

Stefanie Velten, Stephan Abele, Jürgen Seifried,
Julia Warwas (Hg.)



Berufliche Kompetenzen erfassen und fördern

Impulse und Innovationen aus der
Initiative ASCOT+



Berufliche Kompetenzen erfassen und fördern

Impulse und Innovationen aus der Initiative ASCOT+

Stefanie Velten, Stephan Abele, Jürgen Seifried, Julia Warwas (Hg.)

„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.

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

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: 978-3-7639-7009-4

Band 40

Elisabeth Riebenbauer

Kompetenzentwicklung im Masterstudium Wirtschaftspädagogik

Längsschnittstudie zur Unterrichtsplanung im Rechnungswesen

Bielefeld 2022, ISBN: 978-3-7639-7016-2

Band 41

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

Digitale Transformation in der Berufsbildung

Konzepte, Befunde und Herausforderungen

Bielefeld 2022, ISBN 978-3-7639-7137-4

Band 42

Stephan Abele

Problemlösekompetenzen in beruflichen Kontexten

Resultate aus Lehr-Lern-Prozessen sichtbar machen

Bielefeld 2023, ISBN 978-3-7639-7366-8

Band 43

Klaus Beck, Jürgen Seifried (Hg.)

Berufs- und Wirtschaftspädagogik im selbstkritischen Diskurs

Bielefeld 2023, ISBN: 978-3-7639-7605-8

Band 44

Ulrike Weyland, Wilhelm Koschel (Hg.)

Aktuelle Ansätze und Forschungsbefunde zur

beruflichen Bildung im Gesundheits- und Pflegebereich

Bielefeld 2024, ISBN: 978-3-7639-7407-8

Band 45

Gerhard Minnameier (Hg.)

Ordnonomik und Bildung

Verantwortung für die moderne Gesellschaft

Bielefeld 2025, ISBN: 978-3-7639-7740-6

Weitere Informationen finden
Sie unter **wbv.de/wbe**

Stefanie Velten, Stephan Abele, Jürgen Seifried, Julia Warwas (Hg.)

Berufliche Kompetenzen erfassen und fördern

Impulse und Innovationen aus der
Initiative ASCOT+



2025 wbv Publikation
ein Geschäftsbereich der
wbv Media GmbH & Co. KG, Bielefeld

Gesamtherstellung:
wbv Media GmbH & Co. KG
Auf dem Esch 4, 33619 Bielefeld,
service@wbv.de
wbv.de

Umschlagillustration:
Shutterstock.com/Kev Draws
ISBN (Print): 978-3-7639-7888-5
ISBN (E-Book): 978-3-7639-7889-2
DOI 10.3278/9783763978892

Printed in Germany

Diese Publikation ist frei verfügbar zum Download
unter wbv-open-access.de

Diese Publikation ist unter folgender
Creative- Commons-Lizenz veröffentlicht:
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.

Der Verlag behält sich das Text- und Data-Mining nach
§ 44b UrhG vor, was hiermit Dritten ohne Zustimmung
des Verlages untersagt ist.

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.

Die freie Verfügbarkeit der E-Book-Ausgabe dieser Publikation wurde ermöglicht durch ein Netzwerk wissenschaftlicher Bibliotheken und Institutionen zur Förderung von Open Access in den Sozial- und Geisteswissenschaften im Rahmen der *wbv OpenLibrary 2025*.

Die Publikation beachtet unsere Qualitätsstandards für Open-Access-Publikationen, die an folgender Stelle nachzulesen sind:

https://www.wbv.de/fileadmin/importiert/wbv/PDF_Website/Qualitaetsstandards_wbvOpenAccess.pdf

Großer Dank gebührt den Förderern der *wbv OpenLibrary 2025* im Fachbereich *Berufs- und Wirtschaftspädagogik*:

Bundesinstitut für Berufsbildung (BIBB, **Bonn**) | Carl von Ossietzky Universität **Oldenburg** | Eidgenössische Hochschule für Berufsbildung (Hauptsitz **Bern**) | Fachhochschule **Münster** | Fernuniversität **Hagen** | Georg-August-Universität **Göttingen** | Goethe-Universität **Frankfurt am Main** | Hochschule der Bundesagentur für Arbeit (**Mannheim**) | Humboldt-Universität zu **Berlin** | Justus-Liebig-Universität **Gießen** | Landesbibliothek **Oldenburg** | Otto-Friedrich-Universität **Bamberg** | Pädagogische Hochschule **Freiburg** | Staats- und Universitätsbibliothek **Bremen** | TIB **Hannover** | Universitäts- und Landesbibliothek **Darmstadt** | Universitäts- und Landesbibliothek **Münster** | Universitäts- und Stadtbibliothek **Köln** | Universitätsbibliothek **Kassel** | Universitätsbibliothek **Kiel** | Universitätsbibliothek **St. Gallen** | Zentralbibliothek **Zürich**

Inhalt

Vorwort der Reihenherausgebenden	9
Berufliche Kompetenzen erfassen und fördern – Ausgewählte Ergebnisse der Forschungs- und Transferinitiative ASCOT+	11
<i>Klaus Beck</i> ASCOT+: Der lange und schwierige Weg zur Praxis	17
<i>Detlev Leutner</i> ASCOT+: Eine lehr-lernpsychologische Verortung der Projekte	37
Schwerpunkt 1: Technologiebasierte Förderung beruflicher Kompetenzen	47
<i>Tobias Gschwendtner, Stephan Abele, Inga Glogger-Frey, Stefan Hartmann, Peter Hesse, Jödis Kaepernick & Dave Rexhäuser</i> Digitale Interventionen zur Förderung der Diagnosekompetenz bei Kfz-Mechatroniker:innen	49
<i>Eveline Wittmann, Svenja Hill, Philine Krebs, Susan Seeber, Aldin Striković, Julia Warwas, Ulrike Weyland & Larissa Wilczek</i> Ausbildungsbedingungen in der Pflegepraxis – Erste empirische Befunde zu unterschiedlichen Versorgungsbereichen	71
<i>Christiane Spiel</i> Akademische Bildung oder berufliche Bildung – Kommentar zu Schwerpunkt 1 ..	89
Schwerpunkt 2: Problemlösekompetenz als zentrale Zielgröße der beruflichen Bildung	95
<i>Susan Seeber, Hanna Meiners, Eveline Wuttke, Matthias Schumann & Philipp Hartmann</i> Problemhaltigkeit kaufmännischer Prüfungsaufgaben: Eine Analyse mittels Lautem Denken	97
<i>Esther Winther, Viola Deutscher, Fabio Fortunati, Jürgen Seifried & Andreas Rausch</i> Mit Signaturen Szenarios zum Kompetenzerwerb: Die Analyse von Performance Tasks in virtuellen Simulationen	121

Amory H. Danek & Samuel Greiff

Messung der beruflichen Problemlösekompetenz – Kommentar zu
Schwerpunkt 2 145

Schwerpunkt 3: Kompetenzorientierte Gestaltung von Prüfungsaufgaben 153

Esther Winther, Sebastian Schlünkes, Alexander Kohlus, Michael Kerres & Wolfgang Vogel

Zur Qualität beruflicher Abschlussprüfungen: Psychometrik und praxisnahe
Designstandards – Ausgewählte Befunde des Verbundvorhabens ASPE 155

Elmar Dammann, Pia Schäfer & Felix Walker

Schwierigkeitsbestimmende Aufgabenmerkmale in Prüfungsaufgaben im
technischen Bereich am Beispiel des Ausbildungsberufs Mechatroniker:in 175

Andreas Frey

Kompetenzmessung bei Prüfungen in der beruflichen Bildung – Kommentar
zu Schwerpunkt 3 197

Abkürzungsverzeichnis 203

Autor:innenverzeichnis 205

Vorwort der Reihenherausgebenden

Genau vor zehn Jahren durften wir das Vorwort zum Band 32, dem sogenannten „ASCOT-Band“, der Buchreihe „Beruf-Wirtschaft-Ethik“ verfassen. Band 32 bündelt 13 Beiträge zu theoretischen Grundlagen und Ergebnissen aus dem vom BMBF geförderten Forschungsprogramm ASCOT (Technology Based Assessment of Skills and Competences in Vocational Education and Training), die von Klaus Beck, Margarete Landenberger und Fritz Oser zusammengestellt wurden. In sechs, z. T. interdisziplinär zusammengesetzten Forschungsgruppen waren in mehrjähriger intensiver Arbeit elaborierte Verfahren zur technologiebasierten Kompetenzmessung für jeweils zwei Berufe aus dem gewerblich-technischen, kaufmännisch-verwaltenden und medizinisch-pflegenden Berufsfeld entwickelt und validiert worden. Dies hat nicht nur den Erkenntnisstand zur Modellierung und Messung beruflicher Kompetenzen vorangebracht, sondern es konnten einige Projektmitarbeiter:innen promoviert werden, die heute bereits Professuren bekleiden oder an berufsbildenden Schulen unterrichten. Zudem wurden weitere Forschungsaktivitäten zur Kompetenzdiagnostik angestoßen, wie z. B. zu Themen des nachhaltigen Wirtschaftens oder der Kooperation.

Nun freuen wir uns, einen weiteren „ASCOT-Band“ mit dem Titel „Berufliche Kompetenzen erfassen und fördern – Ausgewählte Ergebnisse der Forschungs- und Transferinitiative ASCOT+“ in die Buchreihe aufnehmen zu dürfen. Wie der Titel schon vermuten lässt, schließt dieser Band an das erste ASCOT-Programm an und präsentiert Beiträge zur technologiebasierten Förderung beruflicher Kompetenzen, zur Erfassung von beruflichen Kompetenzen sowie Ansätzen zur Gestaltung von Prüfungsaufgaben. Im Sinne des Constructive Alignment ist eine Zusammenführung von Lehrzielen, Lernangeboten und Diagnostik sinnvoll und liegt auch im Trend von Learning Analytics, wenngleich diese noch ziemlich am Anfang stehen.

Angesichts der erheblichen fachinhaltlichen Varietät und der internen Komplexität der jeweiligen beruflichen Domänen sieht sich die berufliche Lehr-Lern- bzw. Kompetenzforschung allerdings vor erhebliche Herausforderungen gestellt, soll aber zugleich aufgrund eines hohen und teilweise wissenschaftsfernen Erwartungsdrucks mancher Akteur:innen aus der beruflichen Bildungspraxis an unmittelbare Verwertbarkeit von Forschung möglichst schnell und möglichst maßgeschneiderte Konzeptionen liefern. Sowohl die explizite Ausrichtung der Förderinitiative auf eine praktische Verwertung der Ergebnisse als auch die erhebliche Zeit, die bis zur Auflage einer zweiten Förderphase ins Land ging, sind wohl den komplizierten und nicht ohne Weiteres durchschaubaren Governance-Strukturen des beruflichen Bildungssystems geschuldet. Positiv hervorheben lässt sich jedoch, dass das BMBF überhaupt ein weiteres Förderprogramm aufgelegt und damit einigen Forschungsgruppen der ersten Phase die Möglichkeit eröffnet hat, die digitalen Assessmentumgebungen zu Lernumgebungen weiterzuentwickeln. Fließt doch ansonsten noch immer ein erheblicher Teil von BMBF-Mitteln für berufliche Bildung in Entwicklungsprojekte, die nicht selten auf un-

zureichend elaborierten theoretischen Grundlagen basieren und wenig zur Forschung beitragen.

Im Spannungsfeld zwischen Forschungs- und Verwertungsinteressen mussten sich die sechs Projektteams von ASCOT+ behaupten. Ein Blick in die Beiträge offenbart, dass sie diesen Spagat mit sehr viel Engagement und unter Wahrung hoher wissenschaftlicher Standards gemeistert haben. Erneut konnten sich in den Projekten sowohl junge Wissenschaftler:innen qualifizieren als auch relevante Erkenntnisfortschritte ermittelt werden, die skalierbar sind und über die unmittelbare Nutzungsperspektive hinausweisen. Wir danken Stefanie Velten, Stephan Abele, Jürgen Seifried und Julia Warwas für die Zusammenstellung der Beiträge und wünschen allen Rezipient:innen eine anregende und gewinnbringende Lektüre der Forschungsbeiträge.

Frankfurt und Darmstadt im Juli 2025

Gerhard Minnameier und Birgit Ziegler

Berufliche Kompetenzen erfassen und fördern – Ausgewählte Ergebnisse der Forschungs- und Transferinitiative ASCOT+

Einführung in den Sammelband

Als „Forschungs- und Transferinitiative“¹ zielte die vom Bundesministerium für Bildung und Forschung (BMBF) in den Jahren 2019 bis 2023 geförderte Initiative ASCOT+ darauf ab, „bestehende Grundlagen und Ergebnisse (u. a. die validen Testverfahren und -items) aus der Vorgängerinitiative [ASCOT, Laufzeit: 2011–2015, die Verf.] aufzugreifen, einer breiteren Nutzung in der Praxis zuzuführen und in ausgewählte Anwendungs- und Forschungsfelder [...] zu transferieren“ (BMBF, 2018). Das BMBF förderte zu diesem Zweck insgesamt sechs Projekte aus den Berufsfeldern gewerblich-technische, kaufmännische und Gesundheitsberufe. Die Initiative ASCOT+ sollte einen Beitrag dazu leisten,

- Lehr-Lernprozesse in der betrieblichen und berufsschulischen Ausbildung zu verbessern,
- Leistungen von Auszubildenden objektiver festzustellen,
- Akteur:innen in der Ordnungsarbeit bei der kompetenzorientierten Formulierung von Ausbildungsordnungen zu unterstützen sowie
- kompetenzorientierte Prüfungen weiterzuentwickeln.

Dabei steht das „+“ für die Leitidee, zentrale Erkenntnisse aus den Projekten in die Ausbildungs- und Prüfungspraxis zu tragen. Dafür war der Einbezug von Praxispartner:innen bereits in der Phase der Projektkonzeptionen obligatorisch. In Erweiterung zur Vorgängerinitiative ASCOT wurden bei diesem Forschungs- und Entwicklungsprogramm nunmehr also Forschung *und* Transfer in den Blick genommen. Damit wurde u. a. eine Forderung des Wissenschaftsrats aufgegriffen, der anregt, neben den Kernaufgaben der Forschung und Lehre auch vermehrt den Transfer, d. h. die „Vermittlung und Übertragung wissenschaftlicher Erkenntnisse aus allen Wissenschaftsbereichen in Gesellschaft, Kultur, Wirtschaft und Politik“ (Wissenschaftsrat, 2013, S. 26) anzustreben. Ausgehend von vorliegenden Domänenmodellen und Testinstrumenten (vgl. Abele et al., 2021, sowie die Übersicht über die ASCOT-Projekte in Beck et al., 2016) wurden in den verschiedenen Projekten also praxisrelevante und praxistaugliche digitale Instrumente sowie Angebote für die Prüfungspraxis entwickelt. Hierzu wurden bestehende, i. d. R. simulationsbasierte Assessmentverfahren herangezogen und zu Lehr-Lernumgebungen weiterentwickelt. Zudem wurden neue Assessments erarbeitet und flankierend Angebote für Prüfende und Prüfungsaufgabenersteller.

1 ASCOT steht als Akronym für „Technology-based Assessment of Skills and Competences in VET (Vocational Education and Training)“.

lende entwickelt. Einen umfassenden Überblick über die verschiedenen Projekte bietet eine Broschüre für die Zielgruppen der Ausbildungspraxis und des Prüfungswesens (BMBF, 2024). Ergänzende Darstellungen und Einblicke in die Projektarbeit und ihre Resultate sind auf der Website www.ascot-vet.net für jedes Projekt ausgewiesen.

Der nun vorliegende Sammelband stellt weitere ausgewählte Forschungsergebnisse aus den unterschiedlichen Projekten vor. Er baut auf den in ASCOT entwickelten Domänenmodellen auf (vgl. Beck et al., 2016) und knüpft an die Überlegungen von Abele et al. (2021) an, die in den Ausbildungsordnungen und Rahmenlehrplänen curricular verankerten Lehr-Lernziele enger mit Kompetenzmessverfahren und instruktionalen Förderkonzepten zu verzahnen und aufeinander abzustimmen.

Im Fokus des nun vorliegenden Bands stehen nunmehr die Bereiche des betrieblichen und schulischen Lernens während der Ausbildung sowie der Kompetenzmessung im Prüfungswesen. Einige Projekte liefern darüber hinaus Ansätze und Impulse, wie diese beiden bislang eher isoliert betrachteten Bereiche stärker miteinander verzahnt werden können.

Die in den verschiedenen Projekten für unterschiedliche Berufsfelder erarbeiteten Instrumente und Erkenntnisse werden in diesem Sammelband entlang dreier thematischer Schwerpunkte präsentiert: (1) der technologiebasierten Förderung beruflicher Kompetenzen, (2) der Problemlösekompetenz als zentraler Zielgröße der beruflichen Bildung sowie (3) der kompetenzorientierten Gestaltung von Prüfungsaufgaben. Jeder dieser Schwerpunkte umfasst zwei Beiträge aus verschiedenen ASCOT+-Projekten. Wir freuen uns sehr, dass es gelungen ist, zu jedem dieser drei Themenbereiche weiterführende Kommentare und Meinungen von äußerst renommierten Forscher:innen einzuholen, die die ASCOT+-Beiträge (auch aus einer persönlichen Perspektive) würdigen und den Diskurs damit bereichern.

Schwerpunkt 1 wird von *Christiane Spiel* kommentiert. Zu Schwerpunkt 2 äußern sich *Amory Danek* und *Samuel Greiff*. Schwerpunkt 3 schließlich wird durch einen Kommentar von *Andreas Frey* abgerundet. Übergreifend werden die verschiedenen Beiträge durch zwei weitere Kommentare gerahmt. Zur Genese von ASCOT+ äußert sich *Klaus Beck*, und *Detlev Leutner* nimmt eine instruktionspsychologische Einordnung vor. Zur Einstimmung und Orientierung gehen wir kurz auf die Beiträge in den drei Schwerpunkten, die sich jeweils darauf beziehenden Kommentare sowie die den Sammelband rahmenden Kommentierungen, ein.

Zur Genese von ASCOT+

Klaus Beck umreißt zu Beginn des Sammelbands die Entstehungsgeschichte von ASCOT+. Dabei ordnet er das Projekt kritisch in übergeordnete bildungs- und forschungspolitische Kontexte ein und adressiert die in diesem Zusammenhang wichtige Unterscheidung zwischen einem Forschungsprogramm und einer Programmforschung. Zugleich würdigt er die empirischen und methodischen Beiträge von ASCOT+, etwa in der simulationsbasierten Kompetenzdiagnostik. In seinem Kom-

mentar plädiert *Klaus Beck* für eine stärker theoretisch fundierte und wissenschaftsorientierte Ausrichtung zukünftiger Kompetenzforschung.

Eine lehr-lernpsychologische Verortung

Detlev Leutner würdigt die ASCOT+-Projekte im Allgemeinen sowie die in diesem Band präsentierten Beiträge im Besonderen und liefert eine instruktionspsychologische Einordnung anhand zentraler Komponenten der Lehr-Lern-Forschung. *Detlev Leutner* würdigt die theoriebasierte Konzeption und empirische Validierung der digitalen Lern- und Testumgebungen, sieht aber Optimierungsbedarf bei der Berücksichtigung individueller Lernvoraussetzungen. Sein bilanzierender Kommentar betont zudem die Relevanz der ASCOT+-Ergebnisse weit über die berufliche Bildung hinaus – insbesondere für kompetenzorientierte und adaptive Lernsysteme.

Schwerpunkt 1: Technologiebasierte Förderung beruflicher Kompetenzen

Das zentrale Ziel einer dualen Ausbildung besteht in der Entwicklung beruflicher Handlungskompetenzen. *Tobias Gschwendtner, Stephan Abele, Inga Glogger-Frey, Stefan Hartmann, Peter Hesse, Jödis Kaepernick* und *Dave Rexhäuser* untersuchen vor dem Hintergrund dieser Prämisse, welche Effekte digitale Interventionen zur Förderung der Diagnosekompetenz von Kfz-Mechatroniker:innen haben. Drei gezielt variierte Fördermaßnahmen aus dem Projekt „Digitale Diagnostik und Intervention im Kfz-Wesen“ (DigiDIn-Kfz) adressieren die basale und komplexe Rezeptionskompetenz sowie modellbasierte Diagnosestrategien. Es zeigen sich signifikante Lernzuwächse in verschiedenen Facetten der Diagnosekompetenz, während der direkte Diagnoseerfolg in der Simulation nur teilweise gesteigert werden konnte. Die Befunde belegen die Potenziale digitaler Lernumgebungen, verweisen aber auch auf weiteren Forschungsbedarf, u. a. mit Blick auf die langfristigen Wirkungen solcher Interventionen.

Im Beitrag aus dem Projekt „Erweiterte Kompetenzmessung im Gesundheitsbereich“ (EKGe) gehen *Eveline Wittmann, Svenja Hill, Philine Krebs, Susan Seeber, Aldin Striković, Julia Warwas, Ulrike Weyland* und *Larissa Wilczek* der Frage nach, wie sich wahrgenommene Ausbildungsbedingungen in unterschiedlichen Versorgungsbereichen der Pflege auf drei Facetten beruflicher Handlungskompetenz von angehenden Pflegefachkräften auswirken. Dabei zeigen sich signifikante, wenngleich schwache Zusammenhänge insbesondere für negatives Feedback durch zu pflegende Personen sowie für Merkmale der besuchten Pflegeeinrichtung, darunter funktionale Einbindung und Anforderungskomplexität. Zudem unterscheiden sich die wahrgenommenen Ausbildungsbedingungen zwischen den Versorgungsbereichen, wobei die stationäre Altenpflege eher ungünstig bewertet wird. Die Ergebnisse unterstreichen die Bedeutung lernrelevanter Kontextfaktoren der praktischen Ausbildung und weisen auf Gestaltungsbedarf insbesondere im Altenpflegebereich hin.

In ihrem Kommentar nimmt *Christiane Spiel* die beiden empirischen Studien zur Kompetenzentwicklung zum Anlass, um das Spannungsverhältnis zwischen akademischer und beruflicher Bildung grundlegend und kritisch zu reflektieren. Sie hebt hervor, dass viele junge Menschen aufgrund gesellschaftlicher und familiärer Prägungen die berufliche Bildung kaum kennen und deren Lern- und Karrierepotenziale unterschätzen. Mit Blick auf die beiden Studien zeigen sich – trotz bestehender Unterschiede hinsichtlich ihrer Domäne, Zielsetzungen und Forschungsmethoden – doch einige Gemeinsamkeiten. Auf deren Basis schlägt *Christiane Spiel* begründete Vorschläge für die Gestaltung der Berufsbildungspraxis vor, zu denen etwa die Implementierung von gestuften Ausbildungszielen gehört. Abschließend plädiert die Autorin für eine stärkere Sichtbarmachung hybrider Bildungswege, um der strukturellen Schiefelage zwischen akademischer und beruflicher Bildung entgegenzuwirken.

Schwerpunkt 2: Problemlösekompetenz als zentrale Zielgröße der beruflichen Bildung

Susan Seeber, Hanna Meiners, Eveline Wuttke, Matthias Schumann und Philipp Hartmann untersuchen in ihrem Beitrag die Problemhaltigkeit kaufmännischer Prüfungsaufgaben mittels Lautem Denken. Ausgehend von einem von den Beteiligten des Projekts „Technologiebasiertes kompetenzorientiertes Prüfen“ (TeKop) entworfenen Training zur Erstellung von (Prüfungs-)Aufgaben im kaufmännischen Bereich und den dort zu Übungszwecken entwickelten Aufgaben zeigen sie auf, welche Denkprozesse Auszubildende bei deren Bearbeitung durchlaufen. Dabei werden kognitive Prozesse wie Informationsverarbeitung, Strategieeinsatz und Reflexion sichtbar gemacht. Die Studienergebnisse zeigen, dass es den Trainingsteilnehmenden gelungen ist, problemhaltige Prüfungsaufgaben zu entwickeln.

Wie die Gestaltung des Lehrens und Lernens in der Ausbildung mit der Diagnostik von Lernleistungen verknüpft werden kann, zeigen *Esther Winther, Viola Deutscher, Fabio Fortunati, Jürgen Seifried und Andreas Rausch* in ihrem Beitrag zu Signature Scenarios zum Kompetenzerwerb. Sie stellen ihre kaufmännische Bürosimulation LUCA aus dem Projekt „Problem Solving Analytics in Office Simulations“ (PSA-SIM) vor, die schwerpunktmäßig dazu dient, anhand komplexer und prototypischer Arbeitsszenarien, den sogenannten Signature Scenarios, die beruflichen Problemlösekompetenzen kaufmännischer Auszubildender zu fördern. Anhand eines Beispielszenarios wird in dem Beitrag herausgearbeitet, wie neben Indikatoren zum Problemlöseprozess auch sprachlich-argumentative und mathematisch-analytische Kompetenzindikatoren für die Kompetenzmessung genutzt werden können. Die berichteten Befunde verweisen darauf, dass adaptive Hilfen durch simulierte Kolleg:innen die Kompetenzentwicklung positiv beeinflussen und zur Bewältigung kognitiver Hürden beitragen. Die Transformation von Prozessdaten in Bearbeitungsmuster erlaubt weiterführend eine differenzierte Leistungsdiagnostik entlang eines zweifaktoriellen Kompetenzmodells.

Der Kommentar von *Amory Danek* und *Samuel Greiff* analysiert die Beiträge dieser beiden Studien zu Fragen der beruflichen Problemlösekompetenz – eine (nicht nur aus Sicht der Kommentator:innen) Schlüsselkompetenz für die Arbeitswelt von morgen. *Danek* und *Greiff* betonen, dass analytisches Problemlösen in der beruflichen Bildung differenziert messbar ist. Allerdings bleiben Herausforderungen wie psychometrische Qualität und Generalisierbarkeit bestehen. Der Kommentar plädiert für eine stärkere Integration von Querschnittskompetenzen wie Problemlösen in Curricula und Prüfungen sowie für weiterführende Forschung zur Validität simulationsgestützter Assessments.

Schwerpunkt 3: Kompetenzorientierte Gestaltung von Prüfungsaufgaben

Die beiden Beiträge von *Esther Winther*, *Sebastian Schlünkes*, *Alexander Kohlus*, *Michael Kerres* und *Wolfgang Vogel* sowie von *Elmar Dammann*, *Pia Schäfer* und *Felix Walker* widmen sich der Ausgestaltung von Prüfungsaufgaben in zwei unterschiedlichen beruflichen Bereichen. *Winther* et al. stellen das Kernelement aus dem Projekt „Digitale Workbench für kompetenzorientierte Prüfungsaufgaben und Abschlussprüfungen – Assessments for Professional Exams“ (ASPE) vor, die für die Zielgruppe der Prüfungsaufgabenerstellenden konzipiert wurde. Es handelt sich dabei um ein digitales Tool zur professionellen und kompetenzorientierten Erstellung kaufmännischer Abschlussprüfungen, dessen zentrale, auf wissenschaftlichen Erkenntnissen basierende Gestaltungskriterien im Beitrag skizziert werden. Ziel des Projekts war es, die psychometrische Qualität von Prüfungsaufgaben durch die Beachtung von Designkriterien wie Wissenskomplexität, kognitive Aktivierung und funktionale Modellierung zu steigern. Diese Kriterien ermöglichen eine vorausschauende Einschätzung der Aufgabenschwierigkeit und die Umsetzung valider, reliabler und fairer Prüfungen.

Der Beitrag von *Elmar Dammann*, *Pia Schäfer* und *Felix Walker* geht der Frage nach, wie formale Gestaltungsmerkmale von Prüfungsaufgaben wie etwa die Kohärenz zwischen Bild- und Textinformationen die Schwierigkeit dieser Prüfungsaufgaben für angehende Mechatroniker:innen bestimmen. Anhand von Materialien und Daten aus dem Projekt „Technologiebasierte Kompetenzmessung und -förderung in der elektrotechnischen und metalltechnischen Erstausbildung“ (TechKom) wird analysiert, wie sich unterschiedliche Ausprägungen dieser Merkmale spezifisch auf die Bearbeitungsleistung der Prüflinge auswirken. Die Ergebnisse zeigen ein differenziertes Bild mit überwiegend kleinen bis mittleren Effekten, die auf die Notwendigkeit weiterer systematischer Analysen verweisen.

In seinem Kommentar würdigt *Andreas Frey* die beiden von ihm in den Blick genommenen ASCOT+-Projekte „TechKom“ und „ASPE“ als gelungene Beispiele für den Transfer bildungswissenschaftlicher Erkenntnisse in die berufliche Prüfungspraxis. Er betont die Bedeutung psychometrischer Qualität und konzeptueller Klarheit bei der Kompetenzdiagnostik und lobt insbesondere die praktische Nutzbarkeit der ent-

wickelten Instrumente und Modelle. Gleichzeitig weist er auf die zukünftigen Herausforderungen und Potenziale durch den Einsatz Künstlicher Intelligenz hin – sowohl hinsichtlich der sich wandelnden beruflichen Anforderungen als auch der Weiterentwicklung von Prüfungen.

Abschließend kann festgehalten werden, dass die hier exemplarisch gezeigten Befunde der insgesamt sechs ASCOT+-Projekte als Beleg für die Wirkung der Forschungs- und Transferinitiative ASCOT+ gelten können. Ein Blick von außen kann weiterführend dazu beitragen, neue Impulse zu erhalten und Perspektiven zu eröffnen. Die Kommentare bieten diesbezüglich gehaltvolle Einordnungen sowohl der gesamten Initiative ASCOT+ als auch der einzelnen im Sammelband erscheinenden Forschungsbeiträge.

Daher bedanken wir uns sehr herzlich bei den Kommentator:innen, dass sie sich trotz ihrer hohen beruflichen Belastung die Zeit genommen haben, die Beiträge aus den unterschiedlichen ASCOT+-Projekten vor dem Hintergrund ihrer Expertise und ihrer Erfahrung zu reflektieren. Damit haben sie nicht nur den Autor:innen, sondern all jenen wertvolle Anregungen gegeben, die sich mit Fragen der beruflichen Kompetenzdiagnostik und -förderung befassen. Wir danken zudem allen Autor:innen für ihre Beiträge, die allesamt pünktlich geliefert wurden, sodass der ehrgeizige Zeitplan unseres Buchprojekts eingehalten werden konnte. Zudem danken wir den Autor:innen für die konstruktive Aufnahme und Verarbeitung von Hinweisen zu den Beiträgen sowie die kollegiale Zusammenarbeit. Schließlich gilt unser Dank unseren Ansprechpartner:innen von wbv media und dabei insbesondere *Frauke Heilmann* und *Cira Korfmacher* für die gewohnt sehr gute und konstruktive Zusammenarbeit.

Stefanie Velten, Stephan Abele, Jürgen Seifried und Julia Warwas

Bonn, Dresden, Mannheim und Hohenheim, im Juli 2025

Literatur

- Abele, S., Deutscher, V., Nickolaus, R., Rausch, A., Seeber, S., Seifried, J., Walker, F., Weyland, U., Winther, E., Wittmann, E., & Wuttke, E. (2021). Potenziale technologiebasierter Kompetenztests der ASCOT-Initiative aus der Perspektive der Curriculum-Instruction-Assessment-Triade. In K. Beck, & F. Oser (Hrsg.), *Resultate und Probleme der Berufsbildungsforschung. Festschrift für Susanne Weber* (S. 13–42). wbv.
- Beck, K., Landenberger, M., & Oser, F. (Hrsg.) (2016). *Technologiebasierte Kompetenzmessung in der beruflichen Bildung. Ergebnisse aus der BMBF-Förderinitiative ASCOT*. wbv. <https://doi.org/10.3278/6004436w>
- BMBF (2024). *Digitale Tools für die Berufsbildungspraxis – Lehren, Lernen und prüfen mit ASCOT+*. Verfügbar unter https://www.ascot-vet.net/ascot/shareddocs/downloads/files/Praxisbroschuere_ASCOTplus.pdf?blob=publicationFile&v=6
- Wissenschaftsrat (2013). *Perspektiven des deutschen Wissenschaftssystems. Drs. 3228–13*. Verfügbar unter <https://www.wissenschaftsrat.de/download/archiv/3228-13>

ASCOT+: Der lange und schwierige Weg zur Praxis

KLAUS BECK

Zusammenfassung

Die Förderung erziehungswissenschaftlicher Forschung durch das BMBF ist hierzulande längst unverzichtbar geworden. Wichtige Fragestellungen, auch solche der beruflichen Bildung, konnten dank dieser Unterstützung untersucht und beantwortet werden. Ministerien als politische Instanzen folgen aber bei der Mittelvergabe einer eigenen „Logik“, die sich auch an Gesichtspunkten von Aktualität, Themen des öffentlichen Diskurses oder Legitimierungsbedarfen orientiert.

Dieser Beitrag zeichnet ein Stück weit das Zustandekommen der BMBF-Forschungsinitiative ASCOT+ nach, rekonstruiert sie im Zusammenhang forschungslogischer Konzeptualisierungen und erörtert einige Fragen des geforderten Praxistransfers und des weiteren Forschungsbedarfs.

Schlagworte: Berufliche Kompetenzen, Praxistransfer, Programmforschung, Forschungsprogramm

Abstract

The BMBF's funding of educational research has long since become indispensable in Germany. Thanks to this support, it has been possible to investigate and answer important questions, including those relating to vocational education. However, ministries as political authorities follow their own “logic” when allocating funds, which is also based on aspects of timeliness, topics of public discourse or the need for legitimacy.

This article traces the origins of the BMBF research initiative ASCOT+, reconstructs it in the context of research logic conceptualizations and discusses some questions of the required transfer into practice and the need for further research.

Keywords: Professional Competencies, Practice Transfer, Program Research, Research Program

1 Zur Genealogie der „Forschungs- und Transferinitiative ASCOT+“

Wer die Resultate von ASCOT+ angemessen würdigen will, tut gut daran, sich des Kontexts zu vergewissern, aus dem heraus diese vom BMBF finanzierte „Initiative“

erwachsen ist. Tatsächlich verbinden sich mit ihr nämlich nicht lediglich die üblichen Erwartungen an Berichte über Verfahren und Befunde empirischer Forschungsprojekte. Es ranken sich um sie herum auch Hoffnungen und Pläne sowie andererseits Skepsis, Vorbehalte und Zweifel, welche über die unter diesem „Dach“ durchgeführten Vorhaben weit hinausweisen. Um davon einen Eindruck zu gewinnen und zugleich zu verstehen, wie es kommt, dass hier jenseits von Erkenntnisinteressen noch ganz andere Interessen im Spiel sind, lohnt ein Blick zurück auf die bildungspolitisch unruhige Zeit um die vergangene Jahrhundertwende.

Beherrscht wurde sie zunächst ja vor allem vom „Schock“, den die Publikation der für Deutschland unvorteilhaften Resultate der ersten PISA-Studie ausgelöst hatte (Baumert et al., 2001). Ihr zufolge wiesen die 15-jährigen Schüler:innen hierzulande im Vergleich mit vielen anderen Ländern schwache Leistungen auf. Damit stand das Thema „Bildung“ plötzlich im Fokus öffentlicher Aufmerksamkeit und im Zentrum von zum Teil heftigen politischen Auseinandersetzungen, in die sich besonders die zuständigen Bundesländer verwickelt sahen (Althaus et al., 2002). Zugleich jedoch erwies sich das PISA-Projekt als Initialzündung für eine große Zahl von ertragreichen Forschungsaktivitäten, die sich auf die Erklärung der Befunde und auf Möglichkeiten ihrer Beeinflussung durch eine verbesserte pädagogische Praxis richteten.

Für Unruhe sorgte etwa zur gleichen Zeit auch die sogenannte Bologna-Reform im Bereich der akademischen Bildung, die schon 1999 von Bildungsminister:innen aus 29 (heute 49) europäischen Ländern beschlossen, in Deutschland im August 2002 unter der Ägide von Bundesbildungsministerin Edelgard Bulmahn (SPD) in Kraft gesetzt worden war. Mit dieser Maßnahme sollte ein einheitlicher europäischer Hochschulraum geschaffen werden, von dem man sich neben der umstandslosen länderübergreifenden wechselseitigen Anerkennung der akademischen Abschlüsse (Beck, 1995) auch eine Erleichterung der *transnationalen* Mobilität von Studierenden erhoffte, die in Deutschland freilich schon *intranational* als durchaus verbesserungsbedürftig angesehen worden war (Kühl, 2015).

Beide, die PISA-Studie und die Bologna-Reform, befeuerten das Interesse an einer bildungspolitischen Lagebestimmung Deutschlands im internationalen europäischen und darüber hinaus auch im außereuropäischen Feld. Und sie bewegten die Gemüter nicht nur der unmittelbar Beteiligten in ganz erheblichem Maße, die sich, getragen von teilweise starken Emotionen, in Rollen von Befürworter:innen und Gegner:innen, Angeklagten und Verteidiger:innen, Bewahrer:innen und Reformier:innen wiederfanden. Bis heute bestehen diese Differenzen fort, wenngleich der offen ausgetragene Streit abgeebbt zu sein scheint (Winter, 2015; KMK et al., 2024).

In die Zeit der frühen 2000er-Jahre fiel nun in Deutschland auch noch die Vorbereitung und Beschlussfassung (2005) des zuletzt 1981 novellierten Berufsbildungsgesetzes (BBiG). Bei diesem Vorhaben stand das Thema Internationalisierung ebenfalls als ein diskussionsbedürftiges Anliegen auf der Tagesordnung, dies insbesondere unter der Frage, ob berufliche Ausbildungsphasen im Ausland überhaupt und in welchen Anteilen anzuerkennen seien (Klubertz, 2003).

Flankiert wurden solche Überlegungen durch Beiträge aus dem wenige Jahre zuvor installierten mächtigen Europäischen COST-Programm („European Cooperation in Science and Technology“), das von 1998 bis 2002 arbeitete. In dessen Sub-Programmbereich „COST A11“, der von F. Achtenhagen (DE), W. Nijhof (Twente, NE) und F. Coffield (Newcastle, GB) geleitet wurde, waren unter Beteiligung von 16 europäischen Ländern Fragen zu „Flexibility, transferability, mobility as targets of vocational education and training“ bearbeitet worden. Die Förderung für das COST-Programm wurde am Ende der ersten Phase nicht weitergeführt. Seine umfassende Aufgabenstellung freilich, nämlich die einer vergleichenden Gegenüberstellung der vielfältigen, in den europäischen Ländern gewachsenen Berufsbildungskonzeptionen, bestand weiter fort. Dies zeigte sich später auch in weiteren von der EU vorangetriebenen Maßnahmen, nämlich der Einführung des EQF (2008) und des ECVET-Systems (2009), sodass man im Blick auf den Anfang des Jahrhunderts nachgerade von einer eurozentrischen – um nicht zu sagen: eurokratischen – Epoche der Bildungs- und insbesondere der Berufsbildungspolitik sprechen kann.

Es dürfte angesichts dieser Gemengelage europäischer Aktivitäten kaum überraschen, dass nun, mit Nachdruck zunächst in Deutschland, die Frage nach der Möglichkeit aufgeworfen wurde, im Bereich der Berufsbildung ebenfalls eine international vergleichende Studie im Blick auf die dort zu erbringenden bzw. erbrachten Leistungen aufzulegen – ein „Berufsbildungs-PISA“ sozusagen. Und so wurde es zunächst auch offiziell genannt.

Da fügte sich günstig, dass in Göttingen mit dem „Soziologischen Forschungsinstitut“ (SOFI) unter der Leitung von Martin Baethge und dem „Seminar für Wirtschaftspädagogik“ unter der Leitung von Frank Achtenhagen, beide angesiedelt an der Georg-August-Universität Göttingen, die erforderliche wissenschaftliche Kompetenz und Erfahrung versammelt war, die ein solches disziplinenübergreifendes Vorhaben vorantreiben konnte. Nach einer BMWA-seitigen Ausschreibung für ein entsprechendes Forschungsprojekt erfolgte an diese beiden Instanzen der Auftrag zur Erstellung einer „Machbarkeitsstudie für ein Berufsbildungs-PISA“. Es ist diese Studie, aus der – gegen alle Erwartung – das ASCOT-Programm hervorgehen sollte.

Eine interministerielle Arbeitsgruppe aus BMWA, BMBF und BIBB begleitete das Vorhaben, in das im Laufe der Bearbeitung weitere nationale und internationale Wissenschaftler:innen sowie institutionelle Kooperationspartner:innen einbezogen werden konnten (vgl. Baethge et al., 2006, III–IV). Am Ende lag – exemplarisch für sechs Berufe (Kfz-Mechaniker:in, Industriemechaniker:in, Elektroniker:in, Bankkaufleute, Krankenpfleger:in, Fachinformatiker:in) – eine im erforderlichen Detail, von der curricularen Verankerung bis zu den komplizierten Fragen der Stichprobenziehung, vollständig ausgearbeitete „Feasibility Study“ vor (Baethge et al., 2006), an deren Umsetzung acht europäische Länder Interesse bekundet hatten (Dänemark, Deutschland, Finnland, Norwegen, Österreich, Schweden, Schweiz, Slowenien; Baethge et al., 2006, S. 116, S. 130; Achtenhagen & Baethge, 2007; Baethge & Arends, 2008, S. 32; 2009, S. 492).

Doch nun geschah etwas Merkwürdiges. Weil das BIBB als Teil der für die Machbarkeitsstudie eingerichteten Arbeitsgruppe und zugleich als institutioneller Partner für die spätere Steuerung und Durchführung der Vergleichsstudie als künftiger deutscher Projektträger eingebunden war, gehörte es zu seiner Aufgabe, dieses Vorhaben in seinem für solche Angelegenheiten zuständigen Hauptausschuss behandeln zu lassen (§ 92 BBiG). Dort erhoben sich in dessen Sitzung am 13. Dezember 2007 erhebliche Bedenken gegen die Durchführung eines „Berufsbildungs-PISA“ (BIBB, 2007; als Protokollant: Spillner, 2008). Es *„wurde die Sorge geäußert, dass es aufgrund des Untersuchungsdesigns und der vorgesehenen Messinstrumente zu einer Fehleinschätzung der Leistungen der deutschen Berufsbildung im internationalen Vergleich kommen könne“* (Spillner, 2008, S. 55), ja, es drohe die Benachteiligung der deutschen Berufsausbildung (vgl. BIBB, 2007, Nr. 2): Länder mit einer der deutschen vergleichbaren Berufsausbildung seien in Europa in der Minderheit. Bei ihnen wiege eine eher „schulisch-akademische“ Ausbildung vor und gerade auf deren Inhalte zielen die geplanten Berufsbildungs-PISA. Wollte man die Leistungen des deutschen Verfahrens angemessen erfassen, erfordere dies *„jedoch eine Berufsbildungs-PISA-Initiative mit einer anderen inhaltlichen Gewichtung und methodischen Ausrichtung, als es in den bisher ... präsentierten konzeptionellen Vorarbeiten zu erkennen ist“* (ebd. Nr. 3).

In zwei weiteren Sitzungen, nämlich am 18. Dezember 2008 und erneut am 25. Juni 2009 – beide in Reaktion auf die Veröffentlichung von Resultaten eines im Oktober 2008 abgehaltenen AG BFN-Forums (Münk & Schelten, 2010), innerhalb dessen sich starke Stimmen für das VET-LSA erhoben hatten – bekräftigte der Hauptausschuss weiterhin seine ablehnende Haltung. Zwar „begrüßt [er] die erreichten Fortschritte bei der beruflichen Kompetenzmessung. Dennoch sieht er seine damals formulierten Anforderungen bislang als nicht hinreichend erfüllt“ (Spillner, 2009; BIBB, 2009).

Die vehemente Ablehnung dieses Vorhabens insbesondere seitens der Arbeitgebervertretung und auch der Arbeitnehmerseite – also sozusagen „durch die Bank“ – blieb nicht etwa folgenlos. Angesichts der offenbar strittigen Lage verabschiedete sich das BMWA (seit 2005 als BMWi) aus seinem Engagement und das BMBF, in dem, wie sich später zeigte, durchaus Befürworter:innen der Durchführung eines VET-LSA vertreten waren, zog sich zunächst ebenfalls zurück. Offenbar behielt man aber dort das Anliegen im Auge und konnte es mit programmatischer und politischer Unterstützung durch das Göttinger SOFI in neuer und allerdings radikal reduzierter Gestalt lancieren. Beigetragen hat dazu wahrscheinlich auch ein Strategiepapier, das von einer Gruppe aus Berufs- und Wirtschaftspädagog:innen als programmatischer Aufriss für ein Forschungsprogramm zur „Kompetenzdiagnostik in der Berufsbildung“ im Oktober 2009 beschlossen und bekannt gemacht wurde (Seeber et al., 2009). In ihm entfalteten die Autor:innen eine umfassend begründete Konzeption für ein Schwerpunktprogramm, wie es von der DFG hätte gefördert werden können. Dazu kam es jedoch nicht.

Am 15.02.2011 erfolgte die Ausschreibung einer Forschungsinitiative des BMBF unter der Bezeichnung *„Entwicklung und Erprobung von technologieorientierten Mess-*

instrumenten zur Feststellung der beruflichen Handlungskompetenz am Ende der Ausbildung in ausgewählten Berufen auf nationaler Ebene“ (Homepage – Bekanntmachung – BMBF; 06.04.2025). Der Zusatz am Ende ist ein Verweis darauf, dass diese Ausschreibung in einem inhaltlichen Bezug zum ursprünglich geplanten „Berufsbildungs-PISA“ stand.

Als international kommunizierbarer Arbeitstitel diente die Formulierung „Technology-based Assessment of Skills and Competences in Vocational Education and Training“ (Impulse für die Ausbildung der Zukunft | Innovative Kompetenzmessung in einer dynamischen Arbeitswelt, S. 7). Diese Bezeichnung signalisierte ebenfalls eine Anknüpfung an das in der Versenkung verschwundene „Large Scale Assessment in Vocational Education and Training“ (VET-LSA); und sie erfuhr eingängigen Ausdruck in dem hinzugefügten polyvalenten Akronym „ASCOT“, das im SOFI von Lena Arends vorgeschlagen worden war.

Im Verlauf der *Kick-off*-Veranstaltung für ASCOT verlautete seitens des BMBF, dass man mit dieser Initiative den Versuch unternehmen wolle, Forschungsprojekte zu fördern, die auch im Rahmen eines VET-LSA teils vorbereitende, teils weiterführende Fragestellungen hätten verfolgen sollen. Zu ihnen rechne die mit dieser „Initiative“ zu fördernde theoriegeleitete Modellierung von beruflichen Kompetenzen sowie die Erprobung entsprechender Messverfahren. Es sei weiterhin vorgesehen, in zwei anschließenden Schritten diese Instrumente in Felduntersuchungen zur Erklärung und Wirkung der empirisch erfassten Kompetenzen einzusetzen und im dritten Schritt die gesammelten Resultate im Wege von Transferprojekten der Praxis zugänglich zu machen (BMBF, 2014, S. 26).

Von 2011 bis 2015 arbeiteten die geförderten Projekte in Bereichen und Berufen, wie sie in etwa bereits von der Machbarkeitsstudie vorgesehen waren, nämlich im kaufmännisch-verwaltenden, im gewerblich-technischen und im gesundheits- und pflegebezogenen Feld, ergänzt um Studien zu den Entwicklungsbedingungen der zu messenden Kompetenzen im Rahmen der beruflichen Erstausbildung (Beck, Landenberger & Oser, 2016a). Ein wissenschaftlicher Beirat (mit K. Beck, M. Landenberger und F. Oser) begleitete diese Arbeiten und unterstützte im Fortgang des Forschungsprozesses projektübergreifende Treffen für den Austausch zwischen den Arbeitsgruppen.

Bekanntlich profitieren größer angelegte Forschungsprogramme substantiell davon, dass ihre nahtlose Fortführung, auch wenn sie aus verwaltungstechnischen Gründen in Projektperioden organisiert werden müssen, sichergestellt wird, weil andernfalls das umfangreiche und spezialisierte Know-how mit dem Weggang des involvierten Personals verlorengeht. Woran es gelegen hat, dass dies bei ASCOT nicht gelang, lässt sich allenfalls vermuten, aber extern nicht mehr aufklären. Und tatsächlich sind in der Folgezeit viele involvierte Mitarbeitende ausgeschieden oder abgewandert.

Letztlich aufklären lässt sich auch nicht, wie es im BMBF nach drei Jahren dann doch zu einer erneuten Ausschreibung unter dem Namen „ASCOT+“ kam (BMBF, 2018). Mit dieser Etikettierung war immerhin eine klare Anknüpfung an die Vorgängerinitiative signalisiert, die sich insoweit zwar nicht ausschließlich, aber doch vor-

nehmlich an jene Forschenden wandte, die bereits in ASCOT tätig geworden waren (BMBF, 2018, Ziff. 4). Und das ist ausweislich der im vorliegenden Band berichteten Resultate von ASCOT+ – mit hilfreichen Ergänzungen – auch gelungen. Ausdrücklich war es der Wunsch des BMBF, dass bei dieser erneuten Initiative, die von 2019 bis 2023 (also während der Corona-Pandemie; vgl. Velten, 2024, S. 25) arbeitete, der Transfer in die Berufsbildungspraxis ins Zentrum gestellt werde (ebd., 2). Das bedeutete, dass nun die Dissemination der in ASCOT bereits weit entwickelten Messinstrumente unter führender Beteiligung der Wissenschaft betrieben werden sollte (ebd., 3).

Auf die Einrichtung eines wissenschaftlichen Beirats wurde verzichtet. An seine Stelle trat jedoch eine 21-köpfige „Begleitgruppe“ aus Wissenschaft, Wirtschaft und Politik (Partner:innen und Akteur:innen – BMBF ASCOT+), ergänzt durch eine differenzierte Analyse der Transferproblematik mit Blick auf die ASCOT+-Initiative (Rüschhoff & Velten, 2021) und entstanden nun doch unter dem Dach des BIBB, das vom BMBF als „Bewilligungsbehörde“ und „Projekträger“ im Sinne einer „Verwaltungsaufgabe“ (gem. § 90 Abs. 3, Ziff. 1f) eingesetzt worden war. Seine inhaltliche Mitwirkung beschränkte sich jedoch auf eine BIBB-intern gebildete informelle Gruppe, die ASCOT+ unter dem Aspekt „beobachten und begleiten“ sollte, ob es gelingen würde, „den Transfer der entwickelten Instrumente in die ... Praxiskontexte *wahrscheinlicher* zu machen“ (Velten, 2024, S. 20). Am Ende konnten dazu und zu den tatsächlich erreichten Transfererfolgen seitens dieser Gruppe allerdings kaum nähere Aussagen gemacht werden (ebd. S. 25, S. 30–31), auch weil sie sich auf einen „formative(n) Evaluationsansatz“ beschränkt hatte (ebd. 20; vgl. kritisch zu dieser Frage Nickolaus, Ziegler & Abel, 2006, S. 62–72).

In der Vorgehensweise des BMBF kommt mit der Fokussierung auf den Transfer das aufseiten der Politik verständliche Interesse an praktischer Sichtbarkeit der Erfolge von zuvor schon für ASCOT eingesetzten Steuermitteln zum Ausdruck. Andererseits ist aus forschungsstrategischer Sicht bedauerlich, dass der ursprünglich vorgesehene „zweite Schritt“ (vgl. Beck, Landenberger & Oser, 2016b, S. 12), nämlich die programmatische Validierung der bereits erarbeiteten Kompetenzmodellierungen im Rahmen von tragfähigen Kausalstudien, gewissermaßen übersprungen wurde. In seinem Rahmen hätten mittels größerer Stichproben die entwickelten Konstrukte in ihrem Entstehungs- und Wirkungskontext untersucht und als Erklärungs- und Prognoseinstanzen erprobt werden sollen.

2 ASCOT (+) – Forschungsprogramm oder Programmforschung?

Der Blick auf die Genealogie von ASCOT+ vermittelt durchaus und auch zu Recht den Eindruck, dass das Zustandekommen dieses Projekts als eine Folge von letztlich begünstigenden Umständen gedeutet werden kann, auch wenn die Entscheidung, es zu realisieren, sich teils unerwarteten und unvorhergesehenen Entwicklungen verdanken mag. Zugleich wird man jedoch nicht umhin können zu konstatieren, dass diese

Konstellationen ihrerseits nicht etwa als Abfolge eines in irgendeiner Hinsicht mit Notwendigkeit verlaufenen Prozesses verstanden werden dürfen. Vielmehr hätten die jeweiligen kontingenten Bedingungskonstellationen zu jedem Zeitpunkt auch einen anderen Gang der Dinge erlaubt (z. B. Toens & Willems, 2012). Die „realen Antriebe“ (Zabeck, 2013, S. 67 im Anschluss an Reichwein 1925) zu den geschilderten Vorgängen und Beschlüssen dürften letztlich im Wesentlichen dem fortschreitenden europäischen Vereinigungsprozess entsprungen sein, in dessen Verlauf ein gewisser politischer Druck auf die Herstellung eines einheitlich(er)en „europäischen Bildungsraums“ entstanden bzw. bei den beteiligten Agent:innen als solcher empfunden worden war.

Die Einbindung von Wissenschaft und Forschung in derartige Prozesse erfolgt in aller Regel gewissermaßen „auf Abruf“. Wo das politische Handeln auf einen Bedarf an prognostischem, an Steuerungs- oder Ergebniswissen stößt, den es anderweitig nicht befriedigend zu befriedigen vermag, kann es Forschungsprojekte ausschreiben und beauftragen, die entsprechende Informationen beschaffen sollen. Dabei verlaufen die Beziehungen zwischen Politik und Wissenschaft nicht immer so einseitig. Durchaus können auch seitens der Wissenschaft Fragestellungen aufgeworfen und der finanzierenden Politik als untersuchungsbedürftig dargestellt werden. Das geschieht nicht selten auch auf informellen Wegen der Wissenschafts-Politik-Kommunikation. So mag es auch im Falle der ersten PISA-Studien gewesen sein, die unter internationaler Beteiligung mehrerer Wissenschaftler:innen (für Deutschland v. a. Jürgen Baumert) von der OECD unter der Leitung von deren Bildungsdirektor Andreas Schleicher gesteuert wurden. Und ebenso verlief der Kommunikationsprozess – in diesem Fall ausgehend vom Göttinger SOFI – im Blick auf das „Berufsbildungs-PISA“, allerdings bislang eben nicht mit dem seitens der Wissenschaft erwünschten Ausgang. Immerhin aber vergab die in diesem Falle involvierte politische Instanz, das BMBF, wie geschildert, zunächst dann doch die Forschungsaufträge für das ASCOT-Programm und später für das ASCOT+-Programm.

Charakteristisch für ASCOT+ ist nun, dass die durchgeführten Forschungsprojekte Teile einer Programmatik des Geldgebers sind, sich also einem Interesse verdanken, das bildungspolitischen Bedürfnissen entstammt. Dazu ist dem Internet-Auftritt des BMBF „Über ASCOT+“ Folgendes zu entnehmen (Forschungsinitiative ASCOT (2011–2015) – BMBF ASCOT+):

„ASCOT+ hatte das Ziel, digitale Lern- und Messinstrumente für Kompetenzen von Auszubildenden zu entwickeln und in der Praxis zu erproben.“ Und weiter: „ASCOT+ war in den Aktivitäten des Bundesministeriums für Bildung und Forschung (BMBF) zwei Programmen zugeordnet. Es war Teil der Dachinitiative «Berufsbildung 4.0» und stand im Kontext des Rahmenprogramms «Empirische Bildungsforschung»“. Konkreter noch heißt es dazu: „In der Forschungs- und Transferinitiative ASCOT+ brachten verschiedene Partner und Akteure aus Politik, Wissenschaft und Praxis ihre Expertise ein und trugen zum Erfolg bei.“

Eine solche Konstellation bezeichnet man oft als „Programmforschung“ (Schimanke, Veit & Bull, 2011). Diese ist in der Regel gekennzeichnet durch pragmatisch-praktische Zielsetzungen, durch projektimmanente Theorie- und Methodenflexibilität in der Ziel-

erreichung, durch eine ergebnisfokussierte Prozessqualität, eine anwendungs- bzw. verwertungsdienliche Produktformatierung und eher unverbindliche Verlässlichkeitsinformationen. Programmforschung, gelegentlich auch als Ressortforschung oder Anwendungsforschung bezeichnet, folgt demnach einem wissenschafts*externalen* Rationale, das von pragmatischen, verwertungsgesteuerten oder politischen Bedürfnissen angestoßen zu werden pflegt.

Systematisch anders liegen die Dinge bei erkenntnisgetriebenen wissenschaftsimmanenten „Forschungsprogrammen“, wie sie insbesondere durch Imre Lakatos (1970/1974a, b) und Wolfgang Stegmüller (1973; 1986) in der Auseinandersetzung einerseits mit dem Falsifikationismus Poppers (1934/1982; 1974) und andererseits mit der Paradigmenlehre Kuhns (1962/1978; 1974a, b) konzeptualisiert worden sind (für die Erziehungswissenschaft vgl. Alisch, 1995). In ihnen werden Fragestellungen und Untersuchungen aus einer im Zentrum stehenden Hypothese bzw. Hypothesengruppe (ihrem sog. harten Kern; „hard core“) heraus entwickelt und im Sinne einer sog. positiven Heuristik in immer weiterführenden Schritten auf angrenzende Felder und Räume ausgeweitet. Dieser „harte Kern“, die Basishypothese(n), steht nicht zur Disposition; in ihm kommt vielmehr eine Grundüberzeugung über einen Gesamtzusammenhang zum Ausdruck, m. a. W. ein „Fundamentalgesetz“ (Stegmüller, 1986, S. 67); er ist als solcher nicht (direkt) falsifizierbar (wie Lakatos z. B. an der Newtonschen Gravitationstheorie veranschaulicht; 1970/1974a, S. 130).

Gegen Poppers Auffassung, dass Theorien prinzipiell durch empirisch vorfindliche Gegenbeispiele endgültig zu widerlegen seien, wird – grob und knapp, aber überzeugend – eingewandt, dass Theorien selbst angesichts widersprechender einzelner Beobachtungsaussagen weiterhin beibehalten werden können, solange sie, ggf. unter modifizierten ‚Geltungsbedingungen‘ (dem sog. Schutzgürtel um den harten Kern), Prognosen erlauben und darüber hinaus zusätzliche Phänomene erklären können (Lakatos, 1974). Aufgegeben werden Theorien in dieser Sicht erst dann, wenn ihre „progressive“ Erklärungskraft versiegt und sie nur noch gegen widersprechende Befunde verteidigt werden müssen, also quasi „degenerieren“. Das bedeutet zugleich, dass Theorien nicht etwa, wie Kuhn angenommen hat, in einem „revolutionären Akt“ von einer nachwachsenden Wissenschaftler:innengeneration, die neue Ideen ins Spiel bringen will, abgeschafft werden (vgl. insbes. Renn, 2022, S. 609–621). Vielmehr „sterben“ (Popper) sie erst, wenn ihre Leistungsfähigkeit nach und nach erlischt und eine mächtigere Theorie an ihre Stelle treten kann. So wird man etwa im Blick auf den vorliegenden Zusammenhang – etwas pauschal gesprochen – annehmen können, dass die Kompetenzforschung die vorausgegangene Qualifikationsforschung (sensu Mertens, 1974) absorbiert und damit abgelöst hat, während diese ihrerseits als Überwindung der bis in die Mitte des 20. Jahrhunderts wirksamen „Vermögenspsychologie“ Brentanoscher Prägung angesehen werden kann (vgl. auch Herrmann, 1987; Sachs-Hombach, 2001).

Ob Interventionsforschung (vgl. Leutner, 2013), wie sie in einer Transfervariante gerade auch in ASCOT+ praktiziert wird, sich in diesem Zusammenhang eindeutig zuordnen lässt, kann hier nicht im Detail diskutiert werden. Als „Experimentalfor-

schung“ im Sinne eines – mehr oder weniger strengen – *experimentum crucis* würde sie unter das Falsifikationskonzept Poppers fallen. Da sie jedoch konzeptionell stets in je singuläre Kontexte eingebettet ist, kann sie im Sinne von Lakatos als Untersuchung darüber gedeutet werden, welche Hypothesen im Schutzgürtel des jeweiligen Theoriekerns anpassungsbedürftig sind. Wären die Vertreter:innen der sog. Modellversuchsforschung nicht unnötigerweise so erpicht darauf, ihren Aktivitäten mittels eines ganz eigen(tümlich)en „Paradigmas“ zu einem methodologischen Sonderstatus zu verhelfen (zuletzt Sloane, 2024) – ein Standpunkt, der sich im Lichte von Lakatos' Kritik an Kuhn ohnehin als obsolet erweisen dürfte –, so ließen diese sich nahezu zwanglos der Lakatos'schen Sichtweise zuordnen. Einen solchen Status wird man – *cum grano salis* – auch der „Programmforschung“ zuweisen können, soweit sie dem Problem eines theoriefundierten Praxistransfers nachgeht, also eines Verfahrens, dessen Ergebnis allerdings stets auch von den Singularitäten der jeweiligen Settings abhängig ist.

Ohne hier näher auf diese differenzierte Diskussion eingehen zu können, lassen sich – im Rekurs auf Lakatos' Konzeption – zur Einordnung von ASCOT(+) einige programmatische Gesichtspunkte hervorheben. Zunächst ist zu konstatieren, dass ASCOT(+) eingebettet ist in die gegenwärtig weltweit arbeitende Kompetenzforschung, innerhalb derer bereits eine Vielzahl von Kompetenz-Modellierungen vorgelegt worden ist (Mikhridinova, Wolff & van Petegem, 2024; Mulder, 2017; Rothwell, Mozaffari & Hajri, 2025). Sie unterscheiden sich insbesondere darin, ob sie epistemisch holistisch oder kumulativ und ob sie ontologisch latent oder manifest angelegt sind (Baethge et al., 2006; Baethge & Seeber, 2016, S. 15–19; Fleischer et al., 2013, S. 8–10; Hellwig, 2008, S. 119–121; Klieme & Leutner, 2006; Neuweg, 1999; Weinert, 2001). Versucht man, in Begriffen des Forschungsprogramms *sensu* Lakatos zunächst den allgemeinen „harten theoretischen Kern“ zu präparieren, der auch den ASCOT(+)-Projekten innewohnt, so wird man – *grosso modo* – die folgenden zentralen Hypothesen festhalten können (vgl. dazu insbes. Ufer & Leutner, 2017):

Harter Kern (hard core):

- A Kompetenzen sind latente, nicht bewusstseinspflichtige Leistungsdispositionen, die über genetische Dispositionen hinaus intentional oder funktional erworben oder erweitert werden können. [Der Dispositionsbegriff wird hier als theoretischer Term im Carnapschen Sinne eingeführt; vgl. Stegmüller, 1974, Kap. IV, insbes. S. 233–238].
- B Ihre inhaltliche Bestimmung erfolgt unter pragmatischen Relevanzaspekten in jeweils gegebenen historisch-kulturell-gesellschaftlichen Kontexten.
- C Kompetenzen können als Bündel von Leistungsdispositionen unterschiedlicher inhaltlicher Reichweite und Intensität in individuellen Konstellationen und Kombinationen realisiert sein.
- D Kompetenzen sind identifizierbar und messbar.
- E Kompetenzen sind kausale Elemente in der Verhaltensgenese.
- F Die Häufigkeit der Nutzbarkeit einer Kompetenz ist unbegrenzt.

