzen, Yousfi, 2005). Die vergleichsweise geringen Ladungen von FDS5 und FDS6 in M1 könnten Anlass geben, für das Motormanagement und die Beleuchtungsanlage künftig separate Messmodelle zu formulieren. Möglicherweise fallen die Indikatorladungen bei einer system- oder anforderungsbezogenen Homogenisierung höher und auch konsistenter aus, was sich günstig auf die interne Konsistenz auswirken könnte (Green & Yang, 2009). Um dies zu prüfen, sind in Folgestudien systematisch Items nach entsprechenden inhaltlichen Kriterien zu konstruieren.

Angesichts des relativ heterogenen Ladungsmusters von M4 kann gefragt werden, ob ein Retest-Maß zur Reliabilitätsbestimmung nicht geeigneter gewesen wäre als eine Konsistenzanalyse (Lienert & Raatz, 1994, S. 202 f.). Da sich Kfz-Fehlerdiagnosekompetenzen unterrichts- und kommunikationsbedingt zwischen Testungen deutlich entwickeln dürften, brächten Retest-Maße aber vermutlich keine entscheidenden Vorteile (Schermmelleh-Engel & Werner, 2008, S. 116).

Vor dem Hintergrund gängiger Problemlösedefinitionen (s. Greiff et al., 2014) kann der Ansatz insofern kritisiert werden, als die TK-Items nicht alle Anforderungen der Kfz-Fehlerdiagnose abdecken, weshalb letztlich nicht Problemlöse-, sondern nur Teilkompetenzen erfasst werden. Diese theoretische Unterscheidung ist richtig und in pädagogischen Kontexten wichtig. Dort geht es letztlich um die Förderung von Problemlöse- und nicht von Teilkompetenzen. Deswegen ist uns wichtig zu betonen, dass wir mit unserem Ansatz ein diagnostisches Ziel verfolgen: Wir wollen interindividuelle Problemlöseunterschiede reliabel erfassen. Wir behaupten *nicht*, dass die TK-Items das übergeordnete Konstrukt inhaltlich vollständig abbilden. Es ist uns bewusst, dass unser Ansatz wohl v. a. deswegen trägt, weil fachliche Problemlösekompetenzen prozessbezogen aufgebaut werden. Zumindest scheint dies eine gute Erklärung dafür zu sein, dass von Teilkompetenzen auf Gesamtkompetenzen geschlossen werden kann. Trotz dieses kritischen Bewusstseins rechnen wir seitens (beruflicher) Bildungs-

praktiker mit gewissen Widerständen, die durch die Verwendung von TK-Items *und* Problemlöse-Szenarien vielleicht überwunden werden können. Damit weist der Ansatz eine für ein internationales Large-Scale-Assessment in der beruflichen Bildung wichtige Eigenschaft auf: Er wird unterschiedlichen praktischen Anforderungen gerecht und ermöglicht eine auch psychometrisch akzeptable Diagnostik beruflicher Problemlösekompetenzen.

Der vorgelegte Ansatz ist u. E. nicht auf den berufsbildenden Bereich beschränkt, sondern kann prinzipiell überall angewandt werden, wo domänenspezifische Problemlösekompetenzen erfasst werden sollen. Ob der Transfer tatsächlich gelingt, ist in weiteren Studien empirisch zu prüfen.

Obwohl der Ansatz primär für diagnostische Zwecke entwickelt wurde, kann er auch pädagogische Prozesse bereichern: TK-Items bieten die Möglichkeit, individuelle Problemlösebarrieren gezielt zu diagnostizieren und in Interventionen zu bearbeiten. Es wäre interessant, den Effekt einer Teilkompetenzdiagnostik und einer darauf abgestimmten Intervention auf die Entwicklung domänenspezifischer Problemlösekompetenzen zu untersuchen. In Domänen, die sich auf komplexe Systeme beziehen und viele unterschiedliche Anforderungen stellen, dürften Interventionen zu einem Kompetenzzuwachs, aber kaum zu einem Rückgang interindividueller Kompetenzunterschiede führen (Abele, 2014, S. 158). Insofern kann der vorgelegte, auf internen Konsistenzanalysen beruhende Ansatz zur Lösung des Reliabilitätsproblems auch für die Interventionsforschung nützlich sein.

## Literatur

- Abele, S., Greiff, S., Gschwendtner, T., Wüstenberg, S., Nickolaus, R. & Funke, J. (2012). Dynamische Problemlösekompetenz – ein bedeutsamer Prädiktor von Problemlöseleistungen in technischen Anforderungskontexten? *Zeitschrift für Erziehungswissenschaft*, 15, 363–391.
- Abele, S. (2014). *Modellierung und Entwicklung berufsfachlicher Kompetenz*. Stuttgart: Franz Steiner.
- Baethge, M., Achtenhagen, F., Arends, L., Babic, E., Baethge-Kinsky, V. & Weber, S. (2006). *Berufsbildungs-PISA. Machbarkeitsstudie*. Stuttgart: Franz Steiner.
- Bundesinstitut für Berufsbildung. (2010). *Berufsbildungsstatistik der statistischen Ämter des Bundes und der Länder (Erhebung zum 31.12)*, Bundesinstitut für Berufsbildung. Zugriff am 27.05.2014 von <http://www.bibb.de/de/781.htm>.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2010). *Statistik und Forschungsmethoden. Lehrbuch*. Weinheim: Beltz.
- Fischer, A., Greiff, S. & Funke, J. (2012). The process of solving complex problems. *Journal of Problem Solving*, 4, 19–24.
- 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, 39–53.

- Funke, J. (2006). Komplexes Problemlösen. In J. Funke (Hrsg.), *Denken und Problemlösen* (Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie II: Kognition, Bd. 8, S. 375–445). Göttingen: Hogrefe.
- Goldhammer, F., Kröhne, U., Keßel, Y., Senkbeil, M. & Ihme, J. M. (2014). Diagnostik von ICT-Literacy. Multiple-Choice- vs. simulationsbasierte Aufgaben. *Diagnostica*, 60, 10–21.
- Green, S. B. & Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika*, 74, 155–167.
- Greiff, S. (2012). *Individualdiagnostik komplexer Problemlösefähigkeit*. Münster: Waxmann.
- Greiff, S., Kretschmar, A. & Leutner, D. (2014). Problemlösen in der Pädagogischen Psychologie. *Zeitschrift für Pädagogische Psychologie*, 28, 161–166.
- Gschwendtner, T., Abele, S. & Nickolaus, R. (2009). Computersimulierte Arbeitsproben: Eine Validierungsstudie am Beispiel der Fehlerdiagnoseleistungen von Kfz-Mechatronikern. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 105, 557–578.
- Hartig, J., Frey, A. & Jude, N. (2008). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 136–192). Berlin: Springer.
- Hochholdinger, S., Schaper, N. & Sonntag, K. (2008). Ergebnisse der Evaluation betrieblicher E-Learning-Module. In J. Rowold, S. Hochholdinger & N. Schaper (Hrsg.), *Evaluation und Transfersicherung betrieblicher Trainings. Modelle, Methoden und Befunde* (Wirtschaftspsychologie, S. 146–164). Göttingen: Hogrefe.
- Hochholdinger, S. & Schaper, N. (2013). Training troubleshooting skills with an anchored instruction module in an authentic computer based simulation environment. *Journal of Technical Education*, 1, 7–22.
- Kester, L., Kirschner, P. A. & Merriënboer, J. J. (2006). Just-in-time information presentation: improving learning a troubleshooting skill. *Contemporary Educational Psychology*, 31, 167–185.
- Kontogiannis, T. & Moustakis, V. (2002). An experimental evaluation of comprehensibility aspects of knowledge structures derived through induction techniques: a case study of industrial fault diagnosis. *Behaviour & Information Technology*, 21, 117–135.
- Lang, J. W. (2011). Computer-adaptives Testen. In L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Leistungs-, Intelligenz- und Verhaltensdiagnostik* (Enzyklopädie der Psychologie, Serie Psychologische Diagnostik, Bd. 3, S. 403–446). Göttingen: Hogrefe.
- Leutner, D., Fleischer, J., Wirth, J., Greiff, S. & Funke, J. (2012). Analytische und dynamische Problemlösekompetenz im Lichte internationaler Schulleistungsvergleichsstudien. Untersuchungen zur Dimensionalität. *Psychologische Rundschau*, 63, 34–42.
- Lienert, G. A. & Raatz, U. (1994). *Testaufbau und Testanalyse* (5., völlig neu bearbeitete und erweiterte Aufl.). Weinheim: Beltz.
- Moosbrugger, H. & Schermelleh-Engel, K. (2008). Exploratorische (EFA) und konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 305–324). Berlin: Springer.

- Müller, M. & Schelten, A. (2009). Comparative international analysis of occupational tasks and qualification requirements for the labour market and assessment tasks at the end of VET in participating countries. In Bundesministerium für Bildung und Forschung (Hrsg.), *Feasibility study VET-LSA. A comparative analysis of occupational profiles and VET programmes in 8 European countries – International report* (Vocational Training Research, Bd. 8, S. 33–47). Bonn: Herausgeber.
- Muthén, L. K. & Muthén, B. O. (2008). *Mplus* [Computer software]. Los Angeles, CA: Muthén & Muthén.
- Nagy, G. (2005). *Berufliche Interessen, kognitive und fachgebundene Kompetenzen: Ihre Bedeutung für die Studienfachwahl und die Bewährung im Studium*. Unveröffentlichte Dissertation, Freie Universität Berlin.
- Nickolaus, R., Abele, S. & Gschwendtner, T. (2012). Valide Kompetenzabschätzungen als eine notwendige Basis zur Effektbeurteilung beruflicher Bildungsmaßnahmen – Wege und Irrwege. In G. Niedermaier (Hrsg.), *Kompetenzen entwickeln, messen und bewerten* (Bd. 6, S. 537–554). Linz: Trauner.
- Nickolaus, R., Abele, S., Gschwendtner, T., Nitzschke, A. & Greiff, S. (2012). Fachspezifische Problemlösefähigkeit in gewerblich technischen Ausbildungsberufen – Modellierung, erreichte Niveaus und relevante Einflussfaktoren. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 108, 243–272.
- Nickolaus, R., Geißel, B., Abele, S. & Nitzschke, A. (2011). Fachkompetenzmodellierung und Fachkompetenzentwicklung bei Elektronikern für Energie- und Gebäudetechnik im Verlauf der Ausbildung. Ausgewählte Ergebnisse einer Längsschnittstudie. In R. Nickolaus & G. Pätzold (Hrsg.), *Lehr-Lernforschung in der gewerblich-technischen Berufsbildung. Zeitschrift für Berufs- und Wirtschaftspädagogik [Beiheft]*, 25, 77–94.
- Nickolaus, R. & Seeber, S. (2013). Berufliche Kompetenzen: Modellierungen und diagnostische Verfahren. In A. Frey, U. Lissmann & B. Schwarz (Hrsg.), *Handbuch Berufspädagogische Diagnostik* (S. 166–195). Weinheim: Beltz.
- OECD (2013). *PISA 2012 assessment and analytical framework mathematics, reading, science, problem solving and financial literacy*. Paris: OECD.
- Rauner, F., Hitz, H., Spöttl, G. & Becker, M. (2002). *Aufgabenanalysen für die Neuordnung der Berufe im Kfz-Sektor*. Zugriff am 15.05.2014 von <http://www.itb.uni-bremen.de/uploads/media/Kfz-Neuordnung-Bericht4.pdf>.
- R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Zugriff am 15.6.2014 von <http://www.R-project.org/>.
- Renkl, A., Gruber, H., Mandl, H. & Hinkofer, L. (1994). Hilft Wissen bei der Identifikation und Kontrolle eines komplexen ökonomischen Systems? *Unterrichtswissenschaft*, 22, 195–202.
- Ross, C. & Orr, R. (2009). Teaching structured troubleshooting: Integrating a standard methodology into an information technology program. *Educational Technology Research and Development*, 57, 251–265.
- Rost, J. (1996). *Lehrbuch Testtheorie – Testkonstruktion*. Bern: Huber.

- Rowold, J., Hochholdinger, S. & Schaper, N. (Hrsg.). (2008). *Evaluation und Transfersicherung betrieblicher Trainings. Modelle, Methoden und Befunde* (Wirtschaftspsychologie). Göttingen: Hogrefe.
- 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.
- Schaper, N. (1994). *Lernbedarfsanalysen und Trainingsgestaltung bei komplexen Diagnoseaufgaben*. Frankfurt am Main: Peter Lang.
- Schaper, N., Hochholdinger, S. & Sonntag, K. (2004). Förderung des Transfers von Diagnosestrategien durch computergestütztes Training mit kognitiver Modellierung. *Zeitschrift für Personalpsychologie*, 3, 51–62.
- Schermelleh-Engel, K. & Werner, C. (2008). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 114–133). Berlin: Springer.
- Sonntag, K. & Schaper, N. (Hrsg.). (1997). *Störungsmanagement und Diagnosekompetenz*. Zürich: vdf Hochschulverlag.
- Spöttl, G., Becker, M. & Musekamp, F. (2011). Anforderungen an Kfz-Mechatroniker und Implikationen für die Kompetenzerfassung. In R. Nickolaus & G. Pätzold (Hrsg.), *Lehr-Lernforschung in der gewerblich-technischen Berufsbildung. Zeitschrift für Berufs- und Wirtschaftspädagogik [Beiheft]*. 25, 37–53.
- Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen. Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen*. Göttingen: Hogrefe.
- Walter, O. B. (2011). Grundlagen des adaptiven Testens. In L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Leistungs-, Intelligenz- und Verhaltensdiagnostik* (Enzyklopädie der Psychologie, Serie Psychologische Diagnostik, Bd. 3, S. 389–402). Göttingen: Hogrefe.
- Walter, O. & Rost, J. (2011). Psychometrische Grundlagen von Large-Scale-Assessments. In L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Methoden der psychologischen Diagnostik* (Enzyklopädie der Psychologie, Serie Psychologische Diagnostik, Bd. 2, S. 87–149). Göttingen: Hogrefe.
- Wang, J. & Wang, X. (2012). *Structural equation modeling. Applications using Mplus*. Chichester: Wiley.
- Weinert, F. E. (2002). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessung in Schulen* (2. Aufl., S. 17–31). Weinheim: Beltz.
- Yousfi, S. (2005). Mythen und Paradoxien der klassischen Testtheorie (I). *Diagnostica*, 51, 1–11.



# Anhang F

## ***Publikation 6: Erfassung diagnostischer Problemlösekompetenzen mit schriftlichen Key-Feature-Problemen***

Abele, S. (2016). Can Diagnostic Problem-Solving Competences of Car Mechatronics Be Validly Assessed Using a Paper-Pencil Test? *Journal of Technical Education (JOTED)*, 4, 190–211.

### **Eigener Anteil an der Publikation und Begutachtungsverfahren**

Die Publikation resultierte aus einer Teilstudie des Projekts „KOKO Kfz“ der Förderlinie „Technologieorientierte Kompetenzmessung in der Berufsbildung (ASCOT)“ des Bundesministeriums für Bildung und Forschung (BMBF)<sup>54</sup>, an dessen Beantragung ich mitwirkte. In dieser Teilstudie hatte ich wissenschaftliche Leitungsfunktionen und war ich für die Entwicklung der papierbasierten Key-Feature-Probleme, das Studiendesign, die Durchführung sowie die Auswertung verantwortlich. Die Publikation habe ich konzipiert und verfasst, die einschlägigen statistischen Auswertungen stammen ebenfalls von mir. Publikationen im *Journal of Technical Education* durchlaufen ein dreifach-blindes Begutachtungsverfahren.

---

<sup>54</sup> An dem Teilprojekt wirkte ein größerer Kreis von Personen mit, die im Folgenden genannt seien und bei denen ich mich an dieser Stelle bedanke: Dipl.-Gwl. Horst Heinzmann, Sebastian Heilig MA, Dipl.-Gwl. Thomas Schmidt und Thomas Trezbiatowski. Zudem möchte ich den zahlreichen Experten und Schulen für ihre Unterstützung danken.

# Can diagnostic problem-solving competences of car mechatronics be validly assessed using a paper-pencil test?

STEPHAN ABELE

## Abstract

In this study, an authentic paper-based key-feature test for electrotechnical diagnostic problem-solving competence was developed, administered to car mechatronic apprentices ( $N = 206$ ) and validated using diagnostic problem-solving scores. It was hypothesized that the paper-based test scores represent the same construct as the problem-solving scores. The written test turned out to have a relatively high reliability (EAP/PV = .75). Additionally, it was found that the paper-based scores and problem-solving scores were highly correlated ( $r = .76$ ) but represented empirically distinct dimensions. Presumably, the written test especially covered diagnostic knowledge and failed to cover other relevant subcomponents of diagnostic problem-solving competence. It is argued that this unexpected finding might be caused by construct underrepresentation and construct-irrelevant variance of the paper-based key-feature test.

**Keywords:** Professional competence, domain-specific problem-solving, key-feature items, construct representation, construct-irrelevant variance

## Lassen sich diagnostische Problemlösekompetenzen von Kfz-Mechatronikern mit einem Papier-Bleistift-Test valide erfassen?

### Zusammenfassung

In der Studie wurde ein schriftlicher Key-Feature-Test zur Erfassung diagnostischer Problemlösekompetenzen entwickelt, bei einer Stichprobe von Auszubildenden der Kfz-Mechatronik ( $N = 206$ ) eingesetzt und anhand von Scores zum diagnostischen Problemlösen validiert. Es wurde unterstellt, dass die Papier-Bleistift-basierten Testscores dieselbe empirische Dimension abbilden wie die Problemlösescores. Der schriftliche Test erreichte eine relativ hohe Reliabilität (EAP/PV = .75). Gezeigt hat sich zudem, dass die Test- und Problemlösescores eng korrelierten ( $r = .76$ ), letztlich aber empirisch unterscheidbare Dimensionen repräsentierten. Der schriftliche Test erfasste vermutlich v. a. diagnostisches Wissen und „vernachlässigte“ weitere relevante Subkomponenten der diagnostischen Problemlösekompetenz. Dieses unerwartete Ergebnis lässt sich wohl darauf zurückführen, dass der schriftliche Key-Feature-Test das Zielkonstrukt nicht vollständig repräsentierte und zudem Konstrukt-irrelevante Varianz erzeugte.