Dieser Kern ist durch Beobachtungen (d. h. durch Beobachtungssätze) nicht widerlegbar (negative Heuristik; Lakatos, 1970/1974a, S. 129–130). Umgeben wird er von einer Reihe von stützenden und schützenden, aber prüfbareren Hypothesen (positive Heuristik; ebenda, S. 131–134). Diese können modifiziert, preisgegeben, ersetzt oder auch ergänzt werden. Beim gegenwärtigen Forschungsstand dürften sie in etwa so lauten:

Schutzgürtel (protective belt):

1. Die interne Aktivierung und prozessbegleitende Energetisierung einer Kompetenz erfolgt einzeln oder in Kombination durch motivationale, volitionale und emotionale Instanzen.
 - a) Die Stimulierung einer Aktivierungsinstanz für eine Kompetenz kann ihrerseits external oder internal erfolgen.
 - b) Die Aktivierung einer Kompetenz kann unter entsprechenden Umständen verhindert (inhibiert) werden (z. B. Blockaden, Zielkonflikte).
2. Die Intensität der Aktivierung kann die Nutzungsintensität der Kompetenz beeinflussen (Exhaustionsgrad).
3. Die Ausprägungsstärke einer Kompetenz kann skaliert werden (ggf. in Niveaustufen).
 - a) In beruflichen Kontexten können Normwerte curricular an Kriterien oder pragmatisch an zu bestimmenden Vergleichsgruppen orientiert werden.
4. Die inhaltliche Reichweite einer Kompetenz kann interpersonell variieren; sie kann eher eng begrenzt („Inselbegabung“; domänenspezifisch, situationsspezifisch) oder breit erstreckt sein („generalistisch“, domänenübergreifend, situationenübergreifend).
 - a) Das Domänenkonzept bedarf für die Reichweitenbestimmung von Kompetenzen einer hinreichend klaren Abgrenzung.
5. Die Ausprägung einer Kompetenz ist jenseits von Erwerbs-/Lernprozessen intrapersonell *relativ* stabil.
6. Die Leistung einer Kompetenz reicht über den individuellen Erfahrungsbestand hinaus, d. h. sie bewährt sich (erst) in subjektiv neuartigen Kontexten.
7. Kompetenzen können über die längere Frist degenerieren (z. B. bei Nichtnutzung).
8. Je authentischer ein Lehr-Lern-Setting angelegt ist, desto höher ist die Wahrscheinlichkeit von kompetenzbezogenen Erwerbsgewinnen.
9. Je höher die gemessene Kompetenzausprägung ist, desto erfolgreicher kann – *ceteris paribus* – (beruflich) gehandelt werden (Performanz).

Eine solche Rekonstruktion der Kompetenzforschung im Sinne der Lakatos'schen Konzeption ergibt mit Blick auf ASCOT (also den ersten Teil der Förderinitiative), dass die dort verfolgten Fragestellungen sich gut in ein Forschungsprogramm des hier dargestellten Typs einfügen. Ihre Aufgabe bestand ja darin, berufs(ausbildungs)relevante Kompetenzen allererst zu identifizieren, zu modellieren und auf ein Messverfahren abzubilden (s. o. A, B, D; 1, 3, 4).

Dagegen war das Interesse an ASCOT+ darauf gerichtet, unter einer Anwendungsperspektive Veränderungen in der Berufsbildungspraxis zu ermöglichen – also Programmforschung im obigen Sinne. Als „Forschungs- und Transferinitiative“ stand sie unter dem Anspruch, unter Rekurs auf Resultate von ASCOT Verbesserungsvorschläge für die betriebliche bzw. schulische Berufsausbildung zu entwerfen und dazu beizutragen, dass die Kompetenzdiagnostik, wie sie insbesondere in Gestalt von Zwischen- und Abschlussprüfungen praktiziert wird, auf curricular angemessene und verlässlichere Füße gestellt werden kann.

Für diesen als „Transfer“ bezeichneten Prozess hatte das BMBF in Abstimmung mit dem BIBB das folgende Leitkonzept vorgegeben:

„Transfer wird im Rahmen von ASCOT+ definiert als ein kontinuierlicher interaktiver und iterativer Austauschprozess zwischen den Akteuren aus Wissenschaft, Praxis und Politik. Ziel des Transfers ist die Übertragung und Nutzbarmachung von fachlichem, methodischem und technischem Wissen sowie qualitativ hochwertigen Mess- und Lerninstrumenten sowie Instrumenten zur Unterstützung der Prüfungspraxis. Damit soll die Lehr-Lernpraxis sowie das Prüfungswesen und die Ordnungsarbeit in der beruflichen Bildung nachhaltig unterstützt werden.“ (Forschungs- und Transferinitiative ASCOT+ – BMBF ASCOT+; 04.04.2025)

Dieses Verständnis beruht auf den Überlegungen von Rüschhoff und Velten (2021, S. 11), die – eher pauschal – auf die beteiligten „Akteure“ als „Wissenschaft, Praxis und Politik“ Bezug nehmen. Diese sollen, wie es heißt, im Transferprozess eine „kontinuierliche Kollaboration“, möglichst „in Gruppen“ pflegen (ebd., 43). Mit der Bearbeitung solcher Aufgaben verbinden sich allerdings direkte und indirekte Stellungnahmen zur angetroffenen Praxis sowie konkrete Empfehlungen zu Modifikationen von Usancen, die vonseiten wertfreier Wissenschaft einen zumindest virtuellen Rollenwechsel zu wertungs- und entscheidungsbasiertem Gestaltungshandeln implizieren (BMBF, 2024). Der kann und darf sich jedoch allenfalls als zurückhaltende Vorschlagskommunikation äußern, denn zum Eingreifen und Verändern fehlt den nunmehr als Berater:in oder Expert:in agierenden Wissenschaftler:innen die erforderliche Legitimation qua Amt.

Von Verantwortlichkeiten, Zuständigkeiten und Befugnissen, die in solchen Zusammenhängen eine zentrale Rolle spielen können, ist allerdings in der Untersuchung von Rüschhoff und Velten (2021) nicht die Rede, obwohl gerade eine Klärung dieser Aspekte für die Transferpraxis von erheblicher Bedeutung sein dürfte, weil Dissonance oder gar Konflikte am ehesten dort auftreten, wo Gestaltungsentscheidungen zu treffen sind.

Zu Recht jedoch macht die Analyse der beiden Autorinnen darauf aufmerksam, dass in ASCOT+ die Vorstellung von Transfer als „eine einfache Übertragung von dokumentiertem Wissen ... zu kurz greift“ (ebd., 7). Sie weisen darauf hin, dass die involvierten Akteur:innen der Praxis mit „technologiebasierten Instrumenten möglicherweise noch wenig bis keine Erfahrungen gesammelt haben“ (ebd., 10), weshalb der Transfer „auf zweierlei Ebenen geleistet werden (müsse), der Ebene des Wissenstransfers und der Ebene des Technologietransfers“ (ebd.). Tatsächlich sehen sich die poten-

ziellen Rezipient:innen der ASCOT+ – Ergebnisse auf diesen beiden Ebenen ggf. benötigt, entsprechende Lernprozesse zu absolvieren, die für das gesamte Programm sogar erfolgskritisch sein können und insofern beobachtungsbedürftig sind. Auf sie wird noch zurückzukommen sein. Festzuhalten ist an dieser Stelle aber nochmals, dass ASCOT und ASCOT+ sich in der Anlage und in den Anforderungen an die Wissenschaft substanziell voneinander unterscheiden: als Forschungsprogramm das erste, als Programmforschung das zweite.

3 Resultate und Desiderate

Die in ASCOT+ geförderten sechs Projekte haben ihre programmatische Aufgabe, „digitale Lern- und Messinstrumente für Kompetenzen von Auszubildenden zu entwickeln und in der Praxis zu erproben“ (Forschungs- und Transferinitiative ASCOT+ – BMBF ASCOT+; 08.04.2025), unterschiedlich interpretiert. Gemeinsam ist ihnen jedoch, dass sie alle sich auftragsgemäß der beruflichen Ausbildungspraxis zugewandt und einen Transfer wissenschaftlicher Expertise in dieses Feld, teils direkt, teils eher indirekt, angestrebt haben. Dabei waren sie nicht allein darauf verwiesen, in welchem Maße und in welcher Hinsicht sie auf Befunde und Resultate der Vorgängerinitiative ASCOT zurückgreifen konnten. Sie mussten auch in Betracht ziehen,

- welche Optionen ihnen im Praxisfeld für Transferleistungen offenstanden,
- auf welche Kooperationsbereitschaft sie dort stoßen würden,
- wo im vorgegebenen Projektzeitrahmen greifbare Ergebnisse erwartet werden durften bzw. wo mit längeren „Inkubationszeiten“ zu rechnen wäre und
- welche Formen kooperativer Wissenschaft-Praxis-Interaktion jeweils als geeignet und realisierbar einzuschätzen waren.

Während solche Aspekte sich hauptsächlich auf das kontextuelle Umfeld von Transferbemühungen richteten, verbanden sich mit den Vorhaben unter einem übergeordneten Evaluationsgesichtspunkt eine ganze Reihe von Fragen, die sich auf die Substanz und das Resultat des Transfers richten:

- Darf der Gegenstand des Transfers als Anpassung auf die Umstände in der Praxis beliebig modifiziert werden (man denke etwa an die Hinzufügung oder Weglassung schwierigkeitsbestimmender Aufgabenmerkmale)? Wie weit dürfen Veränderungen ggf. gehen, ohne dass statt von Transfer von „ad-hoc-Konstruktionen“ gesprochen werden müsste? Wer bestimmt anhand welcher Kriterien darüber?
- Wie breit ist das als „Praxis“ bezeichnete Feld (geografisch, institutionenbezogen, berufsfeldspezifisch), in dem der Transfer, um als gelungen zu gelten, akzeptiert und praktiziert werden muss?
- Welche „Verweigerungsquote“ im Anwendungsfeld kann als akzeptabel hingenommen werden bzw. wie hoch ist eine zu erreichende Akzeptanzquote anzusetzen?

- Nach welcher Anwendungsdauer einer Neuerung darf man von einem gelungenen Transfer sprechen?
- Sind die erwarteten Effekte des Transfers (Kompetenzzugewinne bei Auszubildenden und bei Autor:innen von Prüfungsaufgaben) eingetreten?

Mit diesen letzteren Fragen hat sich der speziell für ASCOT+ erstellte Transfer-„Leitfaden“ von Rüschhoff & Velten (2021, S. 49) ebenfalls nicht näher auseinandergesetzt. Auch hier besteht jedoch programmspezifischer Präziserungs- und Entscheidungsbedarf (vgl. z. B. Ghadiri, 2023; Otto, Bieber & Heinrich, 2019; Wäfler et al., 2018).

Im Vergabeverfahren der Projektmittel waren zwei thematische Hauptproblemfelder festgelegt worden, denen sich die Projekte, deren individuelle Würdigung an anderer Stelle in diesem Band erfolgt, unter den je besonderen Bedingungen zuwenden sollten, nämlich zum einen die *Lehr-Lern-Prozesse* in der Erstausbildung zu den ausgewählten Berufen und zum anderen die *Diagnostik* der durch sie zu erwerbenden Kompetenzen. Schwerpunkte gesetzt haben (hier die Berichtstitel; zu den Projekttiteln s. Lohmeyer & Velten, 2019)

im Bereich Lehr-Lern-Prozesse die Projekte

„Digitale Interventionen zur Förderung der Diagnosekompetenz bei Kfz-Mechatroniker/-innen“ (DigiDIn-Kfz),

„Mit Signaturen Szenarien zum Kompetenzerwerb: Die Analyse von Performance Tasks in virtuellen Simulationen“ (PSA-Sim) und

„Ausbildungsbedingungen in der Pflegepraxis – Erste empirische Befunde zu unterschiedlichen Versorgungsbereichen“ (EKGe);

im Bereich Diagnostik die Projekte

„Zur Qualität beruflicher Abschlussprüfungen: Psychometrie und praxisnahe Designstandards“ (ASPE),

„Schwierigkeitsbestimmende Aufgabenmerkmale in Prüfungsaufgaben im technischen Bereich am Beispiel des Ausbildungsberufs Mechatroniker/-in“ (TechKom) und

„Problemhaltigkeit kaufmännischer Prüfungsaufgaben: Eine Analyse mittels lautem Denken“ (TeKoP).

Die Konzentration der einzelnen Projekte auf diese Schwerpunkte ist nicht exklusiv. Vor allem die Simulationen (in DigiDIn-Kfz und PSA-Sim) ermöglichen mit ihrem Angebot durchaus eine Doppelnutzung als Lerngelegenheiten einerseits und Diagnosearrangements andererseits und vereinfachen so zugleich auch die berufspädagogisch bedeutsame Messung von Lern- und Entwicklungsgewinnen (vgl. dazu auch Leutner, Hartig & Jude, 2008). Festgehalten werden kann, dass sie alle, gemessen am Stand der jeweils thematisierten Praxis, transferwürdige Erkenntnisse mitteilen oder Neuerungen vorschlagen (s. insbes. ASPE), von deren Rezeption und Implementation Verbesserungen erwartet werden können.

Unter der Frage danach, ob ein entsprechender Transfer gelungen sei, gilt es gemäß der thematischen Ausrichtung auf Lehr-Lern-Prozesse einerseits und auf Diagnostik andererseits, die Aufmerksamkeit zum einen auf Kompetenzgewinne der Auszubildenden zu richten und zum anderen auf Kompetenzgewinne der Diagnostiker:innen, sprich derjenigen, die jene Leistungssituationen – etwa in Form von Prüfungsaufgaben – konstruieren, in denen sich zeigen soll, ob bei den Auszubildenden die durch Verbesserungen angestrebten Kompetenzzugewinne tatsächlich eingetreten sind. Diese beiden – oben im Anschluss an Rüschhoff und Velten (2021) erwähnten – „Ebenen“ stehen mithin in einer systematischen Verbindung zueinander, indem die Qualität auf der letzteren dafür verantwortlich ist, ob für die Veränderungen auf der ersteren verlässliche Informationen verfügbar werden. Und natürlich reicht dieser Anspruch an die Diagnostik in der Praxis deutlich über diesen Zusammenhang hinaus, weil auf der Grundlage der diagnostischen Befunde in Prüfungsverfahren (lebens-)bedeutsame Entscheidungen über berufliche Karrieren getroffen werden.

Irritierend – um es vorsichtig auszudrücken – ist nun jedoch, dass Verbesserungsangebote aus den Projekten in den Feldern, auf die sie zugeschnitten sind, nicht immer auf ein ungeteiltes Rezeptions-, geschweige denn Umsetzungsinteresse zu treffen pflegen (Gräsel 2010, 2019; Nickolaus & Gräsel, 2006; Velten, 2024, S. 22). Im Gegenteil, nicht selten stoßen sie eher auf reservierte Zurückhaltung, die unter anderem auf einer untergründigen Wissenschaftsskepsis oder auch auf einer aufwandsaversen Veränderungsresistenz beruhen mögen. Solche Umstände können Transferbemühungen vor zusätzliche schwierige Probleme stellen, die nochmals deutlich machen, dass die erwähnte Idee eines „linearen Transfers“ von der Wissenschaft in die pädagogische Praxis die Komplexität dieses Problems deutlich unterschätzt hätte (Rüschhoff & Velten, 2021, S. 50; Hartmann & Kunter, 2022). Dies gilt umso mehr, als im Kontext von ASCOT+ nicht allein die Schulen als Rezipient:innen in den Blick zu nehmen waren, sondern daneben vor allem auch die variantenreiche Welt der Ausbildungsbetriebe und die gem. BBiG § 71 „zuständigen Stellen“, also die für die jeweiligen Bereiche stehenden Kammern, die letzteren insbesondere mit ihren Einrichtungen für das Prüfungswesen. Dieser außerschulische Teil der berufspädagogischen Praxis bedarf im Blick auf das Rezeptionsgeschehen gewiss einer besonderen Untersuchung.

Führt man sich das riesige Feld der Berufsausbildung vor Augen, so wird man gewahr, dass mit ASCOT+ nur ein kleiner, wenngleich keineswegs vernachlässigbarer Teil dieses Komplexes ins kompetenzforscherische Visier genommen worden ist. Zwar wird man einige Befunde über die Grenzen der konkret einbezogenen Berufe hinaus ein Stück weit generalisieren können, aber sobald domänenspezifische Inhalte ins Spiel kommen, bedarf es jeweils spezieller Analysen. Bevor man also weitere Transferbemühungen in Angriff nimmt, steht immer noch der oben erwähnte „zweite Schritt“ aus, nämlich die Konsolidierung der im „ersten Schritt“ entwickelten Kompetenzmodelle (vgl. oben Ziff. 1 bis 7 des „protective belt“). Mithilfe der bereits vorliegenden Diagnoseinstrumente bedarf es systematisch angelegter Kausalanalysen, in denen die theoretisch eingeführten Kompetenzkonstrukte hinsichtlich ihrer in Entwicklungs- und Lernprozessen erreichten Ausprägung (Ziff. 8) und vor allem auch hinsichtlich

der ihnen hypothetisch unterstellten Wirkungen in beruflichen Handlungsfeldern untersucht werden (Ziff. 9). Angesichts der erheblichen praktischen Relevanz, die dem berufsbezogenen Kompetenzkonzept sowohl gesamtwirtschaftlich als auch insbesondere hinsichtlich der individuellen Persönlichkeitsentwicklung zukommt, kann die Wichtigkeit einer wissenschaftlich belastbaren Vergewisserung über die mit diesem Konzept verbundenen Wirkungsvermutungen gar nicht überschätzt werden.

Was die Weiterführung eines solchen *Forschungsprogramms* betrifft, so ergeben sich dafür – bedauerlicherweise erst nach einem erneuten *brain drain* aus den abgeschlossenen ASCOT+-Projekten – vielleicht nun doch noch ganz neue Chancen, nachdem die OECD das Thema „Vocational Education and Training“ für sich entdeckt und 2022 in Zusammenarbeit mit WorldSkills eine „PISA-VET-Initiative“ auf den Weg gebracht hat. In ihrer 2024 veröffentlichten Perspektivplanung (OECD, 2024) stützt sie sich jedenfalls ausdrücklich auf die im Hauptausschuss des BIBB siebzehn Jahre zuvor als unzulänglich eingeschätzte deutsche Machbarkeitsstudie.

Literatur

- Achtenhagen, F., & Baethge, M. (2007). Kompetenzdiagnostik als Large-Scale-Assessment im Bereich der beruflichen Aus- und Weiterbildung. In M. Prenzel, I. Gogolin, & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik* (S. 51–70). Zeitschrift für Erziehungswissenschaft, Sonderheft 8/10. VS Verlag. https://doi.org/10.1007/978-3-531-90865-6_4
- Alisch, L.-M. (1995). *Grundlagenanalyse der Pädagogik als strenge praktische Wissenschaft*. Duncker & Humblot. <https://doi.org/10.3790/978-3-428-48377-8>
- Althaus, D., Kraus, J., Gauger, J.-D., & Grewe, H. (2002). *PISA und die Folgen: Neue Bildungsdebatte und erste Reformschritte*. Konrad-Adenauer-Stiftung e. V. Broschüre Nr. 40. Sankt Augustin. Verfügbar unter https://www.kas.de/c/document_library/get_file?uuid=15147722-43a0-63e9-993e-0c44c26da633&groupId=252038 (Zugriff am 10.07.2025).
- Baethge, M., Achtenhagen, F., Arends, L., Babic, E., Baethge-Kinsky, V., & Weber, S. (2006). *Berufsbildungs-PISA. Machbarkeitsstudie*. Franz Steiner.
- Baethge, M., & Arends, L. (2008). International Large Scale Assessment of Vocational Education and Training (VET-LSA). *CESifo DICE Report*, 06(2), 29–35. <https://hdl.handle.net/10419/166936>
- Baethge, M., & Arends, L. (2009). Die Machbarkeit eines internationalen Large-Scale-Assessment in der beruflichen Bildung: Feasibility Study VET-LSA. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 105(4), 492–520. <https://doi.org/10.25162/zbw-2009-0036>
- Baethge, M., & Seeber, S. (2016). Die gemeinsame theoretische und methodische Basis der ASCOT-Projekte. In K. Beck, M. Landenberger, & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung. Ergebnisse aus der BMBF-Förderinitiative ASCOT* (S. 15–31). wbv.

- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J., & Weiß, M. (Eds.) (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Leske + Budrich. https://doi.org/10.1007/978-3-322-83412-6_3
- Beck, K. (1995). Die Äquivalenz von Ausbildungsabschlüssen in Europa als Voraussetzung transnationaler beruflicher Mobilität? Ein erziehungswissenschaftlicher Ansatz zur Lösung des Vergleichsproblems. In K. A. Schachtschneider (Hrsg.), *Wirtschaft, Gesellschaft und Staat im Umbruch. Festschrift der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Friedrich-Alexander-Universität Erlangen-Nürnberg 75 Jahre nach Errichtung der Handelshochschule Nürnberg* (S. 1–16). Duncker & Humblot.
- Beck, K., Landenberger, M., & Oser, F. (Hrsg.) (2016a). *Technologiebasierte Kompetenzmessung in der beruflichen Bildung. Ergebnisse aus der BMBF-Förderinitiative ASCOT*. wbv.
- Beck, K., Landenberger, M., & Oser, F. (2016b). Vorwort der Herausgeber. In K. Beck, M. Landenberger, & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung. Ergebnisse aus der BMBF-Förderinitiative ASCOT* (S. 7–13). wbv.
- BIBB (2007). *Verzeichnis ausgewählter Beschlüsse zur beruflichen Bildung. Nr. 125. Empfehlungen zu einer internationalen Vergleichsstudie in der Berufsbildung („Large-Scale-Assessment“)*. Hauptausschuss des Bundesinstituts für Berufsbildung. Beschluss vom 13. Dezember 2007. Verfügbar unter <https://www.bibb.de/dokumente/pdf/HA125.pdf> (Zugriff am 10.07.2025).
- BIBB (2009). Stand der Arbeiten zu einer internationalen Vergleichsstudie in der Berufsbildung („Large-Scale-Assessment“). Empfehlung der Gruppen der Beauftragten der Arbeitgeber, der Arbeitnehmer und der Länder im Hauptausschuss des Bundesinstituts für Berufsbildung (BIBB) vom 25. Juni 2009. BWP, Beilage zu 4/2009. Verfügbar unter <https://www.bwp-zeitschrift.de/dienst/publikationen/en/1619> (Zugriff am 10.07.2025).
- BMBF (2014). *Impulse für die Ausbildung der Zukunft. Innovative Kompetenzmessung in einer dynamischen Arbeitswelt*. BMBF. Verfügbar unter https://www.ascot-vet.net/ascot/shareddocs/downloads/files/impulse-fuer-die-ausbildung-der-zukunft.pdf?__blob=publicationFile&v=2 (Zugriff am 10.07.2025).
- BMBF (2018). *Richtlinie zur Förderung von Projekten im Rahmen der Forschungs- und Transferinitiative „ASCOT+ – Technologiebasierte Kompetenzmessung in der beruflichen Bildung“*. Bundesanzeiger vom 25.05.2018. Verfügbar unter https://www.bmbf.de/SharedDocs/Bekanntmachungen/DE/2018/05/1686_bekanntmachung.html (Zugriff am 10.07.2025).
- BMBF (2024). *Digitale Tools für die Berufsbildungspraxis. Lehren, Lernen und Prüfen mit ASCOT+*. BMBF. Verfügbar unter https://www.ascot-vet.net/ascot/shareddocs/downloads/files/Praxisbroschuere_ASCOTplus.pdf?__blob=publicationFile&v=6 (Zugriff am 10.07.2025).
- Fleischer, Z., Koeppen, K., Kenk, M., Klieme, E., & Leutner, D. (2013). Kompetenzmodellierung: Struktur, Konzepte und Forschungszugänge des DFG-Schwerpunktprogramms. *Zeitschrift für Erziehungswissenschaft*, 16, 5–22. <https://doi.org/10.1007/s11618-013-0379-z>

- Ghadiri-Mohajerzad, H. (2023). *Transfer von wissenschaftlichem Wissen zur Nutzung digitaler Lern- und Bildungsmedien an Bildungspraxis und -politik. Beiträge zum Wissenstransfer in der Bildungsforschung* (Diss.). Tübingen. Verfügbar unter https://ub01.uni-tuebingen.de/xmlui/bitstream/handle/10900/150752/Manteltext_Mohajerzad.pdf?sequence=3&isAllowed=y (Zugriff am 10.07.2025).
- Gräsel, C. (2010). Stichwort: Transfer und Transferforschung im Bildungsbereich. *Zeitschrift für Erziehungswissenschaft*, 13(1), 7–20. <https://doi.org/10.1007/s11618-010-0109-8>
- Gräsel, C. (2019). Transfer von Forschungsergebnissen in die Praxis. In C. Donie, F. Foerster, M. Obermayr, A. Deckwerth, G. Kammermeyer, G. Lenske, M. Leuchter, & A. Wildemann (Hrsg.), *Grundschulpädagogik zwischen Wissenschaft und Transfer* (S. 2–11). Springer VS. https://doi.org/10.1007/978-3-658-26231-0_1
- Hartmann, U., & Kunter, M. (2022). Mehr Praxis in der Bildungsforschung? Eine Studie zu Praxisperspektiven in Forschungsprojekten. *Bildungsforschung*, 2, 1–24. <https://doi.org/10.25656/01:25469>
- Hellwig, S. (2008). *Zur Vereinbarkeit von Competency-Based Training und Berufsprinzip. Konzepte der Berufsbildung im Vergleich*. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-90981-3>
- Herrmann, T. (1987). Theoriendynamik in psychologischen Forschungsprogrammen. In J. Brandtstädter (Hrsg.), *Struktur und Erfahrung in der psychologischen Forschung* (S. 71–89). W. de Gruyter. https://doi.org/10.1515/9783110858105_71
- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. *Zeitschrift für Pädagogik*, 52(6), 876–903. <https://doi.org/10.25656/01:4493>
- Klubertz, T. (2003). *Zur aktuellen Diskussion über die Novellierung des Berufsbildungsgesetzes*. BIBB. Verfügbar unter https://www.bibb.de/dokumente_archiv/pdf/z3_klubertz_entwicklung-bbig.pdf (Zugriff am 10.07.2025).
- KMK, BMBF, HRK, DAAD, Akkreditierungsrat, fzs, DSW, GEW & BDA (2024). *Die Umsetzung der Ziele des Bologna-Prozesses 2021–2024*. Verfügbar unter https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2024/2024_02_08-Nationaler-Bologna-Bericht-2021-2024.pdf (Zugriff am 10.07.2025).
- Kühl, S. (2015). Die Bologna-Reform in der Bundesrepublik Deutschland. Erklärungen jenseits der Vorstellungen von einem hochschulpolitischen Masterplan. In M. Kneuer (Hrsg.), *Standortbestimmung Deutschlands: Innere Verfasstheit und internationale Verantwortung* (S. 303–323). Nomos. <https://doi.org/10.5771/9783845252841-303>
- Kuhn, T. S. (1962/1978). *Die Struktur wissenschaftlicher Revolutionen*. Suhrkamp.
- Kuhn, T. S. (1974a). Logik der Forschung oder Psychologie der wissenschaftlichen Arbeit? In I. Lakatos & A. Musgrave (Hrsg.), *Kritik und Erkenntnisfortschritt* (S. 1–24). Vieweg. https://doi.org/10.1007/978-3-322-90613-7_1
- Kuhn, T. S. (1974b). Bemerkungen zu meinen Kritikern. In I. Lakatos & A. Musgrave (Hrsg.), *Kritik und Erkenntnisfortschritt* (S. 223–269). Vieweg. https://doi.org/10.1007/978-3-322-90613-7_9

- Lakatos, I. (1970/1974a). Falsifikation und die Methodologie wissenschaftlicher Forschungsprogramme. In I. Lakatos & A. Musgrave (Hrsg.), *Kritik und Erkenntnisfortschritt* (S. 89–189). Vieweg. https://doi.org/10.1007/978-3-322-90613-7_7
- Lakatos, I. (1970/1974b). Die Geschichte der Wissenschaft und ihrer rationalen Rekonstruktionen. In I. Lakatos & A. Musgrave (Hrsg.), *Kritik und Erkenntnisfortschritt* (S. 271–311). Vieweg. https://doi.org/10.1007/978-3-322-90613-7_10
- Leutner, D., Hartig, J., & Jude, N. (2008). Measuring Competencies: Introduction to Concepts and Questions of Assessment in Education. In J. Hartig, E. Klieme, & D. Leutner (Hrsg.), *Assessment of Competencies in Educational Contexts* (S. 177–192). Hogrefe.
- Leutner, D. (2013). Perspektiven pädagogischer Interventionsforschung. In E. Severing & R. Weiß (Hrsg.), *Qualitätsentwicklung in der Berufsbildungsforschung* (S. 17–28). wbv & BIBB. <https://doi.org/10.25656/01:7983>
- Lohmeyer, N., & Velten, S. (2019). Forschungs- und Transferinitiative ASCOT+ startet mit sechs Projekten. *Berufsbildung in Wissenschaft und Praxis (BWP)*, 6/2019, 16–18. Verfügbar unter https://www.bwp-zeitschrift.de/dienst/publikationen/de/10707?publication_search_result%25BfreeTextSearch%25D%3DH%25C3%2583%25C6%2592%25C3%2582%25C2%25B6ft%2520Stefan%26publication_search_result%25BentriesPerPage%25D%3D10%26publication_search_result%25Bsorting%25D%3Dnewest_desc (Zugriff am 10.07.2025).
- Mertens, D. (1974). Schlüsselqualifikationen. *MittAB*, 7, 36–43. Verfügbar unter https://doku.iab.de/mittab/1974/1974_1_MittAB_Mertens.pdf (Zugriff am 10.07.2025).
- Mikhridinova, N., Wolff, C., & van Petegem, W. (2024). Taxonomy of competence models based on an integrative literature review. *Education and Information*, 29, 16997–17033. <https://doi.org/10.1007/s10639-024-12463-y>
- Münk, D., & Schelten, A. (Hrsg.) (2010). *Kompetenzermittlung für die Berufsbildung. Verfahren, Probleme und Perspektiven im nationalen, europäischen und internationalen Raum*. BIBB. Verfügbar unter <https://www.bibb.de/dienst/publikationen/de/6175> (Zugriff am 10.07.2025).
- Mulder, M. (2017). Competence Theory and Research: A Synthesis. In M. Mulder (Ed.), *Competence-based Vocational and Professional Education. Bridging the Worlds of Work and Education. Technical and Vocational Education and Training: Issues, Concerns and Prospects*, 23, 1071–1106. https://doi.org/10.1007/978-3-319-41713-4_50.
- Neuweg, G. H. (1999). *Könnerschaft und implizites Wissen. Zur lehr-lerntheoretischen Bedeutung der Erkenntnis- und Wissenstheorie Michael Polanyis*. Waxmann.
- Nickolaus, R., & Gräsel, C. (Hrsg.) (2006). *Innovation und Transfer. Expertisen zur Transferforschung*. Schneider.
- Nickolaus, R., Ziegler, B., & Abel, M. (2006). Anlage der Expertise und Ergebnisse im Überblick. In R. Nickolaus, & C. Gräsel (Hrsg.), *Innovation und Transfer. Expertisen zur Transferforschung* (S. 9–72). Schneider.
- OECD (2024). *PISA Vocational Education and Training (VET). Assessment and Analytical Framework*. PISA, OECD Publishing. <https://doi.org/10.1787/b0d5aaf9-en>

- Otto, J., Bieber, G., & Heinrich, M. (2019). Aktuelle Desiderata zum systematischen Wissenstransfer und zur Implementation: Einführender Diskussionsbeitrag zu Berichten über transferaffine Forschungsprojekte. *DDS—Die Deutsche Schule*, 111(3), 310–321. <https://doi.org/10.31244/dds.2019.03.05>
- Popper, K. R. (1934/1982). *Logik der Forschung*. Mohr.
- Popper, K. R. (1974). Die Normalwissenschaft und ihre Gefahren. In I. Lakatos, & A. Musgrave (Hrsg.), *Kritik und Erkenntnisfortschritt* (S. 51–57). Vieweg. https://doi.org/10.1007/978-3-322-90613-7_5
- Renn, J. (2022). *Die Evolution des Wissens. Eine Neubestimmung der Wissenschaft für das Anthropozän*. Suhrkamp.
- Rothwell, W., Mozaffari, F., & Al Hajri, A. (2025). A bibliometric overview of competency and capability modeling: research contributions and trends (2000–2024). *Performance Improvement Quarterly*, 37(1). <https://doi.org/10.56811/PIQ-24-0022>
- Rüschhoff, B., & Velten, S. (2021). *Anforderungen an einen erfolgreichen Wissenschafts-Praxis-Transfer. Entwicklung eines Konzepts zur Begleitung der Projekte der Forschungs- und Transferinitiative ASCOT+*. BIBB. Verfügbar unter <https://www.econstor.eu/bitstream/10419/242112/1/1770113797.pdf> (Zugriff am 10.07.2025).
- Sachs-Hombach, K. (2001). Stichwort „Vermögenspsychologie“. In J. Ritter, K. Gründer, & G. Gabriel (Hrsg.), *Historisches Wörterbuch der Philosophie* (S. 727–731). Schwabe Verlag.
- Schimanke, D., Veit, S., & Bull, H. P. (Hrsg.) (2011). *Bürokratie im Irrgarten der Politik: Gedächtnisband für Hans-Ulrich Derlien*. Nomos.
- Seeber, S., Nickolaus, R., Winther, E., Achtenhagen, F., Breuer, K., Frank, I., Lehmann, R., Spöttl, G., Straka, G., Walden, G., Weiß, R., & Zöller, A. (2009). Kompetenzdiagnostik in der Berufsbildung. Begründung und Ausgestaltung eines Forschungsprogramms. *Beilage zur BWP 1/2010*, 1–15. Verfügbar unter <https://www.bwp-zeitschrift.de/de/bwp.php/de/publication/download/6162> (Zugriff am 10.07.2025).
- Sloane, P. F. E. (2024). Quo vadis Berufs- und Wirtschaftspädagogik? Wege aus der Homogenitätsfalle. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 120(2), 169–190. <https://doi.org/10.25162/zbw-2024-0008>
- Spillner, G. (2008). Bericht über die Sitzung des Hauptausschusses am 13. Dezember 2007 in Nürnberg. *BWP 1/2008*, 55–56. Verfügbar unter <https://d-nb.info/103868577X/34> (Zugriff am 10.07.2025).
- Spillner, G. (2009). Bericht über die Sitzung 3/2008 des Hauptausschusses am 18. Dezember 2008 in Bonn. *BWP 1/2009*, 50–51. Verfügbar unter <https://www.bwp-zeitschrift.de/dienst/publikationen/en/1538> (Zugriff am 10.07.2025).
- Stegmüller, W. (1973). *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Band II. Theorie und Erfahrung. Zweiter Halbband. Theoriestrukturen und Theoriendynamik*. Springer. <https://doi.org/10.1007/978-3-662-00224-7>
- Stegmüller, W. (1974). *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Band II. Theorie und Erfahrung. Erster Halbband. Begriffsformen, Wissenschaftssprache, empirische Signifikanz und theoretische Begriffe*. Springer.

- Stegmüller, W. (1986). *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Band II. Theorie und Erfahrung. Dritter Halbband. Die Entwicklung des neuen Strukturalismus seit 1973*. Springer.
- Toens, K., Willems, U. (2012). Kontingenz und Politik – Interdisziplinäre und politikwissenschaftliche Perspektiven. In K. Toens, & U. Willems (Hrsg.), *Politik und Kontingenz* (S. 11–22). Springer VS. https://doi.org/10.1007/978-3-531-94245-2_1
- Ufer, S., & Leutner, D. (2017). Kompetenzen als Dispositionen – Begriffsklärungen und Herausforderungen. In A. Südkamp, & K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften – Theoretische und methodische Weiterentwicklungen* (S. 67–74). Waxmann.
- Velten, S. (2024). *Den Transfer von Anfang an mitdenken: Verwendung eines Transferkonzepts für die wissenschaftliche Begleitung in ASCOT+*. BIBB. Verfügbar unter https://data.pool-bibb.bibb.de/pdfs/Velten_Transferkonzept.pdf (Zugriff am 10.07.2025).
- Wäfler, T., Fischer, K., Kunz, C., & Saric, N. (2018). *Wissenstransfer in Organisationen. Ein anwendungsorientierter Leitfaden zur Implementierung und Evaluation eines Wissenstransfer-Prozesses in Organisationen*. Vdf Hochschulverlag. <https://doi.org/10.3218/3906-1>
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen, & L. H. Salganik (Hrsg.), *Defining and selecting key competencies* (S. 45–65). Hogrefe.
- Winter, M. (2015). *Bologna – die ungeliebte Reform und ihre Folgen*. Verfügbar unter <https://www.bpb.de/themen/bildung/dossier-bildung/204075/bologna-die-ungeliebte-reform-und-ihre-folgen/> (Zugriff am 10.07.2025).
- Zabeck, J. (2013). *Geschichte der Berufserziehung und ihrer Theorie* (2. Aufl.). Eusl.

ASCOT+: Eine lehr-lernpsychologische Verortung der Projekte

DETLEV LEUTNER

Zusammenfassung

In diesem Kommentarbeitrag zum Sammelband werden die sechs Projekte der Forschungs- und Transferinitiative ASCOT+ dahingehend untersucht, inwieweit sie bei der Entwicklung ihrer digitalen Tools und der diesbezüglichen empirischen Forschung die notwendigen Komponenten instruktionspsychologischer Theorien (Lehrzieldefinition und Lehrstoffanalyse, Analyse von Lernvoraussetzungen, Analyse des angestrebten Lernprozesses, Instruktionsdesign sowie Mess- und Evaluationsvorschriften inkl. der Validierung von Instrumenten) angemessen berücksichtigen und umsetzen. Im Ergebnis zeigt sich, dass der Schwerpunkt der Projekte auf den Komponenten „Lehrzieldefinition und Lehrstoffanalyse“ sowie „Mess- und Evaluationsvorschriften“ liegt, während bezüglich der übrigen Komponenten, insbesondere bezüglich der „Analyse von Lernvoraussetzungen“, Optimierungspotenzial zu erkennen ist. Dabei berichtet jedes der Projekte über innovative Ansätze zur Förderung und Überprüfung der Handlungskompetenzen Auszubildender zur Lösung komplexer beruflicher Problemstellungen. Über die Projekte hinweg zeigen sich insgesamt sehr beachtliche Fortschritte in der Entwicklung, Nutzung und empirischen Validierung technologiebasierter Lern- und Testumgebungen, die weit über die berufliche Bildung hinaus relevant erscheinen.

Schlagnvorte: Komponenten instruktionspsychologischer Theorien, Projektanalysen, technologiebasierte Lern- und Testumgebungen

Abstract

In this commentary on the edited volume, the six projects of the ASCOT+ research and transfer initiative are analyzed to determine the extent to which they adequately consider and implement the necessary components of instructional psychology theories (definition of teaching objectives and analysis of teaching content, analysis of learning prerequisites, analysis of the intended learning process, instructional design and measurement and evaluation specifications, including the validation of instruments) in the development of their digital tools and the related empirical research. The results indicate that the projects focus on the components “definition of teaching objectives and analysis of teaching content” and “measurement and evaluation specifications“, while there is potential for optimization with regard to the other components, particularly the “analysis of learning prerequisites“. Each project reports on innovative approaches to fostering and assessing trainees’ competences for solving complex vocational prob-

lems. Across the projects, there has been considerable progress in the development, utilization and empirical validation of technology-based learning and testing environments, which appear to be relevant far beyond vocational education and training.

Keywords: Components of instructional psychology theories, project analyses, technology-based learning and testing environments

Ziel der Forschungs- und Transferinitiative ASCOT+ war es, technologiebasierte Lern- und Testumgebungen für berufliche und berufsübergreifende Kompetenzen von Auszubildenden zu entwickeln und in der Praxis zu erproben. Aus lehr-lernpsychologischer Sicht folgte die Entwicklung der in den Projekten entstandenen digitalen Tools, mit unterschiedlichen Schwerpunktsetzungen, den schon von Snow & Swanson (1992) beschriebenen Komponenten instruktionspsychologischer Theorien (vgl. Leutner & Wirth, 2018):

- (1) Beschreibung des gewünschten Soll- oder Zielzustandes einer Person bezogen auf ein bestimmtes Wissensgebiet (*Lehrzieldefinition und Lehrstoffanalyse*),
- (2) Beschreibung zielrelevanter Ist-Zustände von Personen vor Beginn der Instruktion (*Analyse von Lernvoraussetzungen*),
- (3) Explikation des Prozesses für den Übergang vom Ist- in den Soll-Zustand (*Analyse des angestrebten Lernprozesses*),
- (4) Spezifikation derjenigen instruktionalen Bedingungen, die geeignet sind, den Übergang zu fördern (*Instruktionsdesign*),
- (5) Spezifikation von Verfahren zur Beurteilung des Lernerfolgs und anderer instruktionaler Effekte (*Mess- und Evaluationsvorschriften* inkl. der Validierung von Instrumenten).

In den Projekten geht es um Berufe im gewerblich-technischen Bereich, im kaufmännischen Bereich und im Bereich der Pflege. Für den vorliegenden Sammelband hat jedes der Projekte einen Beitrag geschrieben, in dem jeweils ein spezifischer Schwerpunkt gesetzt wird. In den Beiträgen aus dem ASPE-, dem TechKom- und dem TeKop-Projekt geht es um die Entwicklung von Prüfungsaufgaben und um deren Eigenschaften (mit Schwerpunkt auf Mess- und Evaluationsvorschriften, also der fünften Komponente instruktionspsychologischer Theorien), in den Beiträgen aus dem DigiDIn-Kfz- und dem PSA-Sim-Projekt geht es um die Nutzung komplexer Simulationsszenarien zur Förderung beruflicher Problemlösekompetenz (mit Schwerpunkt auf Lehrzieldefinition und Lehrstoffanalyse, Analyse des angestrebten Lernprozesses und Instruktionsdesign, also der ersten, dritten und vierten Komponente), und im Beitrag aus dem EKGe-Projekt geht es um den Einsatz eines videobasierten, computergestützten Situational-Judgement-Tests zur Analyse von Ausbildungsbedingungen (ebenfalls mit Schwerpunkt auf Mess- und Evaluationsvorschriften, also der fünften Komponente). Zu erkennen ist hier, dass die Projekte vier der fünf Komponenten instruktionspsychologischer Theorien gut berücksichtigt und abgedeckt haben, während die zweite Komponente (Analyse von Lernvoraussetzungen) – bis auf gelegentliche Hinweise in den Diskussionsteilen der Beiträge – kaum zum Tragen gekommen ist.

Die Beiträge im Einzelnen

ASPE. Im Beitrag von Winther et al. aus dem ASPE-Projekt geht es um ein Online-Tool, mit dem Prüfungspersonal kompetenzorientierte Prüfungsaufgaben und Abschlussprüfungen für kaufmännische Berufe erstellen kann. Dabei geht es insbesondere darum, die Aufgaben so zu konstruieren, dass sie weder zu leicht noch zu schwierig sind, um reliable Kompetenzmessung über einen breiten Kompetenzbereich hinweg zu ermöglichen. Zur Konstruktion der Aufgaben wurden drei Schwierigkeitsdimensionen als Design-Kriterien betrachtet: „Art des erforderlichen Vorwissens“, „kognitive Komplexität der zur Lösung erforderlichen kognitiven Prozesse“ (in Anlehnung an die Bloom'sche Lernhierarchie; Bloom et al., 1956) und „Anteil an ablenkenden, lösungshinderlichen, irrelevanten Informationen in der Aufgabenstellung“ (die, so kann man es mit Blick auf die Cognitive-Load-Theorie beschreiben, „extraneous cognitive load“ erzeugen; Sweller, 2010). Im Beitrag wird empirisch gezeigt, dass Fachausschüsse Prüfungsaufgaben unterschiedlicher Schwierigkeit, die nach den Design-Kriterien entwickelt worden waren, übereinstimmend und erwartungskonform im Hinblick auf deren Schwierigkeit beurteilten, was einen relevanten Aspekt der inhaltlichen Validierung der Aufgaben betrifft. Gleiches gilt für die im Rasch-Modell empirisch ermittelte Schwierigkeit der Aufgaben. Darüber hinaus erwiesen sich die Design-Kriterien als geeignet, die Schwierigkeit früherer Aufgaben rückwirkend gut vorherzusagen.

Zusammenfassend ist festzustellen, dass es sich bei dem im ASPE-Projekt entwickelten Online-Tool um einen wünschenswerten, rationalen Ansatz handelt, komplexe Prüfungsaufgaben vorhersagbarer Schwierigkeit über einen breiten Schwierigkeitsbereich hinweg zu konstruieren. Die Rasch-Skalierbarkeit der Prüfungsaufgaben bietet zahlreiche Vorteile, insbesondere die Möglichkeit, die Schwierigkeit der Aufgaben an das erwartete Kompetenzniveau einzelner Personen oder Personengruppen anzupassen, was neben psychometrischen insbesondere auch motivationale Vorteile auf Seiten der zu Prüfenden hat. Optimierungspotenzial ist im Hinblick auf eine stärker kognitionspsychologisch ausgerichtete Ausschärfung der Schwierigkeitsdimension „Kognitive Komplexität“ zu sehen. Bisher ist diese Dimension an die Bloom'sche Taxonomie angelehnt. Alternativ könnte man sich z. B. am ESNaS-Modell orientieren, das vom Institut zur Qualitätsentwicklung im Bildungswesen (IQB) verwendet wird, um das Erreichen der Bildungsstandards in den Naturwissenschaften zu überprüfen (Walpuski et al., 2010). Aus lehr-lernpsychologischer Sicht hat das im Projekt entwickelte Online-Tool nicht nur Potential, die Prüfungspraxis deutlich zu verbessern, sondern insbesondere auch die Wirksamkeit der Ausbildung anzuheben, indem derart entwickelte Aufgaben nicht nur in summativen Abschlussprüfungen, sondern insbesondere auch in formativen Zwischenprüfungen eingesetzt werden. Diesen formativen Zweck erfüllen im allgemeinen Bildungssystem z. B. die VERA-Tests (<https://www.iqb.hu-berlin.de/vera>): Ein Jahr vor Ende eines Bildungsabschnitts wird anhand dieser Tests in der Grundschule und der Sekundarstufe I gemessen, wo die Schüler:innen mit ihrer Kompetenzentwicklung stehen, um erkennbare Kompetenzdefizite rechtzeitig

vor Ende des Bildungsabschnitts noch ausgleichen zu können. Insgesamt ist hier im ASPE-Projekt ein offensichtlich sehr brauchbares Werkzeug zur Entwicklung psychometrisch geeigneter Prüfungsaufgaben entwickelt worden, dessen Nutzbarkeit insbesondere auch für zukünftige Large-Scale-Assessments (LSA) im Bereich der beruflichen Bildung erkennbar ist.

DigiDIn-Kfz. Im Beitrag von Gschwendtner et al. aus dem DigiDIn-Kfz-Projekt geht es um ein simulationsbasiertes Instrument zur Förderung der Kompetenz von Kfz-Mechatroniker:innen, Störungen an Kraftfahrzeugen zu diagnostizieren. In ihrem Beitrag leisten die Autor:innen eine überzeugende Analyse der Diagnosekompetenz im Hinblick auf diagnoserelevantes Wissen, das Lesen und Verstehen von Schaltplänen und Anleitungen sowie das Verstehen und Anwenden geeigneter Diagnosestrategien beim Suchen nach der Ursache einer Störung. Sie berichten eine Reihe empirischer Interventionsstudien, in denen sie untersucht haben, wie sich die Facetten der Diagnosekompetenz fördern lassen. Dabei kamen unterschiedliche, lehr-lernpsychologisch gut begründete instruktionale Maßnahmen zum Einsatz, wie Modeling, Scaffolding, motivationale Anreize, beispielbasiertes Lernen sowie retrospektive und antizipatorische Selbsterklärungs-Prompts. Die Ergebnisse der Interventionsstudien zeigen, dass die untersuchten instrukionalen Maßnahmen im Hinblick auf die Verbesserung der Diagnosekompetenz nur begrenzt wirksamer waren als alternative, lehr-lernpsychologisch weniger gut begründete instruktionale Maßnahmen: Die Auszubildenden lernten generell auf vergleichsweise geringem Niveau, und Lerneffekte zeigten sich eher im Bereich der angeleiteten und weniger im Bereich der eigenständigen Nutzung geeigneter Diagnosestrategien.

Zusammenfassend ist bei diesem Projekt die theoriebasierte Kompetenzanalyse zu würdigen, die Entwicklung eines breit einsetzbaren und niederschwellig in die Ausbildungspraxis transferierbaren Simulationssystems sowie die Entwicklung und Erprobung theoriebasierter Ausbildungskonzepte zur Förderung der Kompetenzfacetten. Optimierungspotenzial ist im Hinblick auf eine Verbesserung der Reliabilität der Kompetenzmessung auszumachen: Grundlegendes Problem ist die Tatsache, dass die freie, nicht durch Anleitung unterstützte Störungssuche an einem Kfz z. T. eine Vielzahl an Problemlöseschritten erfordert, sodass die von einer Person bearbeitbare Anzahl an stochastisch voneinander unabhängigen Testaufgaben sehr begrenzt ist. Hier ist daran zu denken, ggf. Prozessvariablen zu erfassen, die über die einfache Feststellung von Diagnoseerfolg oder -misserfolg hinausgehen (siehe hierzu den Beitrag von Winther et al. aus dem PSA-Projekt). Darüber hinaus stellt sich die Frage, ob die Trainingszeit in den Interventionsstudien ausgereicht hat, um hinreichend starke Lerneffekte generieren und Unterschiede zwischen den untersuchten instrukionalen Maßnahmen aufzeigen zu können. Interessant sind die in der Diskussion angeführten Überlegungen der Autor:innen, mehr Wissen über diagnoserelevante mentale Konzepte und motivationale Zustände der Lernenden haben zu müssen, um die Lernumgebung adaptiv optimieren zu können (siehe die zweite Komponente instruktionspsychologischer Theorien: Analyse von Lernvoraussetzungen).

Aus lehr-lernpsychologischer Perspektive wird in diesem Beitrag aus dem DigiDIn-Kfz-Projekt sehr gut deutlich, wie wichtig es ist, nicht nur Kompetenzen zu analysieren und geeignete Messinstrumente zur Erfassung der Kompetenzen einzusetzen, sondern darüber hinaus auch theoriebasiert Vorstellungen zu entwickeln und zu untersuchen, wie der Erwerb der jeweiligen Kompetenzen gefördert werden kann.

EKGe. Im Beitrag von Wittmann et al. aus dem EKGe-Projekt geht es um den Einsatz eines Situational-Judgement-Tests zur Analyse der Effekte von Ausbildungsbedingungen in der Pflegepraxis. Der am Computer zu bearbeitende, auf Videoszenen basierende Test erfasst drei verschiedene Facetten der beruflichen Kompetenz von Pflegekräften: Kompetenz im Umgang mit Klient:innen (Klient:innenbezogene Kompetenz), Kompetenz im Umgang mit berufsbedingter Belastung (Belastungs- und Bewältigungskompetenz) und Kompetenz im Umgang mit Vertreter:innen anderer Berufsgruppen (interprofessionelle Kooperationskompetenz). In einer Fragebogenstudie erfassten die Autor:innen den Zusammenhang dieser Kompetenzfacetten mit Aspekten der Ausbildungsqualität in Praxisphasen unterschiedlicher Bereiche (ambulanter Pflegedienst, Station im Krankenhaus, Wohnbereich im Altersheim). Im Ergebnis zeigten sich schwache Korrelationen mit einigen der Ausbildungsaspekte, insbesondere mit den Aspekten „Funktionale Einbindung“, „Kommunikation und Konfliktbewältigung“, „Komplexität der Anforderungen“ und „Ausbleibendes negatives Feedback durch Klient:innen“. In der Gesamteinschätzung der Ausbildungsqualität schneidet die Altenpflege schlechter, die ambulante Pflege dagegen deutlich besser ab.

Positiv ist hervorzuheben, dass es in diesem Beitrag um die Rahmenbedingungen der Ausbildung in einem interpersonell besonders herausfordernden Beruf geht und um die Frage, inwieweit diese Rahmenbedingungen die von den Auszubildenden erreichten beruflichen Kompetenzen beeinflussen. Interessant wäre in diesem Zusammenhang, nicht nur den Zusammenhang der Rahmenbedingungen mit dem erreichten Kompetenzniveau, sondern insbesondere auch den Zusammenhang mit dem erreichten Kompetenzzuwachs in den jeweiligen Praxisphasen zu untersuchen. Aus lehr-lernpsychologischer Sicht erscheint der Fokus auf die äußeren Bedingungen, in denen sich berufliche Kompetenzen entwickeln, von großer Bedeutung. Dabei geht es insofern um die vierte Komponente instruktionspsychologischer Theorien (Instruktionsdesign), als zum Instruktionsdesign nicht nur gezielte Instruktion, sondern darüber hinaus auch die Gestaltung der äußeren Bedingungen gehört, unter denen die Instruktion möglichst gute Wirkungen entwickeln kann.

PSA-Sim. Im Beitrag von Winther et al. aus dem PSA-Sim-Projekt geht es um komplexes Problemlösen im Büroalltag auf der Basis von Wissen und Kooperation (z. B. Perspektivenübernahme). Um die erforderlichen kognitiven und sozialen Kompetenzen bei kaufmännischen Auszubildenden zu fördern, wurde in diesem Projekt ein Büro-Simulationssystem entwickelt und eingesetzt. Im Beitrag werden zwei Studien vorgestellt. Die erste Studie untersuchte die Messung der Kompetenzen mit vs. ohne adaptive Hilfe-Prompts durch simulierte Kooperationspartner:innen. Hier zeigte sich, dass das zugrunde liegende zweifaktorielle Kompetenzmodell unter beiden Messbedingungen funktionierte und dass in der Bedingung mit adaptiver Hilfe höhere

Kompetenzen gemessen wurden, wobei die beiden Faktoren hier auch stärker miteinander korrelierten. Die zweite Studie widmete sich dem Problem der mangelnden lokalen stochastischen Unabhängigkeit von Leistungsindikatoren, die in Performance Tasks wie der hier verwendeten Büro-Simulation erfasst werden. Den Autor:innen gelang es, 72 Leistungsindikatoren in 14 sequentielle Bearbeitungsmuster zu transformieren. Als Partial-Credit-Items ließen sich diese 14 Muster zwei Muster-Typen zuordnen: sprachlich-argumentative vs. mathematisch-analytische Muster. Empirisch zeigte sich, dass das zugrunde liegende zweidimensionale Kompetenzmodell die Daten gut anpassen konnte, wobei die sprachlich-argumentativen Kompetenzen höher ausfielen als die mathematisch-analytischen Kompetenzen, was einen Ruf nach fachspezifischer Förderung der Kompetenzen begründet.

Bei diesem Projekt beeindruckt vor allem die Erfassung von Prozessparametern als Indikatoren für erfolgreiches Problemlösen. Interessant wäre es zu untersuchen, ob sich anhand weiterer Simulationssysteme ähnliche Kompetenzwerte messen lassen wie mit der hier eingesetzten Büro-Simulation. Sollte sich dies zeigen lassen, wäre das ein weiterer überzeugender Hinweis auf die Validität der Kompetenzmessung anhand der hier verwendeten Prozessparameter. Aus lehr-lernpsychologischer Sicht hat das Projekt mit seinem Beitrag zur fünften Komponente instruktionspsychologischer Theorien (Mess- und Evaluationsvorschriften) Neuland betreten, indem es einen neuartigen Ansatz zur Kompetenzmessung in einem authentischen, komplexen Problemszenario entwickelt und erfolgreich validiert hat. Darüber hinaus hat sich das Projekt der interessanten Frage gewidmet, ob die Kompetenzmessung in einem solchen Szenario verbessert werden kann, wenn die Auszubildenden adaptive, d. h. situationsangepasste Prompts erhalten, welche geeignet sind, kompetenzbasiertes Verhalten überhaupt erst in Gang zu setzen. Da es hier durch die Hilfestellungen während des Testens möglicherweise zu Lernprozessen kommt, wäre es interessant gewesen, nachfolgend zu prüfen, ob das kompetenzbasierte Verhalten auch ohne die zuvor eingesetzten Prompts weiterhin gezeigt wird. Grundsätzlich erscheint es möglich, Lernen und Testen in ein und demselben Simulationssystem miteinander zu verbinden. Dabei stellt sich allerdings die Frage, wie sich die beim fortlaufenden Lernen und Testen erreichte Kompetenz am Ende reliabel und valide messen lässt (Leutner, 1992, 1993).

TechKom. Der Beitrag von Damman et al. aus dem TechKom-Projekt fokussiert weniger auf Kompetenzen als auf formale (d. h. nicht inhaltliche) schwierigkeitsgenerierende Merkmale von Prüfungsaufgaben zur Kompetenzdiagnostik in der gewerblich-technischen Berufsbildung. Neben der Entwicklung von Förderinstrumenten für analytische und konstruktive Problemlösekompetenzen war die Verbesserung von Prüfungsaufgaben ein weiterer Schwerpunkt des TechKom-Projekts. Die Autor:innen berichten über eine empirische Analyse der Schwierigkeit von PAL-Aufgaben für Mechatroniker:innen. Unter anderem vor dem Hintergrund der Forschung zum Lernen mit Multimedia (Mayer, 2021) wurden, was die inhaltsunabhängige Gestaltung der Prüfungsaufgaben betrifft, die Effekte von Verstößen gegen das Split-Attention-Prinzip, Effekte des Vorhandenseins oder Nichtvorhandenseins von Kohärenzbildungshilfen für die Bild-Text-Integration, Effekte der Lesekomplexität des Aufgabentexts und

Effekte der Anzahl an Fachbegriffen untersucht. Im Ergebnis zeigten sich erwartungsgemäße, aber nur kleine schwierigkeiterhöhende Effekte der Anzahl an Fachbegriffen und von Verstößen gegen das Split-Attention-Prinzip. Im Gegensatz dazu zeigten sich erwartungswidrige Effekte im Zusammenhang mit Kohärenzbildungshilfen zur Text-Bild-Integration: Aufgaben mit und ohne Kohärenzbildungshilfen unterschieden sich nicht in ihrer Schwierigkeit, wobei Aufgaben ohne Bilder generell leichter zu lösen waren als Aufgaben mit Bildern.

Im Hinblick auf dieses Projekt ist würdigend hervorzuheben, dass es sich mit „extraneous cognitive load“, also unnützer kognitiver Belastung, befasst, die durch die kognitive Verarbeitung irrelevanter Informationen in der Aufgabenstellung einer Prüfungsaufgabe entstehen kann. Damit ergänzt es in sinnvoller Weise die von Winther et al. in ihrem Beitrag aus dem ASPE-Projekt berichtete Forschung, indem weitere, potenziell „extraneous cognitive load“-verursachende Aufgabenmerkmale in den Blick genommen werden. In den hier berichteten Ergebnissen aus dem TechKom-Projekt zeigte sich, dass Aufgaben mit Bildern schwieriger waren als Aufgaben ohne Bilder, wobei etwaig vorhandene Kohärenzbildungshilfen zur Text-Bild-Integration keinen Effekt hatten. Auf den ersten Blick scheint dieses Ergebnis dem aus der Multimedia-Forschung bekannten Multimedia-Prinzip zu widersprechen (Mayer, 2021), wonach ein Text mit Bildern im Allgemeinen verständlicher ist als derselbe Text ohne Bilder. Auf den zweiten Blick könnte dieser Schwierigkeitsunterschied jedoch durch andere Eigenschaften der Aufgaben verursacht gewesen sein, wie z. B. inhaltlich höhere Komplexität der Problemstellung, die ein erläuterndes Bild erforderlich macht, um die Problemstellung überhaupt erst verstehen zu können. Dies könnte erklären, warum Aufgaben mit Bildern schwieriger waren als Aufgaben ohne Bilder. Insgesamt werden aus lehr-lernpsychologischer Sicht in diesem Beitrag aus dem TechKom-Projekt interessante Analysen berichtet, deren Ergebnisse für die Gestaltung von Prüfungsaufgaben sinnvoll genutzt werden können.

TeKop. Im Beitrag von Seeber et al. aus dem TeKop-Projekt geht es um die Konstruktion von technologiebasierten Prüfungsaufgaben zur Messung analytischer kaufmännischer Problemlösekompetenz. Im Projekt wurde ein entsprechendes Training entwickelt, und die im vorliegenden Beitrag empirisch zu klärende Frage war, ob die nach erfolgtem Training entwickelten Prüfungsaufgaben tatsächlich hinreichend „problemhaltig“ sind, um Problemlösekompetenz messen zu können. Zur Beantwortung der Frage wurden Auszubildende einem kognitiven Interview unterzogen, d. h. sie wurden gebeten, während der Aufgabenbearbeitung laut zu denken. Im Ergebnis zeigte sich, dass die Auszubildenden bei der Bearbeitung der Aufgaben tatsächlich Barrieren zu überwinden hatten (was kennzeichnend für Problemlösen ist) und dass sie zur Überwindung der Barriere tatsächlich zuvor definierte Problemlösungsschritte durchliefen.

In diesem Beitrag wurde eine wirksame Methode eingesetzt, um die inhaltliche Validität von Prüfungsaufgaben zu untersuchen: Wenn die Aufgaben Problemlösekompetenz messen sollen, dann müssen sie eine nicht durch routiniertes Vorgehen überwindbare Barriere enthalten, und bei der Überwindung der Barriere sollte eine

schrittweise Vorgehensweise erkennbar sein, die für Problemlösen typisch ist (z. B. Situationsanalyse, Hypothesenformulierung, Planung des Vorgehens, Umsetzung und Kontrolle). Bemerkenswert ist, dass sich mithilfe der kognitiven Interviews nicht nur die Validität der Aufgaben belegen ließ, sondern indirekt darüber hinaus auch die Wirksamkeit des Trainings zur Entwicklung solcher Aufgaben. Potential zur Optimierung der kognitiven Interviews könnte ggf. darin bestehen, die Auszubildenden bei der Bearbeitung der Aufgaben zu videografieren, ihnen danach das Video zu präsentieren und sie zu bitten, anhand des Videos ihre Gedankengänge bei der Bearbeitung der Aufgaben zu erläutern. Ein solches kognitives Interview auf Basis der videografierten Aufgabebearbeitung würde die Gefahr reduzieren, dass die kognitiven Prozesse durch die Aufforderung zum lauten Denken während der Aufgabebearbeitung verändert werden. Eine solche Vorgehensweise würde aber nur Sinn machen, wenn im Video hinreichend viele Verhaltensweisen und „Spuren“ der Aufgabebearbeitung zu sehen sind, die dann im Hinblick auf die Überlegungen während der Bearbeitung erläutert werden können. Insgesamt leistet das Projekt aus lehr-lernpsychologischer Sicht einen wichtigen Beitrag zur inhaltlichen Validierung von Aufgaben zur Messung kaufmännischer Problemlösekompetenz.