**Schlüsselwörter:** Berufsfachliche Kompetenz, domänenspezifisches Problemlösen, Key-Feature-Items, Konstruktrepräsentation, Konstrukt-irrelevante Varianz

# 1 Introduction

This paper is about the assessment of diagnostic problem-solving competences of car mechatronics<sup>55</sup>. Diagnostic problem-solving represents an important type of problem-solving (cf. Jonassen 2000) and is relevant in several professional contexts: Physicians have to identify causes of diseases, teachers have to figure out reasons for learning difficulties, technicians and engineers have to diagnose technical defects, etc. According to Shavelson (2010, p. 45), diagnostic problem-solving competence can be defined as a complex ability necessary to master challenging and somehow novel encounters occurring in the workplace. Diagnostic problem-solving competence is a key aspect of professional competence, a relevant subdimension of professional problem-solving competence and an important topic of Vocational Education and Training (e. g., Baethge & Arends 2009). The assessment of diagnostic problem-solving competence is a precondition of the evaluation of corresponding training programs and evidence-based feedback for students.

According to the current state of research, it seems advisable to assess diagnostic problem-solving competences using computer-simulations representing (parts of) the real work environment (e. g., Norcini & McKinley 2007): Such simulations allow for a high degree of validity, standardization and, in comparison to work-based assessments, are much more convenient. Although previous studies on professional problem-solving have documented the immense potential of computer-simulation-based testing (e. g., Rausch et al. 2016; Walker, Link & Nickolaus 2016; Gschwendtner, Abele & Nickolaus 2009), they have also made clear that such assessments imply long development times, high investment costs and also, due to time-consuming introduction into the computer-simulated work environments, high testing times. In contrast, paper-and-pencil-based tests are relatively economical and there are some findings suggesting that such tests might be appropriate to validly measure professional problem-solving competences (e. g., Link & Geißel, 2015). The present paper investigates whether diagnostic problem-solving competences of car mechatronics can be validly assessed using a paper-and-pencil-based (paper-based) test.

## 1.1 Diagnostic problem-solving competence of car mechatronics

### 1.1.1 Problem domain: Electrotechnical diagnostic problems

Diagnostic problems of car mechatronics refer to situations in which the cause(s) of an auto defect (e. g., lighting system defect) has/have to be found. Such situations have the critical attributes of a problem (cf. 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). This definition of a diagnostic problem is in line with Schaafstal, Schraagen & van Berlo (2000, p. 75) but differs from other studies (e. g., Kassirer, Wong & Kopelman 2010, p. 6) since it does *not* cover repair or maintenance.

---

55 The field of car mechatronics covers, among other things, troubleshooting, repair and maintenance of cars (Baethge and Arends 2009).

Previous studies showed that professional problem-solving competence is domain-specific (e. g., van der Vleuten et al. 2010, p. 704). Without any further effort, high performance in a specific problem domain is not necessarily applicable to other problem domains (e. g., Schwartz & Elstein 2011, p. 225). There is evidence documenting the influence of complex (general) problem solving competence on professional competence being quite small (cf. Mainert et al. 2015), especially in comparison to the influence of domain-specific knowledge (cf. Abele et al. 2012). Thus, evidence suggests focusing on a specific problem domain when diagnostic problem-solving competences are measured. There is little consensus, however, on what exactly constitutes a problem domain and how to define it (e. g., Beck 2005, p. 551).

Electrotechnical diagnostic problems of car mechatronics are considered in this work. Diagnostic problems of car mechatronics can be subdivided into mechanical, electrotechnical and electro-mechanical problems. The domain of electrotechnical diagnostic problems consists of problems which require detecting electrotechnical causes of an auto defect. Such causes usually refer to broken electrotechnical components (actuators, sensors, control units, electric motors, lamps, fuses, etc.) and broken wires. Taking into account the international comparative analysis of Baethge and Arends (2009), such problems are (very) important transnational problems in the field of car mechatronics.

### 1.1.2 Electrotechnical diagnostic problem-solving competence

Electrotechnical diagnostic problem-solving competence is defined as the mental basis necessary to identify the electrotechnical cause(s) of an auto defect. It is assumed that this competence comprises several *subcomponents* and refers to successfully organizing a *process*.

Applying the model of scientific problem-solving of Klahr (2000), three sub-processes of the diagnostic problem-solving process can be distinguished: *hypothesis formulation*, *hypothesis testing* and *evidence evaluation*. This process structure has proved to be fruitful in many studies on diagnostic problem solving (e. g., Schaafstal, Schraagen & van Berlo, 2000, pp. 78–79). There are also findings suggesting that *information gathering* is another relevant sub-process (e. g., Roberts, While & Fitzpatrick, 1996).

*Information gathering* relates to collecting information about the symptoms of the car's defect(s), the technical particularities of the car (e. g., brand or vintage), relevant car systems (e. g., using technical information material), etc. In many cases, the diagnostic problem-solving process is supported by computer-based expert systems that provide, among other things, circuit diagrams and information on the location of car components. The results of the information gathering sub-process form the basis for formulating hypotheses on causes of the fault (*hypothesis formulation*). Sometimes the expert system can be used to read out the error-storage, which often indicates the "problem area", narrows down the space of possible causes and (sometimes) provides defect hypotheses. In the next step, the most probable hypothesis is tested (*hypothesis testing*), usually applying electronic test equipment (e. g., multimeters or oscilloscopes). Finally, the test result has to be evaluated (*evidence evaluation*). If the evaluation does

not support the hypothesis, another hypothesis has to be tested, etc. Commonly, the problem-solving process includes several iterations and the order of the sub-processes deviates from the ideal process described here.

The outlined process is illustrated by an example. For illustration purposes, a relatively simple example is used. Facing a malfunctioning air conditioning unit, the technical particularities of the car (e. g.: What kind of air conditioning does the car have?) and the symptoms (e. g.: Is it possible to regulate the temperature?) of the defect have to be determined (*information gathering*). Afterwards, the computer-based expert system could be used to get details on the defect. These details could indicate that the malfunction is related to a specific sensor of the air conditioning system. It could be hypothesized that the sensor is defective, or that the wire connecting the relevant control unit and the air conditioning sensor is broken, etc. (*hypothesis formulation*). The hypothesis of a defective sensor might be tested by an electrotechnical measurement: the measurement of resistance (*hypothesis testing*). If the resistance of the sensor is infinite, the hypothesis is confirmed (*hypothesis evaluation*), i. e., a defective sensor is the cause of the malfunctioning air conditioning. The replacement of the air conditioning sensor is not part of the diagnostic problem-solving, but of the repair. If the resistance value lies in the target area, another hypothesis has to be generated, tested, and so on.

There have been ample studies showing that *diagnostic problem-solving knowledge* is a key subcomponent of diagnostic problem-solving competence (e. g., Walker, Link & Nickolaus 2016; Nickolaus et al. 2012). According to Jonassen and Hung (2006), diagnostic problem-solving knowledge is understood here as an integrative construct made up of several aspects: general domain knowledge (e. g., knowledge of Ohm's law), system knowledge (e. g., knowledge of the structure and function of electrotechnical car systems), device knowledge (e. g., knowledge of electronic test equipment) and experiential knowledge (e. g., historical knowledge of common causes of faults). Strategic knowledge, information gathering skills and measurement skills are other relevant aspects (e. g., Jonassen & Hung 2006). The *strategic knowledge* represents knowledge on how to coordinate and monitor the entire problem-solving process: it is, for example, the basis for decisions on which process step and sub-process should follow another or which hypothesis should be tested first. *Information gathering skills* are, among other things, required to use the computer-based expert system to read out the error-storage, to retrieve circuit diagrams and information on the location of car components. *Measurement skills* are necessary to apply electronic test equipment, i. e. to test error hypotheses. The development, mental transformation and application of this knowledge to specific diagnostic problems depend on *working memory capacity* as well as *causal and analytical reasoning*. It seems obvious that *motivational, volitional and emotional aspects* play a decisive role, too (cf. Sembill, Rausch & Kögler 2013).

## 2 Relation of diagnostic problem-solving competence and relevant paper-based tests

In the study of Walker, Link & Nickolaus (2016), 46 % of the variance of diagnostic problem-solving competence of electronics technicians was explained by a paper-based test score representing domain-specific knowledge. Abele (2014) found a correlation of  $r=.63$  between a paper-based test for diagnostic knowledge and electrotechnical diagnostic problem-solving competence of car mechatronics. Nickolaus et al. (2012) presented an even higher correlation of  $r=.80$  between these measures. It is worth mentioning that these high correlations were achieved even though the paper-pencil tests were not explicitly developed to assess diagnostic problem-solving competence.

The review article of Swanson, Norcini and Grosso (1987) documents the correlation between paper-pencil tests and professional problem-solving competence being low to moderate, but the authors stress that these correlations might be heavily biased by the low reliabilities of the measures. When corrected for reliability, the correlations between paper-based scores and problem-solving competence might be very high (up to  $r=1.0$ ).

Link and Geißel (2015) showed that a specific professional problem-solving competence of electronics technicians<sup>56</sup> can be validly assessed using a paper-pencil test. In this study, the requirements of the paper-based test largely reflected the “reality”, i. e., the authentic requirements occurring in the workplace. Authenticity seems to be one of the key aspects of the validity of an assessment in professional contexts (cf. Tigelaar & van der Vleuten 2014, p. 1251).

So we can see that there is some evidence that paper-based tests might validly represent diagnostic problem-solving competence in case the paper-based test is aligned to assess diagnostic problem-solving competence and has a high degree of authenticity and reliability. Even if the paper-based test does not completely cover the construct, the paper-based scores can provide a convenient basis to estimate the diagnostic problem-solving competence (i. e., the interindividual differences in competence) given the paper-based and the problem-solving scores are highly correlated (close to  $r=1$ ).

## 3 Paper-based key-feature tests: A promising method to assess electrotechnical diagnostic problem-solving competence

A very prominent paper-based method to assess diagnostic problem-solving competences is written simulations (cf. Norcini & McKinley 2007). In such assessments, the test takers are typically confronted with an authentic patient problem and they are asked to

---

<sup>56</sup> The study deals with constructive problem-solving competences, i. e., competences required to programming a programmable logic controller.

answer several questions referring to different aspects of one patient problem (cf. van der Vleuten 1996, p. 44). It is well known that the psychometric quality of the scores resulting from a one-problem-testing is commonly low (e. g., Greiff 2012, p. 72) – especially in terms of reliability. Additionally, such assessments require a comprehensive written introduction to, and explanation of, the patient’s situation; so reading competence plays an important role and the assessment is probably biased by construct-irrelevant variance (cf. Kane 2013).

The paper-based *key-feature method* (cf. Hatala & Norman 2002) focuses on critical steps in solving diagnostic problems and provides the possibility, in comparison to the aforementioned method, to administer more and less text-laden items. Key-feature items consist of a short “stem followed by one or more questions” (cf. Fischer et al., 2005, p. 1) which allows several independent items to be administered. There is evidence that the key-feature approach can result in measures of good psychometric quality (cf. Hryncha, Takahash & Nayer 2014).

It has been well-documented (e. g., Klahr 2000) that successful diagnostic problem solvers are superior in formulating reasonable problem-specific hypotheses (*hypothesis formulation*) and in evaluating evidence resulting from hypothesis testing (*hypothesis evaluation*). Abele, Walker and Nickolaus (2014) showed that electrotechnical diagnostic problem-solving competence in car mechatronics can be validly measured with computer-based key-feature items that refer to using circuit and location diagrams (*information gathering*) and planning how to test diagnostic hypotheses (*hypothesis testing*). On the basis of a research review, Abele (2016) concluded that information gathering, hypothesis formulation, hypothesis testing and hypothesis evaluation are related to problem-solving activities that are critical to the diagnostic problem-solving success. Thus, it is advisable to develop key-feature items covering these critical sub-processes. It is sometimes argued that assessments concentrating on items covering specific aspects of a complex competence (i. e., key-feature items) run the serious risk of construct underrepresentation (e. g., Messick 1994).

In line with the MicroDyn approach (cf. Greiff 2012), it seems, however, possible to obtain an appropriate construct representation if there are sufficient key-feature items for each diagnostic problem-solving sub-process. MicroDyn was developed to assess complex problem-solving competence by using several items for each of the three theoretically distinguished complex problem-solving sub-processes. This approach turned out to be empirically fruitful (e. g., Greiff et al. 2013). Furthermore, there are findings proving that scores of relatively simple paper-based items and complex scenarios are closely correlated ( $r = .75\text{--}.97$ : Swanson, Norcini and Grosso 1987, p. 236; Tigelaar & van der Vleuten 2014, p. 1244) and that complex problems scores “did not add much compared with relatively simple short...scenarios” (cf. van der Vleuten et al. 2010, p. 706).

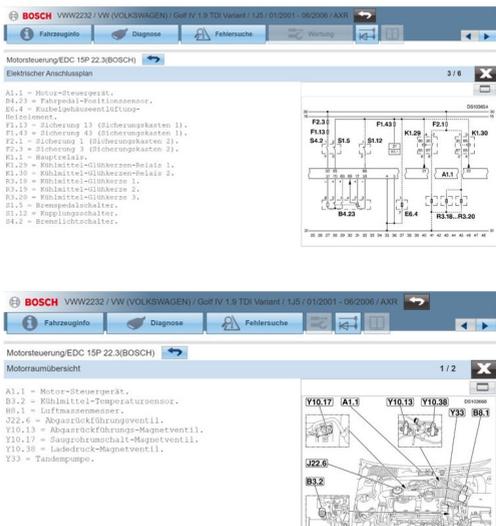
In sum, it seems defensible to assess electrotechnical diagnostic problem-solving competence of car mechatronics using authentic key-feature items that draw on gathering information as well as formulating, testing and evaluating defect hypotheses. In order to achieve an acceptable construct representation, it is advisable to generate sev-

eral key-feature items for each sub-process. Although the paper-based key-feature method is an integral part of medical assessment, this approach has not been used, to the best of my knowledge, in Vocational Education and Training to date.

## 4 Paper-based assessment of diagnostic problem-solving competence

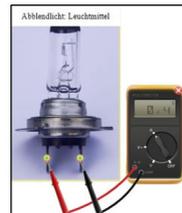
The paper-based assessment was developed in line with the aforementioned argumentation. The assessment is comprised of four booklets: two information booklets containing circuit diagrams and diagrams for the location of car components, one booklet presenting the test stimuli and the questions (key-feature item booklet) and one booklet for the responses. The reason for developing four booklets was to come as close as possible to authentic diagnostic problem-solving. For example, having information booklets and an item booklet require autonomous interactions to gather information depending on the actual status of the problem-solving process. Figure 1 gives screenshots of the information and item booklets.

### Information booklet



### Item stem (item booklet)

Eine Kundin teilt Ihnen mit, dass das Abblendlicht vorne rechts an Ihrem Fahrzeug nicht funktioniert. Zunächst prüfen Sie, ob das Leuchtmittel in Ordnung ist. Unten sehen Sie das Ergebnis dieser Prüfung.



Anschließend prüfen Sie bei eingeschalteter Zündung und eingeschaltetem Abblendlicht, ob das Leuchtmittel mit Spannung versorgt wird (Batterie ist i.O.). Unten sehen Sie das Ergebnis dieser Prüfung.



Woran kann es angesichts der Messergebnisse liegen, dass das Abblendlicht nicht funktioniert? Nennen Sie drei konkrete Fehlerursachen, die Sie messtechnisch überprüfen können.

**Figure 1:** Examples of the information and item booklets (screenshots on the left: technical terms and abbreviated designation of car components, the corresponding circuit diagram and information on their location; screenshot on the right: the upper German text says that the customer complains about a faulty low beam headlamp. Furthermore, it says that the picture with the lamp and voltmeter shows the result of the first measurement. The text in the middle gives context information on the next measuring result, which is shown in the picture below. The text at the bottom includes the question: Considering the measuring results and the specific diagnostic situation, what are possible causes of the defective low beam headlamp? Please give three causes.)

The key-feature items were presented using authentic graphic material. Each item contained short textual descriptions, demanded interpretations of authentic visual stimulus material, referred to a specific problem-solving sub-process and encompassed several problem-solving steps. The amount of text, visual material and problem-solving steps varied between the items. The right part of Figure 2 gives an example of an “average” item. A lot of attention was paid to having only as much text as necessary. The descriptions are purposefully arranged to simulate the “real” sub-process as authentically as possible. Overall, two multiple-choice and 20 short-answer items were developed (table 1).

**Table 1:** Overview of the paper-based key-feature items

ID	Aim and content	Sub-process	Car system
IG1	Localization of the speed sensor	Information gathering	Electronic engine management
IG2	Specifying measuring points at the intake-manifold change-over valve		
IG3	Localization of the exhaust-gas recirculation valve		
IG4	Specifying the abbreviated designation of the vehicle speed sensor		
IG5	Specifying the signal line of the camshaft sensor		
IG6	Specifying the purpose of a specific wire of the electric fuel pump		
IG7	Localization of the pressure-charging valve		
IG8	Specifying the designation of the pressure-charging valve		
HF1	Formulation of hypotheses on causes of a defective fuel temperature sensor	Hypothesis formulation	Electronic engine management
HF2	Formulation of hypotheses on causes of a defective low beam headlamp		Lighting system
HF3	Formulation of hypotheses on causes of a defective spark plug relay		Electronic engine management
HF4	Formulation of hypotheses on causes of a defective engine control unit		Electronic engine management
HF5	Formulation of hypotheses on causes of a defective pressure-charging valve I		Electronic engine management
HF6	Formulation of hypotheses on causes of a defective pressure-charging valve II		Electronic engine management
HT1	Specifying a strategy to test a broken signal line of the camshaft sensor	Hypothesis testing	Electronic engine management
HT2	Selecting electronic test equipment to test the function of the electric starting system		Electronic engine management

(Continuing table 1)

ID	Aim and content	Sub-process	Car system
HT3	Specifying a strategy to test a broken wire of the intake-manifold change-over valve	Hypothesis testing	Electronic engine management
HT4	Specifying strategies to test hypotheses on a defective pressure-charging valve		Electronic engine management
EE1	Evaluation of measuring results and specifying the cause of a defective speed sensor	Evidence evaluation	Electronic engine management
EE2	Specifying the resistance value range of an intact fuse of the low beam headlamp		Lighting system
EE3	Specifying the pulse duty factor of the exhaust-gas recirculation valve using a signal measurement result of an oscilloscope		Electronic engine management
EE4	Evaluation of measuring results and specifying the cause of a defective diesel fuel injector unit		Electronic engine management

Several key-feature items (KFI) were developed for each diagnostic problem-solving sub-process. The *information gathering* KFI required specific information to be given by means of the information booklets (e. g., to find out the location of the speed and reference-mark sensor). The *hypothesis formulation* KFI demanded analysis of a diagnostic situation and formulation of error hypotheses (Figure 1). The *hypothesis testing* KFI were about strategies to test a specific hypothesis or select electronic test equipment in specific test situations. Finally, the *evidence evaluation* KFI demanded an interpretation of given measuring results. The 22 KFI cover the electronic engine management and lighting system.