Zusammenfassende Würdigung der Beiträge

Die sechs Projekte der Forschungs- und Transferinitiative ASCOT+ belegen nachdrücklich große Fortschritte in der erfolgreichen Entwicklung, Nutzung und empirischen Validierung technologiebasierter Lern- und Testumgebungen. Aus lehr-lernpsychologischer Sicht erscheinen diese Fortschritte nicht nur für die berufliche Bildung, sondern darüber hinaus auch für alle jene Lernbereiche relevant, in denen es um den Erwerb und die Überprüfung von Handlungskompetenzen zur Lösung komplexer Probleme geht, also um deutlich mehr als um Wissen und Verstehen im Sinne tradierter Lehrzielhierarchien (z. B. Bloom et al., 1956, oder deren Revision durch Anderson & Krathwohl, 2001).

Im Hinblick auf die schon im Jahr 1992 beschriebenen Komponenten instruktionspsychologischer Theorien (Snow & Swanson, 1992; vgl. Leutner & Wirth, 2018) liegen besondere Stärken der Projekte in der durchweg sehr sorgfältigen Analyse der zu vermittelnden Kompetenzen (*Lehrzieldefinition und Lehrstoffanalyse*), der teilweise sehr gründlichen *Analyse des angestrebten Lernprozesses*, dem durchweg überzeugenden *Instruktionsdesign* sowie der sehr gelungenen Implementation von *Mess- und Evaluationsvorschriften* inkl. empirischer Validierung der entwickelten Instrumente. Hinsichtlich der *Analyse von Lernvoraussetzungen* ist allerdings noch Optimierungspotenzial zu sehen.

Literatur

- Anderson, L. W., & Krathwohl, D. R. (Eds.) (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Bloom, B. S., Engelhart, M. B., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives* (Handbook I: Cognitive domain). Longmans Green.
- Leutner, D. (1992). Das Testlängendilemma in der lernprozess-begleitenden Wissensdiagnostik. *Zeitschrift für Pädagogische Psychologie*, 6, 233–238.
- Leutner, D. (1993). Das gleitende Testfenster als Lösung des Testlängendilemmas: Eine Robustheitsstudie. *Zeitschrift für Pädagogische Psychologie*, 7, 33–45.
- Leutner, D., & Wirth, J. (2018). Instruktionspsychologie. In D. Rost, J. R. Sparfeldt, & S. Buch, *Handwörterbuch Pädagogische Psychologie* (5. Aufl., S. 269–278). Beltz-PVU.
- Mayer, R. E. (2021). *Multimedia Learning*. Cambridge University Press.
- Snow, R. E., & Swanson, J. (1992). Instructional psychology: Aptitude, adaptation, and assessment. *Annual Review of Psychology*, 43, 583–626. <https://doi.org/10.1146/annurev.ps.43.020192.003055>
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22, 123–138. <https://doi.org/10.1007/s10648-010-9128-5>
- Walpuski, M., Kauertz, A., Kampa, N., Fischer, H. E., Mayer, J., Sumfleth, E., & Wellnitz, N. (2010). ESNaS – Evaluation der Standards für die Naturwissenschaften in der Sekundarstufe I. In A. Gehrman, U. Hericks, & M. Lüders, *Bildungsstandards und Kompetenzmodelle* (S. 171–184). Klinkhardt.

Schwerpunkt 1: Technologiebasierte Förderung beruflicher Kompetenzen

Digitale Interventionen zur Förderung der Diagnosekompetenz bei Kfz-Mechatroniker:innen

TOBIAS GSCHWENDTNER, STEPHAN ABELE, INGA GLOGGER-FREY, STEFAN HARTMANN, PETER HESSE, JÖRDIS KAEPERNICK & DAVE REXHÄUSER

Zusammenfassung

In dem vorliegenden Beitrag werden verschiedene digitale Interventionen zur Förderung der Diagnosekompetenz in der Ausbildung von Kfz-Mechatroniker:innen vorgestellt, die im Rahmen des Verbundvorhabens DigiDIn-Kfz durchgeführt wurden. Die Interventionen verfolgen übergreifend das Ziel, die Fähigkeit zu fördern, Störungen an elektrischen und elektronischen Systemen von Kraftfahrzeugen zu diagnostizieren. Sie versuchen dieses Ziel allerdings über die Förderung unterschiedlicher Facetten der Kfz-Diagnosekompetenz zu erreichen. Im Beitrag werden Einblicke in die Herleitung und Definition der Fördergegenstände, die Forschungsmethodik sowie die Gestaltungsmerkmale und die erzielten Effekte der Interventionen gegeben. Insgesamt sind die mit den Interventionen erzielbaren Effekte aus Sicht der beruflichen Ausbildung hoch bedeutsam, wenngleich sich die Erwartungen nicht durchgängig eingelöst haben, beispielsweise an einigen Stellen hinsichtlich des Diagnoseerfolgs.

Schlagworte: Diagnosekompetenz, Lernen mit Computersimulationen, multimediale Lernumgebungen, beispielbasiertes Lernen

Abstract

We report on various digital interventions aiming at promoting diagnostic competences in the training of apprentice automotive technicians, which were implemented as part of the joint project DigiDIn-Kfz. The interventions pursue the overarching goal of promoting the skill of diagnosing faults in electric and electronic car systems. However, in each intervention we try to achieve this goal by aiming at different facets of automotive diagnostic competence. The article provides insights into the derivation and definition of the objects of promotion, the research methodology, the design features, and the effects achieved by the interventions. Overall, the effects achieved with the interventions are highly important from the perspective of vocational training, though the expectations were not consistently fulfilled, for example in some areas with regard to the diagnostic success.

Keywords: Diagnostic Competence, Learning with Computer Simulations, Multimedia Learning Environments, Example-Based Learning

1 Kfz-Diagnosekompetenz

Das Diagnostizieren von Störungen an Kraftfahrzeugen nimmt einen Großteil der Arbeitszeit in Kfz-Werkstätten ein (Spöttl, Becker & Musekamp, 2011) und stellt eine Kernkompetenz von Kfz-Mechatroniker:innen dar (Abele, 2023). Die Förderung dieser Kompetenz ist daher ein wichtiges Ausbildungsziel auf nationaler und internationaler Ebene (Rausch et al., 2024). Aufgrund der zunehmenden Komplexität und Vernetzung von Fahrzeugsystemen, z. B. infolge steigender Sicherheits- und Komfortanforderungen, wird die Bedeutung der Kfz-Diagnosekompetenz vor allem im Bereich elektrischer und elektronischer Fahrzeugsysteme voraussichtlich noch weiter zunehmen. Gleichzeitig sind Diagnosen gerade an solchen Systemen besonders herausfordernd und nur ein vergleichsweise kleiner Teil der Auszubildenden in Deutschland erreicht am Ende der Ausbildung ein zufriedenstellendes Kfz-Diagnosekompetenzniveau (Abele et al., 2014; Güzel et al., 2024; Nickolaus et al., 2012).

Vor diesem Hintergrund präsentieren und diskutieren wir in diesem Beitrag Ergebnisse von Interventionsstudien zur Förderung der Kfz-Diagnosekompetenz. Mit „Kfz-Diagnosekompetenz“ bezeichnen wir die kontextspezifischen kognitiven Leistungsdispositionen, die Individuen in die Lage versetzen, die Ursachen von Störungen der Funktion von Kraftfahrzeugsystemen zu diagnostizieren.

Ein Blick auf den Diagnoseprozess offenbart einige Facetten dieser kontextspezifischen kognitiven Leistungsdispositionen; dieser lässt sich in die folgenden vier Schritte unterteilen (Abele, 2023; Norwig et al., 2021): Zunächst sind (1) diagnoserelevante Informationen mental zu repräsentieren, daran anknüpfend (2) Hypothesen zu potenziellen Störungsursachen zu formulieren und (3) diese Hypothesen sukzessive anhand von Messungen zu prüfen. Im Anschluss daran ist (4) zu bewerten, ob die betreffenden Hypothesen auf die tatsächliche Störungsursache verweisen oder ob sie zu verwerfen und andere Hypothesen zu formulieren und zu untersuchen sind.

Die für eine erfolgreiche Diagnose relevanten Informationen müssen einerseits als **Diagnosewissen** aus dem Langzeitgedächtnis abgerufen, andererseits durch das Studieren verschiedener extern gegebener Dokumente mental modelliert werden. Dafür benötigen Auszubildende **diagnoserelevante Rezeptionskompetenzen**, also spezifische berufssprachliche Kompetenzen, die Auszubildende befähigen, diagnoserelevante Dokumente aufmerksam und verständlich zu rezipieren und adäquate mentale Repräsentationen der darin enthaltenen Informationen zu bilden. Dabei differenzieren wir zwischen einer **basalen** und einer **komplexen diagnoserelevanten Rezeptionskompetenz**. Beide Kompetenzen betrachten wir als Facetten der Kfz-Diagnosekompetenz (s. Abbildung 1; vgl. auch Güzel et al., 2024). Wenn es um die Bewältigung von Anforderungen der „Identifikation, Rezeption und mentale[n] Modellierung lokaler und teilweise verstreuter Informationen in unterschiedlichen alltagsnahen Textarten“ (Norwig et al., 2021, S. 665) geht, verwenden wir den Begriff der **basalen diagnoserelevanten Rezeptionskompetenz**. Solche Anforderungen finden sich bei der Störungsdiagnose bspw. beim Lesen des Werkstattauftrags oder von Einbaulageplänen von Komponenten und der Schritt-für-Schritt-Messanleitungen des computerbasierten As-

sistenzsystems. Wenn es um die Bewältigung von Anforderungen der Informationsgewinnung aus komplexen fachsprachlichen Informationsquellen wie mehrseitigen technischen Schemazeichnungen (beispielsweise elektrische Schaltpläne) geht, verwenden wir den Begriff der **komplexen diagnoserelevanten Rezeptionskompetenz**. Diese umfasst die Identifikation, Rezeption und „mentale Modellierung elektrischer Bauteil- und Systemdarstellungen“ (ebd.) sowie die mentale Modellierung der Funktionslogik der elektrischen bzw. elektronischen Schaltung. Nicht zuletzt vor dem Hintergrund des funktionalen Charakters von Lesen in beruflichen Kontexten (Ziegler et al., 2012) bedeutet mentale Modellierung der Funktionslogik der elektrischen bzw. elektronischen Schaltung ein im Lesen erfolgreiches mentales Probehandeln, bei dem auf Basis von **Diagnosewissen** u. a. konkrete Messstrategien und kontextspezifische Messungen (inkl. Ist-Wert-Erwartungen) entwickelt, gedanklich durchgeführt und reflektiert werden müssen (Abbildung 1).

Die beiden rezeptiven Kompetenzen sind strukturell aufeinander aufbauend zu denken: Im Regelfall sind schwierigere Störungsfälle dadurch geprägt, dass zunächst auf Basis der basalen Rezeptionskompetenz Schritt-für-Schritt-Messanleitungen der computerbasierten Assistenzsysteme rezipiert und abgearbeitet werden. Damit wird der Suchraum eingegrenzt. Daran anschließend – wenn die Messhinweise des Assistenzsystems nicht hinreichend für eine erfolgreiche Störungsdiagnose sind – wird mit Hilfe der komplexen Rezeptionskompetenz die Fehlerursache systematisch weiter eingegrenzt und schlussendlich identifiziert. Bei einfacheren Störungsfällen, in denen die Messhinweise des computerbasierten Assistenzsystems hinreichend zur Fehlersuche sind, ist nur die basale, nicht aber die komplexe diagnoserelevante Rezeptionskompetenz eine notwendige Bedingung für den Diagnoseerfolg.

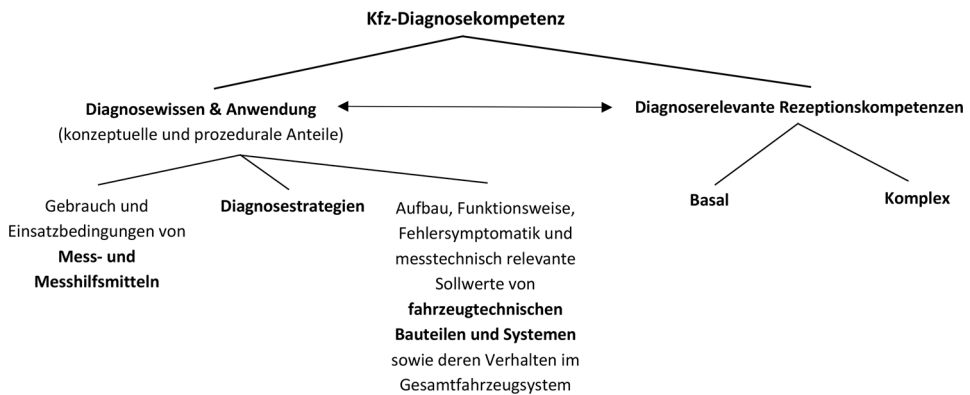


Abbildung 1: Strukturmodell der Kfz-Diagnosekompetenz mit den Kompetenzfacetten Diagnosewissen und diagnoserelevante Rezeptionskompetenz auf der ersten Ebene (der Interaktionspfeil deutet die wechselseitigen Bezüge zwischen Diagnosewissen und diagnoserelevanten Rezeptionskompetenzen an)

Studien zeigen, dass viele Kfz-Auszubildende erhebliche Defizite in diesem Kompetenzbereich aufweisen (Nickolaus et al., 2012; Norwig et al., 2021). Abhängig vom berufsfachlichen Leistungsniveau können bei den Auszubildenden unterschiedliche

Rezeptionsschwierigkeiten auftreten. Basierend auf einem Niveaumodell zur Kfz-Diagnosekompetenz wurde von Nickolaus et al. (2012) festgestellt, dass ca. 17 Prozent der Auszubildenden nicht in der Lage sind, einfachere diagnoserelevante Informationen (z. B. aus Arbeitsaufträgen, Einbaulageplänen) zu verarbeiten und für das Stellen einer Diagnose zu nutzen. Besonders diagnoseschwachen Auszubildenden mangelt es demzufolge an **basaler diagnoserelevanter Rezeptionskompetenz**. Etwa 52 Prozent der Auszubildenden können elektrische Schaltpläne, ein weiteres zentrales Format der diagnoserelevanten Informationsdarbietung, nicht hinreichend rezipieren. Diese Auszubildenden haben Defizite im Bereich der **komplexen diagnoserelevanten Rezeptionskompetenz**. Sie können elektrische Bauteil- und Systemdarstellungen sowie die elektrische Funktionslogik nicht adäquat mental modellieren und sind somit nicht in der Lage, die entsprechenden Informationen für weitere Diagnoseschritte zu nutzen, was sie beim Stellen einer Diagnose oft scheitern lässt (Nickolaus et al., 2012, S. 251 ff.). Man muss davon ausgehen, dass für die Bewältigung von beruflichen Herausforderungen diese zentralen Facetten der Diagnosekompetenz bei ca. der Hälfte der Auszubildenden nicht hinreichend ausgeprägt sind.

Darüber hinaus liefert die genannte Arbeit Hinweise, dass viele Auszubildende in der Ausbildung keine effektiven Kfz-Diagnosestrategien erwerben. **Kfz-Diagnosestrategien** als weitere Facette der Kfz-Diagnosekompetenz (Abbildung 1) werden benötigt, um den Kfz-Diagnoseprozess bzw. die darin auftretenden Anforderungen erfolgreich zu regulieren und die Kfz-Störungsursachen erfolgreich zu identifizieren. Empirisch-analytisch kann zwischen einer computerbasierten, einer fallbasierten und einer modellbasierten Diagnosestrategie unterschieden werden. Die **fallbasierte Strategie**, die bei Abele und von Davier (2019) beschrieben wird, wurde in den hier thematisierten Interventionsstudien nicht gefördert, weil sie über viel Erfahrung entsteht und deren Förderung den zeitlichen Rahmen des Projekts gesprengt hätte. Ein vergleichsweise großer Teil der Auszubildenden beherrscht die **computerbasierte Diagnosestrategie** („geführte Fehlersuche“), bei der eine Schritt-für-Schritt-Diagnoseanleitung aus spezifischen Dokumenten des computerbasierten Assistenzsystems aufgerufen, gelesen und abgearbeitet wird (Abele & von Davier, 2019). Sie erfordert im Wesentlichen die oben genannte basale diagnoserelevante Rezeptionskompetenz. Die **modellbasierte Diagnosestrategie** zeichnet sich dadurch aus, dass Ausschnitte problemrelevanter Kfz-Systeme mental modelliert werden; das relevante Kfz-System muss also verstanden werden. Auf Basis des mentalen Modells werden dann systematisch relevante Hypothesen zu Störungsursachen formuliert und getestet. Im Anschluss an das oben erwähnte Niveaumodell ist anzunehmen, dass die modellbasierte Strategie lediglich von diagnosestarken Auszubildenden beherrscht wird (Nickolaus et al., 2012). Zu vermuten ist ferner, dass die modellbasierte Strategie erst angewandt werden kann, wenn Auszubildende beim Diagnostizieren extern dargebotene Informationen adäquat zur Bildung mentaler Modelle nutzen können, d. h. wenn sie sowohl über basale als auch komplexe diagnoserelevante Rezeptionskompetenzen verfügen (vgl. die in Abbildung 1 dargestellte Wechselwirkung). In der Studie von Abele und von Davier (2019) beherrschten nur ca. 14 Prozent der Auszubildenden die modellbasierte Strategie. Dieser Befund wiegt deshalb besonders schwer, weil die modellbasierte Kfz-Dia-

gnosestrategie im Vergleich zur computer- und fallbasierten Strategie bei sehr unterschiedlichen und anspruchsvollen Störungen mit höherer Wahrscheinlichkeit zum Diagnoseerfolg führt.

Vor diesem Hintergrund wurden im Rahmen des Projekts DigiDIn-Kfz¹ (Digitale Diagnostik und Intervention im Kfz-Wesen) der Forschungs- und Transferinitiative ASCOT+ drei Interventionen zur Förderung der Kfz-Diagnosekompetenz für verschiedene Kompetenzfacetten und Kompetenzausprägungen entwickelt und empirisch evaluiert: für eher diagnoseschwache Auszubildende eine **Intervention zur Förderung der basalen diagnoserelevanten Rezeptionskompetenz** als eine zentrale Voraussetzung der computerbasierten Diagnosestrategie; für Auszubildende des mittleren Leistungsniveaus eine **Intervention zur Förderung der komplexen diagnoserelevanten Rezeptionskompetenz**, die u. a. auf eine adäquate mentale Modellbildung beim Rezipieren elektrischer Schaltpläne abzielte und damit auch ein impliziter Baustein der modellbasierten Strategie ist (beide Interventionen basal und komplex werden integrativ in Abschnitt 3.1 dargelegt); sowie insbesondere für leistungsstarke Auszubildende schließlich eine **Intervention zur Förderung der modellbasierten Kfz-Diagnosestrategie** (Abschnitt 3.2). Bevor im Weiteren die Inhalte, die Gestaltungsmerkmale und die Effekte der Fördermaßnahmen beschrieben werden sowie abschließend eine Diskussion mit dem Ziel geführt wird, praktische Implikationen und weitere Forschungsbedarfe darzustellen (Abschnitt 4), wird nun zunächst mit der Kfz-Computersimulation ein zentrales Instrument beschrieben, das in allen Interventionsstudien eine wichtige Rolle spielte.

2 Kfz-Computersimulation

Die in DigiDIn-Kfz genutzte Kfz-Computersimulation (im Folgenden: Simulation) ist eine digitale Arbeitsumgebung, in der verschiedenste Störungen ausgewählter elektrischer und elektronischer Kfz-Systeme realitätsnah simuliert und diagnostiziert werden können. In den Kfz-Systemen Motorsteuerung, Beleuchtung, Komfortelektronik und Fahrwerk wurden bislang rund 80 Störungsszenarien entwickelt, die zum Teil öffentlich zugänglich sind.² Die Störungsszenarien bilden realistische, häufig auftretende Störungsursachen wie Komponentendefekte, Leitungsunterbrechungen und Übergangswiderstände ab; ein Großteil der implementierten Störungsszenarien entfällt auf die Motorsteuerung.

Im Vergleich zur Diagnose am echten Fahrzeug bietet die mehrfach überarbeitete, erweiterte und validierte Simulation (Gschwendtner et al., 2009; Abele et al., 2016) eine Reihe von Vorteilen: Sie erlaubt ein gleichzeitiges Arbeiten einer großen Zahl von Auszubildenden, wohingegen an einem realen Fahrzeug nur Einzelpersonen oder Kleingruppen diagnostizieren können. Alle Messungen können grundsätzlich zeit- und ortsunabhängig sowie zerstörungsfrei, d. h. ohne das Risiko einer Beschädigung

1 Das Projekt wurde aus Mitteln des Bundesministeriums für Bildung und Forschung gefördert (Förderkennzeichen: 21AP004A-C).

2 <https://kfz-simulation.tools-bpaed.gsw.tu-dresden.de/system>

empfindlicher und teurer Komponenten, durchgeführt werden. Daneben können die verschiedenen Störungsszenarien per Mausklick ausgewählt werden, während am realen Fahrzeug eine mitunter aufwändige Präparation erforderlich ist, um defekte Komponenten, Steuergeräte oder Leitungen zu installieren.

Die Kfz-Systeme wurden mittels Fotos und authentischen Zeichnungen mit hoher Detailtreue abgebildet und sind über einen Auswahlbildschirm erreichbar (Abbildung 2). Nutzende können mit verschiedenen Bildschirmbereichen interagieren. Bei Aktivierung durch Mouseover färben sich diese Bereiche deutlich sichtbar ein und zeigen kurze Textinformationen zur möglichen Interaktion. Nutzende können die Interaktion mit einem Mausklick auslösen und so beispielsweise vergrößerte Ansichten relevanter Bereiche erhalten.

Fahrzeugübersicht / Systemauswahl

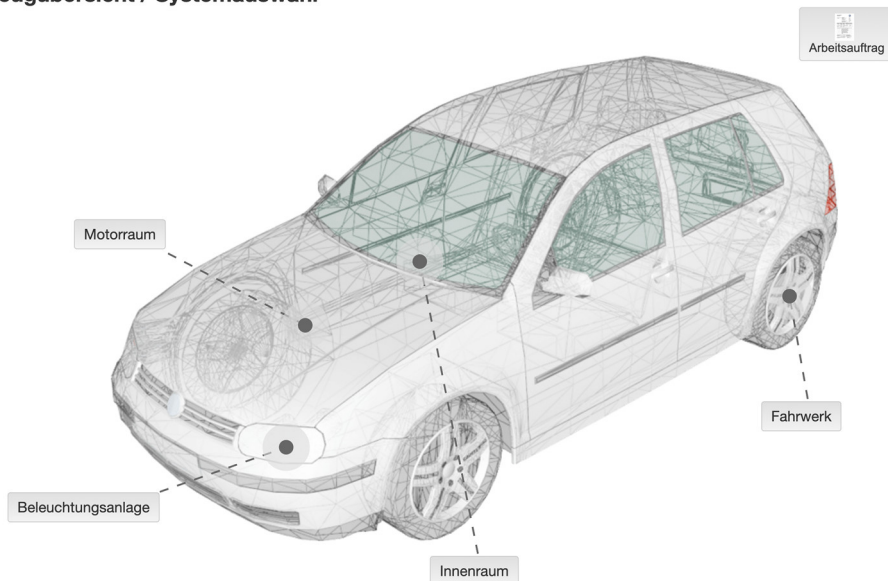


Abbildung 2: Systemauswahl im Übersichtsbildschirm der Kfz-Computersimulation

Über ein Menüband am oberen Bildschirmrand (Abbildung 3) können Nutzende jederzeit auf alle zur Störungsdiagnose notwendigen Arbeitsmittel zugreifen. Als zentrale Informationsquelle zu den Fahrzeugsystemen dient in der Simulation das computerbasierte Assistenzsystem ESI[tronic] 2.0 von Bosch. Die ESI[tronic] wurde aus der Auswahl verfügbarer Systeme bevorzugt, da sie herstellerübergreifend einsatzfähig ist, international genutzt wird und breit sowie tief in der dualen Ausbildung der Kfz-Mechatronik in Deutschland verankert ist. Es wurden die für das simulierte Fahrzeug relevanten Ausschnitte der ESI[tronic] nachgebildet und in das Programm integriert. Nutzende können so Daten zum Aufbau, zur Funktionsweise und zur Störungsdiagnose der verschiedenen Fahrzeugsysteme abrufen. Damit stehen alle für das simulierte Fahrzeug relevanten Informationen wie beispielsweise Einbaulagen, elektrische

Schaltpläne und Prüfanleitungen zur Verfügung. Auch das Auslesen und Löschen des Fehlerspeichers verschiedener Steuergeräte sowie der Abruf ausgewählter Istwerte in der Motorsteuerung (z. B. Kraftstofftemperatur) stehen als realistische Handlungsmöglichkeiten in der Simulation zur Verfügung.

Arbeitsauftrag	Messmittel	Hilfsmittel	ESI[tronic]	FAQ	Notizzettel
	Multimeter 	Cockpit 			
	Oszilloskop 	Sicherungskasten 			
	Strommesszange 	Batterie 			
	Adapterstecker 	Fahrzeug 			
	Buchsenkasten 				

Abbildung 3: Menüband mit ausgeklappten Untermenüs „Messmittel“ und „Hilfsmittel“

Zur Störungsdiagnose können Nutzende akustische, optische und elektrische Prüfungen bzw. Messungen durchführen. Elektrische Messungen können an einprogrammierten Messpunkten durchgeführt werden. Führen Nutzende eine der möglichen elektrischen Messungen (z. B. eine Signalprüfung) durch, erhalten sie Messergebnisse (z. B. Kennlinien im Oszilloskop), die dem aktivierten Störungsszenario entsprechend aus einer Datenbank abgerufen werden.

Zusätzlich können sie in der integrierten ESI[tronic] realitätsnah Informationen gewinnen, um beispielsweise korrekte Messpunkte zu finden oder erhaltene Messergebnisse einem Soll-Ist-Vergleich zu unterziehen. Weitere Arbeitsschritte wie die Erzeugung notwendiger Fahrzeugzustände (z. B. „Zündung eingeschaltet“) oder den Anschluss korrekter Messmittel an entsprechende Messpunkte können Nutzende realitätsnah durchführen. In den folgenden Kapiteln zeigen wir, wie die Simulation zur Kompetenzförderung genutzt wurde.

3 Förderung der Kfz-Diagnosekompetenz

3.1 Basale und komplexe diagnoserelevante Rezeptionskompetenz

Fördergegenstand und Gestaltung der Fördermaßnahmen

In bisherigen Studien ergaben sich bereits erste Hinweise zum Einfluss rezeptiver berufs- und fachsprachlicher (Lese-)Anforderungen auf den Diagnoseerfolg (Nickolaus et al., 2012). Analytisch betrachtet dient die Bewältigung solcher Leseanforderungen

im Diagnoseprozess der Konstruktion mentaler Repräsentationen (Schnotz, 2005). Diesen kommt, wie bereits in Abschnitt 1 dargelegt, eine Schlüsselrolle im Diagnoseprozess zu, da in allen Diagnoseschritten diagnoserelevante Informationen sukzessive gewonnen, aktiviert, mental integriert, modelliert sowie bewertet werden müssen (Abele, 2023).

Vor dem Hintergrund des hohen Prozentsatzes an Kfz-Auszubildenden, die am Ende ihrer Ausbildung über (teilweise sehr) geringe Diagnoseerfolge bei Kfz-Störungen verfügen (Gschwendtner et al., 2009; Nickolaus et al., 2012; siehe auch Abschnitt 1) und der Bedeutung von Rezeptionskompetenzen für den Diagnoseerfolg wurde angenommen, dass Diagnoseerfolge bei Kfz-Auszubildenden wahrscheinlicher werden, wenn berufs- und fachsprachliche Rezeptionskompetenzen gefördert werden.

Als Ausgangspunkt zur Gestaltung zielgerichteter Fördermaßnahmen diente ein Untersuchungszuschnitt, der klären sollte, welche Verhaltensweisen und Wissensbestände eine erfolgreiche Diagnose auszeichnen bzw. welche Barrieren den Auszubildenden im Wege stehen. Dafür wurden die Störungsdiagnoseprozesse von 24 Auszubildenden im 3. Ausbildungsjahr im Rahmen einer qualitativ angelegten Studie von Norwig et al. (2021) unter Nutzung von Cognitive Labs und Think-Aloud-Interviews sensu Leighton (2017) analysiert. In dieser Studie wurden viele Barrieren identifiziert, die einer erfolgreichen Diagnose im Weg stehen: (1) Bei der Informationsgewinnung aus alltagstypischen, also basalen berufssprachlichen Texten, die in Form des Werkstattauftrags und der Schritt-für-Schritt-Messanleitungen des computerbasierten Assistenzsystems³ bei der Störungsdiagnose genutzt werden, zeigten sich u. a. oftmals nicht sicher beherrschte zentrale Lesestile. Dies führt dazu, dass trotz sprachlich präziser Messanleitungen Messungen falsch umgesetzt oder Messergebnisse falsch interpretiert werden. (2) Bei der Informationsgewinnung aus komplexen fachsprachlichen Informationsquellen wie mehrseitigen technischen Schemazeichnungen fehlten oft Wissen zu elektrotechnischen Fachkonzepten oder Kenntnisse über Darstellungskonventionen/Symbolik der hersteller- bzw. diagnosesystemspezifischen elektrischen Schaltpläne.

Ausgehend von den von Norwig et al. (2021) herausgearbeiteten besonders verbreiteten fehlerbehafteten Verhaltensweisen/Wissensbeständen, die von mind. 25% der Proband:innen bei basalen und komplexen diagnoserelevanten Rezeptionsanforderungen gezeigt wurden, entwickelten wir unter Zuhilfenahme zentraler Designideen des 4C/ID-Ansatzes (van Merriënboer, 2020) und Prinzipien multimedialen Lernens (Mayer, 2014) je eine digitale, browserbasierte, multimediale Intervention zur Förderung der basalen und der komplexen diagnoserelevanten Rezeptionskompetenz.

Die beiden Interventionen „basal“ und „komplex“ bestanden jeweils aus einer in der Arbeitsgruppe unter Anwendung (fach-)didaktischer Kriterien diskursiv ausgehandelten Abfolge aus Erklärvideos (inkl. Animationen und Screencasts, die bspw. spezifische Handlungsabläufe in der Simulation darstellen) und direkt darauf bezogener

3 Die in computerbasierten Assistenzsystemen verfügbaren Schritt-für-Schritt-Messanleitungen sind mit teils hoher Dichte an Begriffen, Referenzwerten, Wortlisten (Legenden), zwei- oder dreidimensionalen technischen Darstellungen und Diagrammen versehen.

Übungsaufgaben. In den Erklärvideos wurden die von Norwig et al. (2021) identifizierten besonders häufigen fehlerbehafteten rezeptiven Verhaltensweisen/Wissensbestände durch das Darstellen, Beschreiben und Erklären der richtigen Verhaltensweisen instruktional aufgegriffen und die Probanden damit entsprechend konfrontiert (*Modeling*). Übungsaufgaben mit geschlossenen Antwortformaten dienten dann der gezielten Übung der richtigen Verhaltensweisen und waren nach den Prinzipien der zunehmenden Komplexität und der Kontextvariation gestaltet, teils in einem interaktiven elektrischen Schaltplan oder in der Simulation situiert (Güzel et al., 2024). Nach jeder Antwort erfolgte ein evaluatives Feedback. In Abhängigkeit vom konkreten Antwortverhalten wurden direkt adaptive, gestufte Lösungshilfen (*Scaffolding*) angeboten, im Bedarfsfall wurde eine kommentierte Musterlösung präsentiert. Ferner sollte ein Punkte-Belohnungssystem einen motivationalen Anreiz setzen. Alle umgesetzten Designideen und instruktionalen Prinzipien können in den Originallernmaterialien unter <https://www.tools-bpaed.gsw.tu-dresden.de/nachvollzogen> werden.

Forschungsmethode der Studie

Die beiden Interventionsgruppen „basal“ und „komplex“ wurden in einem Prä-Posttest-Design mit einer Kontrollgruppe verglichen (Güzel et al., 2024). Den Teilnehmer:innen der Intervention „basal“ ($N = 6$ cluster-randomisierte Klassen bzw. 85 Auszubildende) und der Hälfte der Kontrollgruppenteilnehmer:innen ($N = 3$ cluster-randomisierte Klassen bzw. 42 Auszubildende) wurde im Prä- und im Posttest ein selbst entwickelter Test zur Erfassung der basalen diagnoserelevanten Rezeptionskompetenz vorgelegt (DRK basal, 17 Items; Guttman's Lambda-6 (λ -6) im Prätest = .77; λ -6 im Posttest = .87). Die Diagnosekompetenz wurde mit zwei Fehlerszenarien in der Simulation erfasst (Split-Half-Reliabilität im Prätest = .41; Split-Half-Reliabilität im Posttest = .25).⁴

Den Versuchspersonen in der Intervention „komplex“ ($N = 7$ cluster-randomisierte Klassen bzw. 91 Auszubildende) und der anderen Hälfte der Kontrollgruppenmitglieder ($N = 3$ cluster-randomisierte Klassen bzw. 42 Auszubildende) wurde im Prä- und Posttest ein ebenfalls selbst entwickelter Test zur Erfassung der komplexen diagnoserelevanten Rezeptionskompetenz vorgelegt (DRK komplex, 29 Items; λ -6 im Prätest = .90; λ -6 im Posttest = .94), ergänzt um ebenfalls zwei Fehlerszenarien in der Simulation zur Erfassung der Diagnosekompetenz (Split-Half-Reliabilität im Prätest = .08 und im Posttest = .37). Die Auszubildenden bekamen im Posttest jeweils dieselben Items wie im Prätest, allerdings in veränderter Reihenfolge. In der Treatmentphase zu Beginn des 3. Ausbildungsjahres wurden den experimentellen Gruppen jeweils, verteilt auf drei Tage, möglichst ähnliche Lern- und Übungszeiten zugemessen (die Intervention „komplex“ umfasste 360 Minuten, die Intervention „basal“ 270 Minuten und das Treatment der Kontrollgruppe 240 Minuten).

4 Die geringen Reliabilitäten sind unserer Ansicht nach vorwiegend Ausdruck (1) der hohen Anforderungsvielfalt beruflichen Problemlösens (selbst in einem auf das Motormanagement eingeschränkten inhaltlichen Validitätsfenster), aus dem die Fehlerszenarien abgebildet wurden, und (2) der mit nur zwei Fehlerfällen geringen Anzahl an Items für die Skala der Kfz-Diagnosekompetenz.

Die Kontrollgruppe beschäftigte sich während der Treatmentzeit mit den gleichen Fehlerszenarien und Übungsaufgaben der beiden Interventionen und bekam ebenfalls direktes evaluatives Feedback und die Musterlösungen. Das Vorzeigen und Erklären der Musterlösungen wurde durch die Versuchsleiter:innen aus der Forschungsgruppe vorgenommen. Der zentrale Unterschied zu den Treatmentgruppen bestand darin, dass die Kontrollgruppe nicht explizit durch die Erklärvideos über die richtigen rezeptiven Verhaltensweisen instruiert wurde. Die Kontrollgruppe erhielt zudem keine adaptiven Hilfestellungen bei der Bearbeitung der Übungsaufgaben und nahm nicht am Punkte-Belohnungssystem teil.⁵

Die forschungsleitenden Hypothesen waren, ...

- dass alle Versuchspersonen sowohl in den Interventionsbedingungen „basal“ und „komplex“ als auch in den Kontrollgruppen hinsichtlich ihrer diagnoserelevanten Rezeptionskompetenz (H1a) und Diagnosekompetenz (H1b) von den jeweiligen Treatments profitieren und
- dass die Zuwächse in diagnoserelevanter Rezeptionskompetenz (H2a) und Diagnosekompetenz (H2b) für die Interventionsbedingungen „basal“ und „komplex“ im Vergleich zur Kontrollgruppe jeweils signifikant höher ausfallen.

Ergebnisse

Bei der Datenauswertung wurden nur Personen berücksichtigt, die im Treatmentzeitraum zu allen drei Interventionsterminen anwesend waren. Sofern von diesen Personen einzelne Prä- oder Posttestdaten fehlten, wurden diese zunächst mittels multipler Imputation nach dem MICE-Verfahren (Jolani et al., 2015) geschätzt. Anschließend wurden die erreichten Effekte mittels Varianzanalysen mit Messwiederholung geprüft. Im Falle verletzter Voraussetzungen für die Varianzanalyse wurde mit nichtparametrischen Tests (Wald-Statistik nach Noguchi et al., 2012) gerechnet.

Intervention „basal“ im Vergleich zur Kontrollgruppe

Während des Interventionszeitraums verändern sich Treatment- und Kontrollgruppe erwartungskonform sowohl in ihrer basalen diagnoserelevanten Rezeptionskompetenz (Hypothese H1a) als auch in der Diagnosekompetenz (Hypothese H1b) signifikant positiv und praktisch bedeutsam. Der Haupteffekt des Messzeitpunkts auf die basale diagnoserelevante Rezeptionskompetenz war statistisch signifikant ($F[1, 125] = 75.73$; $p < .001$; $\eta^2_p = .08$); ebenso der Haupteffekt des Messzeitpunkts auf die Diagnosekompetenz ($W = 10.08$; $p = .002$). Die Effekte der Intervention auf die diagnoserelevante Rezeptionskompetenz waren mit $d = 0.76$ in der Interventions- und $d = 0.84$ in der Kontrollbedingung nominell stärker als die Effekte auf die Diagnosekompetenz, also den

5 Um sicherzustellen, dass gefundene Effekte auf Spezifika der Lernumgebung und nicht (bzw. nicht allein) auf die allgemeine Auseinandersetzung mit dem Thema und/oder auf das Arbeiten in der Simulation zurückzuführen sind, war es erforderlich, dass sich die Kontrollgruppe in der Treatmentzeit ebenfalls mit dem Thema auseinandersetzt und mit der Simulation arbeitet.

Diagnoseerfolg bei der Bearbeitung simulierter Störungsfälle ($d = 0.30$ in der Interventions- und $d = 0.48$ in der Kontrollgruppe).

Die positiven Entwicklungen sind allerdings zwischen Treatment- und Kontrollgruppe nicht unterschiedlich ausgeprägt (Hypothesen H2a und H2b): Weder für die diagnoserelevante Rezeptionskompetenz ($F[1, 125] = 0.71$; $p = .400$; $\eta^2_p = .00$) noch für die Diagnosekompetenz ($W = 0.82$; $p = .364$) wurden signifikante Interaktionseffekte aus Kompetenzzuwachs und Gruppenzugehörigkeit gefunden. Die computerbasierte Intervention „basal“ ist somit zwar lernwirksam, jedoch erwartungswidrig nicht wirksamer als das Treatment der Kontrollgruppe.

Intervention „komplex“ im Vergleich zur Kontrollgruppe

Ein signifikanter und praktisch bedeutsamer Zuwachs der komplexen diagnoserelevanten Rezeptionskompetenz (Hypothese H1a) wurde nur in der Treatmentgruppe gefunden; die Kontrollgruppe stagniert erwartungswidrig bei nominell sogar leicht abnehmender Tendenz. Der Haupteffekt des Messzeitpunkts ist statistisch signifikant ($F[1, 131] = 34.86$; $p < .001$; $\eta^2_p = .04$); dabei weist der Zuwachs der diagnoserelevanten Rezeptionskompetenz in der Interventionsgruppe eine Effektstärke von $d = 0.70$ auf, während in der Kontrollgruppe praktisch kein Effekt vorhanden ist ($d = 0.08$). Der in Hypothese H2a prognostizierte Interaktionseffekt aus Gruppenzugehörigkeit und Messzeitpunkt wurde erwartungskonform zugunsten der Interventionsgruppe gefunden ($F[1, 131] = 19.79$; $p < .001$, $\eta^2_p = .02$).

Ebenfalls erwartungskonform nimmt die Diagnosekompetenz in beiden Vergleichsgruppen (Hypothese H1b) während der Treatmentzeit signifikant und praktisch bedeutsam mit ähnlich starken Effekten zu (Haupteffekt Messzeitpunkt: $W = 31.33$; $p < .001$; $d_{\text{Intervention}} = 0.60$; $d_{\text{Kontrolle}} = 0.71$). Hinsichtlich der Diagnosekompetenz zeigen sich keine Interaktionseffekte (Hypothese H2b; $W = 0.43$; $p = .510$).

Erreichte Effekte – eine Bilanz

Mit den Befunden kann bilanziert werden, dass sich sowohl diagnoserelevante Rezeptionskompetenzen als auch die Diagnosekompetenz von Kfz-Auszubildenden signifikant und praktisch bedeutsam fördern lassen, wenn sie an relativ kurzen Interventionen (zu denen auch das Treatment der Kontrollgruppe gezählt werden muss) teilnehmen, die gezielt berufs- und fachsprachliche diagnoserelevante Leseanforderungen adressieren. Ein Interaktionseffekt zugunsten der nur in den Interventionen berücksichtigten expliziten Konfrontation der Auszubildenden mit rezeptionsbezogenen Verhaltensweisen, die in einer Vorstudie als besonders häufig fehlerbehaftet identifiziert wurden, scheint allerdings nur in einer der beiden Versuchsgruppen (Intervention „komplex“) und nur in einem der beiden untersuchten Merkmale (DRK komplex) zu existieren.

3.2 Förderung der modellbasierten Diagnosestrategie

Fördergegenstand und Gestaltung der Fördermaßnahmen

In zwei weiteren Interventionsstudien wurden Fördermaßnahmen zur Förderung einer *modellbasierten* Kfz-Diagnosestrategie untersucht. Wie bereits in Abschnitt 1 er-

wähnt, werden bei der modellbasierten Kfz-Diagnosestrategie potentiell störungsrelevante Teile des Kfz-Systems mental modelliert, um eine vorhandene Funktionsstörung nachzuvollziehen und Hypothesen zu möglichen Fehlerquellen zu formulieren. Im Unterschied zur computerbasierten Diagnosestrategie kann sie bei sehr unterschiedlichen und anspruchsvollen Störungen erfolgreich sein. Allerdings können lediglich ca. 14 Prozent der Auszubildenden am Ende ihrer Ausbildung eine modellbasierte Strategie erfolgreich anwenden (vgl. Abschnitt 1). Deshalb wurde im Projekt DigiDIn-Kfz auf theoretischer Grundlage und mit Hilfe von Fachexpert:innen eine modellbasierte Kfz-Diagnosestrategie so beschrieben, dass sie in Lernmaterialien und eine Fördermaßnahme übersetzt werden konnte. Die Strategie sowie die Fördermaßnahme wurden entsprechend den Ergebnissen einer ersten Interventionsstudie (Meier et al., 2022) überarbeitet. Die zunächst entwickelte modellbasierte Kfz-Diagnosestrategie sah vor, dass die Diagnostizierenden zyklisch vier Schritte durchlaufen, bis die Störungursache gefunden wurde: 1 – Hypothese aufstellen, 2 – Messungen zur Hypothesenprüfung planen, 3 – Messung durchführen, 4 – Messergebnisse und Hypothese bewerten (Abbildung 4).

Die überarbeitete Fördermaßnahme wurde durch eine Kürzung und Fokussierung u. a. mit dem Ziel optimiert, die Lernmotivation zu erhöhen. Die zu vermittelnde Diagnosestrategie wurde in enger Zusammenarbeit mit verschiedenen Fachexpert:innen verkürzt und noch stärker auf die Einbindung systemischer Zusammenhänge bei der Störungsdiagnose und die darauf aufbauende Hypothesenbildung fokussiert. Die überarbeitete Strategie (Abbildung 5) fasst die vorherigen vier in drei einmalig zu durchlaufenden Schritten zusammen.

Zur Unterstützung bei der Ausführung der Strategieschritte gaben wir den Auszubildenden Leitfragen an die Hand. Die Leitfragen zur Hypothesenformulierung lauteten:

1. Welche Funktion ist wahrscheinlich gestört?
2. Welche Komponenten sind für diese Funktion relevant?
3. Wie arbeiten die Komponenten zusammen, um die Funktion zu erfüllen?

Die Leitfragen zur Planung der Messungen lauteten:

1. Was genau musst du messen?
2. Wie musst du messen? (An welchen Messstellen musst du messen? Mit welchen Messmitteln musst du messen? Mit welchen Messbereichen musst du messen?)

Zur Strukturierung des Diagnoseprozesses sollten die Auszubildenden lernen, einen Prüfplan entsprechend der Strategie zu nutzen (Abbildung 5). Dieser ist den Auszubildenden so oder in ähnlicher Form aus dem Arbeits- bzw. Prüfungsalltag bekannt. Um die Qualität der Ausführung der Strategie zusätzlich zu erhöhen, erlernten die Auszubildenden drei Regeln beim Diagnostizieren (in der Abbildung in grünen Boxen dargestellt). Beispielsweise lernten sie, dass Vermutungen (d. h. Hypothesen) möglichst gut auf Basis des Verständnisses des betroffenen Kfz-Systems (mentales Modell der Auszubildenden) zu begründen sind.

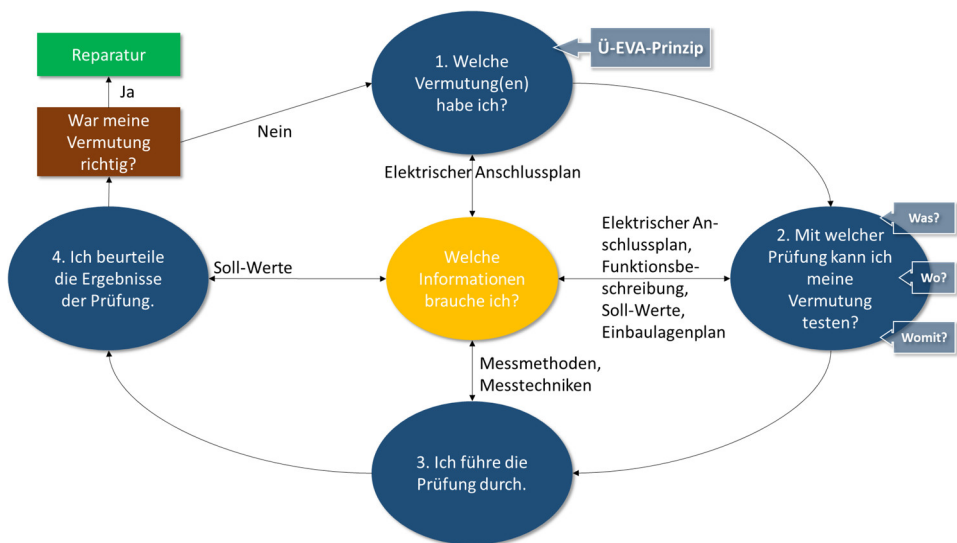


Abbildung 4: Modellbasierte Diagnosestrategie als zyklischer Prozess mit vier Schritten

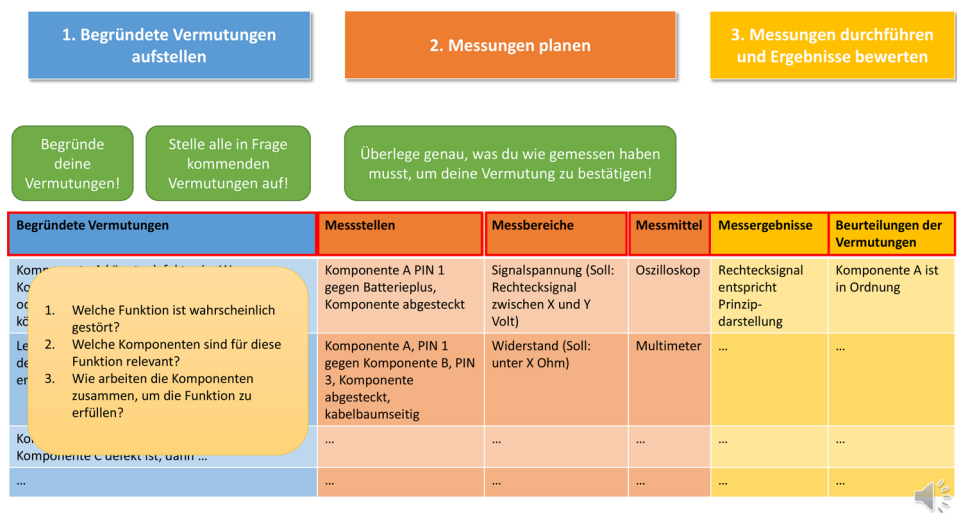


Abbildung 5: Modellbasierte Diagnosestrategie als Dreischritt, bei dem ein Prüfplan ausgefüllt werden soll. Leitfragen strukturieren die Vorgehensweise innerhalb der Schritte.

Sowohl für die Gestaltung der ursprünglichen als auch der überarbeiteten Fördermaßnahme wurde auf Methoden des beispielbasierten Lernens zurückgegriffen, wobei die Computersimulation zum Einsatz kam. Die Strategie wurde den Auszubildenden zunächst in Instruktionsvideos erklärt (1. Fördermaßnahme: Dauer ca. 35 Minuten, überarbeitete Fördermaßnahme: 10 Min). Danach sahen die Auszubildenden videobasierte Modellierungsbeispiele.

Die Fördermaßnahmen unterschieden sich darin, welche Aufgaben zur vertieften Verarbeitung der Modellierungsbeispiele gegeben wurden. In den Interventionsstudien wurden jeweils unterschiedliche Varianten der Aufgaben untersucht, um differenzielle Angebote für Auszubildende mit unterschiedlichen Eingangsbedingungen zu testen. In der ersten ursprünglichen Fördermaßnahme wurde je Diagnoseschritt ein Video gezeigt, in dem ein Experte jeweils zeigt, wie er den Strategieschritt anwendet, und sein diagnostisches Denken laut ausspricht. Der vierschrittige Zyklus wurde mehrmals durchlaufen (Dauer: 90 Min). Dabei bearbeiteten die Auszubildenden Aufgaben, bei denen sie sich den Sinn und Zweck der modellierten Schritte selbst erklären sollten. Solche Aufgaben, sogenannte Selbsterklärungsaufgaben, sind ein wichtiger Bestandteil des Lernens aus Beispielen. Sie helfen dabei, die zugrunde liegenden Prinzipien zu verstehen (Renkl, 2014), hier speziell die sinnvolle Umsetzung der Strategieschritte. Die Strategie ist zyklisch aufgebaut, sodass das Selbsterklären entweder auf vorherige oder auf nachfolgende Schritte im Zyklus bezogen werden konnte.

Retrospektive Selbsterklärungen beziehen sich auf vorherige, kürzlich gezeigte Schritte. Diese Art der Selbsterklärung wurde in Studien häufig untersucht (vgl. prinzipienorientierte Selbsterklärungen). Antizipatorische Selbsterklärungen richten sich hingegen auf nachfolgende Schritte und wurden bisher seltener untersucht. Renkl (1997) identifizierte sie jedoch als typisch für besonders erfolgreiche Lernende.

Antizipatorische Selbsterklärungen beinhalten auch eine retrospektive Komponente, da der nächste Schritt auf Grundlage vorheriger Schritte abgeleitet wird. Deshalb sind sie anspruchsvoller als retrospektive Selbsterklärungen und könnten besonders für leistungsstärkere Auszubildende geeignet sein.

In der überarbeiteten und gekürzten Fördermaßnahme (Meier et al., 2024) wurden die Diagnoseschritte immer direkt nach der instruktionalen Erklärung eines Schrittes modelliert und die Auszubildenden zum Selbsterklären aufgefordert. Der Fokus der Selbsterklärungen verschob sich von einer eher algorithmischen Erklärung der Abfolge von Schritten (und deren Funktion) in der ersten Fördermaßnahme auf das Erklären und somit Verstehen, *wie* die Schritte in ihrer Funktion qualitativ hochwertig ausgeführt werden können (zweite Fördermaßnahme). Die Selbsterklärungsaufforderungen bezogen sich in der überarbeiteten Version auf zwei beispielhaft ausgefüllte Prüfpläne: eine Expertenversion entsprechend dem im Video gezeigten Diagnoseschritt und eine Version eines „Hobbyschraubers“ (d. h. eine qualitativ weniger optimale Version, z. B. waren Hypothesen weniger gut begründet). Diese kontrastierenden Beispiele (Renkl, 2014) sollten entweder nacheinander (sequentiell) in ihrer Qualität erklärt werden oder im Vergleich (kontrastierend). Der direkte Vergleich wurde durch die Aufforderung angeregt, zu erklären, inwieweit der Experte und der „Hobbyschrauber“ die Regel unterschiedlich gut eingehalten hatten. Die Gesamtzeit der gekürzten Fördermaßnahme lag bei ca. 70 Minuten.

Forschungsmethode der Studien

In der ersten Studie zur ursprünglichen Fördermaßnahme (Meier et al., 2022; Meier et al., 2023) wurden in einem Messwiederholungsdesign Effekte der Modellierungsbeispiele und der retrospektiven versus antizipatorischen Selbsterklärungsaufforde-

rungen (SE-Prompts) auf das Wissen zur modellbasierten Diagnosestrategie und den Diagnoseerfolg getestet. Dabei wurde zum einen das *konzeptuelle Wissen über die Diagnosestrategie* mit einer offenen Abfrage erfasst. Zum anderen wurde das Anwenden der Strategie erfasst, indem die Auszubildenden in Bezug auf mehrere diagnostische Situationen einen strategisch sinnvollen nächsten Diagnoseschritt wählen und beschreiben mussten (*gestützte Strategieanwendung*). Der Diagnoseerfolg wurde durch eine freie Diagnose in der Simulation erfasst (*freie Strategieanwendung*). Eine Stichprobe von 67 Auszubildenden wurde experimentell in drei Bedingungen eingeteilt:

1. Experimentalgruppe 1: Modellierungsbeispiele mit retrospektiven SE-Prompts ($n = 21$)
2. Experimentalgruppe 2: Modellierungsbeispiele mit antizipatorischen SE-Prompts ($n = 25$)
3. Kontrollgruppe: offenes Diagnostizieren mit den in den Modellierungsbeispielen verwendeten Störungsszenarien ($n = 21$)

Die Personen aller drei Bedingungen bearbeiteten zunächst die Instruktionsvideos zur Strategie. Im Anschluss wurden den zwei Experimentalgruppen die videobasierten Modellierungsbeispiele gezeigt. Je nach Bedingung wurden sie nach jedem Schritt, auf Grundlage ihres Wissens aus den Instruktionsvideos, zum Erklären des vorangegangenen Schrittes (retrospektiv) oder des nachfolgenden Schrittes (antizipatorisch) aufgefordert. Die Kontrollgruppe erhielt keine Modellierungsbeispiele, sondern bearbeitete die in den Modellierungsbeispielen vorgestellten Störungsszenarien in der Simulation ohne Unterstützung selbstständig.

In der zweiten Interventionsstudie mit der überarbeiteten und gekürzten Fördermaßnahme (Meier et al., 2024) wurden die Effekte von kontrastierenden Modellierungsbeispielen mit Selbsterklärungsprompts auf die gleichen Outcome-Maße wie in der ersten Studie in einem Prä-Post-Design untersucht. Dabei wurden 119 Auszubildende in drei Bedingungen untersucht

1. Experimentalgruppe 1: Modellierungsbeispiele *und* sequenzielle SE-Prompts ($n = 39$)
2. Experimentalgruppe 2: Modellierungsbeispiele *und* kontrastierende SE-Prompts ($n = 43$)
3. Kontrollgruppe: freies (ohne Unterstützung) Diagnostizieren mit den in den Modellierungsbeispielen verwendeten Störungsszenarien ($n = 37$)

Wieder bearbeiteten die Personen *aller* Bedingungen die nun gekürzten Instruktionsvideos zur Strategie. In der Experimentalgruppe 1 wurden die Auszubildenden nach dem Anschauen des Modellierungsbeispiels zum jeweiligen Diagnoseschritt aufgefordert, die beiden Beispiele (Experte vs. „Hobbyschrauber“) nacheinander in ihrer Qualität zu erklären (sequentiell), während sie in der Experimentalgruppe 2 zum vergleichenden Erklären der beiden Beispiele aufgefordert wurden (kontrastierend). Die Kontrollgruppe sollte die Strategie unmittelbar nach den Instruktionsvideos selbstständig (ohne Unterstützung) anwenden, d. h. zum Diagnostizieren die in den Modellierungsbeispielen vorgestellten Störungsszenarien nutzen.

Ergebnisse

In der ersten Studie (Meier et al., 2022; Meier et al., 2023) zeigte sich zunächst, dass die Kontrollgruppe einen signifikant geringeren Prätest-Wert in der freien Diagnose (im Diagnoseerfolg) zeigte als die Experimentalgruppen ($F[1, 65] = 7.727$; $p < .01$), sodass der Vergleich dieser Gruppen in dieser einen Variable im Folgenden nicht betrachtet werden kann. In Bezug auf die Variablen ohne Prätestunterschiede wurden mithilfe einer Varianzanalyse mit Messwiederholung (RM-ANOVA) Haupteffekte des Zeitpunktes gefunden. Über alle drei Gruppen hinweg fanden sich ein großer Effekt für das konzeptuelle Wissen zur modellbasierten Diagnosestrategie ($F[1,65] = 17.938$, $p < .001$, $\eta^2_p = .216$) und ein großer Effekt für die gestützte Strategieranwendung ($F[1,65] = 96.430$, $p < .001$, $\eta^2_p = .597$). Für die freie Strategieranwendung in der Simulation (resp. den Diagnoseerfolg) fand sich ein signifikanter Effekt ($F[1,65] = 7.256$, $p = .009$, $\eta^2_p = .100$). Dieser ist jedoch vollständig auf die außergewöhnlich schwach startende Kontrollgruppe zurückzuführen, die im Posttest deutlich bessere Werte zeigte. Auszubildende erwarben durch die Fördermaßnahme folglich zwar Wissen über die vermittelte Strategie und können sie mit Unterstützung anwenden, zur freien Strategieranwendung (bzw. zum Diagnoseerfolg) gab es jedoch keine eindeutigen Ergebnisse.

Um innerhalb der Experimentalgruppen den Einfluss der Selbsterklärungsprompts näher zu untersuchen, wurde die Eingangskompetenz als Moderator betrachtet. Dabei war eine dreifache Interaktion zwischen dem Zeitpunkt, den beiden Bedingungen und der Eingangskompetenz bezogen auf das konzeptuelle Strategiewissen zu beobachten ($F[1,4] = 4.368$, $p = .043$, $\eta^2_p = .098$). Eine speziell darauf ausgelegte Analyse, bedeutsame Stellen der Interaktion zu identifizieren (Hayes & Matthes, 2009; Montoya, 2019), zeigte, dass Auszubildende mit hoher Eingangskompetenz wie angenommen stärker von antizipatorischen Prompts profitierten.

In der zweiten Studie (Meier et al., 2024) zeigten sich ebenfalls die beiden Haupteffekte des Zeitpunkts auf das Diagnosewissen. So ergaben sich ein großer Effekt für das konzeptuelle Strategiewissen ($F[1,97] = 22.350$, $p < .001$, $\eta^2_p = .187$) und ein großer Effekt für die gestützte Strategieranwendung ($F[1,97] = 28.460$, $p < .001$, $\eta^2_p = .227$). Diesmal ergab sich ebenfalls ein großer Haupteffekt des Zeitpunkts auf den Diagnoseerfolg (freie Diagnose, $F[1,97] = 41.389$, $p < .001$, $\eta^2_p = .299$). Dies liefert Hinweise darauf, dass alle Interventionsversionen das Diagnosewissen und die Diagnosekompetenz der Auszubildenden erhöhten. Zusätzlich konnten signifikante Interaktionseffekte von Zeitpunkt und Modellierungsbeispielen sowohl für das konzeptuelle Strategiewissen ($F[1,97] = 4.030$, $p = .047$, $\eta^2_p = .040$) als auch für die gestützte Strategieranwendung ($F[1,97] = 4.880$, $p = .030$, $\eta^2_p = .048$) gefunden werden, jedoch nicht für die freie Strategieranwendung (Diagnoseerfolg, $F[1,97] = 3.388$, $p = .069$, $\eta^2_p = .034$). Das heißt, konform mit der Theorie des beispielbasierten Lernens (Renkl, 2014) konnten die Auszubildenden durch das Lernen mit Modellierungsbeispielen zu Beginn des Kompetenzerwerbs im (konzeptuellen) Wissen zur Diagnosestrategie und der ersten, unterstützten Anwendung besser gefördert werden als durch das Kontrollgruppenlernen.

Auch in dieser Intervention wurde die Eingangskompetenz erhoben, um die Unterschiede zwischen den Promptarten (sequentiell und kontrastierend) näher zu betrachten. Dabei zeigte sich eine dreifache Interaktion zwischen dem Zeitpunkt, den beiden Bedingungen und der Eingangskompetenz bezogen auf das Strategiewissen ($F[1,61] = 4.171$, $p = .045$, $\eta^2_p = .064$). Eine nähere Analyse zeigte, dass Auszubildende mit höherer Eingangskompetenz (ca. obere Hälfte der Auszubildenden) von kontrastierenden Prompts profitierten (Meier et al., 2024).

Erreichte Effekte – eine Bilanz

Die Untersuchungen zeigten, dass die Fördermaßnahmen bezogen auf das Wissen zur modellbasierten Diagnosestrategie und die gestützte Diagnose wirksam sind bzw. die zweite, überarbeitete Maßnahme auch Effekte auf den Diagnoseerfolg erzielte und somit zukünftig von Ausbilder:innen und Berufsschullehrer:innen in der Ausbildung eingesetzt werden kann. Will man besonders leistungsstarke Auszubildende fördern, ist der Einsatz antizipatorischer und kontrastierender Selbsterklärungsprompts zu empfehlen.

Mit einer kurzen und einmaligen Intervention haben wir eine beachtliche Verbesserung des Wissens zur modellbasierten Diagnosestrategie der Auszubildenden erzielen können. Trotz nachgewiesener Wirksamkeit unserer Fördermaßnahmen ist das Gesamtniveau des Diagnoseerfolgs auch bei der zweiten Messung immer noch als eher gering einzustufen.

4 Diskussion

Unsere Studien zeigen, dass mit den entwickelten Interventionen – zu denen teilweise auch die konkurrierenden, ebenfalls theoriegeleitet und mit entsprechend positiven Effekterwartungen verknüpften Lehr-Lern-Settings der Kontrollgruppen gezählt werden können – oft mittlere bis große Effekte auf Facetten der Kfz-Diagnosekompetenz erzielt werden. Die von uns fokussierten Facetten waren die basale und komplexe diagnoserelevante Rezeptionskompetenz und Anteile des Diagnosewissens und dessen Anwendung, also das konzeptuelle Wissen über die modellbasierte Diagnosestrategie und die Anwendung der modellbasierten Diagnosestrategie.