Since we were aiming for high-quality items, we did not succeed in developing the same number of KFI for each sub-process: During the development, it turned out that it was easy to design authentic information gathering KFI and difficult to design authentic KFI on hypothesis testing and evidence evaluation, because some requirements were especially challenging to implement in a paper-based test. For example, hypothesis testing frequently is related to measurements, which cannot be considered in a paper-based test.

## 5 Research hypothesis

It is assumed that the paper-based test provides valid test score interpretations, meaning that the paper-based scores can be interpreted as indicators of diagnostic problem-solving competence. To test this hypothesis, the electrotechnical diagnostic problem-solving competence of car mechatronic apprentices was assessed and the outlined paper-based test was administered. In line with the hypothesis, evidence for convergent validity was expected, i. e., it was supposed that the paper-based scores and diagnostic problem-solving scores were unidimensional.

## 6 Method

### 6.1 Sample and design

In order to test the hypothesis, 206 car mechatronic apprentices nearing the end of the 3rd year of training were sampled from vocational schools of the federal state of Baden-Württemberg, Germany. The mean age of the apprentices was 21.0 years ( $SD = 3.1$ ) and 6.2 % of the sample were females.

In the present study, a within-subjects-design was used. At the beginning, the apprentices were confronted with the paper-based test in the classroom. Approximately two weeks later, the electrotechnical diagnostic problem-solving competence was measured, administering 4 authentic diagnostic problems in a computer simulation, and 4 authentic diagnostic problems on a car in a garage. To control for position, i. e. exhaustion effects, the problems were administered in a *latin square design* (cf. Frey, Hartig & Rupp 2009, p. 45). For organizational reasons, it was not possible to control for order effects potentially caused by starting with the written test and ending with the authentic problems. The overall testing time was 350 minutes (90 minutes for the paper-based test). Due to the long testing time, the written test was administered to a selection of apprentices ( $N = 121$ ).

Prior to the written test, the sample got a standardized instruction of 10 minutes, including information on the testing time, the number of items, the test and response booklets as well as the information booklets. To prevent cheating, two tests were created, containing the same items in different orders.

The standardized instruction for the computer-based assessment took 30 minutes. The instructor demonstrated the handling of the computer simulation by means of a video projector. Afterwards, the apprentices worked individually on standardized tasks concerning dealing with 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. The introduction to the assessment in the garage took 15 minutes and explained the test setting. In both assessments, the apprentices were acquainted with the response booklet and got instructions on how to prepare the handwritten documentation. They were told to document each relevant problem-solving step (e. g., each electrotechnical measurement result) and to finish the documentation by giving a clear statement on the cause of the diagnostic problem.

### 6.2 Measures

#### 6.2.1 Electrotechnical diagnostic problem-solving competence

The electrotechnical diagnostic problem-solving competence was measured using eight diagnostic problems (table 2). The scores of the computer-based testing and of the testing on the car in the garage were highly correlated ( $r = .94$ , latent) providing strong evidence for convergent validity. This finding justified interpreting the scores as indicators of diagnostic problem-solving competence and integrating them into a single score (cf. Abele & Nickolaus 2016).

The scoring was conducted by two independent raters applying a coding manual to the handwritten documentation of the apprentices and resulted in both dichotomous and polytomous data. In very rare cases of diverging scores, content-oriented discussions produced consensual scoring. For example, a problem was considered solved if the correct cause of a problem (e.g., a broken central locking motor) had been identified, documented and proved by appropriate measurements. In case of correct solutions, the testee got the highest score. If relevant problem-solving steps were *not* documented, lower scores were assigned.

**Table 2:** Overview of the authentic diagnostic problems

ID	Aim and content	Sub-process	Car system
P1	Diagnosing the cause of a defective window regulator lighting	Complete diagnostic problem-solving process	Comfort system: Door control
P2	Diagnosing the cause of a defective car radio		Comfort system: Car radio
P3	Diagnosing the cause of a defective air recirculation flap		Comfort system: Air conditioning
P4	Diagnosing the cause of a defective outflow-temperature sensor		Comfort system: Air conditioning
P5	Diagnosing the cause of a defective central locking motor		Comfort system: Door control
P6	Diagnosing the cause of a defective electric mirror adjustment		Comfort system: Door control
P7	Diagnosing the cause of a defective wheel speed sensor		Brake system
P8	Diagnosing the cause of a defective fresh air fan		Comfort system: Air conditioning

### 6.2.2 Paper-and-pencil-based assessment

The paper-based assessment included 22 categorical items (table 1). Compared to the aforementioned assessment, the key-feature method led to a greatly reduced testing time (from 260 to 90 minutes).

A manual including the correct answers was developed for the scoring. The KFI were scored based on a few keywords. Most of the KFI were dichotomous, although some were polytomous. Regarding the example in figure 1, each correct cause gives one point, implying a maximum score of three.

## 6.3 Statistical procedures

### 6.3.1 Missing data

The paper-based items were administered to 121 of the 206 apprentices. Taking 121 apprentices as reference, 0 to 8.3 % of the item data was missing (item-level missingness), whereby most of the rates of missingness were 0 or very close to 0 (table 4).

Taking the complete sample of 206 apprentices as a reference, the rates of missingness of the problem-solving scores ranged from 11 to 14% (table 3).

Obviously, the missing data was mainly due to a somehow “planned” missing data design (person-level missingness). Thus, large parts of the missing data can be classified as missing completely at random (cf. Enders 2010, p. 22) meaning that the missing data should not bias the statistical parameter estimates (cf. Graham 2009, p. 553). Since the hypothesis was tested using the polytomous Rasch model, the missing data did not lead to a listwise deletion of cases and all item responses could be included to test the hypothesis (cf. Boone, Staver & Yale 2014, p. 380) increasing the statistical power.

### 6.3.2 Analysis of KFI and diagnostic problems

To evaluate the psychometric properties of the KFI and the diagnostic problems, statistical procedures of the classical test theory (CTT) and item response theory (IRT) were carried out. In terms of CTT, the item discrimination, the relative frequency of correct responses (item difficulty) and Cohen’s kappa (inter-rater reliability) were determined. For the estimation of the IRT parameters, the partial credit model of the polytomous Rasch family was applied (cf. Ostini, Finkelman & Nering 2013) generating item difficulties and corresponding estimation errors expressed in the logit metric. In order to evaluate whether the items fit the partial credit model, the item infit (weighted mean square statistic) and outfit (unweighted mean square statistic) were examined (cf. Bond & Fox 2007, p. 137). According to Wilson (2005), infit and outfit statistics ranging from 0.75 to 1.33 indicate an acceptable fit (p. 129). Bond and Fox (2007) point out that  $t$  statistics of the item infit and outfit that are higher than 2 and smaller than -2 indicate a significant misfit (p. 134). Since the  $t$  statistic, however, is fairly sample size sensitive, items were only dropped if the infit/outfit statistics and the related  $t$  statistics deviated from the above-mentioned target ranges (cf. Wilson 2005, p. 129). Finally, the EAP/PV reliability was estimated for both assessments, which can be interpreted similarly to Cronbach’s alpha.

### 6.3.3 Testing and construct validity

The hypothesis of convergent validity was examined in two steps: In the first step, a one-dimensional partial credit model was estimated integrating the diagnostic problems and the paper-based KFI. In the second step, a two-dimensional model was estimated, where the problems and the KFI constitute separate dimensions. Afterwards, the correlation of the two dimensions was inspected and the following analyses were conducted to decide which model fit the data better: First, a chi-square difference test was performed (cf. Reckase 2009, p. 218) with a significance level of 1%. Second, the deviance, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) of the models were compared, where smaller values indicate a better model fit. AIC differences of 10 indicate serious model differences (cf. Burnham & Anderson 2004, p. 271). If the dividend of sample size and model parameter is smaller than 40, also using the CAIC is recommended (cf. Burnham & Anderson 2004, p. 270). The analyses were performed with the ConQuest software (cf. Wu et al. 2007).

## 7 Results

### 7.1 Problem and KFI statistics

table 3 gives the CTT statistics of the problems and the corresponding parameters of the partial credit model and documents the fact that the psychometric quality of the problems was good. Only the inter-rater reliability of problem P8 was remarkably low. Since the other coefficients of P8 were acceptable, this problem was not dropped. The EAP/PV reliability of the measurement is .65, suggesting that the number of problems should be increased in future studies (e. g., Norcini & McKinley 2007, p. 243). It should be noted that, for parameterization reasons, the difficulties of the highest problem scores were constrained, so there were no standard errors available.

**Table 3:** CTT statistics and parameters of the partial credit model of the diagnostic problems

ID	$\kappa$	$f$	$r$	Difficulty	Error	Outfit	Outfit t	Infit	Infit t	N
P1.1	.91	.22	.61	.50	.18	.98	-.2	.99	-.1	184
P1.2		.35		-.25	constrained	.92	-.7	.92	-1.1	
P2.1	.93	.20	.55	.28	.19	.96	-.3	.98	-.1	183
P2.2		.47		-.70	constrained	.96	-.4	.98	-.3	
P3.1	.96	.35	.62	-.66	.16	.96	-.4	.97	-.5	182
P3.2		.31		.19	.17	.98	-.2	.99	-.1	
P3.3		.12		1.40	constrained	.88	-1.1	.95	-.3	
P4.1	.83	.16	.56	-.10	.20	.92	-.7	.98	-.1	183
P4.2		.63		-1.32	constrained	.88	-1.2	.92	-1.2	
P5.1	.97	.25	.46	.98	.18	.97	-.2	1.00	.0	180
P5.2		.09		1.43	constrained	.90	-1.0	1.04	.3	
P6.1	.94	.20	.49	.54	.19	1.04	.4	1.02	.2	177
P6.2		.40		-.54	constrained	1.13	1.2	1.11	1.6	
P7.1	.92	.08	.60	1.67	.27	1.17	1.6	1.01	.1	185
P7.2		.42		-1.46	constrained	.99	-.1	.99	-.1	
P8	.69	.87	.38	-2.02	.23	.87	-1.3	.95	-.3	179

*Notes.*  $\kappa$  = inter-rater reliability;  $f$  = relative frequency of correct responses per category, the frequency of incorrect responses results from 1 minus the sum of the relative frequencies;  $r$  = discrimination: Pearson correlation between the item score and the total raw score; Difficulty = IRT difficulty (logit); error = standard error related to the difficulty estimation.

Table 4 gives the statistics of the paper-based KFI. Here, the discrimination of IG7 and IG8 as well as the Kappa of IG3 are considerably low. But again, these items were not excluded from further analyses due to the good quality of their other statistics. The EAP/PV reliability of the paper-based scores was .75.

**Table 4:** CTT statistics and parameters of the partial credit model of the KFI

ID	$\kappa$	$f$	$r$	Difficulty	Error	Outfit	Outfit t	Infit	Infit t	N
IG1	.97	.42	.27	.39	.20	1.09	.7	1.10	1.6	120
IG2	.85	.95	.35	-3.19	.43	1.09	.8	.93	-.1	120
IG3	.49	.42	.23	.37	.20	1.13	1.0	1.12	1.9	119
IG4	1.0	.88	.25	-2.24	.29	.94	-.4	1.00	.0	121
IG5	.92	.71	.41	-1.00	.21	1.04	.4	.98	-.2	121
IG6	.94	.80	.47	-1.55	.24	.82	-1.5	.89	-.8	121
IG7	.93	.78	.15	-1.39	.23	1.06	.5	1.05	.4	121
IG8	.97	.95	.17	-3.21	.43	.83	-1.3	1.01	.2	121
HF1.1		.31		-.33	.20	.95	-.3	.96	-.5	
HF1.2	.89	.27	.71	.25	.22	.93	-.5	1.03	-.3	111
HF1.3		.13		1.25	constrained	.58	-3.7	.91	-.4	
HF2.1		.20		-1.41	.21	1.01	.1	1.01	.1	
HF2.2	.89	.49	.61	-1.05	.19	1.00	.0	1.00	.1	111
HF2.3		.23		1.05	constrained	.83	-1.3	.97	-.2	
HF3.1		.40	.60	-.61	.20	.96	-.2	.97	-.7	
HF3.2	.92	.32		.40	constrained	.84	-1.2	.93	-.8	111
HF4	.81	.78	.50	-1.45	.24	.78	-1.7	.89	-.9	111
HF5.1		.28	.53	-.74	.21	.97	-.2	.99	-.1	
HF5.2	.86	.52		-.55	constrained	.87	-.9	.91	-1.1	111
HF6.1		.37	.59	-1.35	.19	.94	-.4	.96	-.7	
HF6.2	.84	.49		-.22	constrained	.82	-1.4	.87	-2.0	120
HT1	.97	.73	.39	-1.09	.22	.92	-.6	.95	-.5	121
HT2	.97	.55	.16	-.24	.19	1.14	1.1	1.14	2.2	121
HT3	.94	.20	.20	1.55	.25	1.11	.9	1.11	.8	111
HT4	.92	.38	.43	.54	.20	.92	-.6	.96	-.5	120
EE1	.52	.66	.33	-.74	.20	.94	-.4	.97	-.4	121
EE2	.96	.73	.35	-1.09	.22	1.06	.5	.98	-.1	121
EE3	.97	.63	.20	-.58	.20	1.10	.8	1.07	1.0	121
EE4	.61	.51	.54	-.05	.19	.87	-1.0	.89	-2.0	121

*Notes.*  $\kappa$  = inter-rater reliability;  $f$  = relative frequency of correct responses per category, the frequency of incorrect responses results from 1 minus the sum of the relative frequencies;  $r$  = discrimination: Pearson correlation between the item score and the total raw score; Difficulty = IRT difficulty (logit); error = standard error related to the difficulty estimation.

## 7.2 Evidence for the convergent validity

It was hypothesized that the paper-based scores and the problem-solving scores represent the same construct, so a one-dimensional model should fit the data at least as well as a two-dimensional model. In contrast to this hypothesis, the IRT analyses revealed that the two-dimensional partial credit model outperformed the unidimensional one (table 5).

**Table 5:** Statistics of the one and two-dimensional partial credit model

Model	N	df	Deviance	BIC	AIC	CAIC
One-dimensional	206	46	5,949.1	6,194.2	6,041.1	6,240.2
Two-dimensional	206	46	5,935.7	6,191.4	6,031.7	6,239.4

The AIC documented the better fit regarding the information criteria of the two-dimensional compared to the one-dimensional model ( $\Delta 9.4$ ). Moreover, the chi-square difference test was statistically significant ( $\Delta df = 2$ ;  $\Delta deviance = 13.4$ ;  $p = .001$ ) indicating the superiority of the two-dimensional model. The correlation between the computer-based and paper-based scores was high ( $r = .76$ , latent) but substantially deviated from a perfect correlation.

## 8 Discussion

### 8.1 Summary

The aim of the present study was to investigate whether diagnostic problem-solving competences of car mechatronics can be validly measured using a paper-based test. For that purpose, a paper-based key-feature test for electrotechnical diagnostic problem-solving competence was developed, administered to a sample of car mechatronic apprentices and validated using diagnostic problem-solving scores. It was hypothesized that the paper-based scores would represent the same empirical dimension as the problem-solving scores, i. e., provide valid test score interpretations. Compared to previous studies (e. g., Hryncha, Takahash, & Nayer 2014), the key-feature test resulted in a relatively high reliability. Although the paper-based and problem-solving scores were highly correlated, there was, however, reasonable evidence for their discriminant validity: Model comparisons documented a superior fit of the model in which the paper-based scores and problem-solving scores represented distinct dimensions. Presumably, the written test especially tapped electrotechnical diagnostic knowledge which has been frequently proven to be distinguishable from, but closely correlated with, diagnostic problem-solving competence (e. g., Abele 2014, pp. 57–58; Nickolaus et al. 2012). So the paper-based scores reflected a key subcomponent of diagnostic problem-solving competence but failed to cover some other relevant subcomponents. In contrast to Link and Geißel (2015), the findings of this study suggest that paper-based tests are not com-

pletely suitable to measure professional problem-solving competence. What is the reason for these contradictory findings?

Content validation indicates that the paper-based key-feature test likely suffered *construct underrepresentation*: The written test did not tap an important subcomponent of electrotechnical diagnostic problem-solving competence, namely: measurement skills (i. e., conducting electrotechnical measurements). In a paper-based test, it seems (almost) impossible to authentically consider electrotechnical measurement skills of car mechatronics. Moreover, the key-feature items only covered diagnostic problem-solving sub-processes, i. e., strategic knowledge needed to coordinate the sub-processes and the entire diagnostic problem-solving process was not taken into account. Link and Geißel (2015) did not use key-feature items to assess professional problem-solving competence but rather “holistic” items that covered the relevant subcomponents of their focal construct. Therefore, they probably succeeded in having an appropriate construct representation.