Die Feedbacks von kooperierenden Berufsschulen lassen den Schluss zu, dass die erzielbaren positiven Effekte der Interventionen wahrscheinlich deshalb für einen niederschweligen Transfer in die Ausbildungspraxis vielversprechend sind, weil diese bereits bei relativ geringen Interventionszeiten (zwischen 90 Minuten und 360 Minuten) und innerhalb existierender, leistungsheterogener Berufsschulklassen erzielt wurden. Ergänzend scheint dem Transfer förderlich zu sein, dass ein unterrichtlicher Einsatz der Interventionen prinzipiell – dem Implementationsdesign und den Befunden der Studien entsprechend – im „didaktischen Gießkannenprinzip“ ohne spezifische Maß-

nahmen der Binnendifferenzierung erfolgen kann.⁶ Ähnliches dürfte auch für die betrieblichen und überbetrieblichen Ausbildungsstätten gelten. Weiterhin bestehen aufgrund des vollständig immersiven Selbstinstruktionscharakters der digitalen Interventionen nur geringe Implementationsvoraussetzungen auf Praxisseite. Der „Research-to-Practice-Gap“ (Souvignier & Philipp, 2016, S. 10 ff.) könnte bei hoher Umsetzungstreue und möglicherweise geringen Wirksamkeitseinbußen gering ausfallen. Zudem werden die erfolgversprechendsten Interventionen niederschwellig zum Download angeboten und im Rahmen einer bundesweit verfügbaren Lehrkräftefortbildung ausgebracht (Netzwerk Q 4.0)⁷.

Etwas weniger positiv erscheinen die nur teilweise bedeutsamen Effekte der entwickelten Interventionen auf den Erfolg bei eigenständigen Kfz-Diagnosen in einer Simulation; dieser kann als Globalindikator der Diagnosekompetenz betrachtet werden. Allerdings ist ein solches Befundmuster – ausgeprägte Effekte auf interventionsnahe und ggf. schwächere oder auch teilweise keine Effekte auf interventionsfernere Kriterien – auch nicht untypisch für Interventionsstudien (Rosenshine & Meister, 1994; Souvignier, 2021). Allenthalben wird dann die Hoffnung formuliert, längere Interventionszeiten und um weitere Elemente wie Übungsaufgaben, Aufgaben mit sukzessiven Transferelementen o. Ä. angereicherte Interventionen könnten zu einem Weiterentwickeln, Vertiefen und Automatisieren des erworbenen interventionsnahen Wissens bzw. solcher Fähigkeiten und – parallel dazu oder gar damit kausal verknüpft – zu einer positiven Entwicklung komplexer interventionsferner Kompetenzen führen. Dies erscheint mit Blick auf Abbildung 1 auch theoretisch plausibel: Ggf. wäre eine Intervention nochmals ertragreicher, die gleich mehrere Facetten der Kfz-Diagnosekompetenz und nicht – wie dies in unseren Interventionen umgesetzt wurde – isoliert einzelne Kompetenzfacetten zum Fördergegenstand erhebt. Die Förderbedarfe scheinen in vielen Facetten enorm zu sein.

Ob oder inwieweit sich eine solche Erwartung einlösen könnte, müssen Folgestudien klären, bei denen idealerweise auch Follow-up-Erhebungen einbezogen werden sollten, um die Langfristigkeit der Effekte evidenzbasiert einschätzen zu können. Denn: Über die Kraft kurzfristiger Interventionen hinausgedacht, sollten Interventionen letztlich imstande sein, die Lernenden zu motivieren und zu befähigen, ihre Strategien und Konzepte mit bedarfsgerechtem Scaffolding und Fading allmählich eigenständig weiterzuentwickeln.

Gegebenenfalls wären auch ergänzend Verknüpfungen der multimedialen digitalen Interventionen mit in den Ausbildungsstätten bereits bestehenden und bewährten didaktischen Umsetzungsvarianten möglich und einer Wirksamkeitsanalyse wert. Durch ein solches gezieltes Aufbrechen des Medien- und teilweise auch des Methodenmonismus der digitalen Interventionen könnten auch die motivational herausfordern-

6 Je Treatmentgruppe bzw. -klasse wurde je ein homogenes Treatment für alle Gruppen-/Klassenteilnehmenden umgesetzt („Gießkannenprinzip“); die erzielten Effekte sind Overalleffekte auf Gruppen-/Klassenebene; teilweise existieren auch differentielle Effekte; allerdings scheinen nur wenige Non-Responder:innen in den Gruppen/Klassen zu existieren; d. h. mehr oder weniger scheinen die meisten Teilnehmenden von den Interventionen zu profitieren.

7 <https://netzwerkq40.de/>

den Situationen, von denen uns die Auszubildenden in den Studien teilweise berichten, abgemildert werden.

In Ergänzung zum bisher Diskutierten lassen sich auch weitere potenzielle Gründe für die nur teilweise bedeutsamen Effekte der entwickelten Interventionen auf die Kfz-Diagnosekompetenz ausmachen: So basierten einige unserer Analysen auf linearen Grundannahmen. Da Kompetenzentwicklungen nichtlineare Dynamiken enthalten (Minnameier, 2003), werden auf Basis von Analysen mit linearen Grundannahmen ggf. reale Effekte „weglinearisiert“ oder unterschätzt. Daneben mag die teilweise sehr geringe Reliabilität der Skala zur Messung des Diagnoseerfolgs das Ergebnis zusätzlich (mit)bedingt haben. Folgestudien mit deutlich größeren Stichproben, optimierten Instrumenten und/oder einer A-priori-Diagnostik der Eingangskompetenzen und daran anschließender Zuweisung der Probanden zu experimentellen Gruppen könnten den hier untersuchten Fragestellungen vertieft nachgehen. Damit ließe sich ggf. auch – im Sinne der Theoriebildung – substanzieller klären, welchen Einfluss die propagierten Kompetenzfacetten tatsächlich auf den Diagnoseerfolg als Globalindikator der Diagnosekompetenz haben.

Mehrfach nicht eingelöst haben sich unsere Erwartungen zugunsten der nach (fach-)didaktischen Prinzipien konstruierten computerbasierten Lernumgebungen, die theoretisch dazu imstande sein sollten, Kompetenzfacetten noch wirksamer zu fördern als eine – wenngleich ebenfalls potentiell lernwirksam konstruierte – Kontrollbedingung. Den Gründen dafür versuchen wir aktuell im Rahmen weiterer Forschung nachzugehen. Ausgangspunkt ist die Überlegung, dass wir zwar mittlerweile eine Menge darüber wissen, *welche* diagnoserelevanten Verhaltensweisen *wie* bei der zu fördernden Zielgruppe fachlich inadäquat ausgeprägt sind und *welche* Verhaltensweisen *wie* idealtypisch ausgeprägt sein sollten (was sich bspw. im Formulieren und Instruieren einer modellbasierten Diagnosestrategie zeigt). Wenig wissen wir allerdings beispielsweise über die eigentlichen mentalen Modellierungsprozesse der Auszubildenden oder die ungünstigen motivationalen Zustände infolge unzutreffender Selbsteinschätzungen,⁸ die die fachlich inadäquaten Verhaltensweisen letztlich bedingen. Im Sinne der didaktischen Rekonstruktion (Kattmann et al., 1997) lautet unsere Forschungshypothese, dass wir erst dann, wenn wir die diagnoserelevanten mentalen Modellierungsprozesse und die motivationalen Zustände der Auszubildenden wirklich einordnen und verstehen,⁹ die Chance bekommen, auf den Lernbedarf passgenaue und damit noch lernwirksamere Lernumgebungen zu konstruieren, die dann auch durchgängig plausible Effekte und Interaktionseffekte in Folgestudien aufweisen.

8 Teilweise war zu beobachten, dass Auszubildende die Modellierungsbeispiele nicht vollständig studierten, obwohl sie die modellierte Strategie nicht beherrschten.

9 „(E)inordnen und verstehen“ bedingt an der Stelle, dass es tatsächlich überschaubar viele mentale Modellierungsprozesse (bspw. auf Basis spezifischer Konzepte) gibt, die sich beim Erheben einer Sättigung zuführen lassen und nicht, dass – im Extrem gesprochen – genauso viele unterschiedliche mentale Modellierungsprozesse existieren, wie es Kfz-Mechatroniker:innen gibt.

Literatur

- Abele, S. (2023). *Problemlösekompetenzen in beruflichen Kontexten. Resultate aus Lehr-Lern-Prozessen sichtbar machen*. wbv. <https://doi.org/10.3278/9783763973675>
- Abele, S., & von Davier, M. (2019). CDMs in vocational education: Assessment and usage of diagnostic problem-solving strategies in car mechatronics. In M. von Davier, & Y. S. Lee (Hrsg.), *Handbook of diagnostic classification models* (S. 461–488). Springer International Publishing. https://doi.org/10.1007/978-3-030-05584-4_22
- Abele, S., Walker, F., & Nickolaus, R. (2014). Zeitökonomische und reliable Diagnostik beruflicher Problemlösekompetenzen bei Auszubildenden zum Kfz-Mechatroniker. *Zeitschrift für Pädagogische Psychologie*, 28(4), 167–179. <https://doi.org/10.1024/1010-0652/a000138>
- Abele, S., Behrendt, S., Weber, W., & Nickolaus, R. (2016). Berufsfachliche Kompetenzen von Kfz-Mechatronikern – Messverfahren, Kompetenzdimensionen und erzielte Leistungen. In F. Oser, M. Landenberger, & K. Beck (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der BMBF-Förderinitiative ASCOT* (S. 171–203). wbv.
- Gschwendtner, T., Abele, S., & Nickolaus, R. (2009). Computersimulierte Arbeitsproben: Eine Validierungsstudie am Beispiel der Fehlerdiagnoseleistungen von Kfz-Mechatronikern. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 105(4), 557–578. <https://doi.org/10.25162/zbw-2009-0038>
- Güzel, E., Hartmann, S., Norwig, K., & Gschwendtner, T. (2024). Berufsspezifische Kompetenzen digital erwerben? *Zeitschrift für Pädagogik*, 70(2), 182–201. <https://doi.org/10.3262/ZP2402182>
- Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavior Research Methods*, 41(3), 924–936. <https://doi.org/10.3758/BRM.41.3.924>
- Jolani, S., Debray, T. P., Koffijberg, H., van Buuren, S., & Moons, K. G. (2015). Imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using MICE. *Statistics in Medicine*, 34(11), 1841–1863. <https://doi.org/10.1002/sim.6451>
- Kattmann, U., Duit, R., Gropengießer, H., & Komorek, M. (1997). Das Modell der Didaktischen Rekonstruktion – Ein Rahmen für naturwissenschaftsdidaktische Forschung und Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 3(3), 3–18. <https://doi.org/10.25656/01:31502>
- Leighton, J. P. (2017). *Using Think-Aloud Interviews and Cognitive Labs in Educational Research*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199372904.001.0001>
- Mayer, R. E. (2014). Multimedia instruction. In J. Spector, M. Merrill, J. Elen, & M. Bishop (Hrsg.), *Handbook of Research on Educational Communications and Technology* (S. 385–399). Springer. https://doi.org/10.1007/978-1-4614-3185-5_31

- Meier, J., Spliethoff, L., Hesse, P., Abele, S., Renkl, A., & Glogger-Frey, I. (2022). Promoting car mechatronics apprentices' diagnostic strategy with modeling examples: Development and evaluation of a simulation-based learning environment. *Studies in Educational Evaluation*, 72, 101117. <https://doi.org/10.1016/j.stueduc.2021.101117>
- Meier, J., Hesse, P., Abele, S., Renkl, A., & Glogger-Frey, I. (2023). Better self-explaining backwards or forwards? Prompting self-explanation in video-based modelling examples for learning a diagnostic strategy. *Instructional Science*, 52(2), 613–638. <https://doi.org/10.1007/s11251-023-09651-7>
- Meier, J., Hesse, P., Abele, S., Renkl, A., & Glogger-Frey, I. (2024). Video-based modeling examples and comparative self-explanation prompts for teaching a complex problem-solving strategy. *Journal of Computer Assisted Learning*, 40(4), 1852–1870. <https://doi.org/10.1111/jcal.12991>
- Minnameier, G. (2003). Wie verläuft die Kompetenzentwicklung – kontinuierlich oder diskontinuierlich? In Bundesinstitut für Berufsbildung (Hrsg.), *Berufsbildung im 21. Jahrhundert – Perspektiven für eine globale Gesellschaft: Ergebnisse und Ausblicke (Dokumentation des 4. BIBB-Fachkongresses)* (S. 1–10). BIBB Forum 3, Arbeitskreis 3.4 (separate Seitenzählung).
- Montoya, A. K. (2019). Moderation analysis in two-instance repeated measures designs: Probing methods and multiple moderator models. *Behavior Research Methods*, 51(1), 61–82. <https://doi.org/10.3758/s13428-018-1088-6>
- 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(2), 243–277. <https://doi.org/10.25162/zbw-2012-0016>
- Noguchi, K., Gel, Y. R., Brunner, E., & Konietzschke, F. (2012). nparLD: An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments. *Journal of Statistical Software*, 50(12), 1–23. <https://doi.org/10.18637/jss.v050.i12>
- Norwig, K., Güzel, E., Hartmann, S., & Gschwendtner, T. (2021). Tools to tap into the content of human minds. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 117(4), 658–693. <https://doi.org/10.25162/zbw-2021-0025>
- Rausch, A., Abele, S., Deutscher, V., Greiff, S., Kis, V., Messenger, S., Shackleton, J., Tramonte, L., Ward, M., & Winther, E. (2024). Designing an International Large-Scale Assessment of Professional Competencies and Employability Skills: Emerging Avenues and Challenges of OECD's PISA-VET. *Vocations and Learning*, 17, 393–432. <https://doi.org/10.1007/s12186-024-09347-0>
- Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, 21(1), 1–29. [https://doi.org/10.1016/S0364-0213\(99\)80017-2](https://doi.org/10.1016/S0364-0213(99)80017-2)
- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38(1), 1–37. <https://doi.org/10.1111/cogs.12086>
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, 64(4), 479–530. <https://doi.org/10.3102/00346543064004479>

- Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Hrsg.), *The Cambridge handbook of multimedia learning* (S. 49–69). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816819.005>
- Souvignier, E., & Philipp, M. (2016). Implementation – Begrifflichkeiten, Befunde und Herausforderungen. In M. Philipp & E. Souvignier (Hrsg.), *Implementation von Lesefördermaßnahmen* (S. 9–22). Waxmann.
- Souvignier, E. (2021). Interventionsforschung im Kontext Schule. In T. Hascher, T. S. Idel, & W. Helsper (Hrsg.), *Handbuch Schulforschung* (S. 219–235). Springer VS. https://doi.org/10.1007/978-3-658-24734-8_9-1.
- 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 (S. 37–54). Steiner.
- van Merriënboer, J. J. G. (2020). Das Vier-Komponenten-Instructional-Design (4C/ID) Modell. In H. Niegemann, & A. Weinberger (Hrsg.), *Handbuch Bildungstechnologie* (S. 153–170). Springer. https://doi.org/10.1007/978-3-662-54368-9_8
- Ziegler, B., Balkenhol, A., Keimes, C., & Rexing, V. (2012). Diagnostik „funktionaler Lesekompetenz“. *bwp@ Berufs- und Wirtschaftspädagogik – online*, 22, 1–19.

Ausbildungsbedingungen in der Pflegepraxis – Erste empirische Befunde zu unterschiedlichen Versorgungsbereichen

EVELINE WITTMANN, SVENJA HILL, PHILINE KREBS, SUSAN SEEBER, ALDIN STRIKOVIĆ, JULIA WARWAS, ULRIKE WEYLAND & LARISSA WILCZEK

Zusammenfassung

Ausbildungsbedingungen im Kontext praktischer Ausbildungsanteile sind in der empirischen Berufsbildungsforschung (und auch im Bereich der Pflegeausbildung) ein wichtiger Untersuchungsgegenstand. Vorliegende Befunde zum Zusammenhang zwischen Merkmalen praktischer Ausbildungsanteile und der Handlungskompetenz in der Altenpflege legen dabei nahe, Besonderheiten der Pflegeausbildung in den Blick zu nehmen (Wittmann et al., 2015). Hinzu kommen curriculare Besonderheiten der praktischen Pflegeausbildung: Neben den etablierten Elementen der schulischen Praxisbegleitung und der Praxisanleitung in Pflegeeinrichtungen sind seit der Einführung der generalistischen Pflegeausbildung praktische Einsätze in unterschiedlichen Einrichtungen und Versorgungsbereichen vorgesehen. Damit stellt sich die Frage nach den Lernwirkungen dieser verschiedenen Versorgungsbereiche auf die erreichten Kompetenzen der angehenden Pflegefachkräfte.

Im vorliegenden Beitrag wird vor diesem Hintergrund der Frage nachgegangen, welche Zusammenhänge zwischen wahrgenommenen Ausbildungsbedingungen in Praxiseinrichtungen der Pflegeausbildung einerseits und drei im Projekt EKGe¹ operationalisierten Facetten der berufsfachlichen Kompetenz von Pflegeauszubildenden (der unmittelbar auf die zu pflegende Person bezogenen Kompetenz, der Belastungs-Bewältigungskompetenz und der interprofessionellen Kooperationskompetenz) andererseits bestehen. Dabei interessiert insbesondere, inwiefern sich wahrgenommene Ausbildungsbedingungen zwischen den Versorgungsbereichen unterscheiden.

Zur Beantwortung dieser Frage wurde auf eine Stichprobe von $N = 292$ Pflegeauszubildenden aus der Ausbildung zur/zum Gesundheits- und Krankenpfleger:in, zum/zur Altenpfleger:in und aus der generalistischen Pflegeausbildung im zweiten Ausbildungsjahr zurückgegriffen. Basierend auf vorliegenden und im Projekt entwickelten Skalen wurde ein Fragebogen eingesetzt, der spezifisch auf die skizzierten Ausbildungsbedingungen in der Pflege zugeschnitten ist (Striković et al., 2022). Die Befunde legen nahe, dass zwischen den Versorgungseinrichtungen Unterschiede in der Aus-

¹ EKGe: Erweiterte Kompetenzmessung im Gesundheitsbereich, Förderkennzeichen: 21AP006A-D. Gefördert im Rahmen der BMBF-Forschungs- und Transferinitiative ASCOT+: Technology-based Assessment of Skills and Competencies in VET (URL: www.ascot-vet.net/)

prägung der bedeutsamen Ausbildungsmerkmale bestehen; insbesondere deuten sie darauf hin, dass die Bedingungen im Versorgungsbereich Altenpflegeheim weniger positiv ausgeprägt sind.

Schlagworte: Betriebliche Ausbildungsbedingungen, Praxiseinrichtung, Versorgungsbereiche, personenbezogene Dienstleistungsberufe

Abstract

VET conditions in the context of practical training components at the workplace are an important subject of empirical research in vocational education, including the field of nursing education. The available findings on the relationship between characteristics of practical VET components and competence in geriatric care suggest that the special features of nursing education should be considered (Wittmann et al., 2015). In addition, there are curricular peculiarities of practical nursing training: Practical assignments in different facilities and care areas have been scheduled since the introduction of generalist nursing training. This raises the question of the learning effects of these different care areas on the acquired skills of prospective nursing professionals.

Against this background, this paper examines the question of what relationships exist between perceived training conditions in nursing care training institutions and three facets of nursing trainees' professional competence (competence directly related to the person to be cared for, coping competence and interprofessional collaboration competence) operationalized in the EKGe project. Of particular interest here is the extent to which perceived training conditions differ between the care sectors.

A sample of $N = 292$ nursing students in their second year of training to become health and hospital nurses, geriatric nurses, or generalist nurses was used. Based on existing scales and those developed in the project, a questionnaire was used that was specifically tailored to the outlined training conditions in nursing VET (Striković et al., 2022). The findings suggest that there are differences in the training characteristics between the care facilities. Specifically, training features in nursing homes for the elderly seem to be less favorable than those in other institutional contexts.

Keywords: Company Training Conditions, Practical Facilities, Care Sectors, Personal Service Occupations

1 Einleitung

Ausbildungsbedingungen innerhalb der berufspraktischen Phasen in der Berufsausbildung umfassen vielfältige Struktur- und Prozessqualitäten (z. B. Böhn & Deutscher, 2021; Zimmermann, 1999). Wie diese mit der Kompetenzentwicklung von Auszubildenden verknüpft sind, ist ein wichtiger Untersuchungsgegenstand in der empirischen Berufsbildungsforschung (z. B. Baethge-Kinsky et al., 2016; Dietzen et al., 2011; Geißel, 2009; Hoff et al., 1991; Nickolaus et al., 2010) und speziell in der Forschung zur

Pflegeausbildung (Goller et al., 2019; Wittmann et al., 2015). Vorliegende Befunde aus dem Verbundprojekt TEMA² legen aufgrund der geringen nachweisbaren Zusammenhänge nahe, Ausbildungsbedingungen spezifischer zu erfassen und dazu unter anderem Besonderheiten der Pflege und der Pflegeausbildung gegenüber anderen Berufen und Berufsfeldern verstärkt in den Blick zu nehmen (Wittmann et al., 2015). Hierfür spricht, dass Kompetenzmodellierungen in der Pflege den interaktiven Charakter der Pflegearbeit als besonderes Merkmal hervorheben (z. B. Böhle & Glaser, 2006; Weidner, 1995; vgl. auch Seeber & Wittmann, 2017). Besonders in der Altenpflege wird in diesem Zusammenhang die Notwendigkeit stabiler und dauerhafter Pflegebeziehungen herausgestellt (Bing-Jonsson et al., 2014). Zudem weist die Pflegeausbildung curriculare Besonderheiten auf, indem sie das in der dualen Ausbildung übliche Zusammenwirken zweier Lernorte und ggf. der überbetrieblichen Ausbildung mit der verpflichtenden Festlegung auf unterschiedliche zu absolvierende Versorgungsbereiche „durchkreuzt“: So sind seit Inkrafttreten des Pflegeberufgesetzes und der damit einhergehenden Einführung der generalistischen Pflegeausbildung praktische Einsätze in unterschiedlichen Versorgungsbereichen vorgesehen, insbesondere „Pflichteinsätze in der allgemeinen Akutpflege, in stationären Einrichtungen, der allgemeinen Langzeitpflege in stationären Einrichtungen und der allgemeinen ambulanten Akut- und Langzeitpflege“ (§7 Abs. 1 Pflegeberufgesetz). Pflegeauszubildende müssen im Rahmen der generalistischen Pflegeausbildung also zwingend in unterschiedlichen Versorgungseinrichtungen der Pflegepraxis ausgebildet werden. Damit stellt sich aus Sicht der empirischen Berufsbildungsforschung die Frage nach den Lernwirkungen dieser unterschiedlichen Versorgungsbereiche, deren jeweils geschaffene Ausbildungsbedingungen grundlegend voneinander abweichen können. Darüber hinaus ist gemäß §6 Abs. 3 Pflegeberufgesetz ein „[wesentlicher] Bestandteil der praktischen Ausbildung [...] die von den Einrichtungen zu gewährleistende Praxisanleitung im Umfang von mindestens 10 Prozent der während eines Einsatzes zu leistenden praktischen Ausbildungszeit.“ Der Pflegeschule kommt hierbei explizit die Aufgabe zu, die praktische Ausbildung durch die von ihr in angemessenem Umfang zu gewährleistende Praxisbegleitung zu unterstützen.

Im vorliegenden Beitrag soll eruiert werden, wie sich Lerngelegenheiten in drei zentralen Versorgungsbereichen – der klinischen Versorgung im Krankenhaus, der Versorgung in der stationären Altenpflege und der ambulanten pflegerischen Versorgung – unterscheiden. Dazu wird vor dem skizzierten Hintergrund der Frage nachgegangen, welche Zusammenhänge zwischen wahrgenommenen Ausbildungsbedingungen in Praxiseinrichtungen der Pflegeausbildung einerseits und drei im Projekt EKGe operationalisierten Facetten der berufsfachlichen Kompetenz von Pflegeauszubildenden andererseits bestehen – der unmittelbar auf die zu pflegende Person bezogenen Kompetenz, der Belastungs-Bewältigungskompetenz und der interprofessionellen Kooperationskompetenz. Dabei interessiert insbesondere, welche Rolle pflege-

2 Entwicklung und Erprobung von technologieorientierten Messinstrumenten zur Feststellung der beruflichen Handlungskompetenz in der Pflege älterer Menschen.

spezifischen Ausbildungsmerkmalen zukommt und inwiefern sich wahrgenommene Ausbildungsbedingungen zwischen den Versorgungsbereichen unterscheiden.

Basierend auf vorliegenden und neu entwickelten Skalen wurde im Projekt EKGe ein retrospektiver Fragebogen entwickelt, der spezifisch auf die skizzierten Ausbildungsbedingungen in der Pflege zugeschnitten ist (Striković et al., 2022). In Abschnitt 2 werden die theoretischen Überlegungen und empirischen Vorarbeiten erläutert, die der Entwicklung des Instruments zugrunde liegen. Im anschließenden Abschnitt 3 wird analysiert, wie betriebliche Ausbildungsbedingungen in Zusammenhang mit den drei im Projekt EKGe operationalisierten Facetten der berufsfachlichen Kompetenz von Pflegeauszubildenden stehen. Schließlich wird geprüft, welche Unterschiede in den wahrgenommenen betrieblichen Ausbildungsbedingungen zwischen den Versorgungsbereichen bestehen. Zur Beantwortung dieser Fragen wird auf eine Stichprobe von $N = 292$ Pflegeauszubildenden aus der Ausbildung zur/zum Gesundheits- und Krankenpfleger:in, zum/zur Altenpfleger:in und aus der generalistischen Pflegeausbildung mit Schwerpunkt im zweiten Ausbildungsjahr zurückgegriffen.

2 Entwicklung des Fragebogens zur Erfassung von Ausbildungsbedingungen in der Pflege

Für die Erfassung von Ausbildungsbedingungen in den Praxiseinrichtungen der Pflege kann zunächst auf bewährte Skalen zur Erfassung betrieblicher Ausbildungsqualität zurückgegriffen werden. Einen zentralen Ausgangspunkt für die Erfassung der betrieblichen Ausbildungsqualität bildete das *Learning-Quality-Inventory for In-Company Training in Vocational Education and Training* (VET-LQI) von Böhn und Deutscher (2019; 2021), welches sich aus einer integrativen Inhaltsanalyse von 43 Fragebögen zu Ausbildungsbedingungen speist (Böhn & Deutscher, 2019, S. 55). Allerdings stammen die so verwerteten Erhebungsinstrumente aus dem kaufmännischen und gewerblich-technischen Ausbildungskontext, sodass die Items des VET-LQI nicht hinreichend berufsfeldspezifisch formuliert sind.³ Daher fand für die zur Testkonstruktion herangezogenen Items eine semantische Anpassung an den Pflegekontext statt, z. B. durch Ersetzen von „Betrieb“ durch „Praxiseinrichtung“ oder „Arbeitseinheit“. Es erfolgte zudem eine Augenscheinvalidierung aus pflegeberuflicher Sicht durch zwei Pflegepädagog:innen mit Pflegeausbildung. Darüber hinaus wurden nicht alle Items übernommen; aus testökonomischen Gründen wurden basierend auf Vorergebnissen zur Altenpflege (Wittmann et al., 2015) solche Skalen, die weniger nah an der konkreten Aufgabenausführung von Pflegeauszubildenden lagen – insbesondere Items zur Organisationskultur – nicht in die endgültige Fragebogenversion übernommen.

Um Skalen oder Items in den Fragebogen zu integrieren, die im VET-LQI entweder nicht aufgegriffen werden oder aufgrund ihrer vergleichsweise höheren Pflegespe-

3 Vor diesem Hintergrund stellen beispielsweise Böhn und Deutscher (2021, S. 47; vgl. auch Rausch 2013) fest: „occupation-specific adaptations and additions of items would be helpful in respect of item formulation, so that apprentices can more easily relate to the instrument“.

zifität zu bevorzugen sind, wurde eine systematische Recherche nach weiteren Befragungsinstrumenten zur praktischen Ausbildungsqualität aus dem Pflegekontext durchgeführt (Striković et al., 2022). Die Literatursuche erfolgte angesichts der international gut entwickelten Forschungslandschaft zur *nursing education* in sieben englischsprachigen Datenbanken.⁴ Nach Eignungsprüfung von insgesamt 462 Publikationen, die über die datenbankgestützte Recherche und eine Handsuche gewonnen werden konnten, wurden aus 17 in die Volltextanalyse eingegangenen Studien vier Instrumente identifiziert, die bereits in der Pflegeausbildung erprobt wurden:

- der *Clinical Learning Environment and Supervision plus nurse Teacher* (CLES+T) Fragebogen (CLES+T: Saarikoski et al., 2008), welcher darauf abzielt, die Qualität klinischer Lernumgebungen, der Lernunterstützung durch die Stationsleitung, der Theorie-Praxis-Verknüpfung durch die Lehrkraft und der Kooperation zwischen Lehrkraft und Arbeitseinheit zu erfassen;
- das *Clinical Learning Organizational Culture Survey* (CLOCS: Henderson et al., 2010), welches dazu dienen soll, auf kulturelle Aspekte der Versorgungseinrichtung bezogene Überzeugungen und Annahmen von Pflegenden zu erfassen, die für das Lernen am Arbeitsplatz relevant sind;
- der *Conditions for Work Effectiveness Questionnaire II*⁵ (CWEQ-II: Laschinger, 2012), welcher das Ausmaß erfassen soll, in dem Arbeitnehmer:innen das Gefühl haben, Zugang zu Ressourcen und Unterstützungsstrukturen in ihrem Arbeitsumfeld zu haben;
- der *Survey of Perceived Organizational Support* (SPOS: Eisenberger et al., 1986; vgl. auch Battistelli et al., 2016) zur Erfassung der wahrgenommenen organisationalen Unterstützung.

Die Suche nach geeigneten Fragebögen aus dem nationalen Kontext erfolgte angesichts der geringen empirischen Forschungsverankerung der Pflegepädagogik in Deutschland über eine freie Webrecherche und die Konsultation von Personen, welche durch Expertise in der Erfassung von Ausbildungsbedingungen ausgewiesen sind (Booth et al., 2016; Pettrick & Roberts, 2006). Dies führte zu den folgenden beiden Instrumenten, die zu Teilen oder gänzlich in die Fragebogenkonstruktion eingingen:

- Wenners (2018) Instrument zur Erfassung der Wechselwirkung von Lernortkooperation und Ausbildungsqualität;
- Items aus dem Diagnoseinstrument für gesundheitsförderliche Arbeit (DiGA; Ducki, 2000) und dem Beanspruchungsscreening bei Humandienstleistungen (BHD-FBH; Hacker & Reinhold, 1999), welche pflegetypische Belastungen adressieren.

Zudem wurde ein unveröffentlichtes Instrument zur Erfassung der Praxisanleitungsqualität verwendet, welches an der Technischen Universität München systematisch

4 (i) Medline, (ii) Cochrane Library, (iii) PsycArticles, (iv) ERIC, (v) Web of Science/SSCI, (vi) Education Source und (vii) Scopus.

5 Der CWEQ-II ist mit 21 Items wesentlich testökonomischer als der CWEQ-I ($n = 60$) und wird daher insbesondere für Forschungszwecke eingesetzt.

entwickelt und bereits auf Reliabilität überprüft worden war (Schröter, 2017). Nach dieser heuristisch geleiteten Selektion gingen aus den im nationalen Kontext eingesetzten Fragebögen 120 ggf. berufsfeldspezifisch angepasste Items in den EKGe-Itempool ein. Aus den englischsprachigen Instrumenten CLES+T und CLOCS wurden elf Items übersetzt, semantisch an den nationalen Pflegekontext angepasst und für die Testkonstruktion verwendet, die auch den Bereich der schulischen Praxisbegleitung abbilden sollen.

Dabei ist herauszustellen, dass systematische Überlegungen zu theoretischen Konstrukten für die Modellierung von Ausbildungsqualität im Berufsfeld Pflege nicht existieren, insbesondere solche, die dieses Berufsfeld bzw. die personenbezogenen Dienstleistungsberufe anderen Berufsfeldern gegenüberstellen. Demgegenüber wurde bereits bei Wittmann (2001) die Annahme formuliert, dass Ausbildungsbedingungen sowohl berufs(feld)spezifisch als auch spezifisch für einzelne Kompetenzfacetten differenziert wirken (vgl. auch die Befunde in Wittmann et al., 2015). Vor diesem Hintergrund wurden anhand einer pragmatischen Heuristik weitere angenommene Aspekte der Ausbildungsqualität recherchiert und empirisch umgesetzt, indem Vermutungen angestellt wurden,

- a) dass das eingangs adressierte Merkmal der professionellen interaktiven Arbeit mit und am Menschen Auswirkungen auf die Lernerfahrung aufweist sowie
- b) dass es spezifische förderliche Bedingungen für die Kompetenzfacette der interprofessionellen Kooperationskompetenz geben könnte, z. B. Erfahrungen mit interprofessionellen Kooperationsarrangements in der jeweiligen Praxiseinrichtung.

Zu diesen beiden Punkten wurden sechs Skalen mit insgesamt 27 Items neu konstruiert, da zu den betreffenden Konstrukten keine standardisierten und nach den zentralen Testgütekriterien geprüften Skalen vorlagen. Diese wurden daher auf Basis weiterer, inhaltlich geleiteter Literaturanalysen zum pflegeberuflichen Arbeitskontext ebenfalls an der Technischen Universität München entwickelt (Strikovic et al., 2022):

- Ad a) Zum einen sind die neu konstruierten Skalen „Lernunterstützendes Feedback durch die Arbeitseinheit“ und „Feedback von Patient:innen/Klient:innen/Bewohner:innen“ herauszustellen, die sich an Arbeiten von Giesbers et al. (2015) zum Umgang von Pflegekräften mit Feedback orientieren. Theoretische Grundlage der Item-Konstruktion war also, den Einfluss des interaktiven Charakters der Pflegetätigkeit auf die Kompetenzentwicklung in der Pflegeausbildung zu erfassen. Grund für die Eigenentwicklung dieser Skalen waren fehlende Items und Skalen in traditionellen Erhebungsinstrumenten zur betrieblichen Ausbildungsqualität zu Feedback, welches nicht von Vorgesetzten, Kolleg:innen oder Ausbilder:innen stammt.
- Ad b) Zum anderen basieren weitere neu konstruierte Skalen auf den Erkenntnissen eines weiteren, bislang unveröffentlichten Literaturreviews zur interprofessionellen Kooperation. So wurde vermutet, dass Merkmale wie z. B. die organisationale Hierarchie oder erfahrene interprofessionelle Kooperationsarrange-

ments Zusammenhänge mit der interprofessionellen Kooperationskompetenz aufweisen würden. Auch mit diesen Skalen wird dabei der Aspekt der Arbeit mit und am Menschen adressiert. Dies erfolgt allerdings indirekt, insofern als angenommen wird, dass soziale und institutionelle Barrieren im Austausch über Versorgungsbedarfe zu pflegender Personen sich unmittelbar auf die Kompetenzentwicklung der Auszubildenden niederschlagen könnten.

Mit Abschluss der dargestellten Konstruktionsschritte resultierte ein Pool aus 30 Skalen mit insgesamt 158 Items (Striković et al., 2022), welcher im Rahmen einer Pilotierungsstudie reliabilitätsorientiert optimiert wurde⁶ und in der Endfassung als retrospektiver Kontextfragebogen⁷ im Projekt EKGe zum Einsatz kam. Aufgrund der geringen Stichprobengröße der Pilotierungsstudie von $N = 135$ Probanden konnte eine Prüfung der faktoriellen Struktur erst in der Hauptstudie vorgenommen werden. Hier wurde auf die zuvor genannte Stichprobe (s. Kap. 1) zurückgegriffen, um konfirmatorische Faktorenanalysen entlang folgender Merkmalsbereiche durchzuführen:

- *Aufgabenbezogene Skalen:* i) Komplexität der Anforderung, ii) Entscheidungsspielraum, iii) Arbeitsinhalte, iv) Aufgabenbezogene Kommunikation
- *Organisationsbezogene Skalen:* i) Funktionale Einbindung, ii) Identifikation mit der Praxiseinrichtung, iii) Kommunikation und Konfliktbewältigung, iv) Hierarchie, v) Interprofessionelle Kooperation, vi) Organisationale Ressourcen
- *Feedback-Skalen:* i) Lernunterstützendes Feedback durch die Arbeitseinheit, ii) Positives Feedback durch zu pflegende Personen, iii) Negatives Feedback durch zu pflegende Personen
- *Theorie-Praxis-Skalen:* i) Theorie-Praxis-„Verknüpfung“ durch die Lehrkraft, ii) Kooperation zwischen Schule und Praxiseinrichtung, iii) Theorie-Praxis-Transfer
- *Praxisanleitungs-Skalen:* i) Motivation, ii) Pädagogische und fachliche Kompetenz, iii) Beziehung

Die Fit-Maße (RMSEA, χ^2/DF , CFI, TLI, SRMR) sowie die Faktorladungen der konfirmatorischen Faktorenmodelle liegen in einem akzeptablen bis guten Bereich. Inwieweit diese aus der Perspektive angehender Pflegefachkräfte erfassten Facetten der Ausbildungsqualität in Pflegeeinrichtungen mit den Kompetenzfacetten zusammenhängen, wird im Folgenden geprüft.

3 Empirische Analysen

3.1 Stichprobe, Kompetenztests und Auswertungsverfahren

Die Daten wurden zwischen April und Juli 2021 in 19 Klassen an Pflegeschulen in Nordrhein-Westfalen erhoben. Im Datensatz sind die Daten von 292 Auszubildenden

⁶ Ergänzend wurde inhaltlich überprüft, ob der reduzierte Itemkatalog den semantischen Raum der Ursprungsitems noch abdeckte. Ggf. wurden Items aus inhaltlichen Gründen in der Skala belassen.

⁷ sowie gleichzeitig mit einem Instrument zur prozessbegleitenden Erhebung der Ausbildungsbedingungen

der Pflegeberufe „Altenpflege“ ($n = 166$), „Gesundheits- und Krankenpflege“ ($n = 70$) und der „generalistischen Pflegeausbildung“ ($n = 56$) enthalten, die sich zum Zeitpunkt der Datenerhebung größtenteils im zweiten Ausbildungsjahr befanden. Für sie lagen jeweils Informationen zu Testleistungen für drei Kompetenzfacetten vor, welche jeweils mittels eines videobasierten, computergestützten Situational-Judgement-Tests operationalisiert wurden:

- (1) die unmittelbar auf die zu pflegende Person bezogene Kompetenzfacette, die sich auf die am Pflegeprozess orientierte berufsfachliche Kompetenz in der Pflegebeziehung mit älteren zu pflegenden Personen und deren Angehörigen bezieht (Wittmann et al., 2022),
- (2) die Belastungs-Bewältigungskompetenz für pflegetypische Belastungen, welche die Befähigung zur Bewältigung berufstypischer psychischer Belastungsfaktoren mit dem Ziel der Erhaltung der eigenen psychischen Gesundheit abbildet (Warwas et al., 2023) und
- (3) die interprofessionelle Kooperationskompetenz von Pflegeauszubildenden, die beinhaltet, Wissen über die eigene Rolle und die anderer gesundheitsberuflich Tätiger zu nutzen, um eine angemessene Einschätzung von Pflege- und Versorgungsbedarfen von zu pflegenden Personen zu erreichen (Wittmann et al., 2024).

Die Stichprobe ist in Tabelle 1 dargestellt. Auszubildende haben überwiegend Deutschland als Geburtsland (84,4 %) angegeben und sind überwiegend weiblichen Geschlechts (77,9 %); letzterer Anteil liegt in der Stichprobe nahe am bundesweiten Anteil im Ausbildungsjahr 2021 (75,6 %; Statistisches Bundesamt, 2022). Das Alter der Befragten ist heterogen verteilt; verglichen mit der bundesweiten Verteilung sind die höheren Altersgruppen vor allem in der Altenpflegesubgruppe überrepräsentiert. Eine große Altersheterogenität der Stichproben wurde auch in vorhergehenden Studien beschrieben; sie dürfte auf eine hohe Grundheterogenität in den Schulklassen der Pflegeausbildung hindeuten (Wittmann et al., 2022; 2024).

Ein Teil der Stichprobe befindet sich bereits in der generalistischen Pflegeausbildung. Diese Auszubildenden hatten ihre Ausbildung ca. ein halbes Jahr nach den Auszubildenden im zweiten Ausbildungsjahr der Altenpflege und der Gesundheits- und Krankenpflege begonnen. Tabelle 1 zeigt, dass diese Subgruppe ihren letzten Praxiseinsatz vorwiegend im Wohnbereich im Altenpflegeheim hatte. Insgesamt sind auch aufgrund des hohen Anteils an Auszubildenden im Beruf Altenpflege Praxiseinsätze in diesem Bereich dominant (38,5 %). Stationen im Krankenhaus/Klinikum (29,4 %) sowie in ambulanten Pflegediensten (25,2 %) sind demgegenüber weniger im relevanten Zeitraum absolviert worden. Insbesondere Auszubildende der Gesundheits- und Krankenpflege hatten ihren Praxiseinsatz weit überwiegend in einer Station im Krankenhaus/Klinikum.

Tabelle 1: Untersuchungsstichprobe nach Ausbildungsjahr, Altersgruppe, Geschlecht und Geburtsland (in Prozent)

Pfle- ge- beruf	Ausbildungs- jahr ^a		Altersgruppe			Geschlecht		Geburts- land	Arbeitseinheit im Praxiseinsatz				Ge- samt
	2. AJ	3. AJ	< 21	21–25	> 25	w	m		AP	SK	WA	Sonsti- ges	
APF	80,1	19,9	19,0	34,0	46,9	76,2	23,1	78,9	32,0	12,2	53,7	2,0	56,8
GUK	100,0	0,0	52,2	33,3	14,5	76,8	18,8	94,2	10,1	68,1	0,0	21,7	24,0
GPA	100,0	0,0	26,1	43,5	30,4	84,8	10,4	86,8	26,1	26,1	47,8	0,0	19,2
Ge- samt	88,7	11,3	29,0	35,5	35,5	77,9	19,8	84,4	25,2	29,4	38,5	6,9	100

* Die generalistische Pflegeausbildung begann etwa sechs bis acht Monate nach Beginn der Ausbildung in der Altenpflege und der Gesundheits- und Krankenpflege.

Anmerkungen. *APF* = Altenpflege; *GUK* = Gesundheits- und Krankenpflege; *GPA* = Generalistische Pflegeausbildung; *AJ* = Ausbildungsjahr; *w* = weiblich; *m* = männlich; *AP* = ambulanter Pflegedienst; *SK* = Station im Krankenhaus/Klinikum; *WA* = Wohnbereich im Altenpflegeheim. Für *n* = 30 Auszubildende fehlen Daten zum Geschlecht, zur Altersgruppe und zum Geburtsland.

Zur Untersuchung des Zusammenhangs zwischen Ausbildungsbedingungen in den Pflegeeinrichtungen bzw. -diensten und den erfassten Facetten der beruflichen Handlungskompetenz wurden Pearson-Korrelationen berechnet. Als Testwerte für die Facetten der beruflichen Handlungskompetenz wurden jeweils auf Basis der Item-Response-Theorie über Partial-Credit-Modelle berechnete WLE-Personenfähigkeitsparameter verwendet. Angaben zur betrieblichen Ausbildungsqualität wurden hierbei retrospektiv für die jeweils letzte Praxisphase gemacht und mit den drei Kompetenzfacetten in Beziehung gesetzt.

Grundsätzlich wurde vermutet, dass pflegespezifischere Skalen, d. h. die Feedback-Skalen und die Praxisanleitungs-Skalen, höhere Zusammenhänge mit der Kompetenzentwicklung in pflegespezifischen Kompetenzfacetten aufweisen als allgemeine Skalen zur Ausbildungsqualität. Insbesondere wurde für die Merkmale „Hierarchie“ und Erfahrungen mit „interprofessioneller Kooperation“ angenommen, dass diese aufgrund ihres Rollenbezugs besonders mit der interprofessionellen Kooperationskompetenz zusammenhängen würden. So wäre zu vermuten, dass Hierarchien, insbesondere zwischen Ärzt:innen und Pflegekräften, sich negativ auf interprofessionelles Handeln und eine entsprechende Kompetenzentwicklung auswirken können, weil sie als Barrieren der Zusammenarbeit fungieren und Erfahrungen der Zusammenarbeit im Alltag verhindern. Umgekehrt liegt die Annahme nahe, dass Erfahrungen im Bereich interprofessioneller Zusammenarbeit im Praxiseinsatz (Skala „Interprofessionelle Kooperation“) für die entsprechende Kompetenz förderlich sein dürften.

Analysen zu Unterschieden in den wahrgenommenen Ausbildungsbedingungen in den Praxiseinrichtungen zwischen den Versorgungsbereichen basieren auf einfaktoriellen Varianzanalysen anhand der kategorialen Variablen „Arbeitseinheit im Praxiseinsatz“ mit den Ausprägungen „Ambulanter Pflegedienst“, „Station im Kranken-

haus/Klinikum“, „Wohnbereich im Altenpflegeheim“ sowie „Sonstige“, wozu z. B. Einrichtungen der Rehabilitation zählen. Mangels vorliegender Studien und Plausibilitätsannahmen sind die vorgelegten Analysen hierzu explorativ ausgerichtet.

3.2 Zusammenhang zwischen Ausbildungsbedingungen in Praxiseinrichtungen und Kompetenzfacetten

Die Ergebnisse der Analysen sind in Tabelle 2 enthalten. Für die Belastungs-Bewältigungskompetenz wurden Korrelationen zwischen den Skalen der Ausbildungsbedingungen und erzielten Leistungen in den beiden von Warwas et al. (2023) nachgewiesenen Dimensionen der Belastungs-Bewältigungskompetenz berechnet. Signifikante Korrelationen konnten nur mit der Dimension „Angemessene Auswahl und Begründung von Bewältigungsstrategien“ nachgewiesen werden.

Tabelle 2: Signifikante Pearson-Korrelationen zwischen betrieblichen Ausbildungsbedingungen und den WLE-Personenfähigkeitsparametern der Klient:innen-/Patient:innenbezogenen Kompetenz, der Belastungs-Bewältigungskompetenz (Dimension Angemessene Auswahl und Begründung von Bewältigungsstrategien) und der interprofessionellen Kooperationskompetenz

Skala	Klient:innen/ Patient:innenbezogene Kompetenz	Belastungs-Bewältigungskompetenz – Dimension Angemessene Auswahl und Begründung von Bewältigungsstrategien	Interprofessionelle Kooperationskompetenz
Organisationsbezogene Skalen			
Kommunikation und Konfliktbewältigung	.18**		
Funktionale Einbindung		.13*	.25**
Aufgabenbezogene Skalen			
Komplexität der Anforderungen	.17**	.15*	.26**
Feedback-Skalen			
Negatives Feedback durch zu pflegende Personen (rekodiert)	.17**	.17**	.18**
Praxisanleitungs-Skalen			
Pädagogische und fachliche Kompetenz		.13*	
** $p < 0,01$, * $p < 0,05$, N = 292			

Sämtliche gefundenen Zusammenhänge liegen im niedrigen Bereich. Für alle drei Kompetenzfacetten kann festgestellt werden, dass die rekodierte Skala zum negativen Feedback durch zu pflegende Personen signifikant positiv mit der Leistung im Kompetenztest korreliert. Ebenso weist die Komplexität der Anforderungen mit allen drei Kompetenzfacetten einen signifikant positiven Zusammenhang auf, wobei dieser am stärksten für die interprofessionelle Kooperationskompetenz ausfällt.

Für die unmittelbar auf die zu pflegende Person gerichtete Kompetenzfacette zeigt sich eine weitere signifikant positive Korrelation mit der Kommunikation und Konfliktbewältigung. Keine signifikanten Zusammenhänge liegen für diese Kompetenzfacette mit den Skalen zur Praxisanleitung oder Praxisbegleitung vor.

Für die Strategiewahl und -begründung angesichts stressträchtiger Situationen im Pflegealltag liegt ein signifikant positiver Zusammenhang mit der funktionalen Einbindung in eine Organisationseinheit vor. Zudem ist die Fähigkeit zur Auswahl und Begründung von Bewältigungsstrategien signifikant positiv mit der Einschätzung der pädagogischen und fachlichen Kompetenz der Praxisanleitung verbunden. Keine signifikanten Zusammenhänge bestehen demgegenüber mit den Skalen zur Praxisbegleitung.

Auch für die interprofessionelle Kooperationskompetenz liegt ein signifikant positiver Zusammenhang mit der funktionalen Einbindung vor, wobei dieser etwas höher ausfällt als für die Belastungs-Bewältigungskompetenz. Keine signifikanten Zusammenhänge liegen außerdem – wie bereits bei der unmittelbar auf die zu pflegende Person bezogenen Kompetenzfacette – mit den Skalen zur Praxisanleitung oder Praxisbegleitung vor. Zudem konnten keine statistisch bedeutsamen Zusammenhänge für die Ausbildungsmerkmale gefunden werden, die spezifisch mit Blick auf die interprofessionelle Kooperationskompetenz erfasst worden waren, d. h. die Skalen zur Hierarchie und zur interprofessionellen Kooperation in der Arbeitseinheit.

3.3 Wahrgenommene Ausbildungsbedingungen im Vergleich der Versorgungsbereiche

Unterschiede in den wahrgenommenen Ausbildungsbedingungen der Praxiseinrichtungen zwischen den Versorgungsbereichen wurden im Hinblick auf diejenigen Qualitätsmerkmale geprüft, für welche signifikante Zusammenhänge mit den Kompetenzfacetten bestanden. Tabelle 3 gibt eine Übersicht über die Wahrnehmung dieser Ausbildungsbedingungen in den Praxiseinrichtungen.

Tabelle 3: Mittelwerte und Standardabweichungen der betrieblichen Ausbildungsbedingungen nach Ausbildungsbereich

Arbeitseinheit im Praxiseinsatz	Funktionale Einbindung			Kommunikation und Konfliktbewältigung			Komplexität der Anforderung			Negatives Feed- back durch zu pflegende Perso- nen (rekodiert)			Pädagogische und fachliche Kompetenz		
	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD
Ambulanter Pflegedienst	65	4,24	0,65	63	3,38	1,18	65	3,80	0,71	65	4,59	0,77	64	4,07	0,64
Station im Kranken- haus/Klinikum	77	4,03	0,84	77	3,22	1,13	76	4,05	0,77	76	4,70	0,69	76	4,14	0,77
Wohnbereich im Altenpflegeheim	101	3,89	0,87	95	3,11	0,99	98	3,79	0,82	98	4,53	0,88	95	3,77	0,91
Sonstiges	18	4,22	0,67	18	3,48	1,11	18	3,94	0,94	18	4,72	0,46	18	4,33	0,56
Insgesamt	261	4,04	0,80	253	3,24	1,09	257	3,88	0,79	257	4,61	0,78	253	4,00	0,80

Die Mittelwerte wurden im Weiteren mittels einfaktorieller Varianzanalyse auf Unterschiede zwischen den Praxiseinrichtungen überprüft. Signifikant divergierte hierbei das Ausmaß der funktionalen Einbindung zwischen den Versorgungsbereichen, $F(3, 257) = 2.81$, $p < .05$. Ebenfalls unterschied sich die pädagogische und fachliche Kompetenz, die der Praxisanleitung zugeschrieben wurde, $F(3, 249) = 4.80$, $p < .01$.

Für alle Merkmale ergaben sich im Levene-Test keine Hinweise auf Varianzheterogenität ($p < .05$). Der Tukey-Post-hoc-Test zeigte für die funktionale Einbindung einen signifikanten Unterschied ($p < .05$) zwischen den Praxiseinrichtungen „ambulanter Pflegedienst“ und „Wohnbereich im Altenpflegeheim“ ($-0,34$, 95 %-CI $[-0,67-0,02]$). Für die pädagogisch-fachliche Kompetenz der Praxisanleitung war die Ausbildung in der Praxiseinrichtung „Wohnbereich im Altenpflegeheim“ der „Station im Krankenhaus/Klinikum“ im Post-hoc-Test signifikant ($p < .05$) unterlegen ($-0,37$, 95 %-CI $[-0,68-0,06]$).

Bei beiden Variablen erwies sich der Versorgungsbereich „Wohnbereich im Altenpflegeheim“ also als der Versorgungsbereich mit der geringsten Einschätzung zum jeweiligen Ausbildungsmerkmal. Während die pädagogische und fachliche Kompetenz der Praxisanleitung, welche einen Zusammenhang mit der Dimension „Angemessene Auswahl und Begründung von Bewältigungsstrategien“ aufweist, im Versorgungsbereich „Station im Krankenhaus/Klinikum“ als am ehesten gegeben eingeschätzt wird, fällt in der Einschätzung der Auszubildenden die funktionale Einbindung – ein Merkmal, das sowohl zur genannten Dimension der Belastungs-Bewältigungskompetenz als auch zur interprofessionellen Kooperationskompetenz einen signifikanten Zusammenhang aufweist – im Versorgungsbereich „ambulanter Pflegedienst“ am höchsten aus. Die Wahrnehmung der anderen in Tabelle 3 genannten Ausbildungsbedingungen unterscheidet sich nicht signifikant nach Praxiseinsatz.

4 Diskussion

Das Feedback durch zu pflegende Personen wurde im Projekt EKGe mit dem Ziel erfasst, den spezifischen interaktiven Charakter der Pflegearbeit bei der empirischen Modellierung der Ausbildungsqualität zu berücksichtigen. Die Befunde liefern erste Hinweise darauf, dass diesem eine eigenständige Bedeutung bei der Entwicklung aller drei Kompetenzfacetten zukommen könnte. Der Befund ist gerade auch angesichts der oftmals geringen Zusammenhänge bei retrospektiven Erfassungen betrieblicher Ausbildungsqualität bedeutsam (Rausch, 2012) und deutet an, dass wahrgenommenes negatives Feedback durch zu pflegende Personen nachhaltig wirksam sein und daher in der Pflegeausbildung einer besonderen Beachtung bedürfen könnte. Eine mögliche Interpretation ist, dass das Ausbleiben negativen Feedbacks im Sinne einer negativen Verstärkung als subjektiver Indikator einer gelingenden Pflegearbeit wirksam ist. Andere Skalen zur Erfassung des interaktiven Charakters der Pflegearbeit hingen erwartungswidrig nicht signifikant mit den gemessenen Kompetenzniveaus zusammen; erklärbar sein könnte dies durch den eher distalen Charakter dieser Ausbildungsmerkmale (Wittmann et al., 2015), aber auch durch zeitlich nachgelagerte Wirkungen oder die Vermittlung z. B. über motivationale oder andere Faktoren; dies wäre im Weiteren genauer zu analysieren (Baethge-Kinsky et al., 2016).

Lediglich im Hinblick auf Strategien der Bewältigung pflegetypischer Belastungen ergeben sich außerdem Zusammenhänge zur wahrgenommenen pädagogischen und fachlichen Kompetenz der Praxisanleitung. Als Interpretation bietet sich an, dass Kompetenz aufseiten der Praxisanleitung besonders dort wahrgenommen wird, wo die Praxisanleiter:innen als hilfreich bei der Entwicklung von Strategien zum Umgang mit für die Auszubildenden belastenden Situationen erlebt werden (Krell et al., 2015). Aus diesem Blickwinkel heraus könnten die gefundenen Differenzen zwischen den Versorgungsbereichen bedeutsam sein; insbesondere scheint eine entsprechende Kompetenz des Praxisanleitungspersonals im Bereich „Wohnbereich im Altenpflegeheim“ am wenigsten wahrgenommen zu werden. Gerade auch mit Blick auf den Verbleib von Pflegeauszubildenden deutet sich hier möglicher Handlungsbedarf an: In allen, aber insbesondere in diesem Versorgungsbereich sollte das Bewusstsein der an Ausbildung beteiligten Personen steigen, nicht nur fachlich-methodische Anleiter:innen, sondern auch professionelle Rollenvorbilder im Umgang mit Belastungsfaktoren zu sein.

Für alle Kompetenzfacetten ergibt sich ein positiver Zusammenhang zwischen der wahrgenommenen Komplexität der Anforderungen und dem erreichten Kompetenzniveau. Darüber hinaus ergibt sich für die Strategiedimension der Belastungs-Bewältigungskompetenz sowie für die interprofessionelle Kooperationskompetenz jeweils ein positiver Zusammenhang der funktionalen Einbindung, welche die Einbindung in Aufgaben und Abstimmungen der Arbeitseinheit betrifft. Während Auszubildende sich wiederum im „Wohnbereich im Altenpflegeheim“ am wenigsten eingebunden fühlen, liegen hier besondere Stärken des ambulanten pflegerischen Versorgungsbereichs vor.

Anzumerken ist, dass vergleichbar mit anderen Untersuchungen zur betrieblichen Ausbildungsqualität alle gefundenen Zusammenhänge im niedrigen Bereich liegen (Wittmann et al., 2015). Da keine Korrekturen für eine Alphafehler-Kumulierung vorgenommen wurden, ist außerdem davon auszugehen, dass die gefundenen Signifikanzen die tatsächlichen Zusammenhänge eher überschätzen. Darüber hinaus stehen Analysen zu den längerfristigen Zusammenhängen sowie zu Wechselwirkungen mit personenbezogenen Merkmalen wie der Kompetenz zu Beginn der Praxisphase aus. Schließlich ist darauf hinzuweisen, dass die Stichprobe, welche den Analysen zugrunde lag, nur teilweise bereits aus der generalistischen Ausbildung und hier aus dem Kontext der ersten Implementierung unter Bedingungen der COVID-19-Pandemie entstammt. Vor diesem Hintergrund wären gerade mit Blick auf die Ausbildungsbedingungen in den Praxiseinrichtungen Analysen im Kontext einer etablierten Generalistik erforderlich.

5 Fazit

Mit dem vorliegenden Beitrag werden die spezifischen Ausbildungsbedingungen der Pflege als personenbezogenem Dienstleistungsberuf in den Blick genommen, der die Arbeit mit und am Menschen zum Gegenstand hat. Aufgrund des spezifischen gesetzlichen Rahmens in der praktischen Pflegeausbildung weist der Beruf zudem auch curriculare Besonderheiten auf; mit der curricularen Neuordnung („Generalistik“) sind in diesem Beruf vor allem Pflichteinsätze in unterschiedlichen Versorgungsbereichen vorgesehen.

Die gefundenen Zusammenhänge zwischen negativem Feedback zu pflegender Personen und allen drei gemessenen Kompetenzdimensionen unterstreichen die Bedeutung dieser neu operationalisierten Variablen und insbesondere des für personenbezogene Dienstleistungsberufe wie die Pflege spezifischen persönlichen Bezugs des „Arbeitsgegenstands“, bei allerdings generell niedrigen Zusammenhängen. Insbesondere könnten sie auf das praktische Erfordernis hindeuten, die Wahrnehmung negativen Feedbacks in der Pflegeausbildung aktiv zu bearbeiten. So könnte in den Praxiseinrichtungen der Pflege geprüft werden, inwieweit gezielte Gespräche über negatives Feedback durch zu Pflegenden und den Umgang hiermit hilfreich sind. Insbesondere könnte dies verstärkt zum Gegenstand der Praxisanleitung gemacht werden. Hier wie bei anderen Ausbildungsmerkmalen wären die langfristigen Wirkungen auf die Kompetenzentwicklung im Weiteren differenziert zu analysieren.

Darüber hinaus werden Unterschiede bezüglich der wahrgenommenen Ausbildungsmerkmale zwischen den Versorgungsbereichen deutlich, insbesondere negativer ausfallende Einschätzungen im Bereich der stationären Altenpflege, aber auch besondere Chancen der ambulanten Pflege durch hohe funktionale Einbindung. Eine offene Frage bleibt in diesem Zusammenhang, wie sich die pflegespezifischen sektoralen Übergänge mit ihren unterschiedlichen Merkmalsprofilen auf die Kompetenzentwicklung der Pflegeauszubildenden auswirken. Die vorgelegten Befunde liefern

einen ersten empirischen Anlass dazu, diese Übergänge im Weiteren genauer zu erforschen.

Literatur

- Baethge-Kinsky, V., Baethge, M., & Lischewsky, J. (2016). Bedingungen beruflicher Kompetenzentwicklung: institutionelle und individuelle Kontextfaktoren. In K. Beck, M. Landenberger, & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung. Ergebnisse aus der BMBF-Förderinitiative ASCOT* (S. 265–300). wbv. <https://doi.org/10.3278/6004436w>
- Battistelli, A., Galletta, M., Vandenbergh, C., & Odoardi, C. (2016). Perceived organisational support, organisational commitment and self-competence among nurses: a study in two Italian hospitals. *Journal of Nursing Management*, 24(1), E44–E53. <https://doi.org/10.1111/jonm.12287>
- Bing-Jonsson, P. C., Bjørk, I. T., Hofoss, D., Kirkevold, M., & Foss, C. (2014). Competence in advanced older people nursing: development of 'Nursing older people – Competence evaluation tool'. *International Journal of Older People Nursing*, 10, 1–14. <https://doi.org/10.1111/opn.12057>
- Böhle, F., & Glaser, J. (Hrsg.) (2006). *Arbeit in der Interaktion – Interaktion als Arbeit. Arbeitsorganisation und Interaktionsarbeit in der Dienstleistung*. Springer. <https://doi.org/10.1007/978-3-531-90505-1>
- Böhn, S., & Deutscher, V. K. (2019). Betriebliche Ausbildungsbedingungen im dualen System – eine qualitative Meta-Analyse zur Operationalisierung in Auszubildendenbefragungen. *Zeitschrift für Pädagogische Psychologie*, 33(1), 49–70. <https://doi.org/10.1024/1010-0652/a000234>
- Böhn, S., & Deutscher, V. K. (2021). Development and validation of a learning quality inventory for in-company training in VET (VET-LQI). *Vocations and Learning*, 14(1), 23–53. <https://doi.org/10.1007/s12186-020-09251-3>
- Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic Approaches to a Successful Literature Review* (2nd ed.). Sage.
- Dietzen, A., Velten, S., Schnitzler, A., Schwerin, C., Nickolaus, R., Gönnerwein, A., Nitzschke, A., & Lazar, A. (2011). *Einfluss der betrieblichen Ausbildungsqualität auf die Fachkompetenz in ausgewählten Berufen* (Aqua.Kom). Abschlussbericht. BIBB.
- Ducki, A. (2000). *Diagnose gesundheitsförderlicher Arbeit. Eine Gesamtstrategie zur betrieblichen Gesundheitsanalyse*. vdf.
- Eisenberger, R., Huntington, R., Hutchison, S., & Sowa, D. (1986). Perceived Organizational Support. *Journal of Applied Psychology*, 71(3), 500–507. <https://doi.org/10.1037/0021-9010.71.3.500>
- Geißel, B., Gschwendtner, T., & Nickolaus, R. (2009). Betriebliche Ausbildungsqualität in der Wahrnehmung von Auszubildenden. In E. Wuttke, H. Ebner, B. Fürstenau, & R. Tenberg (Hrsg.), *Erträge und Perspektiven berufs- und wirtschaftspädagogischer Forschung* (S. 9–21). Budrich. <https://doi.org/10.2307/j.ctvdf04n8.4>

- Giesbers, A. P. M., Schouteten, R. L. J., Poutsma, E., van der Heijden, B. I. J. M., & van Achterberg, T. (2015). Feedback provision, nurses' well-being and quality improvement: towards a conceptual framework. *Journal of Nursing Management*, 23(5), 682–691. <https://doi.org/10.1111/jonm.12196>
- Goller, M., Steffen, B., & Harteis, C. (2019). Becoming a nurse aide: An investigation of an existing workplace curriculum in a nursing home. *Vocations and Learning*, 12(1), 67–85. <https://doi.org/10.1007/s12186-018-9209-z>
- Hacker, W., & Reinhold, S. (1999). *Beanspruchungsscreening bei Humandienstleistungen: BHD-System*. Harcourt Test Services GmbH.
- Henderson, A., Creedy, D., Boorman, R., Cooke, M., & Walker, R. (2010). Development and psychometric testing of the Clinical Learning Organisational Culture Survey (CLOCS). *Nurse Education Today*, 30(7), 598–602. <https://doi.org/10.1016/j.nedt.2009.12.006>
- Hoff, E.-H., Lempert, W., & Lappe, L. (1991). *Persönlichkeitsentwicklung in Facharbeiterbiographien*. Huber.
- Krell, J., Worofka, I., Simon, J., Wittmann, E., & Purwins, C. (2015). Herausfordernde Situationen in unterschiedlichen Settings der Pflege älterer Menschen. In T. Tramm, M. Fischer, & C. Aprea (Hrsg.), *Berufliche Lehr-Lernforschung. bwp@ Ausgabe Nr. 28*. http://www.bwpat.de/ausgabe28/krell_etal_bwpat28.pdf
- Laschinger, H. K. S. (2012). Job and career satisfaction and turnover intentions of newly graduated nurses. *Journal of Nursing Management*, 20(4), 472–484. <https://doi.org/10.1111/j.1365-2834.2011.01293.x>
- Nickolaus, R., Rosendahl, J., Gschwendtner, T., Geißel, B., & Straka, G. A. (2010). Erklärungsmodelle zur Kompetenz- und Motivationsentwicklung bei Bankkauffleuten, KFZ-Mechatronikern und Elektronikern. In J. Seifried, E. Wuttke, R. Nickolaus, & P. Sloane (Hrsg.), *Lehr-Lern-Forschung in der kaufmännischen Bildung. Ergebnisse und Gestaltungsaufgabe. Beiheft 23 zur Zeitschrift für Berufs- und Wirtschaftspädagogik* (S. 73–87). Steiner.
- Pettricrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences. A practical guide*. Blackwell Publishing. <https://doi.org/10.1002/9780470754887>
- Rausch, A. (2012). Prozessnahe und retrospektive Erhebungsmethoden der Arbeitsanalyse in der betrieblichen Ausbildung. *Empirische Pädagogik*, 26(2), 247–270.
- Rausch, A. (2013). Task characteristics and learning potentials: Empirical results of three diary studies on workplace learning. *Vocations and Learning*, 6(1), 55–79. <https://doi.org/10.1007/s12186-012-9086-9>
- Saarikoski, M., Isoaho, H., Warne, T., & Leino-Kilpi, H. (2008). The nurse teacher in clinical practice: Developing the new sub-dimension to the clinical learning environment and supervision (CLES) scale. *International Journal of Nursing Studies*, 45(8), 1233–1237. <https://doi.org/10.1016/j.ijnurstu.2007.07.009>
- Schröter, T. (2017). *Vergleich von dezentraler und zentraler Praxisanleitung bei Auszubildenden der Gesundheits- und Krankenpflege* [Unveröffentlichte Master's Thesis; Lehrstuhl für Berufspädagogik der TUM School of Education]. Technische Universität München.

- Seeber, S., & Wittmann, E. (2017). Social competence research: A review. In M. Mulder (Hrsg.), *Competence-based vocational and professional education. Bridging the worlds of work and education* (S. 1029–1050). Springer. https://doi.org/10.1007/978-3-319-41713-4_48
- Statistisches Bundesamt (Destatis) (2022). *Statistik nach der Pflegeberufe-Ausbildungsfinanzierungsverordnung 2021*. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Berufliche-Bildung/Publikationen/Downloads-Berufliche-Bildung/pflegeberufe-ausbildungsfinanzierung-vo-5212401217005.xlsx?__blob=publicationFile&v=5
- Striković, A., Wittmann, E., Warwas, J., Philipps, V., Weyland, U., & Wilczek, L. (2022). Berufsfeldspezifische und prozessnahe Erfassung von Ausbildungsbedingungen im Gesundheitsbereich. In K. Kögler, H.-H. Kremer, & U. Weyland (Hrsg.), *Jahrbuch der berufs- und wirtschaftspädagogischen Forschung 2022* (S. 237–259). Barbara Budrich. <https://doi.org/10.25656/01:26525>
- Warwas, J., Vorpahl, W., Seeber, S., Krebs, P., Weyland, U., Wittmann, E., Wilczek, L., & Striković, A. (2023). Developing and validating an online situational judgment test on the stress coping competence of nursing apprentices. *Empirical Research in Vocational Education and Training*, 15, Article 5. <https://doi.org/10.1186/s40461-023-00145-x>
- Weidner, F. (1995). *Professionelle Pflegepraxis und Gesundheitsförderung. Eine empirische Untersuchung über Voraussetzungen und Perspektiven des beruflichen Handelns in der Krankenpflege*. Mabuse.
- Wenner, T. (2018). Entwicklung eines Instruments zur Erfassung der Wechselwirkung von Lernortkooperation und Ausbildungsqualität. *Journal of Technical Education*, 6(1), 223–237. <https://www.journal-of-technical-education.de/index.php/joted/article/view/123/130>
- Wittmann, E., Weyland, U., Striković, A., Pohley, M., Hill, S., Wilczek, L., Krebs, P., Seeber, S. & Warwas, J. (2024). Interprofessionelle Kooperationskompetenz von Pflegeauszubildenden im Kontext der digitalen Transformation. *Zeitschrift für Pädagogik*, 2024(2), 202–225. <https://doi.org/10.3262/ZP2402202>
- Wittmann, E., Weyland, U., Seeber, S., Warwas, J., Striković, A., Krebs, P., Pohley, M. & Wilczek, L. (2022). Test sensitivity in assessing competencies in nursing education. *Empirical Research in Vocational Education and Training*, 14(3), 1–26. <https://doi.org/10.1186/s40461-022-00131-9>
- Wittmann, E., Weyland, U., Kaspar, R., Döring, O., Hartig, J., Nauerth, A., Rechenbach, S., Möllers, M., Simon, J. & Worofka, I. (2015). Betriebliche Ausbildungsmerkmale und berufsfachliche Handlungskompetenz in der Altenpflege. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 111(3), 359–378.
- Wittmann, E. (2001). *Kompetente Kundenkommunikation von Auszubildenden in der Bank. Eine theoretische und empirische Studie zum Einfluß betrieblicher Ausbildungsbedingungen*. Peter Lang.
- Zimmermann, M., Wild, K.-P. & Müller, W. (1999). Das „Mannheimer Inventar zur Erfassung betrieblicher Ausbildungssituationen“ (MIZEBA). *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 95(3), 373–402.

Akademische Bildung oder berufliche Bildung – Kommentar zu Schwerpunkt 1

CHRISTIANE SPIEL

Zusammenfassung:

Der Kommentar thematisiert das Verhältnis zwischen beruflicher und akademischer Bildung und weist auf mögliche Konsequenzen für die Jugendlichen hin. Zudem werden die beiden Beiträge dieses Sammelbands aus den Projekten DigiDIn-Kfz und EGKe besprochen und Gemeinsamkeiten sowie Unterschiede herausgearbeitet.

Schlagworte: berufliche und akademische Bildung

Abstract:

This commentary addresses the relationship between vocational and academic education and points out possible consequences for young people. In addition, it discusses the two contributions to this edited volume from the DigiDIn-Kfz and EGKe projects and identifies similarities and differences.

Key words: Vocational and academic education

„Rückblickend bedaure ich, dass ich neben der Matura (= Abitur) nicht auch eine berufliche Ausbildung gemacht habe.“ Diese Aussage stammt von mehreren weiblichen Teilnehmerinnen meiner Vorlesung Bildungspsychologie als Ergebnis einer Reflexion ihrer eigenen Bildungskarriere. Es fehle ihnen daher an beruflichen Handlungskompetenzen, die sie für ihre weitere Laufbahn für hoch relevant erachten. Die anschließende Diskussion ergab, dass sowohl die Eltern der Studierenden als auch sie selbst ausschließlich eine akademische Bildung im Blick hatten. Konsequenterweise hatten sie daher ein Gymnasium besucht und mit der Matura abgeschlossen. Viele der anderen Studierenden berichteten über analoge Erfahrungen.

Die Beobachtungen in der Vorlesung sind ohne Zweifel nur anekdotisch. Sie werden aber durch viele Gespräche in anderen Kontexten bestätigt. Die Breite des Spektrums an Ausbildungsmöglichkeiten ist in vielen Familien kaum bekannt. Für Eltern mit akademischer Bildung ist es daher zumeist selbstverständlich, dass auch ihre Kinder eine akademische Bildung erwerben. Der Besuch einer gymnasialen Oberstufe hat dazu auch noch den Vorteil, dass die Entscheidung für die tertiäre Bildung erst mit 18 anstatt mit 14 Jahren getroffen werden muss. Mit 14 Jahren zu wissen, „wohin man will“, ist sehr schwierig. Die Möglichkeit, Matura mit Lehre zu verbinden, ist häufig nicht bekannt.

Die seit vielen Jahren bestehende Forderung von OECD und EU, den Prozentsatz akademisch gebildeter Menschen deutlich anzuheben, hat die Fokussierung auf akademische Bildung in alle Gesellschaftsschichten gebracht und kontinuierlich verstärkt. Die dadurch entstandene Veränderung in der Relation von akademischer und beruflicher Bildung hat jedoch nicht nur positive Effekte (wie von OECD und EU propagiert), sondern auch negative. Ich möchte es mit den Worten von Stefan Wolter beschreiben, einem renommierten Schweizer Bildungsökonom. Er spricht von einer Stellschraube zwischen akademischer und beruflicher Bildung. Wenn die Stellschraube so gestellt ist, dass sehr viele junge Menschen in akademische Bildung gehen und nur wenige in berufliche, schadet das beiden Bereichen: In der akademischen Bildung sind viele überfordert und in der beruflichen Bildung zu wenige ausreichend qualifiziert. Das hat negative Konsequenzen für das einzelne Individuum und die Gesellschaft als Ganze, wie auch die Studie von Gschwendtner et al. zeigt.

Im Folgenden werden die Studien von Gschwendtner et al. (DigiDIn-Kfz) und Wittmann et al. (EGKe) kurz zusammengefasst und diskutiert.

Auszubildende in Kfz-Mechatronik: Interventionen zur Förderung der Diagnosekompetenz (DigiDIn-Kfz)

Zielgruppe der Studie von Gschwendtner et al. waren Auszubildende in Kfz-Mechatronik. Wie die Autor:innen ausführen, erreicht in Deutschland nur ein kleiner Teil der Auszubildenden in Kfz-Mechatronik am Ende der Ausbildung eine zufriedenstellende Kfz-Diagnosekompetenz. Das Diagnostizieren von Störungen an Kraftfahrzeugen ist jedoch eine erforderliche Kernkompetenz von Kfz-Mechatroniker:innen und damit ein sehr wichtiges Ausbildungsziel. Da davon auszugehen ist, dass die Komplexität und Vernetzung der Fahrzeugsysteme künftig noch zunehmen wird, ist eine Förderung der Diagnosekompetenz dringend erforderlich.

Dies stellt jedoch eine gewaltige Herausforderung dar. Denn wie frühere Studien gezeigt haben, sind 17 Prozent der Auszubildenden nicht in der Lage, einfachere diagnoserelevante Informationen zu verarbeiten. Bei diesen muss daher die basale diagnoserelevante Rezeptionskompetenz gefördert werden. Circa 52 Prozent haben Defizite im Bereich der komplexen diagnoserelevanten Rezeptionskompetenz und etwa 86 Prozent sind nicht in der Lage, eine modellbasierte Diagnosestrategie anzuwenden. Das bedeutet im Umkehrschluss, dass nur 14 Prozent das relevante Kfz-System bzw. Ausschnitte davon mental modellieren und basierend darauf Hypothesen formulieren und testen können.

Wie kann bei einer solchen Ausgangslage eine Förderung gestaltet werden, die bewirkt, dass ein substanzieller Prozentsatz der Auszubildenden am Ende der Ausbildung ein zufriedenstellendes Kfz-Diagnosekompetenzniveau erreicht?

Die im Beitrag dargestellte Herangehensweise ist theoretisch und methodisch sehr überzeugend, aber auch sehr aufwendig. Theoretisch fundiert wurden Förderziele, Fördergegenstand und die Gestaltung der Fördermaßnahme an das jeweilige Ausgangsniveau der Diagnosekompetenz angepasst. Die dafür eingesetzte Forschungs-

methodik bestand aus differenzierten Interventionen mit mehreren Treatment- und Kontrollgruppen unter Einsatz einer Kfz-Computersimulation, die bereits in früheren Projekten (z. B. KOKO Kfz der Forschungsinitiative ASCOT) genutzt worden war.

Die Analysen zeigen, dass durch die Interventionen eine Reihe mittlerer bis großer Effekte auf die Kfz-Diagnosekompetenz erzielt werden konnte. Allerdings gab es nur teilweise bedeutsame Effekte auf den Erfolg bei eigenständigen Kfz-Diagnosen in einer Simulation, was laut Autor:innen als Globalindikator anzusehen ist.

Mit Blick auf die oben angeführte Stellschraube nach Wolter, die hier offensichtlich ungünstig gestellt ist, ergibt sich die Frage, ob alle diejenigen, die sich für eine Ausbildung in Kfz-Mechatronik entscheiden, überhaupt in der Lage sind, die erforderliche Kfz-Diagnosekompetenz zu erreichen. Ein hohes Ausmaß an erforderlicher Selbstinstruktion, wie es die Interventionen vorsehen, erfordert u. a. hohe Lernmotivation. Wenn das Kompetenzerleben jedoch gering ist, ist – siehe Selbstbestimmungstheorie – anzunehmen, dass auch die Lernmotivation niedrig ist, ebenso das Wohlbefinden, und es besteht die Gefahr, dass es zu einer Negativspirale kommt. Daran würde vermutlich auch eine längere Intervention nichts ändern, vielmehr könnte diese die Lernmotivation noch mehr reduzieren. Wie die Autor:innen selbst schreiben, konnten solche Aspekte nicht ausreichend berücksichtigt werden; ebenso nicht die Ausbildungsbedingungen, die in der Studie von Wittmann et al. untersucht wurden.

Auszubildende im Pflegebereich: Zusammenhang zwischen Ausbildungsbedingungen und Kompetenzen (EGKe)

Der folgende Abschnitt widmet sich der Studie von Wittmann und Kolleg:innen. Zielgruppe dieser Studie waren Auszubildende im Pflegebereich. Erhoben wurden deren wahrgenommene Ausbildungsbedingungen in der Pflegepraxis, die laut einigen früheren Studien in Zusammenhang mit der Kompetenzentwicklung stehen. Dies veranlasste die Autor:innen, diese Ausbildungsbedingungen detailliert zu analysieren. Zur Erhebung wurde ein sehr umfangreiches retrospektiv orientiertes Fragebogeninventar eingesetzt. Teils wurden bereits vorliegende Skalen verwendet, teils wurden neue Skalen konstruiert, die anhand einer pragmatischen Heuristik recherchiert und empirisch umgesetzt wurden.

Die Stichprobe umfasste Auszubildende aus den drei Pflegeberufen Altenpfleger:in, Gesundheits- und Krankenpfleger:in und generalisierter Pflegeausbildung. Für Letztere sind Einsätze in unterschiedlichen Einrichtungen und Versorgungsbereichen vorgesehen. Von allen drei Gruppen lagen Informationen zu Testleistungen in drei Kompetenzfacetten vor (unmittelbar auf die zu pflegende Person bezogene Kompetenz, Belastungsbewältigungskompetenz, interprofessionelle Kooperationskompetenz), die mittels eines videobasierten, computergestützten Situational-Judgement-Tests operationalisiert wurden. Angaben über Kompetenzniveaus liegen nicht vor und dementsprechend auch keine Informationen über den Prozentsatz der Auszubildenden, die die jeweiligen Niveaus erreicht bzw. nicht erreicht haben.

Die gefundenen Zusammenhänge zwischen Ausbildungsbedingungen und Kompetenzfacetten sind insgesamt ziemlich niedrig; aufgrund der Alphafehler-Kumulierung liegt möglicherweise auch keine statistische Bedeutsamkeit vor. Zwei Ergebnisse sind aus Sicht der Autor:innen inhaltlich bedeutsam: die gefundenen Zusammenhänge zwischen negativem Feedback von pflegenden Personen mit allen drei Kompetenzfacetten und die negativeren Einschätzungen der wahrgenommenen Ausbildungsmerkmale von Auszubildenden im Bereich der stationären Altenpflege im Vergleich zu den beiden anderen Versorgungsbereichen.

Diskussion und Empfehlungen

Ein Vergleich zeigt auf den ersten Blick keine bzw. fast keine Gemeinsamkeiten in den beiden Studien. Es unterscheiden sich die Zielgruppen, die Fragestellungen und damit auch die Methodik, d. h. Intervention versus Querschnittsanalyse von Zusammenhängen.

Bei genauerem Hinsehen lassen sich jedoch einige Gemeinsamkeiten ausmachen. In beiden Fällen ging es um Kompetenzen, die wichtige Ausbildungsziele darstellen: Diagnosekompetenz für die Kfz-Mechatronik, Handlungskompetenzen für den Pflegebereich. In der Interventionsstudie mit den Auszubildenden in der Kfz-Mechatronik sollte deren Diagnosekompetenz gefördert werden. Im Pflegebereich sollte geprüft werden, welche Zusammenhänge zwischen dem Niveau der Handlungskompetenz und Ausbildungsbedingungen bestehen.

Eine weitere Gemeinsamkeit der Studien besteht darin, dass in beiden die Kompetenzen objektiv und realitätsnah erhoben wurden (Simulation, Situational-Judgement-Tests); eine wichtige Voraussetzung dafür, dass von den Studien auf reales Verhalten geschlossen werden kann.

Wie beide Autor:innengruppen schreiben, beschäftigen sich ihre Studien mit Herausforderungen in den jeweiligen Berufsfeldern, die jedoch sehr unterschiedlich gelagert sind. Bei der Kfz-Mechatronik handelt es sich um ein Berufsfeld mit dynamisch zunehmender Komplexität, in dem Digitalisierung und Künstliche Intelligenz immer wichtigere Rollen spielen bzw. spielen werden. Der hohe Anteil an Auszubildenden, die am Ende der Ausbildung kein zufriedenstellendes Kfz-Diagnosekompetenzniveau haben, lässt sich mit hoher Wahrscheinlichkeit auf die steigenden Anforderungen in der Ausbildung zurückführen. Daher wurden von den Autor:innen differenzierte Interventionen zur Förderung der Diagnosekompetenz durchgeführt.

Auch in der Studie mit Pflegeauszubildenden wurden Kompetenzfacetten erhoben, die wichtige Ausbildungsziele repräsentieren. Allerdings gibt es dazu im Artikel keine differenzierten Angaben, weder hinsichtlich Kompetenzstufen noch bzgl. potenzieller Defizite in den Kompetenzfacetten. Der Grund dafür ist ganz offensichtlich der inhaltliche Fokus der Studie, der auf der Analyse der Zusammenhänge zwischen Ausbildungsbedingungen in der Pflegepraxis und den Handlungskompetenzen lag und

nicht auf der Förderung von Kompetenzen. Dennoch wäre es interessant zu erfahren, wie es um die Erreichung der Ausbildungsziele in den Pflegeberufen bestellt ist.

Als weitere Gemeinsamkeit erscheint durchaus plausibel, dass die beiden untersuchten Berufsfelder jeweils prototypisch für eine Vielzahl anderer sind, in denen sich ähnliche Herausforderungen für die Auszubildenden und damit auch die Ausbilder:innen stellen. Erstere Berufsfelder sind eher im technischen Bereich zu finden. Sie sind durch eine hohe Zunahme an Komplexität und damit auch der Anforderungen gekennzeichnet. Weitere Beispiele dafür neben der Kfz-Mechatronik sind Installationstechnik oder Elektrotechnik. Zentrales Ausbildungsziel ist jeweils die Diagnosekompetenz. Das zweite Berufsfeld lässt sich im weiteren Sinn dem Sozialbereich zuordnen. Welche Herausforderungen sich in diesem Bereich generell stellen, darüber kann auf Basis der Studie von Wittmann et al. keine Aussage getroffen werden. Daher beziehen sich die folgenden Empfehlungen nur auf die Kfz-Mechatronik.

Konkret werden einige Anregungen gegeben, wie mit den Herausforderungen in der Ausbildung der Kfz-Mechatroniker:innen bzw. dem weiteren Umfeld der technischen beruflichen Ausbildungen umgegangen werden könnte. Die Anregungen beziehen sich auf die Ziele der Ausbildung, die Zusammenarbeit in den Betrieben und die diagnoserelevanten Dokumente.

In Anbetracht des hohen Prozentsatzes der Kfz-Auszubildenden, die von dem Ausbildungsziel Diagnosekompetenz entfernt sind, und zwar z. T. sehr weit, stellt sich die Frage, ob es überhaupt realistisch und sinnvoll ist, das hohe Ziel der modellbasierten Diagnosestrategie für alle Auszubildenden zu verfolgen. Es würde einen enorm hohen Aufwand erfordern und – falls sich keine Lernerfolge einstellen sollten – vermutlich Frustration und sinkende Lernmotivation mit sich bringen (siehe oben). Und die Wahrscheinlichkeit, das Ausbildungsziel für alle Auszubildenden zu erreichen, erscheint sehr gering. Meine Anregung ist daher, die Ziele der Ausbildung gestuft festzulegen. Das würde bedeuten, dass ein Teil der Auszubildenden explizit nur basale Diagnosekompetenz erwirbt, diese aber gut gefestigt, der andere Teil komplexe, modellbasierte Diagnosekompetenz. Längerfristig könnte durch eine Kombination von Berufstätigkeit und Fortbildungen vermutlich zumindest bei einem Teil der ersten Gruppe eine Anhebung von basaler Diagnosekompetenz zu komplexer erreicht werden. Die Kompetenzförderung sollte generell, d. h. für beide Gruppen, möglichst durch den Einsatz von Kfz-Computersimulationen erfolgen, da diese sich offensichtlich bereits sehr bewährt haben. Allerdings sollten dabei die (wahrgenommenen) Ausbildungsbedingungen, wie Klarheit der Instruktion, Förderung der Lernmotivation etc., berücksichtigt werden (siehe Studie von Wittmann et al.).

Die Zusammenarbeit in Kfz-Werkstätten von Ausgebildeten mit unterschiedlichen Niveaus der Diagnosekompetenz könnte durchaus produktiv sein. Denn damit könnte die Diagnose von Störungen und deren Behebung ohne erwartbare Frustrationen auf allen Seiten (Kfz-Mechatroniker:innen, Werkstattleitung, Kund:innen) erfolgen. Dies ist ohne Zweifel nur in größeren Kfz-Werkstätten möglich. Eine Generalisierung auf Betriebe in anderen technischen Bereichen erscheint durchaus möglich zu sein.

Diagnoserelevante Rezeptionskompetenzen befähigen gemäß Gschwendtner et al. Auszubildende, diagnoserelevante Dokumente zu rezipieren und adäquate mentale Repräsentationen der darin enthaltenen Informationen zu bilden. Wie gut dies gelingt, ist ohne Zweifel nicht nur von den Auszubildenden, sondern auch von der Qualität der diagnoserelevanten Dokumente abhängig, d. h. wie einfach, kompliziert, verständlich, unverständlich diese formuliert sind. Meine eigenen Erfahrungen mit Anleitungen zur Installation oder Nutzung von Geräten sind häufig frustrierend. Da ich diese Erfahrungen mit vielen Personen in meinem Umfeld teile, dürfte es nicht nur an meinen eingeschränkten Rezeptionskompetenzen liegen, sondern auch mit den Anleitungen in den Dokumenten zu tun haben. Ich rege daher eine Analyse und gegebenenfalls Optimierung der diagnoserelevanten Dokumente in den technischen Berufsfeldern an. Eine Generalisierung auf Anleitungen in anderen Bereichen wäre sehr begrüßenswert.

Akademische und berufliche Bildung

Abschließend komme ich wieder auf die Studierenden zurück, die rückblickend bedauerten, neben der akademischen Bildung keine berufsbildende gemacht zu haben. Aus ihrer Sicht lag die Ursache darin, dass in ihren Familien nur akademische Bildung als Option gesehen wurde und nicht auch die berufsbildende. Um dem zu begegnen, ist es offensichtlich erforderlich, die gesamte Breite an Ausbildungsmöglichkeiten sichtbar zu machen, und zwar sowohl für Schüler:innen als auch für die Erziehungsberechtigten. Da die Entscheidung für eine akademische oder berufliche Bildung i. d. R. am Ende der Sekundarstufe I erfolgt, muss davor angesetzt werden, und zwar insbesondere in den Gymnasien. Ein besonderer Fokus sollte dabei auf die mögliche Kombination von akademischer und beruflicher Bildung gelegt werden. Falls diese Kombination künftig von vielen jungen Menschen gewählt wird, gäbe es – siehe Wolter – zwei Stellschrauben und damit drei Gruppen von Auszubildenden: Jugendliche in akademischer Bildung, Jugendliche in Berufsbildung und dazwischen Jugendliche in akademischer und beruflicher Bildung.

**Schwerpunkt 2: Problemlösekompetenz
als zentrale Zielgröße
der beruflichen Bildung**

Problemhaltigkeit kaufmännischer Prüfungsaufgaben: Eine Analyse mittels Lautem Denken

SUSAN SEEBER, HANNA MEINERS, EVELINE WUTTKE, MATTHIAS SCHUMANN & PHILIPP HARTMANN

Zusammenfassung

Durch Veränderungsprozesse beruflicher Anforderungen wird von kaufmännischen Angestellten zunehmend erwartet, dass sie komplexe Probleme lösen können. Während dieses Ziel in Curricula bereits verankert ist (curriculare Ebene), fehlt es an einer entsprechenden Messung dieser spezifischen Facette kaufmännischer Handlungskompetenz in den zentralen Abschlussprüfungen (Wuttke et al., 2022). Vermutlich hängt dies damit zusammen, dass die Förderung von Problemlösekompetenz in der Ausbildung (noch) nicht umfassend implementiert ist (instruktionale Ebene) und daher auch in Abschlussprüfungen (Assessment-Ebene) wenig Berücksichtigung findet. Auch kann es damit zusammenhängen, dass Leitfäden zur Erstellung von Prüfungsaufgaben sehr stark auf die Messung beruflicher Handlungskompetenz abstellen, während Fragen der Problemlösekompetenz dort nicht explizit thematisiert werden. Zudem sind Anforderungen an die Messung von beruflicher Problemlösekompetenz hochkomplex, auch spielt die Förderung und Messung von Problemlösekompetenz in der fachdidaktischen und diagnostischen Ausbildung von Lehrkräften und betrieblichem Ausbildungspersonal nach wie vor eine untergeordnete bzw. keine Rolle. In einer Interventionsstudie wurden daher Lehrkräfte und Auszubildende fortgebildet, um problemhaltige technologiebasierte (Prüfungs-)Aufgaben zu entwickeln und zu evaluieren. Um zu prüfen, ob die im Training entwickelten Aufgaben tatsächlich „Probleme“ für Auszubildende darstellen, wurden im Rahmen des Verbundprojekts TeKoP¹ ausgewählte Aufgaben mithilfe der Methode des lauten Denkens bei Auszubildenden analysiert. Zur Beantwortung der Forschungsfrage wurde eine Laut-Denken-Studie mit 18 Industriekaufleuten und acht Bürokaufleuten durchgeführt. Eingesetzt wurden 16 Aufgaben. Die Auswertungen belegen, dass Auszubildende die Schritte eines Problemlöseprozesses in der Bearbeitung durchliefen. Sie mussten vielfältige Faktoren simultan berücksichtigen und teilweise Ziele abwägen und priorisieren. Bei der Bearbeitung zogen die Auszubildenden Heuristiken aus ähnlichen Fällen heran, sie bildeten Analogien, überprüften die (vorläufigen) Handlungsergebnisse und korrigierten diese teilweise auch wieder. Diese Ergebnisse der Laut-Denken-Studie zeigen, dass es den

¹ TeKoP: Technologiebasiertes kompetenzorientiertes Prüfen, Förderkennzeichen: 21AP001A+B. Gefördert im Rahmen der BMBF-Forschungs- und Transferinitiative ASCOT+: Technology-based Assessment of Skills and Competencies in VET (URL: www.ascot-vet.net/)

Teilnehmenden des Trainings, den Lehrkräften und Auszubildenden, offenbar gelungen ist, weitgehend problemhaltige Prüfungsaufgaben zu entwickeln.

Schlagworte: kaufmännische Problemlöseaufgaben, kaufmännische Problemlösekompetenz, kognitive Validierung, Methode des lauten Denkens

Abstract

Professional requirements are subject to significant changes. In particular, commercial employees are expected to be able to solve complex problems. While this goal has long been reflected in curricula (curricular level), the corresponding measurement of goal achievement in final exams has so far been inadequately implemented (Wuttke et al., 2022). This may be due to the fact that the promotion of problem-solving competence is (still) not sufficiently implemented in education by teachers and trainers (instructional level) and therefore does not play a role in final exams (assessment level). It may also be related to the fact that guidelines for designing examination tasks are strongly focused on measuring vocational action competence, while aspects of problem-solving competence are not explicitly addressed. In addition, the requirements for measuring vocational problem-solving competence are highly complex. Furthermore, the promotion and measurement of problem-solving competence still play a subordinate or no role in the subject-didactic and diagnostic training of teachers and in-company trainers. To answer the research question, a think-aloud study was conducted with 18 industrial clerks and eight office clerks as part of the joint project TeKoP. A total of 16 tasks were used. The analyses show that the trainees went through the steps of a problem-solving process. They had to consider a variety of factors simultaneously and in some cases weigh and prioritize goals. During processing, the trainees applied heuristics from similar cases, formed analogies, checked the (preliminary) outcomes of their actions, and in some cases revised them. These findings from the think-aloud study indicate that the training participants – teachers and trainers – were largely successful in developing problem-based examination tasks.

Keywords: Commercial Problem-Solving Tasks, Commercial Problem-Solving Competence, Cognitive Validation, Think-Aloud Method

1 Einleitung

Berufliche Anforderungen unterliegen gesellschaftlichen, wirtschaftlichen und technologischen Veränderungen. In den letzten Jahren zeichnen sich besonders starke Auswirkungen globaler Entwicklungen und technologischer Veränderungen wie der digitalen Durchdringung der Arbeitswelt ab. Diese führen zu weitreichenden Veränderungen in der Nachfrage nach Berufen, in den Berufsprofilen und Anforderungen an die Beschäftigten. Auch in kaufmännischen Berufen werden Routinetätigkeiten zusehends durch digitale Technologien übernommen. Für die kaufmännischen Beschäftig-

ten verlagert sich die Tätigkeit hin zu anspruchsvolleren Aufgaben im Bereich der Kommunikation mit Kund:innen und Geschäftspartner:innen, in der Bearbeitung von Sonder- und Problemfällen, in zunehmend komplexer Datenauswertung und in der Vorbereitung kaufmännischer Entscheidungen (Geiser et al., 2021).

Daher wird von kaufmännischen Angestellten erwartet, dass sie auch komplexe Probleme erfolgreich bearbeiten können (Seifried et al., 2016). Während sich dieses Ziel schon länger in Curricula widerspiegelt (curriculare Ebene), ist die entsprechende Messung der Zielerreichung in Abschlussprüfungen bislang nur unzureichend umgesetzt, wie Analysen entsprechender Prüfungsaufgaben zeigen (Wuttke et al., 2022). Dies kann daran liegen, dass die Förderung von Problemlösekompetenz in der Ausbildung durch Lehrkräfte und Auszubildende (noch) nicht hinreichend implementiert ist (instruktionale Ebene) und daher auch in Abschlussprüfungen (Assessment-Ebene) (noch) keine Rolle spielt. Es könnte aber auch sein, dass dafür notwendige fachdidaktische und diagnostische Kompetenzen des Bildungspersonals über die Aus- und Fortbildung bisher nicht hinreichend gefördert wurden, sodass kein Problembewusstsein und/oder keine hinreichenden Kompetenzen für die Konstruktion entsprechender Handlungsanforderungen in Lern- und Prüfungsaufgaben vorhanden sind. Deshalb wurde ein Training für Lehrende und Auszubildende zum Erstellen problemhaltiger technologiebasierter Aufgaben entwickelt, durchgeführt und evaluiert (Wuttke et al., 2024).

Die Entwicklung problemhaltiger Aufgaben für den beruflichen Ausbildungs- und Prüfungskontext ist dabei mit einigen Herausforderungen verbunden: Um Problemlösekompetenz valide abzubilden, werden authentische berufstypische Handlungssituationen in den jeweiligen kaufmännischen Geschäftsfeldern in Form von Problemsituationen als Aufgabenstimuli konstruiert. Computerbasierte Szenarien erscheinen dabei besonders gut geeignet, um Authentizität, aber auch Komplexität beruflicher Handlungssituationen abzubilden (vgl. die Beiträge in Beck et al., 2016; Abele et al., 2021). Orientiert an beruflichen Handlungsrealitäten werden berufstypische Anforderungssituationen situiert über digitale Medien abgebildet. Dabei kann computerbasiert ein kaufmännischer Arbeitsplatz mit Zugang zu entsprechenden Werkzeugen (z. B. Software) und Dokumenten nachgebildet werden (Winther, 2010). Mithilfe von Videovignetten und/oder virtuellen Agenten können reale Situationen authentisch simuliert werden, wobei der Grad der Problemhaltigkeit über die Situation als auch über die nutzenden Informationen gesteuert werden kann. Allerdings kommen die Vorteile solcher technologiebasierten Aufgaben nur dann zum Tragen, wenn auch entsprechende mediale Gestaltungsprinzipien berücksichtigt werden (Wuttke et al., 2023, S. 7 f.).

Zugleich sind computerbasierte Assessments mit komplexen Aufgaben- und Problemstellungen herausfordernd, da es bei komplexen Konstrukten wie der kaufmännischen Problemlösekompetenz, die sich auf verschiedene wirtschaftliche Geschäftsbereiche und Handlungsfelder bezieht, nicht immer gelingt, eine hohe Messgenauigkeit abzusichern (z. B. für das Konstrukt der betriebswirtschaftlichen Nachhaltigkeitskompetenz, Seeber et al., 2019). Zugleich sind Aufgabenstellende wie Lehrkräfte und Aus-

bildende nicht immer per se vertraut mit der technologiegestützten Entwicklung und dem Einsatz entsprechender simulativer Testaufgaben (zur Medienkompetenz von Lehrenden und Auszubildenden Meiners et al., 2022). Insofern galt es, in dem Training Lehrende und Auszubildende nicht nur im Bereich der Entwicklung problemorientierter Aufgaben und in der Diagnostik von Problemlösekompetenz zu schulen (Aufgaben- und Diagnosekompetenz), sondern auch die Förderung von (digitaler) Medienkompetenz gleichsam in das Training einzubinden, damit die zu entwickelnden Problemlöseaufgaben eine hohe Authentizität aufwiesen und computerbasiert konstruiert und eingesetzt werden konnten. Dementsprechend verband das Training verschiedene thematische Module², die stets aus Einführungs- und Erarbeitungsphasen sowie aus Phasen der Anwendung, Reflexion und des Transfers bestanden (ein Gesamtüberblick zu den Trainingsbausteinen Wuttke et al., 2024).

Um nun zu prüfen, ob die im Training von den Teilnehmenden entwickelten Aufgaben tatsächlich „Probleme“ für Auszubildende darstellen, also valide mit Blick auf das zu messende Konstrukt sind, wurden ausgewählte Aufgaben mithilfe der Methode des lauten Denkens bei Auszubildenden überprüft. Zentrale Fragestellung des Beitrags ist daher, ob die im Training durch die Teilnehmenden entwickelten Aufgaben Merkmale eines Problemlöseprozesses berücksichtigen, welche sich in den kognitiven Prozessen der Auszubildenden während der Aufgabenbearbeitung widerspiegeln.

2 Theoretische Grundlagen

Das Training adressierte Lehrende und Auszubildende, die Erfahrungen in der Mitwirkung an den paritätisch besetzten Prüfungsausschüssen haben. Dies betrifft Personen, die entweder bereits als Aufgabenerstellende Aufgabenvorschläge bei den Kammern eingereicht oder an mündlichen Prüfungen in Prüfungsausschüssen mitgewirkt haben bzw. aktuell beteiligt sind, oder Lehrkräfte und Auszubildende, die sich auf eine solche Tätigkeit mit dem Training vorbereiten möchten. Für die theoretische Fundierung der Schulung wurde auf die (berufliche) Problemlöseforschung zurückgegriffen. Ein zentrales Kennzeichen für ein Problem, vor dem eine Person steht, ist das Auftreten einer Barriere, die sich zwischen dem (unerwünschten) Anfangszustand und dem erwünschten Endzustand befindet (Wuttke et al., 2023, S. 6–7). Barrieren bei der Bewältigung beruflicher Probleme treten vor allem dann auf, wenn die individuellen Voraussetzungen für die aktuell vorfindlichen beruflichen Anforderungen nicht gegeben sind (Nickolaus, 2016, S. 167) und wenn keine Handlungsschemata und Routinen verfügbar sind, um die Anforderungen erfolgreich zu bewältigen. Neben der Barriere als grundsätzliches Merkmal zur Klassifizierung eines Problems werden verschiedene Komplexitätsmaße unterschieden, die eine Anforderung problemhaltig machen (Bransford &

2 Das modular aufgebaute Training setzt sich aus folgenden Bausteinen zusammen: Baustein (1) thematisiert die Erstellung problemhaltiger Aufgaben, Baustein (2) vermittelt mediendidaktische Grundlagen, (3) beinhaltet eine Technikschi-
lung (ILIAS und Moodle) und (4) nimmt diagnostische Fragestellungen, insb. Gütekriterien der (Prüfungs-)Aufgaben, in
den Blick.

Stein, 1993; Dörner, 1987; Funke, 2003, 2006, 2011; Jonassen & Hung, 2006; Jonassen, 2011; Sembill, 1992; Wuttke et al., 2015; Wuttke & Wolf, 2007). Zu diesen gehören u. a.:

- (1) Die *Klarheit und Eindeutigkeit* des Ausgangszustands und anzustrebenden Endzustands. Diese beinhalten auch die Informationslage und Offenheit bzw. Geschlossenheit von möglichen Lösungen.
- (2) Die (*Eigen*)*Dynamik* einer Situation und deren Auswirkungen auf die Gesamtkonfiguration. Diese ist insbesondere im Falle der Messung dynamischer Problemlösekompetenz zu modellieren (Abele et al., 2012).
- (3) Die *Vernetztheit* der Variablen, die mit erwünschten sowie unerwünschten Neben- und Folgewirkungen einer Problemlösung einhergehen können.
- (4) Zu berücksichtigende *Interessenkonflikte*, denen Auszubildende in den beiden Berufen als künftige kaufmännische Beschäftigte ausgesetzt sein können (z. B. wenig vereinbare Interessen verschiedener Stakeholder oder zwischen Kund:innen oder Lieferant:innen und eigenen betrieblichen Geschäftsinteressen).
- (5) Schließlich stellt sich auch die Frage, ob die Problemlösung und die damit verbundenen Entscheidungen reflektiert und begründet und ggf. Alternativen in Erwägung gezogen oder ermittelt werden müssen.

Je mehr dieser genannten Merkmale auf eine Problemsituation zutreffen und je komplexer die integrierten Variablen und deren Vernetzung sind, desto problemhaltiger sollte eine Aufgabe sein. Die Bearbeitung von Problemen (bzw. problemhaltigen Aufgaben) schließt aufgrund dieser Anforderungen immer den Umgang mit Unsicherheiten und Ungewissheiten und/oder Zielunklarheiten ein.

Studien in beruflichen Kontexten verweisen auf verschiedene Zusammenhänge zwischen Problemlösekompetenz und weiteren kognitiven Kompetenzen sowie nicht-kognitiven Merkmalen, so z. B. auf den Zusammenhang von Wissen in einem beruflichen Handlungsbereich und erfolgreichen Problemlösungen in diesem Bereich (für die diagnostische Problemlösekompetenz in gewerblich-technischen Berufen vgl. Abele et al., 2012; Walker et al., 2016; für Problemlösekompetenz im kaufmännischen Bereich Wuttke et al., 2015). Jonassen (2011) sowie Jonassen und Hung (2006) zeigen zudem jeweils Zusammenhänge zwischen der Arbeitsgedächtniskapazität und Fähigkeiten im schlussfolgernden Denken und dem Problemlöseerfolg auf. Schließlich verweisen Rausch et al. (2015) auf den Einfluss nicht-kognitiver Facetten auf den Problemlöseerfolg (für einen Überblick zum Einfluss kognitiver und nicht-kognitiver Facetten auf eine erfolgreiche Problembearbeitung vgl. Abele, 2023, S. 25–27).

Im Problemlöseprozess selbst werden bestimmte Schritte durchlaufen, in denen verschiedene Facetten der Problemlösekompetenz zum Tragen kommen (Dörner, 1987; Funke, 2011; Seifried et al., 2016; einen Überblick geben Wuttke et al., 2022). In einem ersten Schritt erfolgen die Analyse und Bewertung der Ausgangslage, ggf. die Formulierung von Hypothesen sowie die Abklärung und Formulierung von Zielen. Diese Situationsanalyse umfasst zudem die Informationssuche und -verarbeitung. In einem zweiten Schritt wird die Lösung entwickelt und (in Teilschritten) umgesetzt, die in einem dritten Schritt kontrolliert und reflektiert wird.

3 Methodisches Vorgehen

3.1 Kognitive Interviews

Zur Überprüfung der Problemhaltigkeit der im Training entwickelten und durch die Trainingsteilnehmenden selbst und/oder durch Fachdidaktik-Expert:innen überarbeiteten³ Aufgaben wurden zwei Zugänge genutzt: Zunächst wurden die Aufgaben a priori anhand von Merkmalen, durch die sich problemhaltige Aufgaben auszeichnen, klassifiziert (Wuttke et al., 2022). In einem nächsten Schritt, der im vorliegenden Beitrag im Vordergrund steht, wurden kognitive Interviews, die während der Aufgabenbearbeitung mit Auszubildenden geführt wurden, ausgewertet.

Kognitive Interviews gelten in der empirischen Bildungsforschung als zentrale Möglichkeit, um prozeduralen und dynamischen Merkmalen handlungsleitender Kognitionen auf die Spur zu kommen (Brückner, 2017; Funke & Spering, 2006; Repp, 2022; Sandmann, 2014). Sie spielen eine wichtige Rolle in der kognitiven Validierung von Testaufgaben, um sicherzustellen, ob das mit einem Test zu messende Konstrukt auch hinreichend über die kognitiven Anforderungen der Testaufgaben abgebildet wird. Verhaltensbeobachtungen und die alleinige Beurteilung des kognitiven Anforderungsgehalts der Aufgaben anhand der Lösungshäufigkeiten (auch Teillösungen) oder auch anhand von Bearbeitungszeiten sind wenig zielführend, da keinerlei Aufschlüsse darüber gemacht werden können, wie Proband:innen bei der Bearbeitung vorgegangen sind, welche Prozesse und Strategien sie gewählt haben. Neben der Analyse der Aufgaben zu den Merkmalen der Problemhaltigkeit erwies sich daher ein Zugang über die Verbalisierung der kognitiven Prozesse während der Aufgabenbearbeitung als geeignet. Dabei wurde der Frage nachgegangen, inwiefern kaufmännische Auszubildende bei der Bearbeitung der Aufgaben die Schritte eines Problemlöseprozesses durchlaufen und welche Barrieren im Prozess des Problemlösens auftraten. Auch sollten etwaige Trial-and-Error-Strategien anhand der verbalen Äußerungen erfasst sowie Strategien einer Analogiebildung und des schlussfolgernden Denkens identifiziert werden, die ebenso als Indikatoren für die Problemhaltigkeit der in den Aufgaben modellierten beruflichen Anforderungen betrachtet werden können.

Die kognitiven Interviews fanden in einem Zeitraum von Juni 2021 bis Juni 2022 über die Videokommunikationssoftware Zoom statt, wodurch ein räumlich unbegrenzter Zugang zur Zielgruppe gewährleistet werden konnte. Zusätzlich bot die Software eine integrierte Anwendung zur Aufzeichnung von Gesprächen, die für die Aufnahme und anschließende Transkription der Denkprotokolle hilfreich war. Kaufmännische Auszubildende wurden im Rahmen der kognitiven Interviews aufgefordert, während der Aufgabenbearbeitung, also parallel zum primären Handlungsvollzug, ihre Denkprozesse zu verbalisieren. Die kognitiven Interviews wurden leitfadengestützt von geschulten Testleiter:innen durchgeführt. Bevor die Aufzeichnungen starteten, wurden

3 Die Überarbeitung von ausgewählten Aufgaben unterlag vor allem dem Ziel, Aufgaben zu vervollständigen oder teilweise auch dahingehend zu modifizieren, dass Merkmale von Problemhaltigkeit hinreichend adressiert werden. Diese Anpassungsbedarfe resultierten aus der Bewertung der entwickelten Prüfungsaufgaben anhand des Kriterienrasters zur Beurteilung der Problemhaltigkeit der Aufgaben.

die Auszubildenden über das Ziel der Studie sowie den Ablauf der kognitiven Interviews informiert. Zudem wurde eine Übungsaufgabe vorgeschaltet, um die Auszubildenden mit dem Format vertraut zu machen. Nach der Bearbeitungsphase und einem kurzen Nachgespräch zur Klärung offener Fragen wurden mit einem Fragebogen zusätzliche Informationen über die Auszubildenden erfasst wie soziodemografische und bildungsbiografische Angaben sowie Aspekte zur Ausbildung im Allgemeinen.

Die erhobenen verbalen Daten (Denkprotokolle) wurden anschließend transkribiert und mittels einer gemischt deduktiv-induktiven Herangehensweise einer strukturierenden Inhaltsanalyse nach Mayring (2016) ausgewertet. Das Durchlaufen im Idealfall aller Schritte eines Problemlöseprozesses während der Aufgabenbearbeitung wurde als Indiz dafür gewertet, dass die Aufgabe problemhaltig ist bzw. ihre Bearbeitung Problemlösekompetenz erfordert. Durchlaufen Auszubildende beim Bearbeiten der komplexen Aufgaben die Problemlöseschritte nicht oder nur unvollständig, heißt dies allerdings nicht zwingend, dass die Aufgabe nicht problemhaltig ist. Vielmehr muss berücksichtigt werden, ob und wie der oder die Auszubildende die Lösung erreicht hat, welche Barrieren dabei von den Befragten überwunden werden mussten und welche kognitiven Strategien gewählt wurden, um bei Unsicherheit über den Lösungsweg dennoch zu einer Lösung zu kommen und diese zu reflektieren.

Allerdings kann das ungewohnte Aufgabenformat, insbesondere die Einbettung beruflicher Anforderungen in eine Unternehmenssimulation sowie die geringe Vertrautheit mit technologiebasierten Aufgaben, zusätzliche Barrieren bei der Bearbeitung problemhaltiger Aufgaben erzeugt haben. Umso wichtiger war es, neben dem Durchlaufen der Schritte eines Problemlöseprozesses auch die weiteren, vorn genannten Merkmale der Problemhaltigkeit in die Auswertung der Interviews einzubeziehen. Zusätzlich wurden Störfaktoren und beeinträchtigende Umgebungsbedingungen induktiv am Material herausgearbeitet (z. B. technische Schwierigkeiten, Probleme im Umgang mit der Lernplattform, auf der die Aufgaben eingebettet waren), um diese bei der Interpretation der Daten zu berücksichtigen.

3.2 Stichprobe und Aufgabencharakteristika

Zur Beantwortung der Forschungsfrage wurde eine Laut-Denken-Studie mit 26 kaufmännischen Auszubildenden durchgeführt (vier Industriekaufleute im zweiten Ausbildungsjahr, 14 im dritten Ausbildungsjahr, sieben Kaufleute für Büromanagement im zweiten Ausbildungsjahr, eine im dritten Ausbildungsjahr). Von den Industriekaufleuten waren sechs männlich und 12 weiblich, wobei das Alter zwischen 18 und 22 Jahren lag. Als höchsten allgemeinbildenden Schulabschluss gaben neun angehende Industriekaufleute die allgemeine Hochschulreife, sechs die Fachhochschulreife und drei die mittlere Reife an. Das Alter der angehenden Kaufleute für Büromanagement lag zwischen 19 und 29 Jahren, wobei sechs weibliche und zwei männliche Auszubildende interviewt wurden. Vier Auszubildende verfügten über die allgemeine Hochschulreife als höchsten Schulabschluss, zwei über die Fachhochschulreife und zwei über die mittlere Reife (s. Tabelle 1).

Tabelle 1: Stichprobenbeschreibung

	Industriekaufleute (N = 18)	Kaufleute für Büromanagement (N = 8)
Geschlecht: männlich weiblich	6 12	2 6
Ausbildungsjahr: zweites drittes	4 14	7 1
Durchschnittsalter:	20,44 Jahre	20,88 Jahre
allgemeinbildender Schulabschluss: allgemeine Hochschulreife Fachhochschulreife Mittlere Reife	9 6 3	4 2 2

Eingesetzt wurden 16 problemhaltige Aufgaben zu verschiedenen kaufmännischen Inhaltsbereichen, um auch die Breite beruflicher Anforderungen in kaufmännischen Handlungsfeldern abzubilden (s. Tabelle 2). Bei sechs Aufgaben lag eine hohe inhaltliche und anforderungsbezogene Überschneidung zwischen den beiden Berufen vor. Dadurch wurden die ursprünglich anhand der curricularen Grundlagen für Industriekaufleute ausgearbeiteten Aufgaben in leicht modifizierter Form ebenfalls bei den Kaufleuten für Büromanagement eingesetzt, sodass insgesamt zehn thematische Bereiche mit insgesamt 16 Aufgaben in die kognitiven Interviews einbezogen werden konnten. Im Rahmen der kognitiven Interviews bearbeitete jede Testperson jeweils drei bis vier Aufgaben unter Berücksichtigung der maximalen Bearbeitungszeit von 90 Minuten. Pro Aufgabe liegen zwischen neun und 18 kognitive Interviews vor, wobei die Unterschiede in der Anzahl dem Testheftdesign geschuldet sind.

Tabelle 2: Überblick über die Themen der in der Studie eingesetzten Aufgaben

	Kaufmännischer Themen- bzw. Tätigkeitsbereich	IK	KBM
A1	Auftragsbearbeitung: Kund:innenbeschwerde (Beschaffung & Produktion)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
A2	Vergleichsrechnungen Leasing- und Kaufangebote (Beschaffung & Produktion)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
A3	Berechnung und Interpretation von Finanzkennzahlen (Controlling)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
A4	Produktlebenszyklusanalyse (Marketing & Absatz)	<input checked="" type="checkbox"/>	
A5	Marktsegmentierung (Marketing & Absatz)	<input checked="" type="checkbox"/>	
A6	Preis-Absatz-Funktion (Marketing & Absatz)	<input checked="" type="checkbox"/>	
A7	Optimale Standortermittlung mittels Nutzwertanalyse (Marketing & Absatz)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
A8	Personalbedarfsermittlung (Personal)	<input checked="" type="checkbox"/>	
A9	(Vor-)Auswahl von Bewerbenden für eine ausgeschriebene Stelle (Personal)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
A10	Personalakquise (Personal)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Anmerkung. IK = Industriekaufleute, KBM = Kaufleute für Büromanagement

3.3 Beurteilung des Problemlösegehalts und kognitive Validierung der Aufgaben

Bevor auf das deduktiv-induktiv erarbeitete Kategoriensystem der Laut-Denken-Studie eingegangen wird, soll zunächst kurz das im Training entwickelte Kriterienraster vorgestellt werden, das in einem ersten Schritt für die Beurteilung und Evaluation des Problemgehalts der Aufgaben genutzt wurde.

Kriterienraster zur Beurteilung der entwickelten problemorientierten Aufgaben

Für die Klassifizierung der Aufgaben wurden Merkmale problemhaltiger Aufgaben (siehe Kapitel 2) aus der einschlägigen Literatur entnommen und zu einem Analyseraster mit neun Kategorien⁴ verdichtet, das sowohl bei der Beurteilung des Problemgehalts von Aufgaben genutzt werden kann als auch eine Orientierung für die Entwicklung problemhaltiger Aufgaben darstellt. Dieses Raster wurde beispielsweise auch zur Analyse des Problemgehalts von Prüfungsaufgaben eingesetzt (Wuttke et al., 2022, S. 37–39), aber auch im Training von den Teilnehmenden genutzt, um problemhaltige Aufgaben zu entwickeln und die entwickelten Aufgaben zu evaluieren.

Die Analyse des Problemgehalts der Aufgaben erfolgte durch zwei Kodierende. In einem ersten Schritt stuften beide Kodierenden jeweils die Aufgaben unabhängig voneinander auf einer zwei- bzw. dreistufigen Skala ein (Wuttke et al., 2022, S. 37–39). Im Anschluss daran erfolgte eine Abstimmung zwischen den kodierenden Personen, um Kriterien hinsichtlich ihrer Beschreibung und Trennschärfe zueinander noch einmal zu konkretisieren. Im Anschluss daran erfolgte ein unabhängiges Rating beider Kodierenden. Auf der Grundlage der Beurteilung der im Training durch die Teilnehmenden entwickelten Aufgaben erfolgte ggf. eine Überarbeitung der Aufgaben, damit diese möglichst umfassend den Anforderungen problemhaltiger Aufgaben entsprechen. Erst diese finalen und überarbeiteten Aufgaben wurden in der Laut-Denken-Studie eingesetzt.

Kategoriensystem zur Auswertung der kognitiven Interviews bezüglich der Problemhaltigkeit der Aufgaben

Für die Auswertung der Laut-Denken-Protokolle wurde zunächst anhand des Forschungsstands zum Problemlösen ein deduktives Kategoriensystem entwickelt, das dann im Zuge der Textanalyse der Protokolle induktiv weiterentwickelt wurde. Zunächst wurden Kategorien nach den aus der einschlägigen Literatur gewonnenen (idealen) Problemlöseschritten (siehe oben sowie Wuttke et al., 2022, S. 33 f.) gebildet. Um ein tiefgreifendes Verständnis für das Durchlaufen von Problemlöseprozessen bei der Aufgabenbearbeitung zu erhalten, wurden diese Hauptkategorien im Anschluss daran durch induktiv gebildete Subkategorien erweitert (s. Tabelle 2). Zur Gewährleistung von trennscharfen Kategorien fanden eine wechselseitige Abstimmung und Optimierung sowie Ergänzungen (z. B. der Beschreibungen oder der Ankerbeispiele) zwi-

4 Das Analyseraster besteht aus folgenden Kategorien: (1) Ist-Analyse, (2) Ziel-Analyse, (3) Abgleich und Bewertung der Informationen, (4) Relevanz der Informationen, (5) Lösungsoffenheit, (6) Dynamik, (7) Vernetztheit, (8) Interessenberücksichtigung und (9) Handlungskontrolle und -reflexion. Die fünfserstgenannten Kategorien können dabei dem Komplexitätsmerkmal der *Klarheit und Eindeutigkeit* zugeordnet werden.

schen den Kodierenden statt. Zur Prüfung der Güte der mithilfe des entwickelten und optimierten Kategoriensystems vorgenommenen Kodierungen wurden ca. 30 % der Denkprotokolle von zwei unabhängigen Kodierenden analysiert. Als Koeffizient der prozentualen Übereinstimmung der Kodierung der unabhängigen Kodierenden wurde auf der Ebene der Hauptkategorien jeweils das Cohens Kappa berechnet (Döring & Bortz, 2016, S. 569). Die Koeffizienten (κ) sowie die Beschreibungen der Haupt- und Subkategorien sind dem untenstehenden Kategoriensystem zu entnehmen (s. Tabelle 3).

Tabelle 3: Kriterienraster zur Analyse der kognitiven Interviews

Hauptkategorie Hauptkategorienbeschreibung	Subkategorie Subkategorienbeschreibung	Ankerbeispiel aus den Denkprotokollen
Problem- und Zielanalyse Erkenntnis, dass eine Problemsituation vorliegt, etwa durch den Abgleich des Ist- und Zielzustands (dies kann präzise oder unpräzise geäußert werden). Ggf. erfolgt eine Feststellung, dass Ziele unklar und zunächst zu klären/abzuwägen sind. Feststellung und Erkennen einer Barriere. ($\kappa = .84$)	Problemidentifizierung Bloße Erkenntnis (direkt oder indirekt geäußert) darüber, dass eine Barriere vorliegt, die es zu überwinden gilt. Identifizierung bzw. Wahrnehmung der übergeordneten Problemstellung.	„Und die Aufgabe war was jetzt genau? (2 sek.) Ah okay. Also, im Prinzip Leasing mit Kaufvergleichen.“ (P18, A2_IK)
	Zielpräzisierung Identifizierung von (Teil-)Zielen zur Überwindung der Barriere. Auswahl und Konkretisierung von Kriterien zur Zielerreichung.	„Also würde jetzt einmal gucken bei den Kriterien, die muss ich jetzt ja dann gewichten.“ (P10, A9_IK)
	Zielpriorisierung Sortierung von Teilzielen entsprechend der Relevanz und Realisierbarkeit. Festlegung einer Reihenfolge der Zielerreichung in Teilen oder vollständig bis zur Problemlösung.	„Okay, also hier steht Gewichtung der Kriterien und da schaue ich jetzt mal in das Dokument mit der Nutzwert-Analyse. (...), dann mache ich das jetzt mal zuerst und dann schaue ich mir die ganzen Dateien an.“ (P7, A5_IK)
(komplexe) Informationsverarbeitung Aufnahme, Einsatz und/oder Verwerfung von Informationen sowie Identifizierung und Bewertung von Informationen aus verschiedenen, ggf. irrelevanten, Quellen (z. B. aus integrierten Anhängen und/oder aus der Analyse der Video- und Audiosequenzen für die Situationsdarstellung). Formulieren, Testen und Bewerten von Hypothesen unter Einbindung fachlichen Wissens. ($\kappa = .75$)	Materialidentifikation Sichten und Identifizieren von zur Verfügung gestellten Dokumenten und Materialien (u. a. Text-, Bild-, Video-, Excel-, Audio-Dateien). Erkennen, dass (weiteres) Material zur Problemlösung herangezogen werden könnte, ohne dieses schon zu bewerten.	„Na dann schauen wir mal, was uns hier erwartet. Eine Bestellung. Okay, noch eine Bestellung. Alles klar. BGB. [...] Ja okay. Hopfenglück auf Rädern. Was ist das? Hm. Das ist Lieferung. [...]. Ah. Dann haben wir eine Excel-Tabelle. Kundenbestellungen. Okay. Gut. Und zwei Lieferscheine.“ (P26, A1_KBM)
	Informationsnutzung Erschließen, Abgleichen, Beurteilen und in den Zusammenhang bringen von Informationen. Abrufen und Verknüpfen von fachlichem Wissen, das zur Problemlösung beitragen könnte.	„Der Lieferschein 379. Okay, aber sie hat das Gleiche auch geliefert bekommen. Nee, da ist nämlich der Lieferschein vertauscht worden. Er hat ja den Lieferschein von ihr. Okay, also wenn ich jetzt dann noch mal gucke, was der andere Lieferschein ist. Ja, okay. Sie hätte

(Fortsetzung Tabelle 3)

Hauptkategorie Hauptkategorienbeschreibung	Subkategorie Subkategorienbeschreibung	Ankerbeispiel aus den Denkprotokollen
		nämlich Lieferschein 378 und das ist das, was Herr Meckerbeck hätte bekommen sollen. (...) Okay. Ja, es wurde nämlich vertauscht.“ (P7, A1_IK) „Okay, gehen wir noch mal in die Analyse. Sonja Axt hat also hier eine 3, hat eine ganz klare 1° (unv.) hier aber auch eine 4 oder vielleicht durch die Erfahrung würde ich hier sogar auch eine 5 geben. [2 sek.] Das war gut, das war auch ziemlich gut. [2 sek.] Aber hier fehlt auch das Bild, muss man sagen.“ (P5, A5_IK)
	Umgang mit irrelevanten Informationen Identifizierung und Benennung von irrelevanten Informationen. Informationen werden als Behinderung des zielgerichteten Problemlösens wahrgenommen. Erkenntnis, dass als irrelevant eingestufte Information(en) nicht weiter betrachtet werden.	„BGB, was will ich hier im BGB gucken. Bürgerliches Gesetzbuch, ja. Die Umsetzung folgender Richtlinien. Ja, ja, was wollt ihr mir hiermit sagen? Ja, das brauche ich hier jetzt nicht.“ (P5, A1_IK) „Hopfenglück auf Rädern. (...) okay. Mitgenommen. Also nicht so interessant für mich.“ (P9, A1_IK)
Lösungsfindung Lösungsentwicklung und -durchführung, in der auch verschiedene Perspektiven und mögliche Neben- und Folgewirkungen betrachtet bzw. berücksichtigt und/oder Gedanken um die Lösungsdarstellung verbalisiert werden. ($\kappa = .80$)		„Das ist okay, insgesamt aber ... [unverständlich], dass wir in der Entwicklung, in der Marktentwicklung die neuesten Trends deutlich verpasst haben.“ (P9, A4_IK)
Problemlösekontrolle Hinterfragen, Bewerten und Kontrollieren absolvierter Schritte zur Problemlösung. ($\kappa = .76$)	Zielerreichung Beurteilung, ob und in welchem Maße erfolgte Schritte bereits zur Erreichung von (Teil-)Zielen beitragen. Erkenntnis darüber, dass ein (Teil-)Ziel als final oder vorübergehend erreicht angesehen wird.	„Ich glaube, vier Jahre (unv.) ist günstiger. Und (4 sek.) Haben wir was übersehen? Ich glaube nicht. (...) Ich glaube das war es. Ich denke, mehr habe ich gerade jetzt nicht zu erarbeiten, ich denke, ich hoffe, dass ich nichts falsch gemacht habe. [lacht] (...) Ich glaube, ich bin fertig.“ (P21, A2_KBM)
	Fehleridentifikation Beurteilung, ob und in welchem Maße erfolgte Schritte fehlerbehaftet sind. Identifizierung, Bewertung und Korrektur von Fehlern.	„So Rangfolge, dann ist das hier meine 1. Meine Nummer 2, Nummer 3, Nummer 4, so wieso ist das hier rot? Verstehe ich nicht. (3 sek.) Habe ich irgendwas falsch gemacht? Wahr-

(Fortsetzung Tabelle 3)

Hauptkategorie Hauptkategorienbeschreibung	Subkategorie Subkategorienbeschreibung	Ankerbeispiel aus den Denkprotokollen
	lernen vor dem Hintergrund von (Teil-)Zielen.	scheinlich muss das hier. (3 sek.) Ah, jetzt passt es.“ (P14, A9_IK)
	(Teil-)Zielabbruch Erkenntnis darüber, dass ein (Teil-)Ziel als nicht zielführend angesehen wird. (Teil-)Zielverfolgungen werden abgebrochen, da angenommen wird, dass sie keinen Mehrwert bei der Problemlösung bringen. Abbruch von Teilzielen oder dem gesamten Problemlöseprozess.	„Ich weiß aber nicht, was der Kapitaleinsatz ist, ich weiß es nicht, ich kann es nicht lösen, ich überspringe die Aufgabe, aber ich würde trotzdem einfach etwas reinschreiben.“ (P23, A2_KBM)
Ergebnisbegründung Formulierung von Begründungen für erreichte (Teil-)Ziele oder Problemlösung. Angaben zur Begründung von Ergebnissen, Entscheidungen und Vorschlägen. ($\kappa = .84$)		„Ihr Anspruch ist berechtigt. Wir können die Reklamation anerkennen. Denn die Prüfung Ihrer Bestellung hat ergeben, dass ein systembasierter Fehler aufgetreten ist.“ (P9, A1_IK)

Anmerkung. Kategoriensystem in Anlehnung an Wuttke et al., 2022, S. 37–39; Heyer, 2024

4 Ergebnisse

4.1 Analysen zum Problemlösegehalt der Aufgaben

Die Analyse der Aufgaben anhand des Kriterienrasters zur Beurteilung von deren Problemhaltigkeit zeigt, dass es den Trainingsteilnehmenden gelungen war, Aufgaben zu konstruieren bzw. Aufgabenvorschläge im Überarbeitungsprozess während des Trainings gemeinsam so weiterzuentwickeln, dass diese in der finalen Fassung eine hohe Problemhaltigkeit aufwiesen. Es wurde in den meisten der entwickelten Prüfungsaufgaben eine Situation geschaffen, die eine Barriere für die Zielgruppe, die Auszubildenden, enthält (siehe Kap. 2). Darüber hinaus wurde Problemhaltigkeit in den entwickelten Aufgaben auch dadurch hergestellt, dass die Ausgangssituation des beruflichen Handlungskontextes eher vage formuliert und/oder das Ziel nicht explizit vorgegeben wurde. Dieses Merkmal wurde in ca. der Hälfte der von den Teilnehmenden des Trainings entwickelten Aufgaben umgesetzt. Darüber hinaus mussten im Bearbeitungsprozess der Aufgaben erforderliche Informationen in betrieblichen Datenbanken bzw. Dokumentationen zum Geschäftsvorfall gesucht, bewertet und extrahiert werden. Auch waren irrelevante Informationen enthalten, die hinsichtlich ihrer Bedeutsamkeit für die Problembearbeitung zu bewerten und auszuschließen waren. Damit sollte einerseits die Problemhaltigkeit der Aufgaben sichergestellt werden, andererseits sollte zugleich aber auch dem Kriterium der Authentizität in Bezug auf

berufliche Handlungssituationen entsprochen werden. Die Dynamik als situative Veränderung innerhalb der Aufgabe mit Auswirkung auf die Gesamtkonfiguration wurde hingegen in keiner Aufgabe modelliert. Dies erklärt sich dadurch, dass das Zielkonstrukt, das mit den (potenziellen) Prüfungsaufgaben gemessen bzw. diagnostiziert werden sollte, nicht die dynamische, sondern die analytische Problemlösekompetenz war. Die dynamische Problemlösekompetenz stellt nochmals deutlich höhere Anforderungen an Testdesign, technologische Lösungen zur Administration der Aufgaben und an die Auswertung der erzeugten Problemlösungen (zum dynamischen Problemlösen in beruflichen Handlungsfeldern z. B. Abele et al., 2012), was deutlich über Ziel und Rahmen des Projekts, in dem die Laut-Denken-Studie eingebettet war, hinausgegangen wäre.