Abele, Walker and Nickolaus (2014) concluded that electrotechnical diagnostic problem-solving competence of car mechatronics can be validly measured using *computer-based key-feature* items. Here, it was concluded that the key-feature items are not completely suitable to measure professional problem-solving competence. There are at least three reasons for these somehow contradictory conclusions: (1) *statistical uncertainty*, (2) *construct under-representation* and (3) *construct-irrelevant variance*. (1) The results of Abele, Walker and Nickolaus (2014) were ambiguous. The statistical comparison of a two-dimensional model (dimension 1: computer-based key-feature scores; dimension 2: diagnostic problem-solving scores) to a unidimensional model did not bring clear results: The  $p$ -value of the model comparison was  $p = .04$  indicating that the decision whether the two-dimensional or the unidimensional model fits the data better depends on which significance level is chosen ( $\alpha = .01$  or  $\alpha = .05$ ). Pragmatically, the authors argued that the latent correlation of  $r = .89$  between the two dimensions is high enough to defend the use of computer-based key-feature items to measure diagnostic problem-solving competence. (2) Besides this statistical uncertainty, the contradictory findings might be caused by construct underrepresentation: In the study of Abele, Walker and Nickolaus (2014), the *computer-based key-feature* items had to be mastered by interacting with a computer-simulation reflecting parts of the real work environment of car mechatronics. In the *written* test, the “natural” work environment of car mechatronics could be only simulated rudimentarily. For example, information gathering using information booklets substantially differs from information gathering using the computer-based expert system: The paper-based key-feature items did not (authentically) tap information gathering skills. As mentioned before, the paper-based test did also not allow for including measurement skills, whereas the computer-based key-feature items did. (3) The paper-based test was probably also biased by somehow “artificial” interactions, i. e. construct-irrelevant variance: Although the development of information booklets aimed to come close to authentic information gathering interactions, there seem to be remarkable differences between the interactions connected to the paper-based test and the “reality”.

Since the findings of this paper are bound to the domain of “electrotechnical diagnostic problems” of car mechatronics, the following question arises: What can we learn from this study about whether paper-based (key-feature) tests are appropriate to measure professional problem-solving competences?

## 8.2 Theoretical and practical significance

In view of the study of Link and Geißel (2015), it seems scientifically unjustifiable to principally consider *paper-based* tests as inappropriate to measure professional problem-solving competence, as might be concluded on the basis of other publications (e. g., van der Vleuten et al. 2010, p. 706). The decisive question is whether a test allows an adequate representation of the focal construct and causes construct-irrelevant variance (e. g., Messick 1994). For example, it could be appropriate to measure specific professional problem-solving competences (e. g., designing electric circuits) using paper-based items. There should be, however, many more professional situations in which paper-based tests fall short of completely covering the professional problem-solving competence.

What about *key-feature* items? Generally speaking, authentic “holistic” items should imply a better construct representation and less construct-irrelevant variance than paper-based key-feature items. It should be pointed out, however, that “holistic” items can cause a severe amount of construct-irrelevant variance too – for example, when they require long textual introductions that are not associated with the focal construct. In the end, the question whether, and in which situations and domains, *key-feature* items are appropriate to measure professional problem-solving competences seems to still be open. Undoubtedly, the answer to this question relies on whether rather pragmatic arguments are accepted: In some contexts, latent correlations of  $r \approx .80$  might be sufficient to take key-feature scores as acceptable proxies, i. e. valid indicators of professional problem-solving competence.

The arguments presented indicate that construct representation and construct-irrelevant variance are crucial to validation studies. This spotlights the role of the definition of professional problem-solving competence, which is the basis for judging whether a test is “contaminated” by construct underrepresentation and construct-irrelevant variance. This point is touched on in the *Standards for Educational and Psychological Testing* (cf. AERA, APA & NCME 2014), which highlights the fact that an important concern of validity is the use and interpretation of test scores (cf. AERA, APA & NCME 2014, p. 11). I think it is worth putting more emphasis on the construct definition. The definition of the problem-solving competence (the problem domain, subcomponents of problem-solving competence and problem-solving process) is the decisive reference point of the test score interpretation. It is the theoretical anchor of the validation process. Focusing more on the definition of professional problem-solving competence brings major advantages: It forces scientists to carefully describe and examine the professional problem-solving competence under investigation, increases our knowledge of professional problem-solving competence (e. g., its subcomponents and processes) as well as the precision, practical relevance and testability of corresponding theories.

Practically speaking, the domain-specificity of professional problem-solving competence is (very) challenging: Assuming that there are numerous professional problem domains, numerous different tests of professional problem-solving competence must be developed. There are findings showing that, within one occupational profile, the domain-specific problem-solving competences can be strongly correlated (e. g.,  $r = .77$ ; Walker, Link & Nickolaus 2016). These results could be used to legitimize the generalization of domain-specific test results to other problem domains within one occupational profile. This argument might enormously reduce the number of assessments which are necessary to measure professional problem-solving competences. Its quality, however, strongly depends on the learning opportunities of the testees: If they did not have the opportunity to acquire the diverse domain-specific professional problem-solving competences related to an occupational profile, the generalization fails. Thus, the previous learning experiences of the testees and the organization of their vocational education should be accounted for when domain-specific problem-solving scores are generalized. It might be fruitful to scrutinize the generalizability of domain-specific professional problem-solving scores (cf. Kane 2013) and to find out which measurement error is to be expected when professional problems of different problem domains are sampled to assess the problem-solving competence within one occupational profile. If these studies reveal that, under certain conditions, it is defensible to infer the overall problem-solving competence within one occupational profile from the scores of one or a selection of problem domains, enormous progress would be achieved.

### 8.3 Limitations and outlook

The empirical results of this study rest upon relatively small samples ( $N = 119-185$ ). Penfield (2013) argues that “a sample size of 100 can be viewed as a lower threshold” (p.132) for classical test theory analyses. Following Boone, Staver and Yale (2014), a sample size of 100 might be also sufficient to apply the Partial Credit Model (p. 364). Against this background, it can be assumed that the sample size might not decisively bias the insights presented here.

Another limitation is that the unidimensionality of the paper-based key-feature items (KFI) was not examined, even though the KFI only covered sub-processes. A supplementary confirmatory factor analysis revealed that the item scores referring to different sub-processes can be treated as unidimensional ( $N = 121$ ,  $\chi^2 = 247$ ,  $df = 260.5$ ,  $\chi^2/df = 1.05$ ,  $CFI = .95$ ,  $NNFI = .94$ ,  $WRMR = .95$ ). It should also be noted that the sample size of this confirmatory analysis is relatively small.

As mentioned before, this study is limited to a specific problem domain. Further studies should examine whether the argumentation of this study can be confirmed in other problem domains. Furthermore, studies examining the professional problem-solving process and its mental subcomponents can bring substantial progress and enhance the understanding of professional problem-solving competence. In order to study the problem-solving process, computer-generated log-file data offer interesting options (cf. Greiff et al. 2013).

## References

- Abele, S. (2014). Modellierung und Entwicklung berufsfachlicher Kompetenz in der gewerblich-technischen Ausbildung. [Modeling and development of professional competence in Vocational Technical Education]. Stuttgart: Franz Steiner.
- Abele, S. (2016). Theory of the Diagnostic Problem-solving Process in Professional Contexts and its Evaluation Using Computer-generated Log-file Data. Manuscript submitted for publication.
- Abele, S., Greiff, S., Gschwendtner, T., Wüstenberg, S., Nickolaus, R. & Funke, J. (2012). Dynamische Problemlösekompetenz – ein bedeutsamer Prädiktor von Problemlöseleistungen in technischen Anforderungskontexten? [Dynamic problem solving – an important predictor of problem-solving performance in technical domains?]. *Zeitschrift für Erziehungswissenschaft*, 15, 363–391.
- Abele, S. & Nickolaus, R. (2016). Validität einer Diagnostik berufsfachlicher Problemlösekompetenzen bei Kfz-Mechatronikern mit einer Computersimulation [Validity of a computer-simulation-based assessment of professional problem-solving competence in the domain of car mechatronics]. Manuscript in preparation.
- 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.
- AERA (American Education Research Association), APA (American Psychological Association) & NCME (National Council on Measurement in Education) (2014). Standards for Educational and Psychological Testing. Washington D. C.: American Educational Research Association.
- Baethge, M. & Arends, L. (2009). Feasibility study VET-LSA: A comparative analysis of occupational profiles and VET programmes in 8 European countries – International report. Vocational Training Research: Vol. 8. Bonn: Federal Ministry of Education and Research 22 (BMBF).
- Beck, K. (2005). Ergebnisse und Desiderate zur Lehr-Lern-Forschung in der kaufmännischen Berufsausbildung [Results and desiderata of research on teaching-learning-processes in the field of commercial vocational education and training]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 101(4), 533–556.
- Bond, T. G. & Fox, C. M. (2007). Applying the Rasch Model: Fundamental measurement in the Human Sciences. Mahwah, NJ: Erlbaum.
- Boone, W. J., Staver, J. R. & Yale, M. S. (2014). Rasch analysis in the human sciences. Dordrecht: Springer.
- Burnham, K. P. & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2).
- Enders, C. K. (2010). Applied missing data analysis. New York, NY: Guilford Press.