Die von den Teilnehmenden entwickelten Problemlöseaufgaben wurden – wie bereits oben erwähnt – einer Prüfung anhand des Kriterienrasters, welches auch bereits bei der Beurteilung der Problemhaltigkeit von Abschlussprüfungen verwendet wurde (Wuttke et al., 2022), unterzogen. Der Vergleich der im Training entwickelten Prüfungsaufgaben für die beiden betrachteten kaufmännischen Ausbildungsberufe zeigte, dass die Merkmale (mit Ausnahme des Merkmals Handlungskontrolle und Reflexion) zur Beurteilung des Problemlösegehalts bei den Aufgaben für Industriekaufleute etwas stärker ausgeprägt waren. Schwächere Ausprägungen waren für beide Berufe im Hinblick auf die Vernetztheit von Informationen in den entwickelten Prüfungsaufgaben festzustellen. Hier war es den Teilnehmenden des Trainings in etwa der Hälfte der für die Kaufleute für Büromanagement entwickelten Aufgaben und bei ca. zwei Fünfteln der für die Industriekaufleute vorgeschlagenen Aufgaben gelungen, eine mittlere bis hohe Vernetztheit der Informationen zu erzeugen. Die Modellierung der Ausgangssituation in den Aufgaben wurde ebenso mit unterschiedlicher Komplexität umgesetzt: Vergleichsweise klar strukturierte Ausgangssituationen lagen in etwa der Hälfte der Aufgaben für Kaufleute für Büromanagement vor, während die andere Hälfte eine eher komplexe Analyse der Ausgangssituation erforderte. Bei den Aufgaben für Industriekaufleute galt Letzteres für etwa drei Fünftel der Aufgaben. In der Mehrzahl der Aufgaben waren Interessen verschiedener Akteur:innen zu berücksichtigen (z. B. Kund:innen, Lieferant:innen, Interessen des Unternehmens) und gegeneinander abzuwägen. Die meisten Aufgaben waren in Teilen lösungsoffen konstruiert, d. h. es waren prinzipiell unterschiedliche (Teil-)Lösungen denkbar. Detailliertere Einblicke zu den Ausprägungen der Merkmale getrennt nach Ausbildungsberufen sind der Abbildung 1 zu entnehmen.

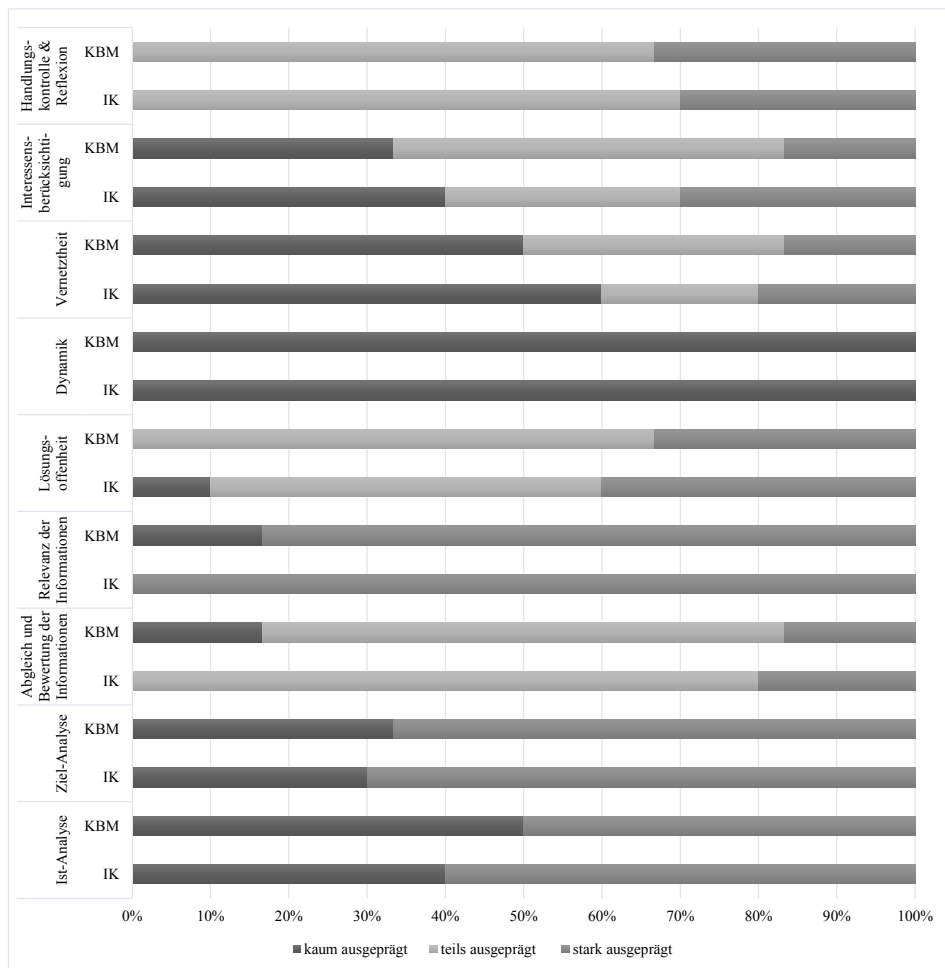


Abbildung 1: Ergebnisse der Aufgabenbeurteilung des Problemlösegehalts

Anmerkung. Die Merkmale der Ist-Analyse, Ziel-Analyse und Relevanz der Informationen wurden dichotom kodiert. IK = Industriekaufleute, KBM = Kaufleute für Büromanagement.

Diese Aufgaben wurden nun in der Laut-Denken-Studie eingesetzt, um zu prüfen, inwiefern die Auszubildenden hier tatsächlich mit problemhaltigen Anforderungen bei der Bearbeitung der Aufgaben konfrontiert sind und inwiefern sie Schritte eines Problemlöseprozesses durchlaufen.

4.2 Ergebnisse der Laut-Denken-Studie zur kognitiven Validierung der problemorientierten Aufgaben

Die Ergebnisse der qualitativen Inhaltsanalyse der Denkprotokolle zeigen, dass die Auszubildenden bei der Bearbeitung der Aufgaben zumeist die für die Bearbeitung der Fälle erforderlichen Problemlöseschritte durchlaufen haben. Dies kann als Indiz interpretiert werden, dass die Aufgaben problemhaltig sind und sich eignen, Problemlösekompetenz bei dieser Zielgruppe zu messen. Dabei ist darauf zu verweisen, dass der Problemlöseprozess nicht zwingend als linear, sondern vielmehr als iterativer Prozessablauf entlang der einzelnen Problemlöseschritte verstanden werden sollte (Rausch et al., 2017, S. 573).

Problem- und Zielanalyse

Die Verbalisierungen der Auszubildenden zur Problem- und Zielanalyse verdeutlichten, dass es sich bei den vorgelegten Aufgaben nicht um Routineaufgaben handelt, für die ein Handlungsschema verfügbar war. Zwar erkannten die Auszubildenden die Thematik der jeweiligen Problemlöseaufgabe, welches durch Aussagen wie beispielsweise „Ok, das habe ich schon einmal gesehen.“ (P23, A2_KBM) verdeutlicht wurde, jedoch wurde nicht direkt eine angemessene Lösungsstrategie verfolgt. Vielmehr untersuchten die Auszubildenden in einem ersten Schritt den Ausgangszustand genauer, um das Problem zu identifizieren. P22 äußerte im Rahmen der Problemidentifizierung z. B. „Ähm, muss ich gerade mal gucken, was genau jetzt gefordert ist. Also, ich weiß gerade nicht, (..), ob da oben irgendwas dann dazu drinsteht.“ (P22, A2_KBM). Es ließen sich Aussagen in den Denkprotokollen finden, die darauf hinweisen, dass die Problemanalyse durch mehrfaches Lesen der Aufgabenstellung gestützt wurde, wie P23 explizit äußerte: „Also, ich muss mir nochmal durchlesen, was jetzt genau ich machen soll.“ (P23, A2_KBM). Auch indirekte Kommentare wie „Und die Aufgabe war was jetzt genau?“ (P18, A2_IK) ließen darauf schließen, dass die Aufgabenstellung nicht sofort erfasst wurde und dass durch ein mehrmaliges Lesen versucht wurde, das Problem einzukreisen und näher zu bestimmen. Die interviewten Auszubildenden strebten dabei an, Hinweise in Materialien zu finden, die Aufschluss über die Zielsetzung der Aufgabe geben.

Je nach Ausgestaltung der Problemsituation innerhalb der Aufgabe (siehe Merkmalsausprägung der Aufgabenbeurteilung in Kap. 4.1) wurden nach der Erkenntnis über das Vorhandensein einer Barriere verschiedene Strategien der Lösungsfindung verfolgt. Auszubildende verbalisierten dafür angestrebte (Teil-)Ziele und priorisierten diese, indem sie eine Rangfolge bildeten. P7 äußerte sich in diesem Zusammenhang wie folgt: „Und jetzt schaue ich mir mal wieder an, was ich als Nächstes machen muss. Ja, genau, ich soll eine Antwort verfassen. Das lese ich jetzt mal, lese ich mir jetzt mal durch. Okay, Frau Axt. Ich gucke jetzt mal, ob ich den Erstbesten nehmen sollte oder die beste Kandidatin.“ (P7, A9_IK). Zu erkennen ist, dass P7 das Vorgehen systematisierte, indem schrittweise Teilziele festgelegt, anhand von Kriterien konkretisiert („Erstbesten“ oder „die beste Kandidatin“) und im Anschluss verfolgt wurden.

(Komplexe) Informationsverarbeitung

Um präzierte und priorisierte Ziele zu verfolgen und Lösungsalternativen abzuwägen, nutzten die Auszubildenden bei der Aufgabenbearbeitung zur Verfügung gestellte Materialien. Einige Auszubildende identifizierten zunächst erst einmal, welche Materialien und Informationen bereitstanden. P5 verbalisierte dies beispielsweise durch diese Aussage: „Gucken wir uns doch mal die Bestellung an, was haben wir denn?“ Aus Verbalisierungen von anderen Auszubildenden ließ sich schließen, dass im Rahmen der Informationssuche simultan eine bewertende Einschätzung der Informationen erfolgte, sodass eine Klassifikation der Informationen bezüglich ihrer Relevanz für den Problemlöseprozess vorgenommen wurde. Dies wurde zum Beispiel in der Äußerung von P3 deutlich: „Das hilft mir jetzt noch nicht so viel weiter, dass es unserer (...) Seite. Okay, das hilft mir jetzt auch nicht viel weiter. Kundenbestellungen. Ah, das ist hier wieder so eine Excel-Datei. Aha, da sehe ich jetzt alle Bestellungen von uns. Okay. Lieferschein, das ist der Eigentliche.“

Die Einbindung fachlichen (Vor-)Wissens ist für den Problemlöseprozess unverzichtbar. Durch Verbalisierung von Auszubildenden ließ sich zum Teil ableiten, dass einige Themen der Aufgaben, insbesondere jene, die mathematische Modellierungen und Berechnungen erforderlich machten, als herausfordernd wahrgenommen wurden. Das Folgende äußerte P14 unmittelbar nach dem Durchsehen der Informationsmaterialien: „Also ich versuche irgendwie die Lösungen hinzukriegen. Also erstmal alles aufmachen. Ach du liebe Zeit, Verfahren zur Investitionsrechnung.“

Zusätzlich griffen Auszubildende bei der Aufgabenbearbeitung auf bereits bekannte Heuristiken und Schemata zurück, was durch Verbalisierungen eines strukturierten Vorgehens ersichtlich wurde. In der Problemlöseaufgabe zur Auftragsbearbeitung einer Kund:innenbeschwerde äußerte P23 in diesem Zusammenhang: „So, und jetzt gucken wir mal, was hier überhaupt falsch gelaufen ist. Äh, ok, also (...) die Bestellung von ihm war die Bestellung 78452. Er hat zwei helle Hopf (unv.) bestellt. [...] Ähm, so das war die Heidi Heidenreich. Die hat das bestellt. Und hat auch das gleiche bezahlt. Ok. Also ist mit der Rechnung schon mal alles super. [...] Das muss ich abgleichen, ob das stimmt. Bestellung 78452 78452. Kundennummer? Auch richtig. Was war das in der Lieferscheinnummer. 378 und 379. Ok, da ist wahrscheinlich was verkehrt gelaufen. [...] Ok, dann ist da ein – Ok. Da ist ein, Vertausch.“ (P23, A1_KBM)

Lösungsfindung

Nicht losgelöst von der zuvor thematisierten komplexen Informationsverarbeitung ist die Entwicklung und Umsetzung von Lösungsansätzen zu verstehen. In Abgrenzung zur Zielformulierung und -priorisierung äußerten die Auszubildenden ihre tatsächliche Planaus- bzw. -durchführung. So erwähnte P23 beispielsweise bei der Durchführung der Nutzwertanalyse zur Standortermittlung (A7) das Folgende: „Ok, ich vergebe nun Gewichtungsfaktoren. Ah, ok. Gewichtet, wie wichtig ich was finde für die Veranstaltung. Dann gebe ich die Punkte dafür und rechne die Gewichtungsfaktoren mal die Punkte und jetzt habe ich den Gesamtwert.“ (P23, A7_IK)

Problemlösekontrolle

Zur schlussendlichen Beurteilung, ob und inwiefern das aufgabenbezogene Problem gelöst wurde, ist das Hinterfragen und Bewerten der durchgeführten Schritte zur Problemlösung nötig. In allen Denkprotokollen der Auszubildenden konnten Verbalisierungen einer solchen Problemlösekontrolle gefunden werden. Während des Notierens des Ergebnisses innerhalb der Aufgabenplattform erwähnte P7: „Und genau jetzt ergibt sich, dass der erstbeste Bewerber meiner Nutzwertanalyse (lacht). Ähm, ja Martin ist. Den Nachnamen kann ich jetzt nicht aussprechen (lacht). Dann kommt Frau Axt und dann kommt Frau Schreiner. Genau, das passt.“ (P7, A9_IK)

Interessant war diese Äußerung auch dahingehend, dass zunächst das Ergebnis des Problemlöseprozesses wiederholt und im Anschluss bewertet wurde. Wurde hingegen festgestellt, dass Fehler aufgetreten sind oder eine Lösung nicht zielführend erschien, nahmen die Auszubildenden Korrekturen vor (z. B. „Aber warum? Nur 50.000 und hier nur 30.000? Äh, das kann irgendwie nicht sein. Irgendwie ist das komisch. Habe ich irgendwas übersehen? Scheint so zu sein.“ (P21, A2_KBM)). Ergänzend dazu erwähnten Auszubildende auch Kontrollschritte, ohne dabei einen inhaltlichen Bezug zur Aufgabe herzustellen, wie P18 in der Aufgabe zur Personalakquise: „Genau, einmal zur Kontrolle. Okay. Ah, das haben wir.“ (P18, A10_IK)

Insbesondere bei der Aufgabe A2 zur Investitionsrechnung haben die Auszubildenden nach wiederholten Versuchen, eine Problemlösung zu erreichen, den Bearbeitungsprozess abgebrochen. Aus den Äußerungen der Auszubildenden wurde ersichtlich, dass der Abbruch auf fehlendes Fachwissen zurückzuführen war (z. B. „Deshalb würde ich die Aufgabe, glaube ich, einfach überspringen, weil ich weiß eh nicht, wie man rechnet.“ (P14, A2_IK); „Ja, ich weiß nicht ganz, was ich jetzt machen soll. Kann ich auch einfach weiterklicken, weil ich die Frage nicht beantworten kann?“ (P23, A2_KBM)).

Ergebnisbegründung

In einem engen Zusammenhang mit der Problemlösekontrolle und -reflexion steht auch die Begründung der Entscheidung, die zur Lösung des Problems geführt hat. Dabei wurden neben der Abwägung von potenziellen Risiken und Chancen auch die Stakeholder-Interessen berücksichtigt. Nachdem im Rahmen der Aufgabe A9 die Beurteilung der Bewerbenden als abgeschlossen eingestuft wurde, äußerte sich P5 wie folgt: „Ja, sie sollte eingeladen werden. Wo haben wir das denn? Gut, weil geeignete Vorkenntnisse, Kontakt mit Kunden vorhanden sind.“ (P5, A9_IK)

Als Begründung erwähnte P5 die Vorkenntnisse und konkretisierte sie mit Blick auf Erfahrungen im Bereich des Kontakts mit Kund:innen. Als ein Beispiel für die Berücksichtigung von Interessen der Stakeholder erwähnte P8, dass eine Person gut ins Team passen würde („Ja, also die Bewerbung an sich fand ich einfach gut. Und ich würde behaupten, dass er gut zum Team passt.“ (P8, A9_KBM)). Dagegen wurden durch P9 mit der nachfolgenden Äußerung auch Bedenken, vor allem Risiken einer Entscheidung aufgrund von Bewerbungsunterlagen, geäußert: „Also ich vertraue nicht,

dass der Lebenslauf real ist. Es ist mein persönliches Gefühl, dass mir sagt, dass das nicht gut ist.“ (P9, A9_IK)

Die Ergebnisse dieser Studie erlauben es zudem, Faktoren zu ermitteln, die den Problemlöseprozess negativ beeinflusst haben und die teilweise erklären, warum Problemlöseprozesse nicht erfolgreich verliefen. Aus den Aussagen der Auszubildenden war beispielsweise zu entnehmen, dass das Thema einer Aufgabe (vermeintlich⁵) noch nicht behandelt worden war (z. B. Finanzierung). Gleichzeitig gab es Teilnehmende, die sich an das Thema erinnerten, aber nicht mehr das erforderliche Wissen parat hatten, um eine Lösung zu finden (z. B. „also ich weiß, dass es dafür glaube ich eine Formel gibt [...] aber [...] ich weiß sie leider nicht mehr.“ (P12, A2_IK)). Auch kam es in Einzelfällen zu technischen Problemen (etwa mit dem Office-Programm Excel), die den Prozess störten. Zudem führte das ungewohnte technologiebasierte Arbeiten bei einzelnen Teilnehmenden bei bestimmten Aufgaben zu zusätzlichen Herausforderungen. Letzteres wurde beispielsweise an den Äußerungen von Person P11 (IK, Nachgespräch) deutlich, die sagte: „Also ich glaube, hätte ich das jetzt in Papierform vor mir gehabt, dann wäre es anders gelaufen. Weil dann hätte ich die Tabelle eher erkannt. [...] Aber das lag jetzt eher so an dem PDF-Format. Das hat man erst öffnen müssen, um zu sehen, was das alles für Anlagen hat.“ Des Weiteren gab es in kleinerem Ausmaß einschränkende Rahmenbedingungen während des Lösungsprozesses. Dazu zählten beispielsweise Ablenkungen im Arbeitsumfeld, die durchaus realistische Arbeitsbedingungen darstellen wie ein Klopfen an der Tür, die aber unter den Bedingungen der Verbalisierung des Lösungsprozesses als deutlich störender empfunden wurden. Auch traten vereinzelt Zeitmanagementprobleme auf.

5 Diskussion

Insgesamt bestätigen die Analysen, dass das Training insofern erfolgreich war, als es den Lehrenden der berufsbildenden Schulen und den Auszubildenden in den Betrieben gelungen war, im Zuge des Trainings problemhaltige Prüfungsaufgaben zu entwickeln. In den kognitiven Interviews bestätigte sich über die Äußerungen der Auszubildenden, dass diese für die Lösungsfindung typische Problemlöseschritte durchlaufen haben, bei denen diverse Barrieren zu überwinden waren. Auch mussten die Auszubildenden bei der Bearbeitung der Probleme Ziele herausarbeiten und priorisieren sowie komplexe Informationen identifizieren, bewerten und verknüpfen.

Die kognitiven Interviews selbst wie auch deren Auswertung unterliegen jedoch auch verschiedenen Limitationen. Im Unterschied zu anderen qualitativen Interviewformaten zeichnet sich die Methode des lauten Denkens dadurch aus, dass mentale und kognitive Prozesse während der Bearbeitung von Aufgaben geäußert werden

5 Eine Prüfung dieser Aussage der Auszubildenden war nicht möglich, da den Teilnehmenden zugesichert wurde, dass die Daten vertraulich behandelt werden. Bei der Themenwahl der Aufgaben fand eine Orientierung an den curricularen Vorgaben statt, wonach davon ausgegangen werden konnte, dass die jeweiligen Lernfelder/Themen von der Zielgruppe bereits behandelt sein müssten. Abweichungen davon sind jedoch denkbar.

(Knoblich & Öllinger, 2006, S. 692). Damit gehen durchaus auch einige Nachteile einher, denn eine solche Verbalisierung während der Bearbeitung kann zur Verlangsamung von kognitiven Prozessen führen, die den Lösungsprozess beeinträchtigen könnte. Auch ergeben sich durch die Äußerungen während der Bearbeitung längere Test- und Belastungszeiten der Auszubildenden, die sich gleichfalls negativ auf die Lösung der Aufgaben bzw. der Probleme auswirken könnten (Fox et al., 2011). Zudem ist nicht auszuschließen, dass die Schritte des Problemlöseprozesses nicht vollständig in den einzelnen Denkprotokollen zu finden sind. Die Auszubildenden könnten beim Problemlösen die Schritte zwar durchlaufen, diese jedoch trotz Aufforderung nicht notwendigerweise auch in vollem Umfang verbalisiert haben. Dies könnte einerseits dem Umstand geschuldet sein, dass Problemlöseaufgaben als so komplex gelten, dass keine vollständige Verbalisierung aller Denkprozesse parallel zur Bearbeitung erfolgen kann. Andererseits könnte ein Erklärungsansatz darin bestehen, dass das laute Denken den kognitiven Denkprozess beeinflusst, sodass Auszubildende unbewusst auf Verbalisierungen verzichteten (Schnell, 2016, S. 29). Demzufolge enthielten die Denkprotokolle teilweise unvollständige bzw. keine Aussagen zu einzelnen oder mehreren Problemlöseschritten.

Auch die Stichprobe ist mit Limitationen verbunden, denn bei der Akquise von kaufmännischen Auszubildenden wurden gezielt ausgewählte Unternehmen im Umfeld der beteiligten Hochschulen kontaktiert, die in beiden Berufen ausbilden. Damit umfasste die Stichprobe besonders engagierte Unternehmen und bildet vermutlich die Bandbreite an Unternehmen, die Industriekaufleute und Kaufleute für Büromanagement ausbilden, nur unzureichend ab. Dies zeigte sich insbesondere an der kleinen Stichprobe bei den KBM, bei denen laut Datenbank DAZUBI (BIBB) auch Auszubildende mit Hauptschulabschluss zu finden sind, die in unserer Stichprobe nicht enthalten waren. Hier kommen vor allem zwei Einflussfaktoren für diese Verzerrung zum Tragen: Es wurden in der Akquise für die Studie überwiegend größere Unternehmen erreicht, die eine hohe Ausbildungsplatznachfrage haben und Bewerber:innen auf Ausbildungsplätze entsprechend ihrer Vorbildung selektieren konnten. Die Statistiken belegen hier, dass Bewerber:innen mit niedrigem Schulabschluss eher in kleinen Unternehmen Ausbildungsmöglichkeiten erhalten, während leistungsstarke Jugendliche leichter Zugang zu den begehrten Ausbildungsplätzen in größeren Unternehmen finden. Darüber hinaus könnten die Unternehmen für die Teilnahme an der Studie auch gezielt Auszubildende vorgeschlagen haben, die sie tendenziell als leistungsstark oder leistungsstärker im Vergleich zu anderen einschätzten.

Die Aufgaben wurden überdies in einer quantitativen Studie eingesetzt. Dabei deuten die Lösungshäufigkeiten der Auszubildenden gleichfalls an, dass es sich um eher anspruchsvolle und problemorientierte Aufgaben handelt. Jedoch stehen interpretative Verknüpfungen beider Studienergebnisse noch aus.

Aufgrund der hier gezeigten Herausforderungen bei der Konstruktion und Auswertung von komplexen Prüfungsaufgaben, die berufliche bzw. kaufmännische Problemlösekompetenz messen, zeichnet sich ab, dass aufgrund des Forschungsstands und der in den Abschlussprüfungen in naher Zukunft umsetzbaren Lösungen eine

Messung dynamischer Problemlösekompetenz noch erheblich entfernt liegt. Dies gilt insbesondere auch mit Blick auf die Rechtssicherheit der Prüfungen. Daher zielten die hier entwickelten und getesteten Prüfungsaufgaben zunächst auf die Messung analytischer und nicht dynamischer Problemlösekompetenz in kaufmännischen Handlungsfeldern ab. Letzteres sind jedoch notwendige Erweiterungen der kaufmännischen Problemlöseforschung, zumal aktuelle technologische Entwicklungen die Möglichkeiten der Messung dynamischer Problemlösekompetenz deutlich erweitern können. Allerdings stellen sich hier auch Herausforderungen in der Bewertung der Lösungen und in der Sicherstellung der Reliabilität des zu messenden Konstrukts. Die Messung dynamischer kaufmännischer Problemlösekompetenz ist sicherlich eine Erweiterung der Forschungsperspektiven auf Problemlösekompetenz, die in künftiger Forschung stärker einzubeziehen wäre.

Literatur

- Abele, S. (2023). *Problemlösekompetenzen in beruflichen Kontexten Resultate aus Lehr-Lern-Prozessen sichtbar machen*. wbv. <https://doi.org/10.3278/9783763973675>
- Abele, S., Deutscher, V. K., Nickolaus, R., Rausch, A., Seeber, S., Sembill, D., Seifried, J., Walker, F., Weyland, U., Wittmann, E., & Wuttke, E. (2021). Potenziale technologiebasierter Kompetenztests der ASCOT-Initiative aus der Perspektive der Curriculum-Instruction-Assessment-Triade. In K. Beck, & F. Oser (Hrsg.), *Resultate und Probleme der Berufsbildungsforschung* (S. 13–42). wbv. <https://dx.doi.org/10.3278/6004788w>
- 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. <https://doi.org/10.1007/s11618-012-0277-9>
- Beck, K., Landenberger, M., & Oser, F. (Hrsg.) (2016). *Technologiebasierte Kompetenzmessung in der beruflichen Bildung. Ergebnisse aus der BMBF-Förderinitiative ASCOT*. wbv.
- Bransford, J., & Stein, B. S. (1993). *The ideal problem solver. A guide for improving thinking, learning, and creativity* (2. Aufl.). Freeman.
- Brückner, S. (2017). *Prozessbezogene Validierung anhand von mentalen Operationen bei der Bearbeitung wirtschaftswissenschaftlicher Testaufgaben*. Verlag Empirische Pädagogik.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation*. Springerverlag. https://doi.org/10.1007/978-3-642-41089-5_5
- Dörner, D. (1987). *Problemlösen als Informationsverarbeitung* (3. Aufl.). Kohlhammer.
- Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological bulletin*, 137(2), 316–344. <https://doi.org/10.1037/a0021663>
- Funke, J. (2003). *Problemlösendes Denken* (Einführungen und Allgemeine Psychologie, 1. Aufl.). Kohlhammer. <https://doi.org/10.17433/978-3-17-022830-6>
- Funke, J. (2006). Komplexes Problemlösen. In J. Funke (Hrsg.), *Denken und Problemlösen* (Bd. 8, S. 375–446). Hogrefe.

- Funke, J. (2011). Problemlösen. In T. Betsch, J. Funke, & H. Plessner (Hrsg.), *Denken – Urteilen, Entscheiden, Problemlösen. Allgemeine Psychologie für Bachelor* (S. 137–199). Springer.
- Funke, J., & Spering, M. (2006). Methoden der Denk- und Problemlöseforschung. In J. Funke (Hrsg.), *Denken und Problemlösen* (S. 647–744). Hogrefe.
- Geiser, P., Seeber, S., Busse, J., Schumann, M., Lange, A., Weber, S., Achtenhagen, F., Hiller, F., Zarnow, S., & Hackenberg, T. (2021). Kompetenzfacetten in digitalisierten kaufmännischen Arbeitsplatzsituationen – Eine vergleichende Perspektive von Auszubildenden und Lehrkräften. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 117, 2021/4, 630–657. <https://doi.org/10.25162/zbw-2021-0024>
- Heyer, K. L. (2024). *Validierung von Problemlöseprozessen bei der Bearbeitung digitaler Aufgaben – Eine Analyse von Think-Aloud-Protokollen mit kaufmännischen Auszubildenden*. [Unveröffentlichte Masterarbeit]. Georg-August-Universität Göttingen.
- Jonassen, D. H., & Hung, W. (2006). Learning to troubleshoot: A new theory-based design architecture. *Educational Psychology Review*, 18(1), 77–114. <https://doi.org/10.1007/s10648-006-9001-8>
- Jonassen, D. H. (2011). Supporting Problem Solving in PBL. *Interdisciplinary Journal of Problem-Based Learning*, 5(2), 95–112. <https://doi.org/10.7771/1541-5015.1256>
- Knoblich, G., & Öllinger, M. (2006). Die Methode des Lauten Denkens. In J. Funke, & P. A. Frensch (Hrsg.), *Handbuch der Allgemeinen Psychologie – Kognition* (S. 691–696). Hogrefe.
- Mayring, P. (2016). *Einführung in die qualitative Sozialforschung* (6. Aufl.). Beltz.
- Meiners, H., Hartmann, P., Niegemann, H., Seeber, S., Wuttke, E., & Schumann, M. (2022). Digitale Medienkompetenz als Voraussetzung für die Erstellung von Prüfungsaufgaben. In S. Schumann, S. Seeber, & S. Abele (Hrsg.), *Digitale Transformation in der Berufsbildung: Konzepte und empirische Befunde* (S. 123–144). wbv. <https://doi.org/10.3278/978376397138>
- Nickolaus, R. (2016). Barrieren bei der Bewältigung berufsfachlicher Aufgaben. Ausgewählte Ergebnisse aus quantitativen und qualitativen Analysen und ihr didaktisches Potential. Editorial. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 112, 2016/2, 167–183.
- Rausch, A., Kögler, K., Frötschl, C., Bergrab, M., & Brandt, S. (2017). Problemlöseprozesse sichtbar machen: Analyse von Logdaten aus einer computerbasierten Bürosimulation. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 113(4), 569–594. <https://doi.org/10.25162/zbw-2017-0024>
- Rausch, A., Seifried, J., Wuttke, E., Kögler, K., Egloffstein, M., Warwas, J., & Wolf, K. (2015). „Und wo bleiben die Emotionen?“ – Zur Berücksichtigung nicht-kognitiver Faktoren bei der Förderung und Diagnose von Problemlösekompetenzen. Schneider Verlag.
- Repp, A. (2022). *Kognitive Prozesse von Studierenden im Nachhaltigkeitsmanagement: Eine Think-Aloud-Studie zur kognitiven Validierung simulationsbasierter Testaufgaben*. Cuvillier Verlag.

- Sandmann, A. (2014). Lautes Denken – die Analyse von Denk-, Lern- und Problemlöseprozessen. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 179–188). Springer. https://doi.org/10.1007/978-3-642-37827-0_15
- Schnell, C. (2016). „Lautes Denken“ als qualitative Methode zur Untersuchung der Validität von Testitems. Erkenntnisse einer Studie zur Diagnose des ökonomischen Fachwissens von Schülerinnen und Schülern der Sekundarstufe I. *Zeitschrift für ökonomische Bildung*, 5, 26–49.
- Seeber, S., Michaelis, C., Repp, A., Hartig, J., Aichele, C., Schumann, M., Anke, J. M., Dierkes, S., & Siepelmeyer, D. (2019). Assessment of Competencies in Sustainability Management: Analyses to the Construct Dimensionality. *Zeitschrift für Pädagogische Psychologie*, 148–158. <https://doi.org/10.1024/1010-0652/a000240>
- Seifried, J., Rausch, A., Kögler, K., Brandt, S., Eigenmann, R., & Schley, T. (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). wbv.
- Sembill, D. (1992). *Problemlösefähigkeit, Handlungskompetenz und emotionale Befindlichkeit. Zielgrößen forschenden Lernens*. Hogrefe.
- 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. <https://doi.org/10.1186/s40461-016-0034-z>
- Winther, E. (2010). *Kompetenzmessung in der beruflichen Bildung*. wbv. <https://dx.doi.org/10.3278/6004148w>
- Wuttke, E., Seeber, S., Meiners, H., Hartmann, P., Turhan, L., Niegemann, H., Schumann, M., & Geiser, C. (2024). Technologiebasiertes und problemorientiertes Prüfen in kaufmännischen Berufen. Konzeption eines Fortbildungskonzeptes für Lehrer*innen und Ausbilder*innen. *HLZ – Herausforderung Lehrer*innenbildung*, 7(1), 259–282. <https://doi.org/10.11576/hlz-6733>
- Wuttke, E., Seeber, S., Geiser, C., & Turhan, L. (2022). Zur Problemhaltigkeit von Aufgaben in kaufmännischen Abschluss- und Zwischenprüfungen – Ergebnisse aus Aufgabenanalysen. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 118(1), 25–52. <https://doi.org/10.25162/zbw-2022-0002>
- Wuttke, E., Seeber, S., Meiners, H., & Turhan, L. (2023). Entwicklung und Evaluation eines Trainingskonzeptes für die Gestaltung problemhaltiger technologiebasierter Lern- und Prüfungsaufgaben. In M. Hommel, C. Aprea, & K. Heirichs (Hrsg.), *Netzwerke – Strukturen von Wissen, Akteuren und Prozessen in der beruflichen Bildung. Digitale Festschrift für Bärbel Fürstenau zum 60. Geburtstag*. bwp@ Profil 8. https://www.bwpat.de/profil8_fuerstenau/wuttke_et_al_profil8.pdf

- Wuttke, E., Seifried, J., Brandt, S., Rausch, A., Sembill, D., Martens, T., & Wolf, K. D. (2015). Modellierung und Messung domänenspezifischer Problemlösekompetenz bei angehenden Industriekaufleuten. Entwicklung eines Testinstruments und erste Befunde zu kognitiven Kompetenzfacetten. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 111(2), 189–207. <https://doi.org/10.25162/zbw-2015-0013>
- Wuttke, E., & Wolf, K. D. (2007). Entwicklung eines Instrumentes zur Erfassung von Problemlösefähigkeit – Ergebnisse einer Pilotstudie. *Europäische Zeitschrift für Berufsbildung*, 2(41), 99–118.

Mit Signature Scenarios zum Kompetenzerwerb: Die Analyse von Performance Tasks in virtuellen Simulationen

ESTHER WINTHER, VIOLA DEUTSCHER, FABIO FORTUNATI, JÜRGEN SEIFRIED & ANDREAS RAUSCH

Zusammenfassung

Im Verbundvorhaben PSA-Sim¹ wurde mit der LUCA Office Simulation eine authentische Büro-Simulation entwickelt, die flexibel einsetzbar ist und neben der Gestaltung von Lehr-Lern-Prozessen auch das Assessment von Lernleistungen ermöglicht. Wesentliches Element des Assessments sind sogenannte Signature Scenarios, die zentrale Inhalte der kaufmännischen Grundausbildung abbilden. Die Modellierungen der Szenarien basieren auf umfassenden Domänenanalysen, und die Szenarien sind in ein Modellunternehmen eingebettet. Die Signature Scenarios weisen eine Bearbeitungsdauer von etwa 30 bis 45 Minuten auf. Die Auszubildenden werden dabei mit authentischen Arbeitsszenarien konfrontiert, für deren Lösung berufliche und berufsübergreifende Kompetenzen der kaufmännischen Domäne benötigt werden, die auch in der realen Handlungssituation zum Erfolg führen würden. Der Beitrag zeigt exemplarisch, wie Auszubildende mit den Signature Scenarios in LUCA arbeiten, welche Leistungen sie dabei zeigen und welche Rückschlüsse mit Blick auf den Kompetenzerwerb und die Förderung von kaufmännischen Kompetenzen möglich sind. Der Beitrag basiert auf IRT-Analysen des Signature Scenarios „Lieferantenauswahl“, das bei 1.748 Auszubildenden in Baden-Württemberg und Nordrhein-Westfalen eingesetzt wurde. Im Folgenden berichten wir über die Befunde zweier Studien: In Studie 1 wird der Problembearbeitungsprozess analysiert und in Studie 2 erfolgt eine individuelle Leistungsdiagnostik. Die Ergebnisse zeigen zweierlei: (1) Die Komplexität der Signature Scenarios stellt vor allem zu Beginn der Ausbildung eine Hürde dar, die durch den Einsatz von (teil-)automatisierten Hilfen, zum Beispiel in Form von Prompts, überwunden werden kann. (2) Die Bewältigung der Anforderungssituation setzt sowohl einen sprachlich-argumentativen als auch einen mathematisch-analytischen Zugang voraus, die sich in einem zweidimensionalen IRT-Modell plausibel abbilden lassen. Die Item- und Personenparameter erlauben eine angemessene Unterscheidung zwischen den Testpersonen, sodass interindividuelle Leistungsunterschiede innerhalb dieser Zugänge analysiert werden können.

¹ Das Verbundprojekt „Problem Solving Analytics in Office Simulations“ (PSA-SIM) ist Teil der Initiative ASCOT+ (siehe www.ascot-vet.net) und wurde durch das Bundesministerium für Bildung und Forschung (BMBF) gefördert (Förderkennzeichen: 21AP008A [Teilprojekt Mannheim] bzw. 21AP008B [Teilprojekt Duisburg-Essen]).

Schlagworte: Kaufmännische Bürosimulation, komplexe berufliche Handlungssituationen, Scoringverfahren, Kompetenzdiagnostik

Abstract

In the project PSA-Sim, the LUCA office simulation was developed as an authentic learning and assessment environment that can be used flexibly and, in addition to supporting teaching and learning processes, also enables the assessment of learning performance. An essential element of the assessment are so-called signature scenarios, which depict the central content of basic commercial training. The Signature Scenarios last around 30 to 45 minutes to complete and are based on a model company. Learners are confronted with authentic problem scenarios that require skills that would also lead to success in a real-life situation. We use an example scenario to show how trainees work with the Signature Scenarios in LUCA, how they perform and what conclusions can be drawn with regard to skills acquisition and the promotion of commercial skills.

The contribution is based on IRT analyses of the signature scenario “Supplier selection”, which was used with 1,748 trainees in Baden-Württemberg and North Rhine-Westphalia. We report on two studies: Study 1 analyzes the problem-solving process and study 2 describes individual performance diagnostics using performance tasks. Two findings are particularly noteworthy: (1) The complexity of the signature scenarios represents a hurdle, especially at the beginning of the training, which can be overcome by using (partially) automated aids, for example in the form of prompts. (2) To meet the requirements, trainees need linguistic-argumentative and mathematical-analytical skills, which can be plausibly depicted in a two-dimensional IRT model. The item and person parameters allow an appropriate differentiation between the test subjects so that inter-individual differences in performance can be analyzed within these approaches.

Keywords: Office Simulation, Complex Professional Situations, Scoring Methods, Competence Diagnostics

1 Die Bürosimulation LUCA als Lern- und Assessmentumgebung

1.1 Ausgangslage

Jüngere Forschungsprogramme auf nationaler und internationaler Ebene verweisen darauf, dass die Digitalisierung der Arbeit in Kombination mit New Work branchenübergreifend zu einem Wandel der Arbeits- und Produktionsprozesse führt (z. B. Gallego & Kurer, 2022; Hasenbeck, 2019; Spöttl & Windelband, 2021). Die aktuellen, insbesondere digital und KI-getriebenen Transformationen haben weitreichende Auswirkungen auf bestehende Organisationsstrukturen und -prozesse sowie auf Kundenbeziehungen und Geschäftsmodelle, die entsprechend der Digitalisierung von Produkten, Dienstleistungen und Prozessen neu ausgerichtet werden (Sczogiel et al., 2019).

Arbeit sowie Arbeitsplätze verändern sich und damit auch die Kompetenzanforderungen als Grundlage für berufsadäquates Handeln (Brötz et al., 2014; Lane & Williams, 2023). In solch einem komplexen Umfeld gewinnen u. a. die Fähigkeit, (auch) komplexere Aufgaben oder Probleme zu bewältigen (z. B. Rausch et al., 2021), aber auch Teamarbeit und Kollaboration (z. B. Schlicht, 2019) an Bedeutung. Eine Forderung in diesem Zusammenhang ist, die Lern- und Assessmentprozesse in der beruflichen Bildung stärker kompetenzorientiert auszurichten (z. B. Anselmann et al., 2022; Klotz & Winther, 2012), Auszubildende zur Auseinandersetzung mit unternehmensspezifischen Entscheidungen in komplexen und authentischen Szenarien anzuregen und somit letztlich auch die Attraktivität der beruflichen Bildung zu steigern (Dohmen et al., 2023). Dabei sollten kompetenzorientierte Lern- und Assessmentprozesse reale Herausforderungen der beruflichen Praxis widerspiegeln und aktuelle Entwicklungen in der Arbeitswelt adäquat berücksichtigen. Vor diesem Hintergrund erscheint es zielführend, reale berufliche Arbeitsaufgaben unter Rückgriff auf (fach-)didaktische Überlegungen in authentische Arbeitsszenarien zu transformieren. Solche Arbeitsszenarien lassen sich – je nach deren Komplexitätsgrad und dem Kompetenzstand der Lernenden – als Metaprobleme (Jonassen, 2000) oder problemhaltige Aufgaben auffassen (Rausch et al., 2021).

Die Übersetzung beruflicher Realität in Lern- und Assessmentprozesse können effektiv über computerbasierte Simulationen erfolgen. Zum einen können simulationsbasiertes Lernen und Prüfen den Kompetenzerwerb in vielfältiger Weise fördern. Die Spannweite reicht hier von der Aneignung fachspezifischen Wissens und Routinen über die Entwicklung von Problemlösungs-, kritischen Denk- und selbstregulatorischen Fähigkeiten bis hin zur Förderung von Einstellungen, Interessen und Selbstwirksamkeit sowie deren Diagnostik in Bezug auf die jeweilige Domäne (z. B. Chernikova et al., 2020; 2024). Zum anderen ermöglichen computerbasierte (Arbeitsplatz-) Simulationen ein hohes Maß an Authentizität hinsichtlich verschiedener Umgebungs- und Arbeitsmerkmale, sodass reale Anforderungen für Lern- und Assessmentprozesse hinreichend gut abgebildet werden können. Sie gelten daher als geeignete Annäherungen an die berufliche Realität (Codreanu et al., 2020; Duchatelet et al., 2022). Vor dem skizzierten Hintergrund wurde für die kaufmännische Domäne im Projekt PSA-Sim die Bürosimulation LUCA als Lern- und Assessmentumgebung entwickelt. LUCA ist eine webbasierte Umgebung, die authentische und individualisierte Lern- und Assessmentprozesse ermöglicht (Deutscher et al., 2022; Gentner et al., 2024; Rausch et al., 2021; Seifried et al., 2021). Nachfolgend werden die unterschiedlichen Herausforderungen zur Gestaltung kompetenzorientierter Lern- und Assessmentprozesse aus theoretischer Sicht beschrieben, um daraus dann die empirischen Studien begründen zu können. Beide hier exemplarisch berichteten Studien beziehen sich auf das Signature Szenario „Lieferantenauswahl“, da dieses als Querschnittsaufgabe in zahlreichen kaufmännischen Rahmenlehrplänen verankert ist (Deutscher et al., 2022).

1.2 Designprinzipien für die Gestaltung der Signature Scenarios

Der Kompetenzerwerb kaufmännischer Auszubildender ist an Lerngelegenheiten in Schule sowie am Arbeitsplatz gebunden und abhängig davon, wie diese Gelegenheiten instruktional und über Assessmentprozesse begleitet werden (Deutscher & Winther, 2018; Seifried et al., 2021). Im Rahmen des Projekts PSA-Sim wurden verschiedene sogenannte Signature Scenarios entwickelt, die berufsfachliche Kompetenz entlang des Modells beruflicher Handlungskompetenz erfassen (Rausch et al., 2021). Handlungskompetenzen umfassen im vorliegenden Fall die selbstständige und fachkundige Suche und Verarbeitung von Informationen während der Planungs- und Anwendungsphase, ein hierauf basiertes, begründetes Treffen beruflicher Entscheidungen sowie die Kommunikation beruflicher Entscheidungen und Reflexionsprozesse. Die entwickelten Signature Scenarios sollen exemplarisch die didaktischen Möglichkeiten der Bürosimulation aufzeigen und Lehrkräften die Möglichkeit eröffnen, die Lernumgebung unmittelbar zu nutzen (Deutscher et al., 2022; Gentner et al., 2024; Rausch et al., 2021; Seifried et al., 2021). Sie basieren auf umfassenden Domänenanalysen und sind zudem das Ergebnis eines Modellierungsprozesses, der sich an den Prinzipien des *Four-Components-Instructional-Design-Ansatzes* (4C-ID) nach Kirschner und van Merriënboer (2018) orientiert. Die Grundidee des 4C-ID-Ansatzes besteht darin, dass sich komplexe Lehr-Lern-Arrangements über vier Komponenten beschreiben lassen: (a) *Learning tasks* sind klar beschriebene und auf spezifische Lehr-Lernziele abgestimmte Problemszenarien. Sie bieten eine Auseinandersetzung mit authentischen und komplexen Fällen, die auf den Erwerb beruflicher Kompetenzen abzielen. (b) *Supportive information* adressiert die (potenziell) problemhaltigen Aspekte eines Szenarios. Diese Materialien bieten Lernenden die Möglichkeit, sich das für die Problemlösung notwendige Wissen selbstständig anzueignen. Sie verdeutlichen, wie Domänenexpert:innen die Probleme lösen, und stehen den Lernenden während des gesamten Problemszenarios zur Verfügung. (c) *Procedural information* bezieht sich auf die Routineaspekte eines Problemszenarios wie bspw. Rechenalgorithmen, die für die Problembearbeitung vorausgesetzt werden (bspw. Währungsrechnen im Rahmen einer Lieferantenauswahl). Diese Informationen werden idealerweise zielgerichtet bereitgestellt, wenn Lernende sie tatsächlich benötigen. (d) *Part-task practice* schließlich meint die Bereitstellung kleinschrittiger Übungen zu klar abgrenzbaren Handlungsschritten im Rahmen der komplexen Problemlösung (bspw. die Bereitstellung einer Reihe von kleineren Aufgaben zum Währungsrechnen). Entsprechende Aufgaben können angeboten werden, um notwendiges Vorwissen zu aktivieren oder in der Problembearbeitung neu erworbenes Wissen zu routinisieren. Während leistungsstärkere Lernende von problembasierten Ansätzen profitieren, benötigen insbesondere schwächere Lernende gezielte Unterstützung („*Scaffolding*“) bei der Bewältigung von Lernhürden (Kirschner et al., 2006; siehe auch „*Aptitude-Treatment-Interaction*“). Entsprechend der hier skizzierten Designüberlegungen ist die LUCA Office Simulation als adaptives Lernsystem konzipiert, das durch die Darbietung kognitiver, non-kognitiver und metakognitiver Lernprompts die Lernaktivitäten zielgerichtet und adaptiv unterstützt (Deutscher et al., 2022). Prompts stellen gezielte Interventionen dar, um Lernende zur Reflexion oder

zur Produktion in Bezug auf einen spezifischen Lerngegenstand anzuregen (Devolder, 2012; Pintrich, 2000).

In der Simulation werden die Auszubildenden mit authentischen Problemszenarien konfrontiert, für deren Lösung berufliche Handlungskompetenzen benötigt werden, die auch in der realen Handlungssituation eingesetzt werden. Die Signature Scenarios weisen eine Bearbeitungsdauer von etwa 30 bis 45 Minuten auf und basieren auf einem Modellunternehmen (in unserem Fall einem Fahrradhersteller). Ein Szenario umfasst dabei ein Bündel an Dokumenten (i. d. R. Geschäftsdokumente wie Briefe, Rechnungen, Aufträge etc.), die mit für kaufmännische Arbeitsplätze üblichen Software- und Kommunikationstools wie einem E-Mail-Client, einem Tabellenkalkulations- und Textverarbeitungsprogramm mit jeweils grundlegenden Funktionen, einer Enterprise-Resource-Planning-(ERP)-Software mit Recherchefunktionen (read only), einem Notizblock sowie einem Taschenrechner bearbeitet werden. Ein Szenario kann zusätzliche Hilfen in Form optionaler E-Mails (mit Anhängen) enthalten, die den Teilnehmenden während der Bearbeitung nur dann angezeigt werden, wenn bestimmte, im Vorfeld definierbare Bedingungen erfüllt sind (bspw. wenn ein relevanter Datensatz im ERP-System nach zehn Minuten noch nicht geöffnet wurde oder ein erwartetes Zwischenergebnis in der Tabellenkalkulation nach 15 Minuten noch nicht enthalten ist). Die Abbildungen 1 und 2 geben einen Einblick in die Funktionalitäten der LUCA Office Simulation sowie die für die Bearbeitung eines Szenarios bereitgestellten Informationen und Dokumente.

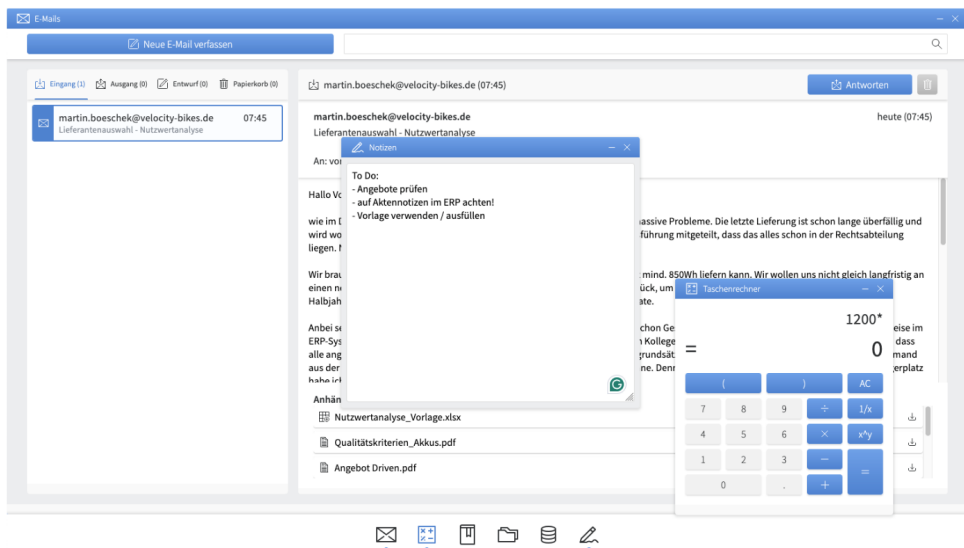


Abbildung 1: User Interface der LUCA Office Simulation

The screenshot displays a Microsoft Excel spreadsheet titled "Nutzwertanalyse_Vorlage.xlsx". The spreadsheet is used for a supplier selection analysis. It features a table with columns for various criteria and their weighted scores. The criteria include "Auswahlkriterium", "Gewichtung", "Angebotswert", "Punkte", "gewichtete Punkte", "Angebotswert", "Punkte", "gewichtete Punkte", "Angebotswert", "Punkte", "gewichtete Punkte", and "Angebotswert". The table compares three suppliers: JINSHU GONGSI LTD., VeloCity Bikes SE, and Steiger & Söhne. The analysis is based on a weighted scoring system where "gut" (good) is 3, "mittelmäßig" (average) is 2, "schwach" (weak) is 1, and "unannehmbar" (unacceptable) is 0. The spreadsheet also shows a table of offers from different suppliers, including details like "Angebot vom", "Angebotswert", and "Lieferzeit".

Abbildung 2: Beispieldokumente des Signature Scenarios/der Performance Tasks „Lieferantenauswahl“

1.3 Konstruktion beruflicher Assessments

Im Rahmen von Assessmentprozessen dienen die Signature Scenarios als Performance Tasks (Lane & Stone, 2006). Performance Tasks sind komplexe, zusammenhängende Aufgaben, die es den Lernenden ermöglichen, ihre Leistungsfähigkeiten (und deren spezifische Facetten) auf verschiedene Weise zu demonstrieren. Hierfür weisen Performance Tasks vielfältige Formate auf (Argumentationen, Visualisierung etc.); in der beruflichen Bildung stellen sie oft auf das Erzeugen von Arbeitsprodukten (Kalkulationen, Charts etc.) ab. Bei der Konstruktion von Performance Tasks ist eine Orientierung am GRASPS-Modell (Wiggins & McTighe, 2005) zielführend. Das Akronym GRASPS steht für *Goal* (eine Aufgabe muss das zu lösende Problem oder die Herausforderung benennen), *Role* (eine Aufgabe muss erklären, welche Perspektive die Lernenden im Assessment-Szenario einnehmen und was von ihnen erwartet wird), *Audience* (eine Aufgabe muss angeben, für wen die Lernenden das Problem lösen und wen sie von der Gültigkeit und dem Erfolg ihrer Lösung für das Problem überzeugen müssen), *Situation* (eine Aufgabe muss den Kontext beschreiben sowie kontextbeeinflussende Faktoren enthalten, die die Lösung des Problems erschweren), *Product, Performance and Purpose* (eine Aufgabe muss auf ein zu erstellendes Produkt oder eine spezifische Leistung sowie auf einen übergeordneten Zweck abstellen) sowie *Stan-*

dards for Success (eine Aufgabe muss die Standards vorgeben, die erfüllt werden müssen, um die Aufgabe erfolgreich lösen zu können).

Ein zentraler Aspekt bei der Konstruktion (beruflicher) Assessments besteht darin, das Validitätsparadoxon zwischen hoher Inhaltsvalidität einerseits und der Messung konstrukt-irrelevanter Merkmale andererseits (bspw. Sprachkompetenz, Komplexität der Testumgebung, Kontextualisierung, Cognitive Load etc.) zu reduzieren (Minkley et al., 2021; Mislevy, 2019). Diese Problematik wird dadurch verstärkt, dass mittels Performance Tasks zumeist inhaltlich eng definierte Konstruktausschnitte erhoben werden (Seifried et al., 2021). Im Projekt wurde versucht, dieser Schwierigkeit dadurch zu begegnen, dass die Performance Tasks unterschiedliche didaktische Zugänge adressieren, um die Konstruktbreite – im Sinne einer möglichen Vergleichbarkeit von Leistungen über verschiedene Anforderungssituationen – des benötigten Fähigkeitsspektrums zur Anforderungsbewältigung zu erhöhen.

Für das Assessment wurde die Performance Task „Lieferantenauswahl“ ausgewählt, um verschiedene Facetten des Lernens, des Arbeitens und der Diagnostik mittels der LUCA Office Simulation zu beschreiben. Das Thema der Lieferantenauswahl hat eine hohe curriculare und praktische Relevanz im Bereich der kaufmännischen Berufsausbildung. Inhaltlich lässt sich das Szenario in den Wertschöpfungsprozessen von Unternehmen und hier insbesondere im Bereich des Einkaufs verorten. Dabei sind die Auszubildenden aufgefordert, aus mehreren Angeboten einen geeigneten Lieferanten auszuwählen und diese Auswahl zu begründen. Die Testpersonen sichten zunächst verschiedene Angebote und erstellen anschließend eine Nutzwertanalyse zur Bewertung der Angebote, auf deren Basis sie eine schriftliche Begründung ihrer Entscheidung an eine fiktive vorgesetzte Person formulieren.

1.4 Zwischenfazit

Vor dem Hintergrund der obigen Ausführungen lässt sich Folgendes festhalten:

- (1) Das Signature Szenario „Lieferantenauswahl“ als Lernumgebung berücksichtigt die Aspekte des 4C-ID-Ansatzes wie folgt:
 - Berufliches Lernen findet weitgehend in realen Situationen statt – nämlich in Geschäftsprozessen, an denen Dritte beteiligt sind und deren Lösungen für Dritte von Relevanz sind: *learning tasks u. a. mittels Unternehmenssimulationen*.
 - Berufliches Lernen ist komplex, multiperspektivisch und sozial; es setzt an authentischen Informationen, Materialien und Interaktionen an: *supportive information u. a. durch Erfahrungsberichte und Expert:inneneinschätzungen*.
 - Berufliches Lernen erfolgt entlang praktischer Beispiele, die immer auch allgemeine Kompetenzen und Heuristiken adressieren, die in den Aufgaben aktualisiert werden (können): *procedural information u. a. durch Lösungsalgorithmen und Vorlagen*.
 - Berufliches Lernen benötigt Wiederholung und Anleitung durch Expert:innen, um sowohl die kognitiven als auch sozialen Teilaspekte beruflicher Aktivität zu stärken: *part-task practice u. a. durch computergestützte Agenten*.

- (2) Die Performance Task „Lieferantenauswahl“ basiert auf einem Bündel von Test-items und berücksichtigt die Aspekte des GRASPS-Modells wie folgt:
- Das Ziel (*Goal*) der Performance Tasks besteht darin, dass die Lernenden
 - in ihre Rolle (*Role*) als Mitarbeitende im Einkauf
 - dem Abteilungsleiter (*Audience*) einen Vorschlag für den bestmöglichen Lieferanten von Fahrradakkus mit spezifischer Leistungsanforderung präsentieren und über Abteilungsgrenzen hinweg diskutieren,
 - sodass er eine unternehmerisch sinnvolle Entscheidung (*Situation*) treffen kann.
 - Das Endergebnis (*Product*) ist eine kommentierte Nutzwertanalyse (unter Verwendung gängiger Office-Software),
 - die erstens alle relevanten Einflussfaktoren der Lieferantenauswahl in die Kalkulation mit einbezieht, die zweitens weiterführende Informationen, die von Kolleg:innen zur Verfügung gestellt werden, berücksichtigt und die drittens in den Kommentierungen auf verschiedene Abteilungsperspektiven eingeht und die getroffene Entscheidung vor diesem Hintergrund auch begründet (*Standards of Success*).

2 Ausgewählte Teilstudien

2.1 Stichprobe und Assessmentdesign der beiden Teilstudien

Im Rahmen des Projekts wurden mehrere Teilstudien durchgeführt, die vor allem das Arbeiten mit der LUCA-Umgebung und die Möglichkeiten der individuellen Leistungsdiagnostik in den Blick genommen haben. Um diese Aspekte empirisch zu untersuchen, wurde die „Lieferantenauswahl“ im Querschnittsdesign eingesetzt (s. Abbildung 3). Insgesamt nahmen 1.748 Auszubildende an 39 kaufmännischen Berufsschulen aus Baden-Württemberg und Nordrhein-Westfalen teil. Davon waren 914 Testpersonen Industriekaufleute, 816 Bürokaufleute und 18 besuchten das Wirtschaftsgymnasium. Die Auswahl der Berufsschulen erfolgte randomisiert. Die Stichprobe repräsentiert kaufmännische Auszubildende gut – das Alter beträgt im Mittel 21,4 Jahre ($SD = 3,32$) und 48 Prozent der Testpersonen sind männlich. Aus dieser Gesamtstichprobe wurden für die unterschiedlichen Forschungsfragen Teilstichproben gezogen. Das Assessmentdesign ist so angelegt, dass (1) Vergleiche zwischen unterschiedlichen Testumgebungen und -modi sowie (2) die Testung von Messinvarianz sowie verschiedene Prüfungen der Konstruktvalidität möglich sind.

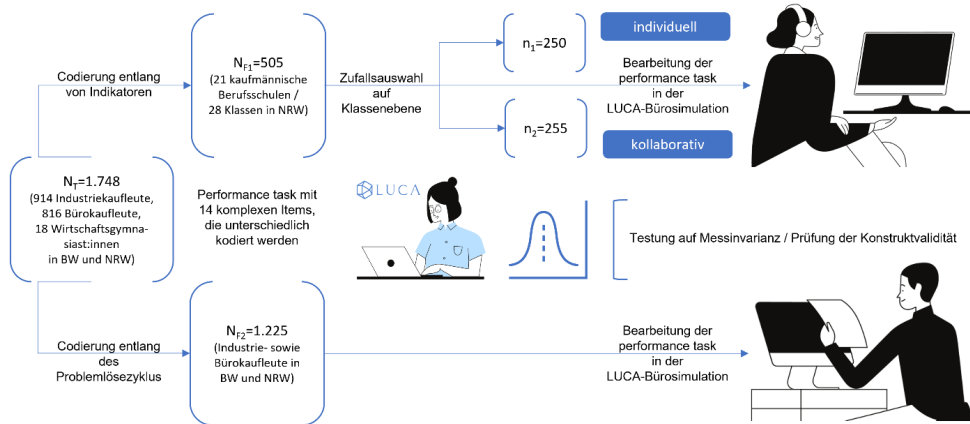


Abbildung 3: Assessmentdesign

2.2 Teilstudie 1: Mit der LUCA Office Simulation (kollaborativ) arbeiten

Ein Fokus im Projekt wurde auf die Bedien- und Nutzbarkeit der LUCA Office Simulation gelegt. Hierbei ging es vor allem darum, komplexes, kollaboratives Problemlösen empirisch zu beschreiben. Nachfolgend wird eine Studie vorgestellt, die folgende Forschungsfrage beantwortet:

Inwiefern hängt die Qualität der Aufgabenbearbeitung (kognitive und soziale Teilleistungen) davon ab, ob Auszubildende die Anforderungssituation allein oder unterstützt durch automatisierte Hilfestellungen bearbeiten?

Zur Beantwortung der Forschungsfrage wurden zwei Testumgebungen konstruiert: In der ersten Testumgebung ist die „Lieferantenauswahl“ von den Auszubildenden allein zu bewältigen; in der zweiten Testumgebung werden die Auszubildenden im Bearbeitungsprozess durch Hilfestellungen, die Kolleg:innen oder die Ausbildungsleitung repräsentieren, automatisiert unterstützt. Die automatisierten Hilfestellungen zielen sowohl auf die kognitiven als auch sozialen Teilleistungen der Aufgabenbearbeitung ab. So können die Testteilnehmenden durch eine simulierte Kontaktaufnahme eines Kolleg:innen mit diesem interagieren. Je nach gewählter Antwortoption werden unterschiedlich detaillierte Prompts dargeboten (s. Abbildung 4).

Es ist vor diesem Hintergrund von Interesse, welchen Einfluss die Gestaltung der Testumgebung auf die kognitiven und sozialen Teilleistungen der Auszubildenden ausübt. Um diese Frage zu beantworten, wurde die „Lieferantenauswahl“ bei einer Teilstichprobe von $N_{FI}=505$ Industriekaufleuten eingesetzt. 255 Testpersonen haben die Testumgebung mit (teil-)automatisierten Hilfestellungen und 250 Proband:innen das individuelle Assessment durchlaufen (vgl. Abbildung 3, oberer Strang).

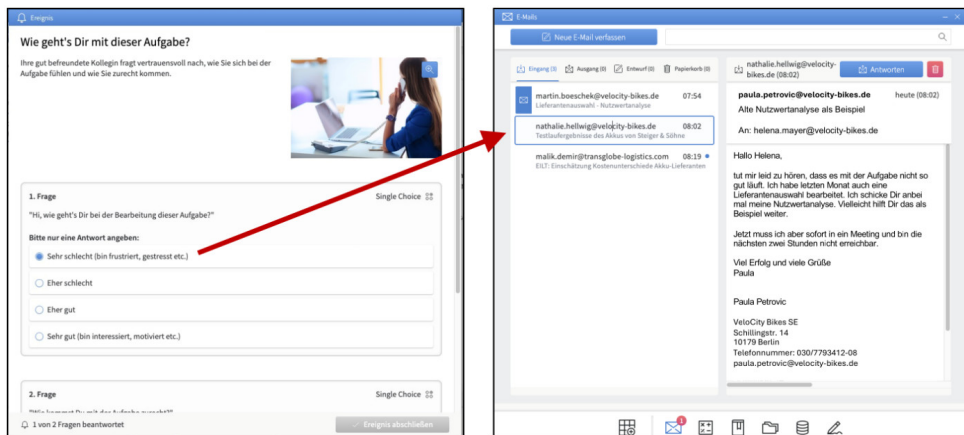


Abbildung 4: Simulierte Interaktion und Antwortverhalten des computerbasierten Agenten

Der eingesetzte Test zur Beschreibung von Facetten des Arbeitens in der LUCA Office Simulation umfasst 14 komplexe Testitems, die entlang eines Indikatorenmodells, das zwischen sozialen und kognitiven Teilleistungen im Lösungsprozess komplexer Probleme differenziert, entwickelt wurden. Das Indikatorenmodell geht auf Hesse und Kolleg:innen (2015) zurück, die Problemlösen sowohl als kognitiven als auch sozialen Prozess beschreiben. Tabelle 1 zeigt die ex ante formulierten Indikatoren, die einem Rating unterzogen wurden (Qualitätskriterien sowie Nennungen).

Tabelle 1: Kognitive und soziale Indikatoren der Leistungsbewertung

Kognitive Teilleistungen im Rahmen der Aufgabenbearbeitung (vier komplexe Items)	Soziale Teilleistungen im Rahmen der Aufgabenbearbeitung (elf komplexe Items/zehn nach empirischer Prüfung)
Aufgabenregulierung <ul style="list-style-type: none"> Umfang des Screenings relevanter Informationen für die Nutzenanalyse Umfang des Screenings der verschiedenen Angebote Qualität der aus dem Screening der Angebote gewonnenen Informationen Kommunikation der modifizierten Nutzwertanalyse mittels E-Mail 	Soziale Regulierung <ul style="list-style-type: none"> Gründe für die Wahl eines Lieferanten (Qualität und Quantität der Argumente) Zielgruppenspezifische Präsentation der Auswahl (angepasste Auswahl relevanter Gründe, angepasste Begrüßung und Anrede) Perspektivübernahme <ul style="list-style-type: none"> Abweichende Gewichtungsfaktoren für unterschiedliche Abteilungen und Zielvorgaben (Qualität und Quantität der Argumente) begründen. Partizipation <ul style="list-style-type: none"> Wechselseitige Begründung der Lieferantenauswahl in Abhängigkeit der verschiedenen Zielvorgaben

Zur besseren Anschaulichkeit ist nachfolgend in Abbildung 5 die Grundidee des Indikators „Perspektivenübernahme“ exemplarisch dargestellt. Die Auszubildenden sind aufgefordert, die „Lieferantenauswahl“ aus der Perspektive einer spezifischen Abteilung des Unternehmens (randomisierte Zuweisung) zu bearbeiten und anschließend die verschiedenen Abteilungsperspektiven in einem kollaborativen Aushandlungsprozess miteinander abzugleichen.

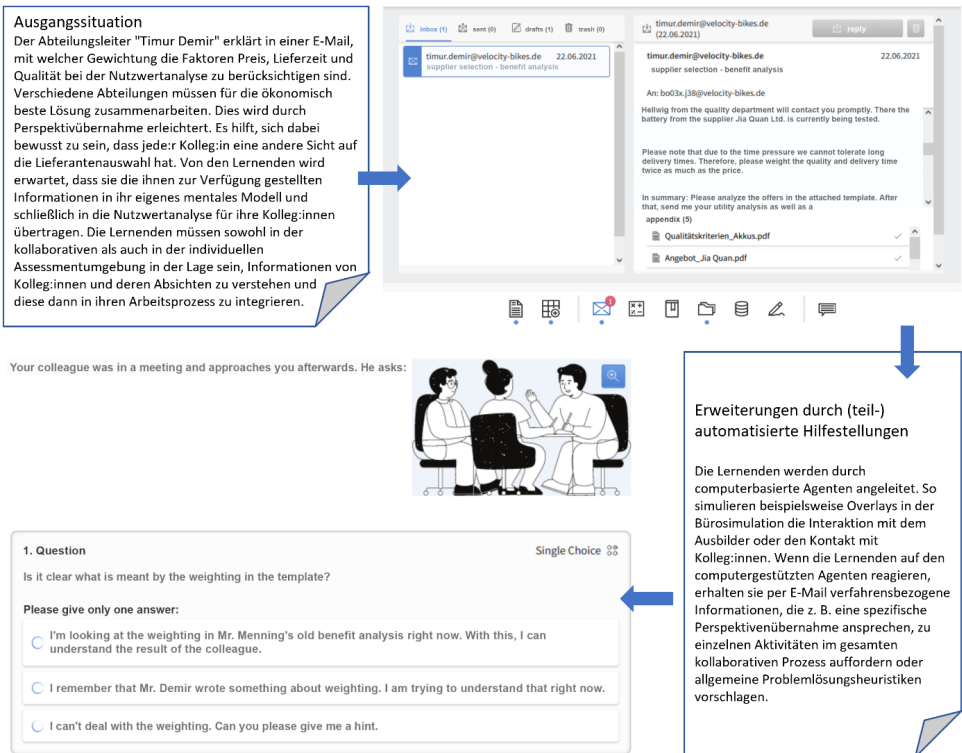


Abbildung 5: Der Indikator „Perspektivübernahme“

Die kognitiven und sozialen Teilleistungen, die beim Arbeiten mit der LUCA Office Simulation gezeigt werden, lassen sich über die hier aufgeführten Indikatoren erfassen (siehe hierzu Paeßens & Winther, 2024). Am Beispiel der Perspektivenübernahme – als soziale Teilleistungsfacette – wurde auf einer dreistufigen Skala bewertet, für welche Hilfestellung sich die Auszubildenden entscheiden (keine [= 0] bis umfangreiche Hilfestellung [= 2]) und wie sich die gegebene Hilfestellung auf die gezeigte Performanz (Gewichtung nicht berücksichtigt [= 0] bis Gewichtung korrekt berücksichtigt [= 2]) auswirkt. Beide Bewertungsinformationen werden zu einem Partial-Credit-Item zusammengefasst.

Die – vor allem mit Rückbezug auf Hesse et al. (2015) – angenommene zweidimensionale Struktur des Problemlöseprozesses ist in Abbildung 6 modellhaft dargestellt. Bei der Dimensionierung wird berücksichtigt, dass kognitive und soziale Teilleistungen zwei unterschiedlich latente Dimensionen im Problemlöseprozess repräsentieren, jedoch nicht unabhängig voneinander sind.

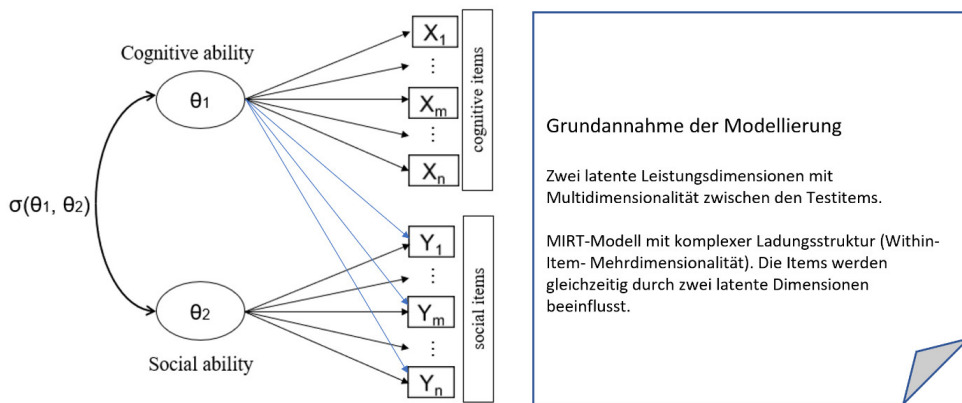


Abbildung 6: Schematische Darstellung des zweidimensionalen Rasch-Modells (in Anlehnung an Hartig & Höhler, 2009)

Zur Untersuchung der Konstruktvalidität von Tests mit mehreren Dimensionen können multidimensionale IRT-Modelle (MIRT) herangezogen werden (Embretson, 1980; Field, 2013; Santelices & Caspary, 2009; Wilson & Moore, 2011). Diese Modelle ermöglichen es u. a., die Fähigkeiten der Lernenden für die beteiligten Dimensionen getrennt zu bewerten. Die im vorliegenden Fall angenommene zweidimensionale Modellierung geht davon aus, dass die kognitive Leistung die Teilkompetenzen repräsentiert, die notwendig sind, um die Performance Task zu bewältigen. Die soziale Leistungsdimension dagegen repräsentiert spezifische, überfachliche Kompetenzen, die zusätzlich für die erfolgreiche Bearbeitung von kollaborativen Problemlöseaufgaben benötigt werden. Eine vergleichende Bewertung der Leistungsfähigkeiten in unterschiedlichen Testumgebungen setzt zudem voraus, dass die Tests ein Konstrukt psychometrisch äquivalent messen. Um dies zu prüfen, werden Messinvarianz-Testungen genutzt. Im vorliegenden Fall wurde die Messinvarianz auf Basis von drei verschiedenen konfirmatorischen Mehrgruppen-Faktorenanalysen (MG-CFA) mit dem Statistikpaket lavaan getestet. Es wurde zwischen Personen, die das Signature Scenario alleine oder kollaborativ bearbeitet haben, unterschieden. Die MG-CFA ohne gruppenübergreifende Gleichheitsrestriktionen (konfigurales Modell) zeigt eine gute Passung ($CFI = 0,975$, $RMSEA = 0,053$, $SRMR = 0,061$) gemäß Kline (2010). Auch die MG-CFA mit Gleichheit der Faktorladungen zwischen den Gruppen (metrisches Modell) weist eine akzeptable Passung auf ($CFI = 0,963$, $RMSEA = 0,061$, $SRMR = 0,077$). Partiiell metrische Messinvarianz wird erreicht, wenn ein Item der sozialen Dimension (hier: ein Unteritem des Indikators ‚Soziale Regulierung‘) aus den Analysen ausgeschlossen wird ($CFI = 0,966$, $RMSEA = 0,059$, $SRMR = 0,078$). Zusammenfassend lässt sich damit festhalten, dass die Variante, mit der die Testumgebung durchlaufen wird, keinen Einfluss auf das zu erfassende Konstrukt hat.

Neben der Überprüfung der Messinvarianz ist weiterhin von Relevanz, ob sich die angenommene mehrdimensionale Struktur der Leistungsdimensionen bestätigt. Hierzu werden ein eindimensionales und ein mehrdimensionales Rasch-Modell auf

die Testergebnisse angewandt und die Anpassungen dieser Modelle verglichen. Die Befunde sind in Tabelle 2 zusammengefasst. Es zeigt sich, dass das zweidimensionale Modell mit einer signifikant kleineren Modellabweichung, einer größeren Log-Likelihood, einem kleineren AIC und einem kleineren BIC besser zu den Daten passt als das eindimensionale Modell. Das ex ante angenommene theoretische Konstrukt, das zwischen kognitiven und sozialen Teilleistungen bei einer komplexen Problembearbeitung differenziert, kann somit als statistisch plausibel für die weiteren Analysen angenommen werden.

Tabelle 2: Modellanpassungsstatistiken für das ein- und mehrdimensionale Modell

Modell	loglike	Deviance	Npars	AIC	BIC	Chisq	df	p
uni-dimensional	-2697.273	5394.545	12	5418.545	5469.240	187.222	2	<.001
zwei-dimensional	-2603.662	5207.323	14	5235.323	5294.467			

Mit Blick auf die latenten Personenfähigkeiten wird davon ausgegangen, dass die Lernenden in der Testumgebung, in der automatisierte Unterstützungen angeboten werden, sowohl in der kognitiven als auch in der sozialen Dimension signifikant höhere Fähigkeiten zeigen. Dies liegt zum einen an den Scaffolding-Maßnahmen, die die implementierten automatisierten Hilfestellungen bei Bedarf im Verlauf der Testbearbeitung bieten. Zum anderen wird angenommen, dass durch die automatisierten Hilfestellungen und die damit auch implementierte bzw. simulierte soziale Komponente die Teilleistungsdimensionen stärker korrelieren und damit beispielsweise Schwächen in der kognitiven Teilleistung durch soziale Teilleistungen zumindest partiell (und insgesamt gesehen besser als bei der Variante der individuellen Bearbeitung) ausgeglichen werden können.

Zur Prüfung der Abhängigkeiten werden die Korrelationen zwischen den beiden Dimensionen und die jeweiligen Reliabilitäten geschätzt. Die Befunde sind Tabelle 3 zu entnehmen. Sie lassen sich wie folgt zusammenfassen: In der Testumgebung ohne Hilfestellung korrelieren die Teilleistungen mit .34 moderat, wohingegen in der Testumgebung mit (teil-)automatisierter Hilfestellung mit .54 ein großer Effekt vorliegt. Dieser Befund stützt die Überlegung, dass der Zusammenhang zwischen kognitiven und sozialen Teilleistungen in der Testumgebung mit angebotener Hilfe stärker als in der Testumgebung ohne Hilfestellung ausgeprägt ist.

Tabelle 3: Korrelationen und Reliabilitäten der Teilleistungen sowie die durchschnittlichen Personenfähigkeiten

	EAP-Reliabilität		Fähigkeiten der Person (Logits)		
	Korrelationen zwischen den Dimensionen	Kognitive Teilleistung	Soziale Teilleistung	Kognitive Teilleistung	Soziale Teilleistung
ohne Hilfestellung	0.34	0.690	0.591	-0.075	-0.208
mit Hilfestellung	0.54	0.634	0.681	1.044	0.188

Die EAP-Reliabilitäten der kognitiven und sozialen Teilleistungen reichen von .591 bis .690 (siehe Tabelle 3), was im Rahmen von Performance-based Assessments als ausreichend zu betrachten ist. Folgende Effekte lassen sich identifizieren: Die Auszubildenden, die die Testumgebung mit (teil-)automatisierten Hilfestellungen durchliefen, weisen sowohl in der kognitiven als auch in der sozialen Dimension höhere Teilleistungen auf. Mittels ANOVA wurde geprüft, ob die mittleren Leistungsunterschiede signifikant sind. Die Zuweisung der Testumgebung hat einen signifikanten Einfluss sowohl auf die kognitive Teilleistung ($F(1.503) = 38,41$, $p < .001$, part. $\eta^2 = .071$, $n = 505$) als auch auf die soziale Teilleistung ($F(1.503) = 7,64$, $p < .001$, part. $\eta^2 = .015$, $n = 505$). Zur Bewertung der praktischen Relevanz wird das partielle η^2 in die Effektgröße f nach Cohen (1992) umgerechnet. Für die kognitive Teilleistung beträgt die Effektgröße $f = .276$ und entspricht einem mittleren Effekt; für die soziale Teilleistung zeigt sich mit $f = .123$ ein mittlerer Effekt.

Die skizzierten Befunde lassen sich wie folgt zusammenfassen: Im Testsetting ohne Hilfestellung müssen die Auszubildenden die Komplexität der beruflichen Anforderung allein bewältigen; im Testsetting mit Hilfsangeboten hingegen erhalten sie gezielte Unterstützung. Wichtig dabei ist die Feststellung, dass diese Unterstützung im vorliegenden Fall keine klassische Lösungshilfe ist, sondern authentische Kooperationssituationen (im Sinne automatisierter Prompts) simuliert. Um beide Testsettings sinnvoll vergleichen zu können, wurde ein Indikatormodell gewählt, das kognitive und soziale Komponenten des komplexen Problemlösens (vgl. Hesse et al., 2015) ausdifferenziert. Komplexes Problemlösen – ob individuell oder in Kooperation – erfordert das für das Problemlösen notwendige Wissen (hier: Aufgabenregulation) und soziale Kompetenzen (hier: Perspektivenübernahme, soziale Regulation und Partizipation). Es zeigt sich, dass sich kognitive und soziale Teilleistungen durch computerbasierte Agenten im Sinne simulierter Kooperation fördern lassen, während gleichsam die Abhängigkeit der Teilleistungen untereinander zunimmt. Die Ergebnisse geben erste Hinweise darauf, dass Lernende durch modellierte Zusammenarbeit besser mit komplexen Aufgaben umgehen können und so die geforderten Aufgabenstandards eher erreichen. Die hierzu vorgestellten Analysen machen deutlich, dass die Nutzung des Indikatorenmodells für das Messen von Problemlösefähigkeiten insbesondere dann sinnvoll ist, wenn der Bearbeitungsprozess selbst Teil des Konstrukts ist.

2.3 Teilstudie 2: Individuelle Leistungsdiagnostik mit der LUCA Office Simulation

Im Projektverlauf wurden unterschiedliche empirische Analysen zur Erfassung berufsfachlicher Leistungen durchgeführt. Hierbei ging es vor allem darum, Teilfacetten beruflicher Kompetenz empirisch zu beschreiben. Das Szenario „Lieferantenauswahl“ ist für sich allein nicht als psychometrischer Kompetenztest konzipiert. Über Performance Tasks wird die Erfassung des Problemlösens als komplexes Phänomen in einer authentischen und inhaltsreichen Testumgebung ermöglicht (Seifried et al., 2021), aber aus psychometrischer Perspektive ist die Kompetenzdiagnostik auf Basis komplexer Performanzen in authentischen Arbeitsszenarien insbesondere mit der Herausfor-

derung des Validitätsparadoxons verbunden (siehe oben). Im Projekt wurde versucht, dieser Schwierigkeit dadurch zu begegnen, dass die Performance Tasks unterschiedliche didaktische Zugänge – sprachlich-argumentativ und mathematisch-analytisch – adressieren, um eine ausreichende Konstruktbreite zu gewährleisten (Winther, 2010; siehe auch Klotz & Winther, 2012). Vor diesem Hintergrund wird nachfolgend eine Studie vorgestellt, die folgende Forschungsfrage beantwortet:

Welche berufsfachlichen Leistungen lassen sich mittels Performance Tasks identifizieren und welche Rückschlüsse lassen sich aus der individuellen Leistungsdiagnostik für instruktionale Prozesse ziehen?

Zur Beantwortung dieser zweiten Forschungsfrage ist es von Relevanz, dass die Performance Tasks eine für diagnostische Zwecke hinreichende psychometrische Güte aufweisen. Gerade für didaktische Gestaltungsfragen sind solche Befunde zu nutzen, die (1) aus verlässlich messbaren Test- und Itemformaten abgeleitet werden können (siehe hierzu Wilson, 2004) und über die sich (2) Leistungsdaten und/oder Lernerfolge unterrichtsbezogen attribuieren lassen (Polikoff, 2010). Vor diesem Hintergrund wurden für die nachfolgenden Analysen die Rohdaten der Assessmentumgebungen sekundäranalytisch verarbeitet. Eine zentrale Herausforderung bei dieser Analyse bestand darin, ein Scoringmodell zu finden, das die mittels Performance Task erhobenen Leistungsdaten (hierfür standen 72 Einzelitems zur Verfügung) so aufbereitet, dass die Items – die in hohem Umfang (66 von 72 Einzelitems) Verletzungen der für die Anwendung von Verfahren der Item-Response-Theorie zentralen Annahme der lokalen Unabhängigkeit aufwiesen – neu zusammenfasst. Hierzu wurde für die Einzelitems, die den Grenzwert zur Verletzung lokaler Unabhängigkeit überschritten haben, inhaltslogisch geprüft, inwieweit sich diese in sequenzielle Bearbeitungsmuster (LCS-Patterns) überführen lassen. Im Ergebnis stehen durch Aggregation der ursprünglich 72 Einzelitems 14 komplexe, inhaltliche Bearbeitungsmuster – LCS-Patterns in Form von Partial-Credit-Items, basierend auf dem Problemlösezyklus nach Wuttke et al. (2015) – zur Verfügung, die sich über fachliche Zugänge – sprachlich-argumentative und mathematisch-analytische Bearbeitungs- und Lösungsmuster – systematisieren und strukturieren lassen (Fortunati & Winther, in Vorbereitung).

Die Differenzierung in sprachlich-argumentative und mathematisch-analytische Teilfacetten berufsfachlicher Leistungen lässt sich durch Studienergebnisse zum Einfluss des allgemeinen schulischen Vorwissens und hierbei insbesondere der Deutsch- und Mathematikfähigkeiten auf berufliche Kompetenzen begründen (Klotz, 2015): Der Einfluss mathematischer Grundkompetenzen auf berufsfachliche Kompetenzen ist empirisch für einige Ausbildungsberufe – mit jedoch tendenziell geringen Effektstärken – belegt (z. B. Geißel et al., 2013; Rosendahl & Straka, 2011). Demgemäß ist mathematisches Vorwissen als zentrale Voraussetzung für berufliche Sozialisation und auch als gängiges Kommunikationsmittel in zahlreichen Berufen zu verstehen (Klieme et al., 2000). Sowohl kaufmännisch-curriculare Analysen (z. B. Preiß, 2005; Winther, Sangmeister & Schade, 2013) als auch empirische Befunde bestätigen, dass mathe-

matische Anforderungen insbesondere der Wirtschaftsdidaktik immanent sind (z. B. Beziehungs- und Größenverständnis betriebswirtschaftlicher Zahlen, ökonomische Optimierungsprobleme, Gebrauch von Tabellen und Grafiken etc.). Sprachliches Vorwissen ist für kaufmännisches Arbeiten gleichfalls relevant: Betriebliche Unternehmensprozesse zeichnen sich durch einen hohen Kommunikationsgrad in deutscher und englischer Sprache sowohl innerhalb der Unternehmensbereiche als auch extern gegenüber Kund:innen und Partner:innen aus (Klotz, 2015).

Die mathematischen und sprachlichen Vorwissens- und Bearbeitungslogiken bei der Problemlösung in den Fokus zu rücken, hat den Vorteil, dass neben Informationen zu den einzelnen Prozessschritten der Problembearbeitung auch Aussagen darüber getroffen werden können, welche kognitiven Bearbeitungsprozesse zur Problembewältigung benötigt werden. Informationen hierzu ermöglichen (1) eine gezieltere Steuerung von didaktischen Interventionen und sie können (2) Hinweise darauf geben, inwieweit eine Familiarität zwischen der modellierten Testaufgabe und der unterrichtlichen Realität besteht. Berufliche Handlungssituationen lassen sich curricular über einen *sprachlich-argumentativen* oder *mathematisch-analytischen Zugang* modellieren (Klotz, 2015; Winther, 2010; Winther et al., 2016). Der sprachlich-argumentative Zugang bündelt die Nutzung von text- und bildsprachlichen Kenntnissen sowie das notwendige Verständnis von grundlegenden ökonomischen Konzepten in wirtschaftlichen Kontexten, die in eine komplexe berufliche Anforderung eingebracht werden müssen. Der mathematisch-analytische Zugang fasst solche Teilkompetenzen zusammen, die auf die Bewältigung von Rechenoperationen mit wirtschaftlichen Bezügen sowie auf die Interpretation und das Zustandekommen betriebswirtschaftlicher Realwerte abzielen.

Zur Prüfung der Dimensionalität der Performance Tasks wird der zweidimensionalen Modellspezifikation über fachliche Zugänge ein eindimensionales inhaltsbezogenes Modell entgegengestellt. In die Modellprüfung sind Daten von $N_{F2}=1.225$ Industrie- und Bürokaufleuten eingeflossen (vgl. Abbildung 3, unterer Strang). In Tabelle 4 ist der Vergleich der Modellfit-Indizes dargestellt.

Tabelle 4: Modellanpassungsstatistiken für das ein- und das zweidimensionale Modell

Modell	Deviance	Npars	AIC	BIC	Chisq	df	p
uni: inhaltsbezogen	27251.746	27	27305.46	27444.129			
zweidimensional: fachliche Zugänge	26756.395	29	26814.395	26963.029	495.351	2	<.001

Die Modellfit-Statistik zeigt, dass das fachzugangsspezifische, zweidimensionale Modell einen niedrigeren Deviance-Wert aufweist als das eindimensionale und somit einen für den Datensatz besseren Fit aufweist. Die Korrelation zwischen den Zugängen ist mit .28 recht gering und deutet ebenfalls auf die Plausibilität des zweidimensionalen Modells hin (s. Tabelle 5).

Tabelle 5: Korrelationen und Reliabilitäten der Zugänge sowie durchschnittliche Personenfähigkeiten

EAP-Reliabilität		Fähigkeiten der Person (Logits)		
Korrelationen zwischen den Zugängen	Sprachlich-argumentativ	Mathematisch-analytisch	Sprachlich-argumentativ	Mathematisch-analytisch
0.280	0.678	0.684	−0.049 (SE = 0.028)	−0.451 (SE = 0.035)

Die Reliabilitäts- und Personenfähigkeitenwerte sind in Anbetracht der geringen Anzahl der Testitems (14 Items insgesamt; 9 davon für den sprachlich-argumentativen Zugang) zufriedenstellend. Die Item-Infit-Werte liegen alle im Grenzbereich von $0.85 \leq \text{wMNSQ-Wert} \leq 1.15$ (Adams & Khoo, 1996; Kastberg et al., 2001). Hierbei streuen die Itemschwierigkeitsparameter im Bereich von -1.078 und 1.792 mit einer Spannweite von 2.870 . Der Mittelwert der Itemschwierigkeitsparameter ist aufgrund der Modellspezifikation auf 0 fixiert ($M = 0$; $SD = 1.591$). Die Personenparameter (WLE) streuen bei beiden fachlichen Zugängen breiter als die Itemschwierigkeiten. Der Mittelwert des Personenfähigkeitsparameters des sprachlich-argumentativen Zugangs beträgt -0.049 ($SE = 0.028$) und beim mathematisch-analytischen Zugang -0.451 Logits ($SE = 0.035$). Die Personenfähigkeitsparameter der Zugänge unterscheiden sich signifikant mit einer mittleren Effektstärke ($t = -8.430$, $df = 2486$, $p < 0.001$, $r = 0.215$).

Die Befunde der Studie 2 lassen sich wie folgt einordnen: Leistungsdaten mittels Performance Tasks zu generieren, ist voraussetzungsvoll. So verweisen Bennett et al. (2016) darauf, dass Items zur Erfassung von Problemlösekompetenzen mittels Performance Tasks besonders anfällig für die Verletzung lokaler Unabhängigkeit sind. Im PSA-Sim-Projekt konnte gezeigt werden, dass durch inhaltliche Aggregation die psychometrischen Eigenschaften von Performance Tasks so verbessert werden können, dass IRT-Verfahren zur Prüfung der Konstruktvalidität zur Anwendung kommen können (siehe auch Seifried et al., 2020). Im Rahmen des PSA-Sim-Projektes wurde – analog zu anderen Projekten in der kaufmännischen Bildung (siehe hierzu u. a. Forschungsinitiative ASCOT sowie das Projekt ECON2022) – ein zweidimensionales Modell, das zwischen sprachlich-argumentativen und mathematisch-analytischen Fähigkeitsbündeln differenziert, untersucht. Für die kaufmännische Bildung verfestigt sich das Bild, dass das Fähigkeitsniveau, mathematisch-analytische Zugänge zur Lösung komplexer, authentischer Probleme zu nutzen, geringer ausgeprägt ist als die sprachlich-argumentative Fähigkeitsstruktur. Dieser Befund unterstreicht erneut die Bedeutung einer kontextsensitiven, im Sinne einer eng an beruflichen Situationen orientierten Didaktik bei der Förderung berufsbezogener, mathematischer Kompetenzen (Hering et al., 2021). Gerade vor dem Hintergrund, dass kaufmännischer Kompetenzerwerb vorrangig sprachlich aufgebaut wird, wächst die Bedeutung einer mathematisch-analytischen Instruktionspraxis, um Auszubildende für ein umfangreiches Portfolio beruflicher Anforderungen zu qualifizieren (siehe hierzu die curricularen Analysen von Fortunati & Winther, 2021).

3 Schlussbemerkungen

Der vorliegende Beitrag greift aus dem PSA-Sim-Projekt exemplarisch zwei Fragestellungen heraus, die empirisch untersucht wurden. Grundlage dieser Studien ist das Signature Scenario „Lieferantenauswahl“, das unter verschiedenen Forschungsperspektiven eingesetzt werden konnte. Die Herausforderung bestand darin, berufliches Problemlösen zu erfassen. Obwohl das Problemlösen als Forschungsgegenstand umfangreich bearbeitet wurde, liegen für das Erheben des beruflichen Problemlösens nur wenige Arbeiten vor (Abele, 2023). Seifried et al. (2020) differenzieren beim Erhebungsansatz zwischen übergreifenden, dekontextualisierten, eng umrissenen Problemstellungen, wie sie zumeist in Large-Scale-Assessments (u. a. OECD, 2013, 2015, 2019) erhoben werden, und Performance-based-Assessments. Diese ermöglichen die Erhebung des Problemlösens als komplexes Phänomen in authentischen und inhaltsreichen Testumgebungen. Im Hinblick auf die psychometrische Güte realisieren die Aufgaben eines Performance-based-Assessments eine hohe inhaltliche und ökologische Validität bei schwacher Reliabilität. Eine weitere Herausforderung besteht darin, dass die Aufgaben/Testitems eines Performance-based-Assessments häufig lokal voneinander abhängig sind (Christensen et al., 2017). Für die Performance Task „Lieferantenauswahl“, die im Rahmen des PSA-Sim-Projekts als Assessmentumgebung eingesetzt wurde, zeigen sich die Einschränkungen in der psychometrischen Qualität ebenso. Um die psychometrische Qualität zu steigern, wurden verschiedene Analyseansätze entwickelt, die mit unterschiedlichen Scoring-Verfahren arbeiten:

- Zur Beschreibung verschiedener Bearbeitungsmodi (mit und ohne teilautomatisierter Hilfestellung) wurde ein Indikatorenmodell entwickelt, das berufliches Problemlösen in kognitive und soziale Teilleistungen zerlegt.
- Zur Beschreibung verschiedener fachlicher Zugänge, unter denen ein berufliches Problem konstruiert und den Testpersonen zur Bearbeitung angeboten wird, wurden komplexe, inhaltliche Bearbeitungsmuster identifiziert und als Partial Credits modelliert.

Beide Scoring-Verfahren geben Hinweise auf Ausschnitte des beruflichen Kompetenzerwerbs. Die über das Indikatorenmodell gewonnenen Erkenntnisse machen deutlich, wie anspruchsvoll der Umgang mit komplexen Problemen ist. Gezielte Scaffolding-Angebote, die über einen computerbasierten Agenten authentische Kooperationssituationen (und keine klassischen Lösungshilfen) simulieren und in den Testverlauf eingebunden sind, unterstützen Auszubildende dabei, ihre Leistungsfähigkeiten zu zeigen. Das Scoring mittels verdichteter Handlungsmuster fokussiert hingegen auf fach- bzw. domänensystematische Zugänge. Es kann gezeigt werden, dass die Lösung eines komplexen beruflichen Problems in der kaufmännischen Domäne sowohl sprachlich-argumentative als auch mathematisch-analytische Fähigkeitsbündel voraussetzt. In Analogie zu anderen Studien im kaufmännischen bzw. ökonomischen Bereich lässt sich festhalten, dass die mathematisch-analytischen Fähigkeiten eine stringendere Förderung benötigen.

Literatur

- Abele, S. (2023). *Problemlösekompetenzen in beruflichen Kontexten: Resultate aus Lehr-Lern-Prozessen sichtbar machen*. wbv. <https://doi.org/10.3278/9783763973675>
- Adams, R. J., & Khoo, S. T. (1996). *Quest-Interactive test analysis system*.
- Anselmann, S., Harm, S., & Faßhauer, U. (2022). Input from the grassroots level - Reflecting challenges and problems for VET professionals in Germany. *International Journal for Research in Vocational Education and Training*, 9(2), 239–268. <https://doi.org/10.13152/IJRVET.9.2.5>
- Bennett, R. E., Deane, P., & van W. Rijn, P. (2016). From cognitive-domain theory to assessment practice. *Educational Psychologist*, 51(1), 82–107. <https://doi.org/10.1080/00461520.2016.1141683>
- Brötz, R., Annen, S., Kaiser, F., Kock, A., Noack, I., Peppinghaus, B., Schaal, T., & Tiemann, M. (2014). *Abschlussbericht Forschungsprojekt „Gemeinsamkeiten und Unterschiede kaufmännisch-betriebswirtschaftlicher Aus- und Fortbildungsordnungen“*. Bundesinstitut für Berufsbildung. Verfügbar unter https://www2.bibb.de/tools/fodb/pdf/eb_42202.pdf (Zugriff am 10.07.2025).
- Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: A meta-Analysis. *Review of Educational Research*, 90(4), 499–541. <https://doi.org/10.3102/0034654320933544>
- Chernikova, O., Holzberger, D., Heitzmann, N., Stadler, M., Seidel, T., & Fischer, F. (2024). Where salience goes beyond authenticity. A meta-analysis on simulation-based learning in higher education. *Zeitschrift für Pädagogische Psychologie*, 38(1–2), 15–25, <https://doi.org/10.1024/1010-0652/a000357>
- Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q3: Identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement*, 41(3), 178–194. <https://doi.org/10.1177/0146621616677520>
- Codreanu, E., Sommerhoff, D., Huber, S., Ufer, S., & Seidel, T. (2020). Between authenticity and cognitive demand: Finding a balance in designing a video-based simulation in the context of mathematics teacher education. *Teaching and Teacher Education*, 95, Article 103146. <https://doi.org/10.1016/j.tate.2020.103146>
- Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, 1(3), 98–101. <https://doi.org/10.1111/1467-8721.ep10768783>
- Deutscher, V., Seifried, J., Rausch, A., Thomann, H., & Braunstein, A. (2022). Die LUCA Office Simulation in der Lehrerinnen- und Lehrerbildung – didaktische Design-Empfehlungen und erforderliche Lehrkompetenzen. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach, & M. Stock (Hrsg.), *Digital Literacy in der beruflichen Lehrer:innenbildung: Didaktik, Empirie und Innovation* (S. 107–121). wbv.
- Deutscher, V., & Winther, E. (2018). Instructional sensitivity in vocational education. *Learning and Instruction*, 53, 21–33. <https://doi.org/10.1016/j.learninstruc.2017.07.004>

- Devolder, A., van Braak, J., & Tondeur, J. (2012). Supporting self-regulated learning in computer-based learning environments: systematic review of effects of scaffolding in the domain of science education. *Journal of Computer Assisted Learning*, 28(6), 557–573. <https://doi.org/10.1111/j.1365-2729.2011.00476.x>
- Duchatelet, D., Jossberger, H., & Rausch, A. (2022). Assessment and evaluation of simulation-based learning in higher education and professional training: An introduction. *Studies in Educational Evaluation*, 75, 101210. <https://doi.org/10.1016/j.stueduc.2022.101210>
- Dohmen, D., Bayreuther, T., & Sandau, M. (2023). *Monitor Ausbildungschancen 2023. Gesamtbericht Deutschland*. Bertelsmann Stiftung.
- Embretson, S. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, 45, 479–494. <https://doi.org/10.1007/BF02293610>
- Field, J. (2013). Cognitive validity. In A. Geranpayeh, & L. Taylor (Ed.), *Examining listening: Research and practice in assessing second language listening* (pp. 77–151). Cambridge University Press.
- Fortunati, F., & Winther, E. (2021). Ein Curriculum genügt nicht. Wie aus neuen Inhalten gute Instruktionsprozesse werden (können). *Berufsbildung*, 188, 31–35.
- Gallego, A., & Kurer, T. (2022). Automation, digitalization, and artificial intelligence in the workplace: implications for political behavior. *Annual Review of Political Science*, 25(1), 463–484. <https://doi.org/10.1146/annurev-polisci-051120-104535>
- Geißel, B., Nickolaus, R., Stefanica, F., Härtig, H., & Neumann, K. (2013). Die Relevanz mathematischer und naturwissenschaftlicher Kompetenzen für die fachliche Kompetenzentwicklung in gewerblich-technischen Berufen. In R. Nickolaus, J. Retelsdorf, E. Winther, & O. Köller (Hrsg.), *Zeitschrift für Berufs- und Wirtschaftspädagogik: Bd. 26. Mathematisch-naturwissenschaftliche Kompetenzen in der beruflichen Erstausbildung: Stand der Forschung und Desiderata* (S. 39–66). Franz Steiner Verlag. <https://doi.org/10.25162/9783515104104>
- Gentner, S., Ludwig, S., Braunstein, A., Mayer, C., Deutscher, V., Rausch, A., & Seifried, J. (2024). The digital office simulation LUCA from the perspective of teachers and learners: first findings of usability analyses. In Cedefop and OECD (2024), *Apprenticeships and the digital transition: modernising apprenticeships to meet digital skill needs*. Publications Office of the European Union (S. 102–117). Cedefop reference series, 125. <https://data.europa.eu/doi/10.2801/074640>
- Hartig, J., & Höhler, J. (2009). Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation*, 35(2–3), 57–63. <https://doi.org/10.1016/j.stueduc.2009.10.002>
- Hasenbeck, F. (2019). *Macht die Digitalisierung alles komplexer?* Roundtable. Verfügbar unter <https://www.abitur-und-studium.de/Blogs/Fraunhofer-Gesellschaft/Roundtable-Macht-die-Digitalisierung-alles-komplexer> (Zugriff am 10.07.2025).
- Hering, R. von, Rietenberg, A., Heinze, A., & Lindmeier, A. (2021). Nutzen Auszubildende bei der Bearbeitung berufsfeldbezogener Mathematikaufgaben ihr Wissen aus der Schule? Eine qualitative Untersuchung mit angehenden Industriekaufleuten. *Journal für Mathematik-Didaktik*, 42(2), 459–490. <https://doi.org/10.1007/s13138-021-00181-8>

- Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In E. Care, & P. E. Griffin (Eds.), *Assessment and teaching of 21st century skills: Educational assessment in an information age* (S. 37–56). Springer. https://doi.org/10.1007/978-94-017-9395-7_2
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48(4), 63–85. <https://doi.org/10.1007/BF02300500>
- Kastberg, D., Cummings, L., Ferraro, D., & Perkins, R. C. (2021). *Technical report and user guide for the 2018 Program for International Student Assessment (PISA)*. National Center for Education Statistics.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1
- Klieme, E., Baumert, J., Köller, O., & Bos, W. (2000). Mathematische und naturwissenschaftliche Grundbildung: Konzeptuelle Grundlagen und die Erfassung und Skalierung von Kompetenzen. In J. Baumert, W. Bos, & R. Lehmann (Hrsg.), *TIMSS/III Dritte Internationale Mathematik- und Naturwissenschaftsstudie — Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* (S. 85–133). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-322-83411-9_4
- Kline, R. (2023). *Principles and Practice of Structural Equation Modeling* (5. Edition). The Guilford Press.
- Klotz, V. K. (2015). *Diagnostik beruflicher Kompetenzentwicklung: Eine wirtschaftsdidaktische Modellierung für die kaufmännische Domäne*. Springer Gabler. <https://doi.org/10.1007/978-3-658-10681-2>
- Klotz, V. K., & Winther, E. (2012). Kompetenzmessung in der kaufmännischen Berufsausbildung: Zwischen Prozessorientierung und Fachbezug. Eine Analyse der aktuellen Prüfungspraxis. *Betriebs- und Wirtschaftspädagogik online (bwp@)*, 22(8), 1–16.
- Lane, S., & Stone, C. A. (2006). Performance assessment. In R. Brennan (Ed.), *Educational Measurement* (pp. 387–431). Praeger Pub Text.
- Lane, M., & Williams, M. (2023). *Defining and classifying AI in the workplace*. OECD Social, Employment and Migration Working Papers. No. 290, OECD Publishing, Paris. Verfügbar unter <https://www.oecd-ilibrary.org/deliver/59e89d7f-en.pdf?itemId=/content/paper/59e89d7f-en&mimeType=application/pdf> (Zugriff am 10.05.2025).
- Minkley, N., Xu, K. M., & Krell, M. (2021). Analyzing relationships between causal and assessment factors of cognitive load: associations between objective and subjective measures of cognitive load, stress, interest, and self-concept. *Frontiers in Education*, 6, Article 632907. <https://doi.org/10.3389/feduc.2021.632907>
- Mislevy, R. J. (2019). Advances in measurement and cognition. *The ANNALS of the American Academy of Political and Social Science*, 683(1), 164–182. <https://doi.org/10.1177/0002716219843816>
- OECD. (2013). *The survey of adult skills*. OECD. <https://doi.org/10.1787/9789264204027-en>
- OECD. (2015). *Adults, Computers and Problem Solving*. OECD. <https://doi.org/10.1787/9789264236844-en>

- OECD. (2019). *The survey of adult skills*. OECD. <https://doi.org/10.1787/f70238c7-en>
- Paeßens, J., & Winther, E. (2024). Kollaboration, aber sinnvoll: Wie sich mit komplexen Problemszenarien berufliche und kollaborative Kompetenzen erfassen lassen. *Zeitschrift für Pädagogik*, 2, 162–181. <https://doi.org/10.3262/ZP2402162>
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). Academic Press. <https://doi.org/10.1016/B978-012109890-2/50043-3>
- Polikoff, M. S. (2010). Instructional sensitivity as a psychometric property of assessments. *Educational Measurement: Issues and Practice*, 29, 3–14. <https://doi.org/10.1111/j.1745-3992.2010.00189.x>
- Preiß, P. (2005). Entwurf eines Kompetenzkonzepts für den Inhaltsbereich Rechnungswesen/Controlling. In P. Gonon, F. Klauser, R. Nickolaus, & R. Huisinga (Hrsg.), *Kompetenz, Kognition und neue Konzepte der beruflichen Bildung* (S. 67–85). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-322-86895-4_4
- Rausch, A., Deutscher, V., Seifried, J., Brandt, S., & Winther, E. (2021). Die web-basierte Bürosimulation LUCA – Funktionen, Einsatzmöglichkeiten und Forschungsausblick. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 117, 372–394. <https://doi.org/10.25162/zbw-2021-0017>
- Rosendahl, J., & Straka, G. A. (2011). *Effekte personaler, schulischer und betrieblicher Bedingungen auf berufliche Kompetenzen von Bankkaufleuten während der dualen Ausbildung: Ergebnisse einer dreijährigen Längsschnittstudie*. Institut Technik und Bildung. <https://media.suub.uni-bremen.de/handle/elib/2900>
- Santelices, M., & Caspary, K. (2009). Entwicklung eines multidimensionalen Maßes für akademisches Engagement. *Zeitschrift für angewandte Messtechnik*, 10(4), 371–393.
- Schlicht, J. (2019). *Kommunikation und Kooperation in Geschäftsprozessen: Modellierung aus pädagogischer, ökonomischer und informationstechnischer Perspektive*. wbv.
- Sczogiel, S., Schmitt-Rueth, S., Malapally, A., & Williger, B. (2019). *Future Digital Job Skills: Die Zukunft kaufmännischer Berufe*. IHK Nürnberg für Mittelfranken.
- Seifried, J., Brandt, S., Kögler, K., & Rausch, A. (2020). The computer-based assessment of domain-specific problem-solving competence—A three-step scoring procedure. *Cogent Education*, 7(1). <https://doi.org/10.1080/2331186X.2020.1719571>
- Seifried, J., Gentner, S., Brandt, S., Braunstein, A., Deutscher, V., Gorshid, G. D., Ludwig, S., Mayer, C., Rausch, A., & Winther, E. (2021). Flexibel einsetzbare Lehr- und Lernplattform für kaufmännische Bildung: LUCA Office Simulation. *VLB-Akzente: berufliche Bildung in Bayern*, 30, 9–11.
- Spöttl, G., & Windelband, L. (2021) The 4th industrial revolution – its impact on vocational skills. *Journal of Education and Work*, 34(1), 29–52. <https://doi.org/10.1080/13639080.2020.1858230>
- van Merriënboer, J. J. G., & Kirschner, P. (2018). *Ten steps to complex learning: A systematic approach to four-component instructional design* (3rd Edition). Routledge/Taylor & Francis. <https://doi.org/10.4324/9781315113210>
- Wiggins, G. P., & McTighe, J. (2005). *Understanding by Design* (2nd Ed.). Pearson Education, Inc. Print.

- Wilson, M. (2004). *Constructing measures: an item response modeling approach*. Routledge.
- Wilson, M., & Moore, S. (2011). Building out a measurement model to incorporate complexities of testing in the language domain. *Language Testing*, 28(4), 441–462. <https://doi.org/10.1177/0265532210394142>
- Winther, E. (2010). *Kompetenzmessung in der beruflichen Bildung*. wbv. <https://doi.org/10.3278/6004148w>
- Winther, E., Sangmeister, J. & Schade, A. K. (2013). Zusammenhänge zwischen allgemeinen und beruflichen Kompetenzen in der kaufmännischen Erstausbildung. In R. Nickolaus, J. Retelsdorf, E. Winther & O. Köller (Hrsg.). *Mathematisch-naturwissenschaftliche Kompetenzen in der beruflichen Erstausbildung. Stand der Forschung und Desiderata. Zeitschrift für Berufs- und Wirtschaftspädagogik*, Beiheft 26 (S. 139–160). Steiner.
- Winther, E., Seeber, S., Festner, D., Sangmeister, J., & Liedtke, M. (2016). Large-scale Assessments in der kaufmännischen Berufsbildung – Das Unternehmensassessment ALUSIM (CoBALIT). In K. Beck, M. Landenberger, & F. Oser (Hrsg.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der BMBF-Förderinitiative ASCOT* (S. 55–74). wbv.
- Wuttke, E., Seifried, J., Brandt, S., Rausch, A., Sembill, D., Martens, T., & Wolf, K. D. (2015). Modellierung und Messung domänenspezifischer Problemlösekompetenz bei angehenden Industriekaufleuten – Entwicklung eines Testinstruments und erste Befunde zu kognitiven Kompetenzfacetten. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 111(2), 189–207. <https://doi.org/10.25162/zbw-2015-0013>

Messung der beruflichen Problemlösekompetenz – Kommentar zu Schwerpunkt 2

AMORY H. DANEK & SAMUEL GREIFF

Zusammenfassung

Vor dem Hintergrund sich wandelnder Anforderungen der Arbeitswelt und zunehmender Bedeutung von „employability skills“ diskutiert der Kommentar von Danek und Greiff die Relevanz beruflicher Problemlösekompetenz. Der Kommentar bezieht sich dabei auf zwei Verbundvorhaben der ASCOT+-Initiative, nämlich TeKoP (Technologiebasiertes kompetenzorientiertes Prüfen, siehe Seeber et al., in diesem Band) sowie PSA-Sim (Problem Solving Analytics in Office Simulations: Winther et al., in diesem Band). Im Projekt TeKoP wurde ein Kategoriensystem zur Bewertung der Problemhaltigkeit kaufmännischer Prüfungsaufgaben entwickelt und mittels Laut-Denken-Protokollen die Problemlösestrategien von Auszubildenden analysiert. Im Projekt PSA-Sim ging es um die Entwicklung und Evaluation der Bürosimulation LUCA als authentisches, computerbasiertes Testformat zur Erhebung beruflicher Problemlöseleistung. Beide Studien leisten wichtige Beiträge zur Erfassung komplexer Problemlöseprozesse im Ausbildungskontext und adressieren aktuelle Herausforderungen in der Kompetenzmessung. Quasi als Ausblick plädieren wir für eine stärkere Integration domänenübergreifender Kompetenzen in Ausbildung und Prüfung sowie für eine Weiterentwicklung psychometrisch fundierter, digitaler Testformate.

Schlagnote: Problemlösekompetenz, Employability Skills, Performance-based Assessment, Digitalisierung von Prüfungsformaten

Abstract

Against the backdrop of changing demands in the world of work and the growing importance of employability skills, the commentary by Danek and Greiff discusses the relevance of vocational problem-solving competence. The commentary draws on two collaborative projects within the ASCOT+ initiative: TeKoP (Technology-Based Competence-Oriented Assessment; see Seeber et al., in this volume) and PSA-Sim (Problem Solving Analytics in Office Simulations; see Winther et al., in this volume). In the TeKoP project, a categorization system was developed to evaluate the problem content of commercial examination tasks, and think-aloud protocols were used to analyze apprentices' problem-solving strategies. The PSA-Sim project focused on the development and evaluation of the LUCA office simulation as an authentic, computer-based test format for assessing vocational problem-solving performance. Both studies make important contributions to capturing complex problem-solving processes in vocational edu-

cation and address current challenges in competence assessment. As a forward-looking perspective, we advocate for the stronger integration of cross-domain competencies into training and assessment, as well as the continued development of psychometrically sound, digital test formats.

Keywords: Problem-Solving Competence, Employability Skills, Performance-Based Assessment, Digitalization of Assessment Formats

Einleitung

Es scheint unstrittig, dass sich die Arbeitsplätze der Zukunft schneller verändern werden als bisher. So wird es immer schwieriger, abzuschätzen, welche Fähigkeiten junge Menschen, die heute ausgebildet werden, denn tatsächlich in 10 bis 20 Jahren benötigen werden. Eine mögliche Lösung ist, sich vermehrt auf sogenannte „employability skills“ zu konzentrieren. Damit sind grundlegende Fähigkeiten gemeint, die nötig sind, um auf einem wandelbaren Arbeitsmarkt erfolgreich zu bestehen. Im aktuellen PISA-VET Framework werden Lesefähigkeit (literacy), Problemlösen (problem solving), die Fähigkeit, Aufgaben angemessen zu erledigen (task performance), und Kooperationsfähigkeit (collaboration) zu diesen Querschnittskompetenzen gezählt (OECD, 2024a). Der vorliegende Kommentar befasst sich mit der beruflichen Problemlösekompetenz.

Nicht nur im beruflichen Kontext gilt: Menschen sind hervorragende Problemlöser:innen. Vermutlich ist diese ausgeprägte Problemlösefähigkeit neben der Sprache der entscheidende evolutionäre Vorteil des Menschen gegenüber jeder anderen Spezies. Dabei ist Problemlösefähigkeit in vielen Situationen ausgesprochen hilfreich. Nicht nur im Alltag, sondern auch im Berufsleben wird man regelmäßig mit neu auftauchenden Problemen konfrontiert. Manchmal kann dazu auf vorhandene Routinen und Verhaltensmuster zurückgegriffen werden oder bereits erworbenes Wissen direkt angewandt werden. Häufiger jedoch ist die Entwicklung neuer Lösungswege gefragt, zum Beispiel basierend auf Transferprozessen. Auch wenn Problemlösefähigkeit immer schon (über)lebenswichtig war, so scheinen die sogenannten „Megatrends“ des 21. Jahrhunderts, wie die Entwicklung künstlicher Intelligenz, die grüne und die digitale Transformation, den Druck noch erhöht zu haben (z. B. Frey & Osborne, 2017). In der modernen Arbeitswelt nehmen Routinen ab, da sie von digitalen Assistenten übernommen werden. Oft ist es nicht mehr ausreichend, nur vorhandenes Wissen anzuwenden und Aufgaben nach einem vertrauten Schema abzuarbeiten. Stattdessen wird Problemlösekompetenz benötigt. Hinzu kommt die Tatsache, dass Wissen, sei es auch noch so spezifisch, heutzutage prinzipiell überall und sofort verfügbar ist – oft nur „einen Klick entfernt“. Dies trifft selbst auf Expert:innenwissen zu, welches bisher speziell ausgebildeten Menschen vorbehalten war. Daher scheint es zumindest vordergründig so, dass Wissen weniger wichtig wird.

Es stellt sich also die Frage, welche Auswirkungen die genannten Entwicklungen auf die berufliche Ausbildung haben. Zunächst muss während der Ausbildung die Möglichkeit geschaffen werden, sich überfachliche Kompetenzen anzueignen. Des

Weiteren erscheint unter Prüfungsaspekten eine Abkehr von reiner Abfrage von Wissen notwendig. Wenn Querschnittskompetenzen wie Problemlösen unterrichtet und gemessen werden sollen, erfordert dies die Entwicklung einer neuen Generation von Testaufgaben (OECD, 2024b). Im vorliegenden Kommentar sollen zwei Forschungsprojekte diskutiert werden, die sich dieser Herausforderung stellen: Das erste Projekt (TeKoP: Technologiebasiertes kompetenzorientiertes Prüfen, siehe Seeber et al., in diesem Band) beschäftigt sich mit der Frage, wie die Problemhaltigkeit von Prüfungsaufgaben überprüft werden kann. Das zweite Projekt (PSA-Sim: Problem Solving Analytics in Office Simulations: Winther et al., in diesem Band) analysiert den Problembearbeitungsprozess in einem computerbasierten Szenario, der LUCA-Büro-Simulation. Schwerpunkt ist dabei die Messung der Problemlösekompetenz von kaufmännischen Auszubildenden.

„Problemhaltigkeit kaufmännischer Prüfungsaufgaben: Eine Analyse mittels Lautem Denken“ von Seeber et al.

Die fortschreitende Digitalisierung verändert auch die kaufmännische Arbeitswelt. Routinetätigkeiten werden durch digitale Technologien vereinfacht oder können zum Teil von digitalen Agenten übernommen werden. Damit geht eine Verlagerung der ursprünglichen kaufmännischen Tätigkeit hin zu anspruchsvolleren Aufgaben einher. Vor diesem Hintergrund erscheint es unstrittig, dass Absolvent:innen einer beruflichen Ausbildung nach deren Abschluss in der Lage sein sollten, komplexe berufliche Probleme zu lösen. Wie oben besprochen, zählt Problemlösefähigkeit zu den immer wichtiger werdenden „employability skills“ (OECD, 2024a). Dementsprechend ist ein Ausbildungsziel in kaufmännischen Curricula die Fähigkeit, komplexe Probleme zu lösen. Aktuell fehlt es jedoch an einer adäquaten Messung der Problemlösekompetenz in den Abschlussprüfungen. Dies kann einerseits daran liegen, dass die Förderung von Problemlösekompetenz auf der instruktionalen Ebene kaum eine Rolle spielt, oder aber daran, dass Lehrkräfte und Auszubildende nicht über die nötigen Kompetenzen verfügen, um geeignete Testaufgaben zu entwickeln. Die vorliegende Studie von Seeber et al. (2025) setzt an dem zweiten Problem an. Sie basiert auf einem von Wuttke et al. (2022) entwickelten Training für Lehrkräfte, das diese zum Erstellen problemhaltiger technologiebasierter Aufgaben befähigen soll. Der Aspekt „technologiebasiert“ ist wichtig, da sich vor allem computerbasierte Szenarien dafür eignen, die Komplexität beruflicher Handlungssituationen realitätsnah abzubilden (vergleiche auch Winther et al., in diesem Band). Außerdem lässt sich mit ihnen Transfer gut messen.

Vor dem Hintergrund, dass die Entwicklung problemhaltiger Aufgaben für den beruflichen Ausbildungskontext dringend notwendig ist, leistet die vorliegende Studie von Seeber et al. einen wichtigen Beitrag zur Methodik. Inhaltlich wurde die analytische Problemlösekompetenz von kaufmännischen Auszubildenden gemessen. Hauptziel war die Beantwortung der Frage, inwieweit die in dem Training entwickelten

Aufgaben tatsächlich problemhaltig sind, also mehr als nur den Abruf von Wissen erfordern. Die Autor:innen wählten dazu folgende Herangehensweise:

- (1) Analyse der Aufgaben bezüglich ihrer Problemhaltigkeit: Als Maß für Problemhaltigkeit wurden klassische Eigenschaften komplexer Probleme (Funke, 2011; Greiff et al., 2015) herangezogen, nämlich Eigendynamik, Vernetztheit und Polytelie (Interessenkonflikte) sowie als zusätzlicher Punkt die Strukturiertheit der Aufgabe (nach Jonassen, 2000). Dann wurde überprüft, wie viele und welche dieser Merkmale auf die erstellten Aufgaben zutrafen und diese a priori nach Wuttke et al. (2022) klassifiziert.
- (2) Auswertung von kognitiven Interviews: Dies diente der Identifizierung von Problemlösestrategien wie z. B. „Trial-and-Error“ oder Analogiebildung. Im Rahmen der über Zoom geführten Interviews wurden die Testpersonen aufgefordert, ihre Denkprozesse zu verbalisieren.

Seeber et al. (2025) adressieren in ihrem Beitrag zwei grundsätzliche methodische Herausforderungen, vor denen die Problemlöseforschung steht: Die erste ist die Auswahl geeigneter Aufgaben, die über den reinen Abruf von gespeichertem Wissen hinausgehen und stattdessen von den Probanden das Durchlaufen eines Lösungsprozesses erfordern. Es wurden bereits verschiedene Taxonomien von Problemen vorgeschlagen (z. B. Dörner, 1976; Weisberg, 1995). Zur aktuellen Frage nach der Testung analytischer Problemlösekompetenz im beruflichen Ausbildungskontext wurde auf Wuttkes Aufgaben-Klassifikationssystem (2022) zurückgegriffen. Dabei wurde ein Analyseraster zur Beurteilung des Problemgehalts von Aufgaben entwickelt, das auch für andere Aufgaben eingesetzt werden kann. Dieses Kategoriensystem kann als wichtiger erster Schritt zur Entwicklung eines normierten Aufgabenpools gewertet werden.

Zweitens ist es prinzipiell schwierig, während des Problemlösens stattfindende Denkprozesse zu messen. In der Problemlöseforschung ist seit einiger Zeit eine methodische Abkehr vom reinen Erfassen der Lösungsraten und Antwortzeiten hin zu kontinuierlichen „trace methods“ wie lautem Denken, Messung der Augenbewegungen oder Videoanalyse zu beobachten, welche idealerweise im Rahmen multimodaler Messungen parallel eingesetzt werden (Azevedo & Gašević, 2019). Diese Verfahren geben einen tieferen Einblick, wie Proband:innen eine Aufgabe tatsächlich bearbeiten, und erlauben auch Rückschlüsse auf den zeitlichen Verlauf des Problemlöseprozesses, bspw. ob die Aufgabe schrittweise gelöst wurde oder eher plötzlich (z. B. Bilalić et al., 2021). Dabei ist vor allem bei der Methode des lauten Denkens zu beachten, dass nicht alle kognitiven Prozesse verbalisierbar sind, da manche eher unbewusst ablaufen, wie z. B. im Falle von plötzlichen Aha!-Momenten, wo die Lösung scheinbar aus dem Nichts auftaucht und weniger analytische Anteile vorhanden sind (Kounios & Beeman, 2014). Die vorliegenden Aufgaben wurden für die Messung analytischer Problemlösekompetenz konstruiert und sind auch nur durch analytisches Vorgehen lösbar. Insofern erscheint die methodische Entscheidung für eine Laut-Denken-Studie schlüssig, weil zu erwarten ist, dass die Proband:innen tatsächlich Auskunft über ihren Problemlöseprozess geben können.