- Fischer, M. R., Kopp, V., Holzer, M., Ruderich, F. & Jünger, J. (2005). A modified electronic key feature examination for undergraduate medical students: validation threats and opportunities. *Medical Teacher*, 27, 1–6.
- 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.
- Graham, J. W. (2009). Missing data analysis: making it work in the real world. *Annual Review of Psychology*, 60, 549–576.
- Greiff, S. (2012). *Individualdiagnostik komplexer Problemlösefähigkeit*. Münster: Waxmann.
- Greiff, S., Wüstenberg, S., Holt, D. V., Goldhammer, F. & Funke, J. (2013). Computer-based assessment of Complex Problem Solving: concept, implementation, and application. *Educational Technology Research and Development*, 61, 407–421.
- Gschwendtner, T., Abele, S. & Nickolaus, R. (2009). Computersimulierte Arbeitsproben: Eine Validierungsstudie am Beispiel der Fehlerdiagnoseleistungen von Kfz-Mechatronikern [Computer-simulation-based work samples: A validation study taking troubleshooting performance of car mechatronics as an example]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 105, 557–578.
- Hatala, R. & Norman, G. R. (2002). Adapting the key features examination for a clinical clerkship. *Medical Education*, 36, 160–165.
- Hryncha, P., Takahash, S. G. & Nayer, M. (2014). Key-feature questions for assessment of clinical reasoning: a literature review. *Medical Education*, 48, 870–883.
- 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.
- Kane, M. J. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73.
- Kassirer, J., Wong, J. & Kopelman, R. (2010). *Learning clinical reasoning* (3rd ed.). Baltimore, MD: Lippincott Williams & Wilkins.
- Klahr, D. (2000). Scientific discovery as problem solving. In D. Klahr (Ed.), *Exploring science. The cognition and development of discovery processes* (pp. 21–39). Cambridge, MA: MIT Press.
- Link, N. & Geißel, B. (2015) Konstruktvalidität konstruktiver Problemlösefähigkeit bei Elektronikern für Automatisierungstechnik. [Construct validity of the constructive problem-solving competence of electronics technicians for automation technology]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 111, 208–221.
- Mainert, J., Kretschmar, A., Neubert, J. C. & Greiff, S. (2015). Linking complex problem solving and general mental ability to career advancement: Does a transversal skill reveal incremental predictive validity? *International Journal of Lifelong Education*, 34(4).
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23.

- Nickolaus, R., Abele, S., Gschwendtner, T., Nitzschke, A. & Greiff, S. (2012). Fachspezifische Problemlösefähigkeit in gewerblich technischen Ausbildungsberufen – Modellierung, erreichte Niveaus und relevante Einflussfaktoren. [Domain-specific problem-solving competence in industrial-technical professions – Modelling, achieved levels and relevant predictors]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 108, 243–272.
- Norcini, J. J. & McKinley, D. W. (2007). Assessment methods in medical education. *Teaching and Teacher Education*, 23, 239–250.
- Ostini, R., Finkelman, M. & Nering, M. (2013). Selecting among polytomous IRT models. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment* (pp. 285–304). New York, NY: Routledge.
- Penfield, R. D. (2013). Item analysis. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J. I. C. Hansen, N. R. Kuncel, S. P. Reise & M. C. Rodriguez (Eds.), *Test theory and testing and assessment in industrial and organizational psychology: Vol. 1. APA handbook of testing and assessment in psychology* (pp. 121–138). Washington, DC: American Psychological Association.
- 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
- Reckase, M. D. (2009). *Multidimensional item response theory*. Dordrecht: Springer.
- Roberts, J., While, A. E. & Fitzpatrick, J. (1996). Exploring the process of data acquisition: Methodological challenges encountered and strategies employed. *Journal of Advanced Nursing*, 23, 366–372.
- Schaafstal, A., Schraagen, J. M. & van Berlo, M. (2000). Cognitive task analysis and innovation of training: the case of structured troubleshooting. *Human Factors*, 42.
- Schwartz, A. & Elstein, A. S. (2011). 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.
- Sembill, D., Rausch, A. & Kögler, K. (2013). Non-cognitive facets of competence: Theoretical foundations and implications for measurement. In O. Zlatkin-Troitschanskaia & K. Beck (Eds.), *From Diagnostics to learning success – Proceedings in vocational education and training*. Rotterdam: Sense.
- Shavelson, R. J. (2010). On the measurement of competency. *Empirical Research in Vocational Education and Training*, 2(1), 41–63.
- Swanson, D. B., Norcini, J. J. & Grosso, L. J. (1987). Assessment of clinical competence: Written and computer-based simulations. *Assessment and Evaluation in Higher Education*, 12(3).
- Tigelaar, D. E. & van der Vleuten, C. P. (2014). Assessment of professional competence. In S. Billet, C. Harteis & H. Gruber (Eds.), *International Handbook on Research in Professional and Practice-based Learning* (pp. 1237–1270). Dordrecht: Springer.

- van der Vleuten, C. (1996). The assessment of professional competence: Developments research and practical implications. *Advances in Health Science Education*, 1, 41–67.
- van der Vleuten, C. P., Schuwirth, L. W., Scheele, F., Driessen, E. W. & Hodges, B. (2010). The assessment of professional competence: building blocks for theory development. *Best Practice & Research Clinical Obstetrics and Gynaecology*, 24, 703–719.
- Walker, F., Link, N. & Nickolaus, R. (2016). A multidimensional structure of problem-solving competencies of electronics technicians for automation technology. *Empirical Research in Vocational Education and Training*, 8, 1–16. doi:10.1186/s40461-016-0034-z
- Wilson, M. (2005). *Constructing Measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wu, M. L., Adams, R. J., Wilson, M. R. & Haldane, S. A. (2007). *ACER ConQuest Version 2.0: Generalized item response modelling software*. Camberwell, VIC: ACER Press.



# **Autor**

Prof. Dr. Stephan Abele ist Inhaber der Professur für Berufspädagogik an der Technischen Universität Dresden. Seine Forschungsschwerpunkte sind die Messung und Entwicklung beruflicher Kompetenzen, berufliches Lehren und Lernen sowie die Professionalisierung berufsbildender Lehrkräfte. Er untersucht mit computerbasierten Verfahren u. a. den Prozess der Kfz-Störungsdiagnostik sowie die Entwicklung, Messung und Förderung der Kfz-Diagnosekompetenz.

Kontakt: [stephan.abele@tu-dresden.de](mailto:stephan.abele@tu-dresden.de)



 Wirtschaft – Beruf – Ethik, 41  
2022, 268 S., 49,90 € (D)  
ISBN 978-3-7639-7137-4  
E-Book im Open Access

Stephan Schumann, Susan Seeber, Stephan Abele (Hg.)

## Digitale Transformation in der Berufsbildung

Konzepte, Befunde und Herausforderungen

Mit Blick auf den gesellschaftlichen Fortschritt und die wirtschaftliche Leistungsfähigkeit muss der digitale Rückstand im deutschen Bildungssystem aufgeholt werden. Dabei spielt die Berufsbildung aufgrund ihrer engen Verbindungen zur Wirtschaft eine Schlüsselrolle in der digitalen Transformation. In den Beiträgen des Sammelbandes werden berufsspezifische Aspekte und Querschnittsthemen vorgestellt, bei denen digitale Medien auf inhaltlicher und methodischer Ebene neue Handlungsmöglichkeiten eröffnen: Organisation der Ausbildung, Professionalisierung der Lehrenden, digital gestützte Prüfungsaufgaben, Zusammenarbeit der Lernorte, Veränderungen der Didaktik durch den Einsatz digitaler Medien wie Learning Analytics Dashboards, Augmented Reality und Virtual Reality, Individualisierung von Lernprozessen, Diagnose von Lernständen, Inklusion und berufliche Teilhabechancen durch digitale Technologien sowie Formen politischer Sozialisation und Handlungsfähigkeit als Aufgabe der beruflichen Bildung. Der Sammelband schließt mit einem Blick in die Zukunft und Ideen für die Berufsbildende Schule 2030.

[wbv.de/wbe](http://wbv.de/wbe)

Um berufliche Probleme zu lösen, müssen Individuen über Problemlösekompetenzen verfügen. Diese Kompetenzen ermöglichen es ihnen, den Problemlöseprozess zielgerichtet zu koordinieren und ein erwünschtes Ergebnis zu erreichen. Sie werden in der beruflichen Ausbildung erworben und sind nicht direkt beobachtbar. Was aber macht diese Kompetenzen aus? Wie können sie sichtbar gemacht und gemessen werden?

Damit beschäftigt sich der Berufspädagoge Stephan Abele am Beispiel der diagnostischen Problemlösekompetenz in der Kfz-Mechatronik. Sein Buch enthält sechs Studien sowie deren theoretische Grundlegung und Diskussion. In den Studien wird u. a. untersucht, welche Bedeutung der Umgang mit Komplexität für das Lösen beruflicher Probleme hat und ob Problemlösekompetenzen mit Papier-Bleistift-Tests genauso aussagekräftig gemessen werden können wie mit einer Computersimulation.

Die Reihe **Wirtschaft – Beruf – Ethik** widmet sich ökonomischen und ethischen Fragen im Kontext der beruflichen Aus- und Weiterbildung sowie der Berufs- und Unternehmenskultur.

Gerhard Minnameier (Professor für Wirtschaftsethik und Wirtschaftspädagogik an der Goethe-Universität Frankfurt am Main) und Birgit Ziegler (Professorin für Berufspädagogik an der Technischen Universität Darmstadt) geben die Reihe gemeinsam heraus.



ISBN: 978-3-7639-7366-8