Das hier angewandte zweistufige Verfahren ist in der Gesamtschau positiv zu bewerten. Nur die Aufgaben, die im ersten Schritt anhand des Kriterienrasters als ausreichend problemhaltig klassifiziert wurden, wurden in der Laut-Denken-Studie eingesetzt. In der Studie konnte dann anhand von Denkprotokollen überprüft werden, inwieweit bei der Bearbeitung tatsächlich ein Problemlöseprozess durchlaufen wird. Noch zu ergänzen wäre eine Quantifizierung, wie viele der ausgewerteten Aufgaben nun letztlich als problemhaltig anzusehen sind. Als Ergebnis der Laut-Denken-Studie werden die verschiedenen Schritte des Problemlöseprozesses aufgelistet und anhand wörtlicher Zitate aus den Denkprotokollen illustriert. Es bleibt jedoch unklar, bei wie vielen und welchen Aufgaben das Vorhandensein des jeweiligen Aspekts (z. B. eine Problem- und Zielanalyse) aufgrund der Daten tatsächlich erkennbar war. In der Diskussion wird erwähnt, dass die Denkprotokolle „teilweise unvollständige bzw. keine Aussagen zu einzelnen oder mehreren Problemlöseschritten“ enthielten, dies wird jedoch nicht weiter quantifiziert. Die Aufgaben wurden auch in einer quantitativen Studie eingesetzt. Die Verknüpfung der vorliegenden qualitativen Daten und Analysen mit deren Ergebnissen wäre sehr interessant und könnte weiterführende Rückschlüsse ermöglichen.

„Mit Signature Scenarios zum Kompetenzerwerb: Die Analyse von Performance Tasks in virtuellen Simulationen“ von Winther et al. (2025)

Mit diesem Projekt leisten Winther et al. (in diesem Band) ebenfalls einen Beitrag zur Messung kaufmännischer Kompetenzen von Auszubildenden, in diesem Fall über computerbasierte Szenarien. Hierbei wurde die LUCA-Büro-Simulation verwendet (entwickelt von Rausch et al., 2021), welche basierend auf dem 4C-ID-Designansatz (van Merriënboer & Kirschner, 2018) konzipiert wurde. LUCA besteht aus authentischen Arbeitsszenarien, beispielsweise sollen Auszubildende aus mehreren Angeboten einen geeigneten Lieferanten auswählen und ihre Entscheidung begründen. Die Büro-Simulation bietet zur Bearbeitung der Aufgaben eine Auswahl typischer Software-Tools eines kaufmännischen Büroarbeitsplatzes. Ein Szenario kann zusätzliche Hilfen in Form optionaler E-Mails enthalten, die den Teilnehmenden nur dann angezeigt werden, wenn bestimmte Bedingungen erfüllt sind, z. B. wenn eine wichtige Excel-Datei nach fünf Minuten noch nicht geöffnet wurde. Somit ist der Aspekt der Eigidynamik erfüllt, welcher eine wichtige Eigenschaft komplexer Probleme darstellt.

Winther et al. (2025) untersuchten die Frage, ob das Vorhandensein solcher Hilfestellungen einen Einfluss auf die Bearbeitung des Szenarios hat. Anhand einer großen Stichprobe von Auszubildenden wurde die Leistung im Szenario Lieferantenauswahl in zwei Testumgebungen verglichen, einmal mit Hilfestellung (durch teilautomatisierte Prompts, die beispielsweise eingehende E-Mails von Kolleg:innen simulieren) und einmal ohne. Wie erwartet zeigte die Gruppe mit Hilfestellung höhere Leistungen, sowohl in der kognitiven als auch in der sozialen Dimension (mittlere Effektstärken).

Die klare theoretische Fundierung im 4C-ID-Ansatz (van Merriënboer & Kirschner, 2018) ist eine Stärke der vorliegenden Studie. Außerdem adressiert sie eine Forschungslücke, da es nur wenige Arbeiten zur Erhebung beruflichen Problemlösens gibt (Abele, 2023). Im Unterschied zu dekontextualisierten Aufgaben, wie sie häufig bei Large-Scale-Assessments verwendet werden (z. B. OECD, 2021), ermöglichen Performance-based Assessments die Erfassung von beruflichem Problemlösen in seiner ganzen Komplexität (Seifried et al., 2020). Bei dem in der vorliegenden Studie verwendeten Szenario sind mehrere positive Aspekte zu nennen: Die Einbettung in eine inhaltsreiche Testumgebung anhand eines konkreten Modellunternehmens (z. B. Fahrradhersteller) ermöglicht einen Bezug zur realen Arbeitswelt. Die Bearbeitungsdauer von etwa 30–45 Minuten pro Szenario stellt einen vertretbaren Zeitaufwand dar, um Problemlösen im Kontext von Lern- und Assessmentprozessen in der beruflichen Bildung zu erfassen.

Wie dies auch bei anderen Performance-based Assessments häufig der Fall ist, zeigten sich bei der LUCA-Büro-Simulation jedoch Einschränkungen der psychometrischen Qualität, weil die einzelnen Items lokal voneinander abhängig waren. Als Maßnahme zur Steigerung der psychometrischen Güte schlagen Winther et al. (2025) daher vor, mit unterschiedlichen Scoring-Verfahren zu arbeiten. Dies führten sie exemplarisch in der dargestellten Studie durch.

Zusammenfassung und Ausblick

Die hier vorgestellten Projekte leisten einen Beitrag zur Messung der beruflichen Problemlösekompetenz, einem Bereich, der aktuell zu wenig beforscht wird. Seeber et al. (2025) stellen ein Kategoriensystem zur Verfügung, welches zur Beurteilung des Problemgehalts von Aufgaben eingesetzt werden kann. Dies wird hilfreich für die Entwicklung zukünftiger Aufgaben sein, wobei Anstrengungen in Richtung eines normierten Aufgabenpools unternommen werden sollten. Winther et al. (2025) zeigen anhand eines computerbasierten Szenarios, der LUCA-Büro-Simulation, dass im Rahmen von Performance-based Assessments die Erfassung von beruflichem komplexem Problemlösen in authentischer Testumgebung möglich ist.

Abschließend ist zu bemerken, dass die Integration von „employability skills“ (wie z. B. Problemlösekompetenz) im Sinne von PISA-VET (OECD, 2024a) in die aktuellen Forschungsbemühungen sehr zu begrüßen ist. Hieraus abgeleitete Erkenntnisse werden die Akteur:innen im Bildungswesen dazu befähigen, junge Menschen in ihrer beruflichen Ausbildung adäquat auf zukünftige Anforderungen vorzubereiten. Offen bleibt die Frage, wie sowohl die berufliche Ausbildung als auch die dazugehörige Forschung Schritt halten kann, wenn es darum geht, betriebliche Prozesse und die Komplexität des Arbeitslebens abzubilden. Simulationen wie LUCA sind als wichtiger erster Schritt zu werten, müssen jedoch noch weiter ausgebaut werden. Auch bezüglich ihrer unzureichenden psychometrischen Güte besteht weiterer Forschungsbedarf.

Literatur

- Abele, S. (2023). *Problemlösekompetenzen in beruflichen Kontexten: Resultate aus Lehr-Lern-Prozessen sichtbar machen*. wbv. <https://doi.org/10.3278/9783763973675>
- Azevedo, R., & Gašević, D. (2019). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. *Computers in Human Behavior*, 96, 207–210. <https://doi.org/10.1016/j.chb.2019.03.025>
- Bilalić, M., Graf, M., Vaci, N., & Danek, A. H. (2021). The temporal dynamics of insight problem solving – restructuring might not always be sudden. *Thinking & Reasoning*, 27(1), 1–37. <https://doi.org/10.1080/13546783.2019.1705912>
- Dörner, D. (1976). *Problemlösen als Informationsverarbeitung*. Kohlhammer.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280. <https://doi.org/10.1016/j.techfore.2016.08.019>
- Funke, J. (2011). Problemlösen. In T. Betsch, J. Funke, & H. Plessner (Hrsg.), *Denken – Urteilen, Entscheiden, Problemlösen* (S. 137–199). Springer.
- Greiff, S., Fischer, A., Stadler, M., & Wüstenberg, S. (2015). Assessing complex problem-solving skills with multiple complex systems. *Thinking & Reasoning*, 21(3), 356–382. <https://doi.org/10.1080/13546783.2014.989263>
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48(4), 63–85. <https://doi.org/10.1007/BF02300500>
- Kounios, J., & Beeman, M. (2014). The cognitive neuroscience of insight. *Annual Review of Psychology*, 65(1), 71–93. <https://doi.org/10.1146/annurev-psych-010213-115154>
- OECD (2021). *The Assessment Frameworks for Cycle 2 of the Programme for the International Assessment of Adult Competencies*. OECD. <https://doi.org/10.1787/4bc2342d-en>
- OECD (2024a). Employability Skills. In *PISA Vocational Education and Training (VET): Assessment and Analytical Framework: Vol. Chapter 7*. OECD. <https://doi.org/10.1787/b0d5aaf9-en>
- OECD (2024b). *PISA Vocational Education and Training (VET): Assessment and Analytical Framework*. OECD. <https://doi.org/10.1787/b0d5aaf9-en>
- Rausch, A., Deutscher, V., Seifried, J., Brandt, S., & Winther, E. (2021). Die web-basierte Bürosimulation LUCA – Funktionen, Einsatzmöglichkeiten und Forschungsausblick. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 117(3), 372–394. <https://doi.org/10.25162/zbw-2021-0017>
- Seifried, J., Brandt, S., Kögler, K., & Rausch, A. (2020). The computer-based assessment of domain-specific problem-solving competence – A three-step scoring procedure. *Cogent Education*, 7(1), 1719571. <https://doi.org/10.1080/2331186X.2020.1719571>
- van Merriënboer, J. J. G., & Kirschner, P. (2018). *Ten steps to complex Learning: A systematic approach to four-component instructional design* (3rd ed.). Routledge.
- Weisberg, R. W. (1995). Prolegomena to theories of insight in problem solving: A taxonomy of problems. In R. J. Sternberg, & J. E. Davidson (Eds.), *The nature of insight* (S. 157–196). MIT Press.

Wuttke, E., Seeber, S., Geiser, C., & Turhan, L. (2022). Zur Problemhaltigkeit von Aufgaben in kaufmännischen Abschluss- und Zwischenprüfungen – Ergebnisse aus Aufgabenanalysen. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 118(1), 25–52. <https://doi.org/10.25162/zbw-2022-0002>

Schwerpunkt 3: Kompetenzorientierte Gestaltung von Prüfungsaufgaben

Zur Qualität beruflicher Abschlussprüfungen: Psychometrik und praxisnahe Designstandards – Ausgewählte Befunde des Verbundvorhabens ASPE¹

ESTHER WINTHER, SEBASTIAN SCHLÜNKES, ALEXANDER KOHLUS, MICHAEL KERRES & WOLFGANG VOGEL

Zusammenfassung

Im Verbundvorhaben wurde mit der ASPE-Workbench eine Umgebung konzipiert, mit der (1) die Gestaltung kaufmännischer Abschlussprüfungen entlang von Designstandards erfolgt und (2) die psychometrische Qualität neu konzipierter Abschlussprüfungen gesichert wird. Wesentliches Merkmal der Workbench ist der Praxisbezug. Dies bedeutet zum einen, dass die Designkriterien so in digitale Templates überführt werden, dass sie für die Prüfungserstellenden direkt handlungswirksam werden. Zum anderen werden die konzipierten Prüfungsaufgaben einem Rating unterzogen, das ihren (antizipierten) Schwierigkeitsgrad vorhersagt. Beide Aspekte haben direkten Einfluss auf die Qualität der Prüfungen, da sie (1) eine handlungs- und kompetenzorientierte inhaltliche Gestaltung (Stärkung der Konstruktvalidität) und (2) eine angemessene Schwierigkeitsverteilung, die zwischen den unterschiedlichen Leistungsfähigkeiten der Auszubildenden gut diskriminieren kann (Stärkung der Reliabilität), garantieren.

Der Beitrag beschreibt die theoretischen Bezüge hinter den Designkriterien und zeigt auf, wie sich die kompetenzorientierte Gestaltung kaufmännischer Abschlussprüfungen nahtlos in die internationalen Diskurse um lerner:innenzentrierte und authentische Assessments einbinden lässt. Zudem werden empirische Befunde aus Reliabilitätsstudien berichtet, aus denen ein praxisnahes Rating abgeleitet wurde, das Prüfungserstellende ohne ausgewiesene statistische Expertise in die Lage versetzt, a priori die Schwierigkeit einer Prüfungsaufgabe zu prognostizieren. Diese Schwierigkeitsprognostik ist in hohem Maße qualitätssichernd, da sie für die einzelne Prüfung eine angemessene Verteilung der Prüfungsaufgaben entlang der zu erwartenden Kompetenzstandards ermöglicht und über die Zeit vergleichbare Gesamtschwierigkeiten in den Prüfungsjahrgängen erzeugt.

Schlagworte: kaufmännische Abschlussprüfungen, Designstandards, Aufgabenschwierigkeit, authentische Assessments

1 Das Verbundprojekt „Digitale Workbench für kompetenzorientierte Prüfungsaufgaben und Abschlussprüfungen – Assessments for Professional Exams“ (ASPE) ist Teil der Initiative ASCOT+ (siehe www.ascot-vet.net) und wurde durch das Bundesministerium für Bildung und Forschung (BMBF) gefördert (Förderkennzeichen: 21AP007A [Teilprojekt Duisburg-Essen] bzw. 21AP007B [Teilprojekt Aka Nürnberg]).

Abstract

In the joint research project, the ASPE Workbench was designed to provide an environment that (1) ensures the design of commercial final examinations in accordance with design standards and (2) ensures the psychometric quality of newly designed final examinations. A key feature of the workbench is its practical relevance. On the one hand, this means that the design criteria are transferred into digital templates in such a way that they become directly effective for the examiners. On the other hand, the designed examination tasks are subjected to a rating that predicts their (anticipated) level of difficulty. Both aspects have a direct influence on the quality of the examinations, as they guarantee (1) an action- and competence-oriented content design (strengthening construct validity) and (2) an appropriate distribution of difficulty that can effectively discriminate between the different abilities of the trainees (strengthening reliability).

The paper describes the theoretical references behind the design criteria and shows how the competence-oriented design of final commercial examinations can be seamlessly integrated into the international discourses on learner-centered and authentic assessments. In addition, empirical findings from reliability studies are reported, from which a practice-oriented rating was derived that enables examiners without proven statistical expertise to predict the difficulty of an exam task a priori. This difficulty prognosis is highly effective in ensuring quality, as it enables an appropriate distribution of exam questions for the individual exam along the expected competency standards and generates comparable overall difficulty levels over time in the exam years.

Keywords: Commercial Final Examinations, Design Standards, Task Difficulty, Authentic Assessments

1 Die Professionalisierung des Aufgaben- und Prüfungserstellungsprozesses

1.1 Ausgangslage und Zielsetzung

Das Prüfungswesen im dualen Ausbildungssystem in Deutschland ist eine zentrale Säule zur Sicherstellung der Qualität und Validität beruflicher Qualifikationen. Prüfungsaufgaben, die eine Gesamtprüfung bilden, dienen dabei als entscheidendes Instrument zur Messung der fachlichen und beruflichen Kompetenzen der Auszubildenden und bestimmen somit maßgeblich über deren Berufsqualifikation. Die Qualität dieser Prüfungsaufgaben ist von entscheidender Bedeutung: Nur wenn die Aufgaben präzise und psychometrisch valide gestaltet sind, können sie als zuverlässige Grundlage für die Beurteilung des Kompetenzniveaus und die Eignung der Auszubildenden im beruflichen Kontext dienen. Die lernortübergreifende Berufseingangsprüfung wird für kaufmännische Ausbildungsberufe in einem Großteil der Bundesländer von der zentralen Aufgabenstelle für kaufmännische Abschluss- und Zwischenprüfungen (AkA) organisiert. Die AkA ist hierbei verantwortlich für die Vorbereitung und Erstel-

lung der Prüfungen, die objektiv und valide die in den Ausbildungsordnungen und Rahmenlehrplänen vorgeschriebenen Lernziele abprüfen. Gemeinsam mit Expert:innen der Aka-Fachausschüsse (drittelparitätische Besetzung aus Arbeitgebervertreter:innen, Ausbilder:innen sowie Lehrkräften) werden die Prüfungen in einem komplexen Erstellungsprozess konstruiert und mehrfach auf ihre Plausibilität sowie mit Blick auf die Gütekriterien (inhaltliche und externe Validität, Reliabilität und Trennschärfe) geprüft. Der Erstellungsprozess stellt hohe Anforderungen an die beteiligten Akteur:innen, die zu einem signifikanten Teil ehrenamtlich tätig sind. Einen solchen Prozess, der sich über Jahrzehnte bewährt hat und der vor allem durch gewachsene Partizipationsstrukturen eine hohe Qualität des Prüfungssystems garantiert, weiter zu professionalisieren und in Teilen zu digitalisieren, war eine der großen Herausforderungen des Verbundes. Hierbei ist vor allem zu berücksichtigen, dass die Ergebnisfeststellung der lernortübergreifenden Berufseingangsprüfungen ein prüfungsrechtlicher Verwaltungsakt ist, bei dem die Justiziabilität (Objektivität und Transparenz) jederzeit zu gewährleisten ist. Hinzu kommt, dass die Prüfungen mit hohen Teilnehmendenzahlen zu einheitlichen Terminen durchzuführen sind und vor diesem Hintergrund Kriterien der Ökonomie, Effizienz und Kontinuität eine wesentliche Rolle spielen.

Die Arbeiten im Verbundvorhaben ASPE sind vor diesem Hintergrund einzuordnen; das Verbundvorhaben ASPE baut als Transfervorhaben konsequent auf Befunden auf, die in vorangegangenen Projekten erzeugt wurden und in direkter Beziehung zu den Berufseingangsprüfungen der Aka stehen. So zeigten sich zusammengefasst mit Blick auf die psychometrische Qualität der Abschlussprüfungen (u. a. Winther, 2010; Klotz & Winther, 2012) folgende Befunde, die in Vorläuferprojekten bearbeitet werden konnten:

- Objektivitätsmängel durch Korrektur-Bias-Effekte,
- Validitätsmängel im Hinblick auf die curriculare Abbildung sowie
- Reliabilitätsmängel, da sich nur mittlere Leistungsfähigkeiten verlässlich schätzen lassen.

Die Defizite in der Psychometrik ließen sich durch eine stringente Ausrichtung an einer kompetenzorientierten Aufgaben- und Prüfungsgestaltung auflösen. Die Kompetenzorientierung als Leitbild der Prüfungserstellung zu nutzen, bietet verschiedene Vorteile:

- 1) Die Beschreibung vollständiger Arbeitstätigkeiten (in authentischen Geschäftsprozessen) über sinnvoll zusammenhängende Aufgaben bzw. Aufgabenbündel (Geschäftsvorfälle) erhöht die Authentizität der Prüfungsaufgaben und stellt damit sicher, dass die Auszubildenden sich inhaltlich mit dem Prüfungsinhalt identifizieren können.
- 2) Die fachliche Bündelung von Aufgaben erlaubt die Konstruktion von Kompetenzmodellen. Hierdurch wird ein kriterienorientiertes Prüfverfahren möglich, über das plausibel beschrieben werden kann, wie Bestehens- und/oder Leistungsgrenzen inhaltlich begründet gesetzt werden sollten.
- 3) Die einzelnen Aufgabenschwierigkeiten, die in eine Gesamtprüfung einfließen, lassen sich bereits bei der Aufgabenentwicklung abschätzen (Möglichkeit der

Vermeidung von Decken- und Bodeneffekten), sodass die Diagnostik mittels Prüfung verbessert wird.

Gerade der Punkt 3) hat sich als besonderes wirksam erwiesen. Es wurden daher Designprinzipien entwickelt und empirisch geprüft, die eine gute Prognostik der Aufgabenschwierigkeiten befördern. Folgende Befunde lassen sich hier zusammenfassen (u. a. Winther, 2010; Winther & Klotz, 2016; Deutscher & Winther, 2018; Winther et al., 2023):

Die Schwierigkeit kompetenzorientierter Aufgaben lässt sich über drei Ebenen modellieren:

- Die funktionale Modellierung bewertet Prüfungsaufgaben dahingehend, wie anspruchsvoll es ist, die zur Lösung notwendigen Schritte aus der zugrunde liegenden Anforderungssituation zu extrahieren. Die Aufgabenkonstruktion wird hier auf **Kontextebene** beurteilt.
- Die inhaltliche Wissenskomplexität der Domäne bewertet Prüfungsaufgaben hinsichtlich des zu ihrer Lösung benötigten (Fach-)Wissens. Die Aufgabenkonstruktion wird hier auf **Inhaltsebene** beurteilt.
- Die Art der kognitiven Aktivierung bewertet Testaufgaben mit Blick auf die zu ihrer Lösung eingeforderten Leistungsfähigkeiten der Auszubildenden. Die Aufgabenkonstruktion wird hier auf kognitiver **Prozessebene** beurteilt.

Vor dem Hintergrund dieser Befunde war die zentrale Annahme des Verbundvorhabens ASPE, dass eine Professionalisierung des Aufgaben- und Prüfungserstellungsprozesses entlang der drei Ebenen der Aufgabenschwierigkeit die AkA dabei unterstützt, die psychometrische Qualität der Prüfungsaufgaben (und der Gesamtprüfung) zu erhöhen und die am Prozess beteiligten Akteur:innen zu entlasten (s. Abbildung 1).



Abbildung 1: Projektziele des Verbundvorhabens ASPE

Die Rahmung des Professionalisierungsprozesses wurde durch die digitale ASPE-Workbench gewährleistet. Die Workbench ist ein Professionalisierungs-Toolset, das verschiedene Formate zur Weiterentwicklung der Prüfungspraxis enthält.

1.2 Die digitale Workbench-ASPE als Professionalisierungs-Toolset

Die ASPE-Workbench ist ein Online-Tool für das Prüfer:innen-Ehrenamt und übersetzt modelltheoretische und psychometrische Kompetenzannahmen in Geschäftsprozesse und zugehörige Aufgaben – also wie Modelle und psychologische Messungen genutzt werden können, um in Geschäftsprozessen die Leistung (Fähigkeiten, Eigenschaften oder Kenntnisse) einer Person zu bewerten. Die Workbench hilft dabei, Prüfungsaufgaben für kaufmännische Ausbildungsberufe kompetenzorientiert und modellkonform zu konstruieren und schlüssig in eine Gesamprüfung zu überführen. Sie bildet den gesamten Workflow zwischen der zentralen Aufgabenstelle für kaufmännische Abschluss- und Zwischenprüfungen (AkA) und den Aufgabenentwickler:innen ab.

Die Workbench hat den Anspruch, ein sich permanent weiterentwickelndes Tool zu sein, das integrierte Informations- und Weiterbildungsangebote, verbesserte Kommunikationskanäle sowie größtmögliche Transparenz im Workflow zwischen den Aufgabenentwickelnden und den verantwortlichen Prüfungsbeauftragten anbietet und auf die Bedarfe ehrenamtlicher Prüfender anpasst (s. Abbildung 2).

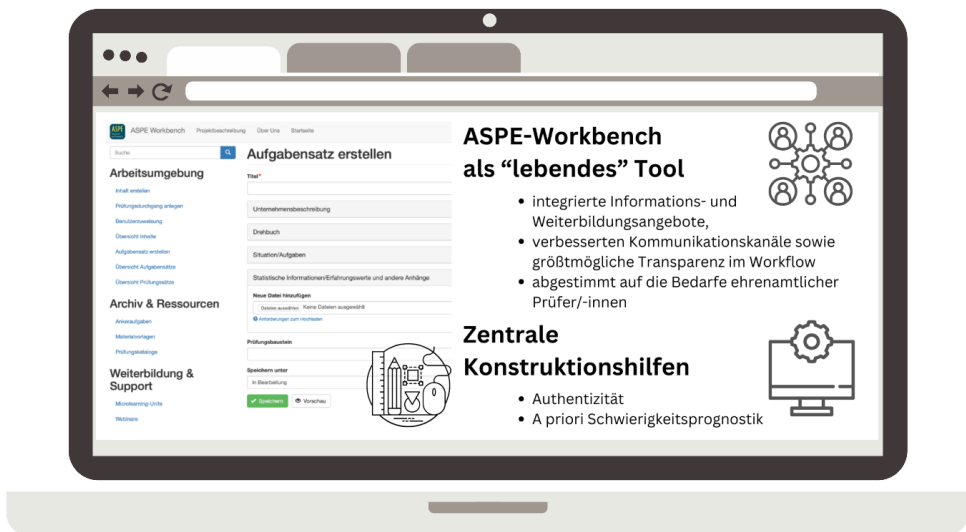


Abbildung 2: ASPE-Workbench als zentraler Projektoutput

Die ASPE-Workbench umfasst folgende Funktionalitäten:

- Im Bereich „Arbeitsumgebung“ finden die Aufgabenentwickelnden unter dem Reiter „Zugewiesene Inhalte“ eine Übersicht über die Inhalte, welche von ihnen entwickelt werden sollen. Neben der Erstellung von Prüfungsaufgaben werden

die Ehrenamtlichen auch gebeten, weitere Prüfungsteile, wie beispielsweise Unternehmensbeschreibungen, zu erstellen. Unter dem Reiter „Inhalt erstellen“ sind deshalb unterschiedliche Eingabemasken wie Unternehmensbeschreibungen, Drehbücher, Situationen und Prüfungsaufgaben zu finden. Unter dem Reiter „Alle Inhalte“ wird den User:innen eine Übersicht über die eigenen sowie von anderen Prüfungserstellenden freigegebenen Inhalte angezeigt. Durch den Zugriff auf freigegebene Inhalte anderer Prüfungserstellender kann Kooperation und Zusammenarbeit entstehen. Hier kann zwischen der Berechtigung zur Einsicht und Kommentar oder Überarbeitung gewählt werden, d. h. die Erstellenden können entscheiden, ob sie den Prüfungsinhalt lediglich zur Ansicht und Kommentar oder zur Überarbeitung für eine ausgewählte andere Person freigeben. In den Reitern „Aufgabensatz“ und „Prüfungssatz“ kann Einsicht in die von den Prüfungsverantwortlichen erstellten „Aufgaben- oder Prüfungssätze“ genommen werden. Dies führt zu mehr Transparenz im gesamten Prüfungserstellungsprozess.

- Im Bereich „Archiv und Ressourcen“ finden sich unter den Reitern „Ankeraufgaben“, „Materialvorlagen“ und „Prüfungskataloge“ unterschiedliche Inhalte, die die Erstellung von Prüfungsinhalten unterstützen. „Ankeraufgaben“ sind exemplarische, nach wissenschaftlichen Kriterien erstellte Musteraufgaben, unter „Materialvorlagen“ finden sich Vorlagen zur Erstellung authentischen Materials zu einer Prüfungsaufgabe. Die „Prüfungskataloge“ ermöglichen eine Übersicht über die Prüfungsbereiche mit einer Gliederung nach Gebieten, Funktionen, Fragenkomplexen, Themenkreisen und Beispielen für betriebliche Handlungen, welche die Grundlage für die Entwicklung von Prüfungen bildet.
- Im Bereich „Weiterbildung und Support“ können die Aufgabenerstellenden auf „Microlearning Units“, „Tutorials“ sowie auf ein „Seminar“ zugreifen. In den „Microlearning-Units“ werden die schwierigkeitsgenerierenden Merkmale einer Prüfungsaufgabe thematisiert, in den „Tutorials“ mit Hilfe von Screencasts und PDF-Materialien in die Benutzung der digitalen Workbench eingeführt, im „Seminar“ wird die Möglichkeit eines Selbststudiums zum Thema „Erstellung kompetenzorientierter Prüfungsaufgaben“ gegeben.

Die Workbench wurde auf Basis von User Cases erstellt. Hierbei wurde deutlich, dass die ehrenamtlichen Prüfungserstellenden insbesondere bei der Konstruktion authentischer Prüfungsaufgaben sowie bei der a-priori-Bewertung der Schwierigkeitsgrade von Aufgaben (digitale) Unterstützung suchen. Insbesondere im Hinblick auf diese Herausforderungen bietet die Workbench automatisierte Algorithmen und Designkriterien an, die die Gesamtqualität der Prüfung erhöhen.

2 Designkriterien zur Qualitätssicherung

2.1 Das Denken in Aufgabenschwierigkeiten

Kompetenzorientierung als Leitbild der Aufgaben- und Prüfungserstellung bedingt einen Paradigmenwechsel; die Denkrichtung, aus der Aufgaben konstruiert werden, ändert sich und orientiert sich an den folgenden Leitfragen für eine kompetenzorientierte Aufgaben- und Prüfungs konstruktion:

- 1) Über welche Aufgaben bzw. Aufgabenbündel können die Auszubildenden ihr Potenzial, kompetent in beruflichen Situationen zu handeln, zeigen?
- 2) Was sind die empirisch beschreibbaren Eigenschaften dieser Aufgaben? Oder anders formuliert: Welche Aufgabenmerkmale erschweren oder vereinfachen die Lösung für die Auszubildenden systematisch?

Die Frage 1) adressiert die Praxis- und Handlungsorientierung als Grundanforderung der Aufgabenerstellung. Kompetenzorientierte Aufgaben fordern von Auszubildenden, berufsrelevante Probleme zu lösen und dabei ihre Fähigkeiten und Fertigkeiten gezielt anzuwenden. Für die Konstruktion der Prüfungsaufgabe ist hierbei eine Situationsvorgabe mit Praxisbezug und Adressatenorientierung tragend, die durch authentische Materialien unterstützt wird. Die Auszubildenden sind in die Situation eingebunden und werden aufgefordert, spezifische Handlungselemente der Situationsbearbeitung anzusteuern und Lösungen zu generieren. Diese Grundanforderung ist eng an das Konzept der Authentizität gebunden (Winther et al., 2023). Es zeigt sich, dass in der Praxis Misskonzeptionen der Authentizität dominieren; es erfolgt regelmäßig eine Gleichsetzung zwischen Realität und Authentizität – dies erschwert den Zugang zu kompetenzorientierten Konstruktionslogiken und setzt Konstruktionshürden unverhältnismäßig hoch. In der ASPE-Workbench werden zur Überwindung dieser typischen „Wissenschafts-Praxis-Diskrepanz“ Blaupausen von authentischen Materialien bereitgestellt, die den Konstruktionsprozess direkt anleiten und den Unterschied zwischen realen Arbeitsplatzsituationen und authentischen Situationsaufgaben als Baustein einer kompetenzorientierten Aufgabe herausstellen. Hierbei sind die Aufgaben typischerweise so gestaltet, dass Auszubildende in eine authentische Arbeitssituation eingebettet werden – ein Konzept, das den Kern der *Situiertheit* ausmacht. Die Situiertheit hat ihre Verankerung in der *Situated Learning Theory*, die davon ausgeht, dass Lernen und Wissenserwerb durch authentische Kontexte und Interaktionen gefördert werden (Lave & Wenger, 1991). Es wird herausgestellt, dass die Bewältigung situierter Aufgaben die Anwendung von Fertigkeiten und Fähigkeiten in einem dynamischen Arbeitskontext voraussetzt (Bjørnavold, 2000; Rychen & Salganik, 2003; CEDEFOP, 2003).

Frage 2) berührt den Konstruktionsprozess direkt. Hierbei wird zunächst angenommen, dass jeder inhaltliche Aspekt sowohl über eine einfache als auch eine schwierige Aufgabe modelliert und geprüft werden kann. Der Unterschied besteht darin, wie die Situationsvorgabe sowie die in ihr notwendigen Handlungselemente zur Abbildung des Lösungsraums konstruiert sind und welchen Einfluss sie auf das gezeigte Lösungsverhalten haben (werden).

Kompetenzorientierte Aufgaben bestehen im Kern folglich aus zwei übergeordneten Bestandteilen: einer beruflichen Situation und dem in dieser zu zeigenden Verhalten. Diese beiden Faktoren sind entscheidend, um Authentizität zu gewährleisten und das Verhalten der Auszubildenden im Kontext ihrer beruflichen Fähigkeiten zu prüfen.

- Situation: Die Aufgabe ist so konzipiert, dass sie eine berufliche Realität simuliert, beispielsweise durch die Nachbildung eines Geschäftsprozesses oder einer anderen typischen Arbeitssituation. Diese Situierung fördert die Authentizität der Aufgabe, die ein zentrales Gütekriterium für die prognostische Validität darstellt. Denn nur authentische Aufgaben ermöglichen es, Rückschlüsse auf die tatsächliche berufliche Leistung der Auszubildenden zu ziehen (Winther, 2010).
- Verhalten: Das Verhalten, das von den Auszubildenden zur Lösung der Aufgabe gezeigt werden muss, basiert auf den individuellen Fähigkeiten und erlernten Fertigkeiten.

Auf beide Bestandteile kann durch Designkriterien der Aufgabenkonstruktion bewusst Einfluss genommen werden; diese Designkriterien modellieren gleichsam die Schwierigkeit der Aufgabe. Hierbei gilt: Die Schwierigkeit einer Aufgabe ist keine inhärente Eigenschaft des Aufgabenmaterials, sondern entsteht erst in der Interaktion zwischen Aufgabe und Bearbeitendem; die Schwierigkeit einer Aufgabe wird erst im Kontext der kognitiven Kapazität, Vorkenntnisse und des aktuellen Wissensstandes der Auszubildenden als „schwer“ oder „leicht“ wahrgenommen. Damit wird eine (ordinale) Stufung der Schwierigkeit unterstellt. Für das in ASPE gewählte Modell wird jedes Designkriterium in drei Ausprägungen definiert, wodurch sich jede Prüfungsaufgabe in einer 3×3 -Schwierigkeitsmatrix verorten lässt (s. Abbildung 3).

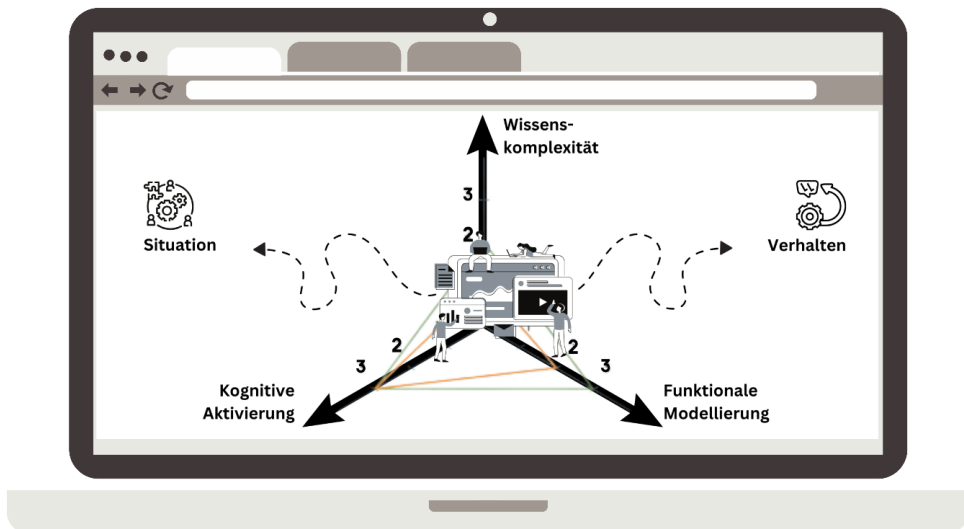


Abbildung 3: Schwierigkeitsbeschreibende Designkriterien

Nachfolgend werden die einzelnen Kriterien beschrieben.

2.1.1 Designkriterium „Wissenskomplexität der Domäne – Spezifität“

Das Designkriterium „Wissenskomplexität der Domäne“ beschreibt die Art und Tiefe des Wissens, das zur Bewältigung einer authentischen Aufgabe erforderlich ist. Dieser Faktor analysiert, welche und wie viele Wissensbereiche einer Domäne zur Bearbeitung einer Aufgabe relevant sind und ob diese Wissensbereiche in einer hierarchischen Beziehung zueinanderstehen. **Das Designkriterium zielt folglich auf das zugrunde liegende inhaltliche Anforderungsprofil des zu entwickelnden Items ab.** In Anlehnung an Gelman & Greeno (1989) wird zwischen domänenverbundenem und domänenspezifischem Wissen unterschieden. Je domänenspezifischer ein Item entwickelt wird, desto spezifischeres kaufmännisches Wissen bezüglich der Inhalte oder Kenntnisse über domänenspezifische Arbeitsweisen und Methoden sind notwendig, um eine Aufgabe lösen zu können. Unspezifische Items greifen auf generisches Wissen und Fähigkeitsstrukturen zurück, die jedoch eine Voraussetzung für das Lösen domänenverbundener Problemstellungen darstellen. Ziel dieses Kriteriums ist es, zu erklären, warum der Abruf von allgemein verfügbarem, domänenverbundenem Wissen in manchen Fällen leichter oder schwieriger sein kann als der Abruf von domänenspezifischem Wissen, das nur in einem Fachkontext erworben wird. Der Transfer von allgemeinen, domänenverbundenen in domänenspezifische Kompetenzen kann maßgeblich von der Kontextualisierung der Aufgabenstellung abhängen (hierzu u. a. Hering et al., 2020; 2021). Arbeiten von Deutscher et al. (2022) weisen darauf hin, dass Aufgaben, die spezifisches Fachwissen (domänenspezifisches Wissen) erfordern, als schwieriger empfunden werden. Die Studien im Rahmen des Verbundvorhabens ASPE zeichnen ein differenziertes Bild: Es kann beispielsweise gezeigt werden, dass etwa mathematische Aufgaben, die den korrekten Einsatz in einer zuvor in der Schule gelernten Formel verlangen, herausfordernder sein können als eine domänenspezifische Aufgabe, die auf ein grundlegendes Berufswissen (z. B. Angebotsrecht) abzielt. Dies weist darauf hin, dass nicht nur die Art des Wissens, sondern auch seine Abrufbereitschaft (d. h., wie leicht und sicher es aus dem Gedächtnis abgerufen werden kann) die Schwierigkeit beeinflusst. Wissen, das routinemäßig angewendet wird, ist häufig abrufbarer und sicherer enkodiert als selten abgerufenes Wissen (Baddeley, 2012). Zudem zeigt sich, dass die Schwierigkeit einer Aufgabe steigt, wenn Wissen aus mehreren Wissensgebieten der Domäne kombiniert werden muss (Winther, 2010; hierzu auch Sweller, 1988; Sweller et al., 1998). Dies unterstreicht, dass der Schwierigkeitsgrad einer Aufgabe nicht nur durch den Wissensbereich, sondern auch durch die erforderliche Kombination von Wissensteilbereichen beeinflusst wird.

Zur systematischen Einordnung der „Wissenskomplexität der Domäne“ und deren Auswirkung auf die Aufgabenschwierigkeit wird das folgende dreistufige Modell definiert:

- Stufe 1: Die Aufgabe kann mit allgemein verfügbarem Wissen und Grundfähigkeiten wie Lesen oder Rechnen gelöst werden. Die Wissensanforderungen sind minimal und umfassen Kenntnisse, die im schulischen Kontext erworben und routinemäßig genutzt werden. Diese Aufgaben greifen auf das alltägliche, domä-

nenverbundene Wissen zurück, das als allgemein verfügbar gilt. Diese Stufe charakterisiert Anforderungen, die keine spezifischen, unterrichtsbezogenen Vorkenntnisse zur Lösung eines Items voraussetzen.

- Stufe 2: Hier wird spezifisches Wissen aus einem bestimmten Bereich abgefragt, das zuvor erlernt und enkodiert wurde. In der Berufsausbildung ist dies häufig domänenspezifisches Wissen, das nicht zu den allgemein zugänglichen Wissensressourcen gehört, sondern nur durch gezielte Ausbildung erworben wurde. Die Aufgabe erfordert klar definierte Wissensinhalte, ohne die eine Lösung nicht möglich ist. Für diese Stufe ist die Verfügbarkeit von Wissen, das auch außerhalb des spezifischen Berufskontextes liegen kann, aber spezifisch und erlernt sein muss, wesentlich.
- Stufe 3: Diese Stufe umfasst vertieftes oder spezialisiertes Wissen, das entweder auf vorherigem (Fach-)Wissen aufbaut oder spezifische Nischenkenntnisse umfasst. Dies kann Wissen umfassen, das hierarchisch strukturiert ist und aufeinander aufbauende Ebenen von Wissen erfordert. Solches Wissen wird regelmäßig nur in bestimmten beruflichen Kontexten erworben. Für diese Stufe ist von Bedeutung, dass sie auf Anforderungen abzielt, für deren Lösung domänenspezifische Wissenskombinationen einzubringen sind.

Für einen gelingenden Aufgabenerstellungsprozess ist entscheidend, diese Stufen möglichst zufallsfrei identifizieren zu können. So kann sich eine Aufgabe als leichter erweisen, wenn sich herausstellt, dass nur minimaler Wissensabruf erforderlich ist. Andererseits kann eine Aufgabe komplexer werden, wenn die Aufgabenerstellenden übersehen, dass die Lösung vertiefte oder spezialisierte Wissensinhalte einfordert.

2.1.2 Das Designkriterium „Kognitive Aktivierung – Verständnistiefe“

Das Designkriterium „Kognitive Beanspruchung – Verständnistiefe“ beschreibt das Maß, in dem sich die Auszubildenden bei der Lösungsdarbietung inhaltlich mit der Thematik auseinandersetzen müssen. Die Spannbreite reicht vom isolierten Erkennen, Wiedergeben und Benennen von Wissen bis hin zur Analyse, Evaluation und strategischen Entscheidungsfindung als Lösungsstrategie (Bloom, 1956). Dieses Kriterium bezieht sich ausschließlich auf die Tiefe der kognitiven Aktivierung, die zur Präsentation einer Lösung erforderlich ist; es beschreibt, inwiefern das Wissen im Lösungsprozess verknüpft, reflektiert und begründet wird. **Das Designkriterium zielt folglich auf das kognitive Aktivierungspotenzial des zu entwickelnden Items ab und ist ein Indikator für den Grad der zur Lösung eingeforderten Verständnistiefe.** Je mehr kognitive Ressourcen für das Bearbeiten eines Testitems benötigt werden, desto komplexere kognitive Prozesse werden durch diese adressiert. Um kognitive Prozesse für die Itementwicklung modellieren zu können, werden Taxonomien benötigt (hier u. a. für ASPE in Anlehnung an Marzano & Kendall, 2007; Kendall & Marzano, 2007). Die in der Literatur verfügbaren Taxonomien folgen den Vorschlägen Tylers (1946):

„The most useful form for stating objectives is to express them in terms which identify the kind of behavior to be developed in the student and the content [...] in which the behavior is to operate“ (Tyler, 1946, zit. n. Marzano & Kendall, 2007, S. 17).

Für die Aufgabenkonstruktion bedeutet dies, dass sich mithilfe der Taxonomien Lern- und Arbeitsanforderungen durch die Art des Wissens und die zum Tragen kommenden kognitiven Prozesse beschreiben lassen, auf die die Aufgabe hin ausgerichtet ist. Die Schwierigkeit einer Aufgabe ergibt sich aus der Komplexität der kognitiven Kontrollprozesse, die die Auszubildenden bei der Aufgabenbewältigung aktivieren müssen. Die Komplexität wird bestimmt (a) über den angesprochenen kognitiven Prozess selbst und (b) über den Grad der Vertrautheit der zugrunde liegenden Anforderung.

Zur Strukturierung der „Kognitiven Aktivierung“ in Bezug auf die Aufgabenschwierigkeit wird das folgende dreistufige Modell definiert:

- Stufe 1: Diese Stufe erfordert lediglich das isolierte Nennen, Wiedergeben oder Erkennen einer Lösung. Hier genügt es, eine Antwort ohne weitere Erklärungen zu liefern, z. B. indem ein einzelner Grund für die Ungültigkeit eines kaufmännischen Angebots genannt wird. Ein weiteres Beispiel wäre, wenn in einer Mathematikaufgabe nur Zahlen in eine vorgegebene Formel eingesetzt werden müssen. Das Verständnis für den inhaltlichen Zusammenhang oder die Notwendigkeit einer Erklärung entfällt; es geht nur um das reine Abrufen von Fakten oder das Anwenden einfacher Schritte in einer vertrauten Anforderungssituation.
- Stufe 2: Auf dieser Stufe muss die Lösung nicht nur genannt, sondern auch erläutert werden. Hier wird eine Verständnistiefe verlangt, bei der Zusammenhänge hergestellt und Begründungen gegeben werden. Die Lösung muss so dargelegt werden, dass eine außenstehende Person die Relevanz und die Beziehung zwischen den Wissenselementen nachvollziehen kann. Ein Beispiel im kaufmännischen Kontext wäre die Erklärung, warum das Einhalten einer Angebotsfrist zu einem rechtsgültigen Kaufvertrag führt.
- Stufe 3: Diese Stufe fordert eine noch tiefere Auseinandersetzung. Die Prüflinge müssen den Sachverhalt analysieren, die einzelnen Elemente zerlegen und evaluieren, um eine strategische Entscheidung zu treffen oder eine Lösung fundiert zu begründen. Im kaufmännischen Kontext könnte dies bedeuten, dass die Prüflinge abwägen, ob eine Angebotsfrist gesetzt werden sollte, indem Vor- und Nachteile ins Verhältnis gesetzt und im Zusammenhang der eigenen Unternehmensstrategie analysiert werden.

Der Aufgabenerstellungsprozess profitiert dann von diesem Designkriterium, wenn der Fokus auf der Frage liegt, was die Auszubildenden zur Lösungserbringung leisten müssen. Im Praxisfeld liegen allgemeine Kenntnisse im Umgang mit kognitiven Taxonomien vor, die sich primär aber auf Formulierungen im Itemstamm („Benenne drei Faktoren [...] Begründe Deine Auswahl [...]“) konzentrieren. Im Konstruktionsprozess können diese Formulierungen eine Orientierung bieten; sie sollten aber immer entlang der tatsächlich geforderten Verständnistiefe bewusst genutzt werden. Die Verständnistiefe misst, wie intensiv die Auszubildenden die Wissenselemente in die Lösungsfindung einbeziehen, verknüpfen und reflektieren. Oder anders formuliert: Welchen Grad an kognitiver Aktivität fordert die Aufgabe ein?

2.1.3 Das Designkriterium „Funktionale Modellierung – Komplexität“

Das Designkriterium „Funktionale Modellierung – Komplexität“ beschreibt die Struktur und Architektur einer Aufgabe. Eine Aufgabe mit niedriger Komplexität ermöglicht einen direkten Lösungsweg ohne Ablenkungen, während eine komplexe Aufgabe zusätzliche kognitive Operationen erfordert, die den Arbeitsfluss der Prüflinge behindern (hierzu u. a. Ansätze der Cognitive Load Theory; Sweller, 1988; für die didaktische Modellierung auf Itemebene vgl. Winther, 2010). Die Prüflinge müssen die Aufgabe unter Umständen in Teilschritte zerlegen und mit unerwarteten Interferenzen und Distraktoren umgehen, die den eigentlichen Lösungsweg verdecken. **Das Designkriterium zielt folglich auf das zugrunde liegende funktionale Anforderungsprofil des zu entwickelnden Items ab.** Hierbei wird zwischen kohärenten Handlungsabläufen und solchen mit zusätzlichen Interferenzen unterschieden. Kohärente Aufgaben bestehen aus logischen, aufeinander abgestimmten Schritten, die zur Lösung gehören und damit keine ablenkenden Operationen erfordern. Beispielsweise wäre das Schreiben eines Geschäftsbriefs als kohärenter Ablauf nicht als komplex zu bewerten, da die Schritte wie Anrede, Einleitung, Hauptteil und Abschluss klar strukturiert sind und keine Störung des Arbeitsablaufs darstellen. Das funktionale Anforderungsprofil erhöht sich erst dann, wenn Interferenzen oder Distraktoren eingebaut sind, die von den Auszubildenden zusätzliche Operationen einfordern. Ein Beispiel ist eine Rechenaufgabe, bei der zur Lösung Einheiten umgerechnet oder unterschiedliche Informationen berücksichtigt werden müssen. Solche zusätzlichen Anforderungen erhöhen die Schwierigkeit einer Anforderungssituation, da sie den Aufwand zur Lösungs(weg)findung dadurch steigern, dass die Prüflinge ihre Aufmerksamkeit auf unerwartete und möglicherweise irrelevante Informationen lenken müssen und diese als solche zu identifizieren sind.

Zur Strukturierung der „funktionalen Modellierung“ und ihrer Auswirkung auf die Aufgabenschwierigkeit wird das folgende dreistufige Modell definiert:

- Stufe 1: Die Aufgabe verlangt keine Modellierungsleistung; die Lösung kann direkt dargeboten werden, ohne dass die Auszubildenden bei der Lösungsfindung auf ablenkende Informationen oder zusätzliche Operationen achten müssen.
- Stufe 2: Die Aufgabe enthält einen einzelnen Distraktor, der die Bearbeitung erschwert und eine zusätzliche kognitive Operation erfordert. Ein solcher Distraktor zwingt die Auszubildenden, kurzzeitig eine zusätzliche Informationsverarbeitung vorzunehmen, bevor sie zur eigentlichen Lösungsfindung zurückkehren.
- Stufe 3: Die Aufgabe enthält mehrere unterschiedliche Distraktoren, die jeweils eine separate kognitive Operation auslösen und in Sequenz bearbeitet werden müssen. Jeder Distraktor erfordert eine neue Operation, sodass der Aufwand zur Lösungs(weg)findung sowie die Lösungskomplexität erheblich steigen.

In der Aufgabenkonstruktion kann dieses Designkriterium gezielt dazu genutzt werden, die Schwierigkeit einer Aufgabe kleinschrittig zu justieren, indem Distraktoren und ablenkende Informationen eingebettet werden. Hierbei ist darauf zu achten, dass

tatsächliche Interferenzen erzeugt und kognitive Zusatzoperationen notwendig werden – ohne diese Effekte bleibt ein Anstieg des Anforderungsniveaus aus.

2.2 Designkriterien in der Anwendung

Zur Darstellung der Plausibilität der Designkriterien werden nachfolgend zwei exemplarische Aufgaben präsentiert, die anhand der Kriterien bewertet und dann im Hinblick auf ihre empirische Schwierigkeit analysiert werden (s. Abbildung 4).

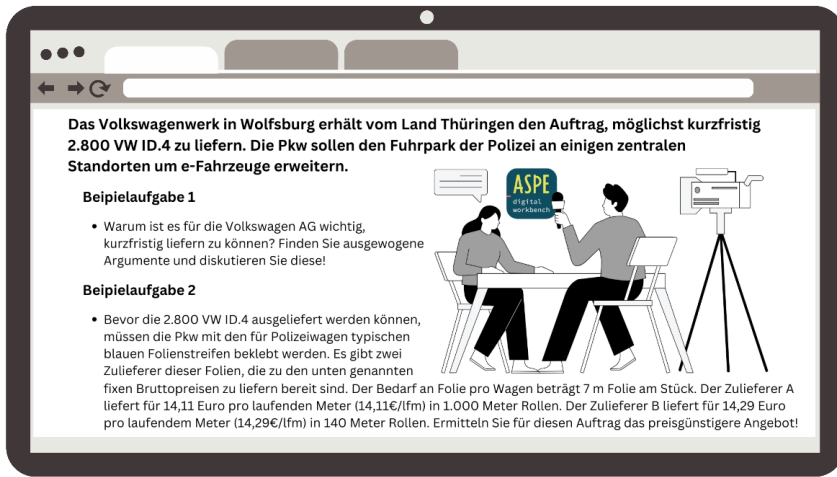


Abbildung 4: Beispielaufgaben zur Veranschaulichung der Designkriterien

Die Beispielaufgaben sind so gewählt, dass sie allgemein verständlich sind. Sie repräsentieren Anforderungen, die sich im Curriculum der Ausbildungsberufe Industriekaufmann/-frau sowie Kaufmann/-frau für Spedition und Logistikdienstleistungen identifizieren lassen und gleichsam exemplarisch für den kaufmännischen Eingangunterricht stehen.

2.2.1 Schwierigkeitsbewertung der Aufgaben durch Fachausschüsse

Die Beispielaufgaben wurden zur Veranschaulichung der Designkriterien im Rahmen eines ASPE-Ankeraufgaben-Events eingesetzt. Das Ankeraufgaben-Event hatte zum Ziel, allgemeingültige Standards der Prüfungsaufgaben festzulegen. Der Prüfungserstellungsprozess ist vor allem unter dem Gesichtspunkt der Justiziabilität standardisiert. Mit dem ASPE-Projekt werden designkritische Standards (die o.g. modellierbaren Schwierigkeitsparameter) ergänzt. Hierfür wurden aus Perspektive der Wissenschaft konstruierte, kompetenzorientierte Beispielaufgaben exemplarisch durch Akteur:innen des Aufgaben- und Prüfungserstellungsprozesses hinsichtlich ihrer Schwierigkeit bewertet. Am Ankeraufgaben-Event haben Mitglieder aus zwei Fachausschüssen (Industrie: $n_{IK}=12$; Spedition: $n_{SK}=9$) teilgenommen.

Alle Teilnehmenden haben die Aufgaben individuell entlang der 3×3 -Schwierigkeitsmatrix geratet. Tabelle 1 zeigt eine Übersicht der erzeugten Ratings, differenziert nach Fachausschüssen. Es zeigt sich zweierlei: Zum einen sind die Ratings in den Fachausschüssen überwiegend homogen; die Einschätzungen in den zwei Berufsgruppen unterscheiden sich nicht signifikant. Zum anderen kommen die Mitglieder der Fachausschüsse mit dem Ratingsystem hinreichend gut zurecht; die Interdecoder-Reliabilitäten – hier Krippendorffs Alpha – weisen akzeptable Werte auf ($\alpha \geq 0,67$). Der große Vorteil des Krippendorff-Kriteriums liegt darin, dass die Angabe des 95 %-Konfidenzintervalls Aufschluss über die Präzision der Reliabilitätsmessung gibt (Hayes & Krippendorff, 2007; Krippendorff, 2004).

Tabelle 1: Rating und empirische Befunde zu den Beispielaufgaben

Designkriterien und Aufgabenschwierigkeit	Industrie ($n_{IK} = 12$)		Spedition ($n_{SK} = 9$)	
	Beispiel 1 mean/modus	Beispiel 2 mean/modus	Beispiel 1 mean/modus	Beispiel 2 mean/modus
Wissenskomplexität	2.87/3.00	1.95/2.00	2.91/3.00	2.06/2.00
Kognitive Aktivierung	2.92/3.00	2.39/2.00	2.88/3.00	2.42/2.00
Funkt. Modellierung	2.21/2.00	2.45/2.00	2.25/2.00	2.32/2.00
Erwartete Schwierigkeit (Darstellung als Summenscore mit min = 3 und max = 9)			Beispiel 1	8.00
			Beispiel 2	6.00
Krippendorffs Alpha			Beispiel 1	.80 / [.73;.87]
(Darstellung des Koeffizienten und des 95 %-Konfidenzintervalls)			Beispiel 2	.75 / [.71;.79]

Die Beispielaufgabe 1 wird entlang der 3×3 -Schwierigkeitsmatrix mit einer 8 geratet. Die Designkriterien „Wissenskomplexität in der Domäne“ und „Kognitive Aktivierung – Verständnistiefe“ werden auf der höchsten Stufe (= 3) verortet; die funktionale Modellierung wird auf der Stufe 2 eingeschätzt. Die Fachausschussmitglieder stellen bei ihrem Rating in Rechnung, dass eine fachlich einschlägige Antwort auf Wissen basiert, das hierarchisch strukturiert sein muss, um zwischen Binnen- und Außenverhältnissen bei einer Auftragsabwicklung zu differenzieren (hohe Wissenskomplexität). Diese Differenzierung ist notwendig, um ausgewogene Argumente zu finden und angemessen diskutieren zu können. Die Prüflinge müssen hierzu nicht nur die Interdependenzen der Auftragsabwicklung, sondern auch die beteiligten Marktpartner:innen kennen. Hierzu ist es hilfreich, das Unternehmen in einem komplexen ökonomischen System verorten zu können (hohe kognitive Aktivierung). Dass die Auftragsabwicklung einen staatlichen Akteur enthält, ist eine Information, die die Aufgabenbearbeitung erschwert (mittlere funktionale Modellierung).

Die Beispielaufgabe 2 wird in allen drei Designkriterien mit 2 eingeschätzt. Das Urteil der Fachausschussmitglieder basiert darauf, dass die Bewältigung der Aufgabe mathematische Grundfertigkeiten voraussetzt, die kontextualisiert an die Auftragsabwicklung gebunden sind. Diese Grundfertigkeiten müssen erlernt und dekontextualisiert in einem neuen – hier kaufmännischen – Anwendungskontext zur Verfügung

stehen, um in die Lösung eingebracht werden zu können (= mittlere Wissenskomplexität). Gleichsam stellt die Aufgabe die Auszubildenden vor die Herausforderung, unterschiedliche Bezugsgrößen (Längenmaße, Preisangaben) miteinander ins Verhältnis zu setzen (= mittlere kognitive Aktivierung) und zu erkennen, dass ausgewiesene Preisvorteile durch ungünstige Mengeneinheiten aufgezehrt werden (= mittlere funktionale Modellierung).

2.2.2 Schwierigkeitsbewertung der Aufgaben auf Basis einer Einpunktmessung

Den Schwierigkeitsratings im Rahmen des ASPE-Ankeraufgaben-Events lassen sich empirische Schwierigkeiten gegenüberstellen. Die Beispielaufgaben wurden als Einpunktmessung im Oktober 2020 zur Vorwissensdiagnostik bei 340 Industriekaufleuten zu Ausbildungsbeginn eingesetzt. Hiervon sind 162 weiblich und 178 männlich. Das Alter der Proband:innen liegt zwischen 17 und 30 Jahren. 29 Prozent der Lernenden verfügen zu Beginn ihrer Ausbildung bereits über eine kaufmännische Vorbildung (höhere Handelsschule, Wirtschaftsgymnasium, vergleichbare duale Ausbildung im kaufmännischen Bereich).

Die Einpunktmessung zur Vorwissensdiagnostik repräsentiert über insgesamt 27 Testitems eine Logitskala von -4.135 bis 3.505 Logits. Die Items sind gleichmäßig zwischen den Skalenendpunkten – mit einer leichten Häufung im mittleren Schwierigkeitsbereich – verteilt. Die Verteilung der Personenfähigkeiten der Vorwissensdiagnostik ist mit einem Populationsmittelwert von $-.869$ (.127) und einer Standardabweichung von 2.373 beschrieben. Die WLE-Reliabilität ist mit $.644$ adäquat. Im Hinblick auf die Qualität der Items ist zusammenfassend festzustellen, dass die gewichteten Abweichungsquadrate (weighted MNSQ; Wright & Masters, 1982) für alle Items innerhalb der zulässigen Grenzen von $.75$ bis 1.33 (Adams & Khoo, 1996; Bond & Fox, 2001; Chang, Lin & Lin, 2007) liegen und symmetrisch um den exakten Fit von 1.00 verteilt sind. Dabei variieren die Infit-Schätzer der Vorwissens-Skala zwischen $.92$ und 1.23 . Vor dem Hintergrund dieser Psychometrik lassen sich die exemplarischen Aufgaben entsprechend verorten und analysieren (s. Tabelle 2). So steht die Beispielaufgabe 1 für ein stark am Curriculum ausgerichtetes Item. Ziel der Aufgabe ist es, zu diskutieren, warum eine fristgerechte Lieferung für das produzierende Unternehmen von entscheidender Bedeutung ist. Die Beispielaufgabe erfasst damit das konzeptuelle Verständnis der Lernenden im Hinblick auf die Innen- und Außenbeziehungen eines Unternehmens im Rahmen der Auftragsabwicklung. Dieser Lerninhalt wird im kaufmännischen Eingangsunterricht intensiv vermittelt. Es ist folglich zu vermuten, dass diese Aufgabenstellung Lernende mit kaufmännischer Vorbildung systematisch bevorteilt und von Lernenden, die noch keine kaufmännische Vorbildung aufweisen, schwer zu lösen ist. Die Beispielaufgabe 2 kann hingegen auf Basis einfacher mathematischer Prozeduren gelöst werden. Sie ist über den Itemstamm zwar formal an die Auftragsabwicklung gebunden, kann aber losgelöst von dieser bearbeitet werden. Vor diesem Hintergrund ist anzunehmen, dass Lernende mit kaufmännischer Vorbildung keine Vorteile bei der Aufgabenlösung haben.

Tabelle 2: Empirische Analyse der Beispielaufgaben

	Schwierigkeit in Logit	DIF-Schätzer	Standardfehler	Geratete Referenz
Beispiel 1	3.505	-.774	.247	8 [3;9]
Beispiel 2	1.475	-.335	.166	6 [3;9]

In der empirischen Analyse zeigt sich, dass die Beispielaufgabe 1 – wie erwartet – eine schwierige Aufgabe ist (Itemschwierigkeit 3,505 Logits). Auszubildende mit spezifischen kaufmännischen Vorkenntnissen haben substanzielle Vorteile, diese Aufgabe lösen zu können. Die Itemschwierigkeit unterscheidet sich zwischen Auszubildenden mit Vorkenntnissen und jenen ohne um 1,548 Logits; der DIF-Effekt für dieses Item ist groß. Die empirischen Daten repräsentieren das Rating und die Diskussion der Fachausschüsse.

Gleiches gilt für die Beispielaufgabe 2. Die Beispielaufgabe 2 ist im Vergleich zur Beispielaufgabe 1 eine leichtere Aufgabe (Itemschwierigkeit 1,475 Logits). Obwohl die Aufgabenbewältigung kein spezifisches Fachwissen voraussetzt, haben Auszubildende mit kaufmännischen Vorkenntnissen substanzielle Vorteile bei der Aufgabebearbeitung. So weist die Beispielaufgabe für Lernende mit kaufmännischer Vorbildung im Vergleich zu Lernenden ohne kaufmännische Vorbildung einen um 0,670 Logits kleineren Schätzer der Itemschwierigkeit auf. Das von den Fachausschüssen als mittelschwere Aufgabe geratete Beispiel zeigt auch in der empirischen Analyse Kennwerte einer mittleren Itemschwierigkeit auf.

2.2.3 Prognostische Validität schwierigkeitsbeschreibender Designkriterien

Im Ergebnis korrespondieren die Befunde des Ratings in den Fachausschüssen mit denen der empirischen Analyse. Hierbei konnten jedoch nur einzelne Aufgaben aus verschiedenen curricularen Inhaltsbereichen geratet und entsprechend diskutiert werden. Um Aussagen darüber treffen zu können, ob die schwierigkeitsbeschreibenden Designkriterien die empirisch erzeugte Schwierigkeit prognostizieren, wurden sämtliche Abschlussprüfungssätze der Jahre 2014 bis 2018 (jeweils Sommer- und Winterprüfung) für die Ausbildungsberufe Industriekaufmann/-frau sowie Kaufmann/-frau für Spedition und Logistikdienstleistungen kodiert. Insgesamt sind 507 Aufgaben ($n_{IK} = 198$; $n_{SK} = 309$) in die Analyse eingeflossen. Empirisch wurde im Rahmen eines Regressionsmodells geprüft, welchen Erklärungsanteil die einzelnen schwierigkeitsbeschreibenden Designkriterien an der empirischen Schwierigkeit einer Prüfungsaufgabe haben. In die Analyse sind Kodierungen der Designkriterien von drei Rater:innen eingeflossen; die Interrater-Reliabilität wurde mit Krippendorffs Alpha ermittelt und liegt zwischen 0,65 und 0,85. Items mit einem geringeren Übereinstimmungswert als 0,67 wurden konsensual nachkodiert. Es wurden Interrater-Effekte sowohl auf Gesamtpfungsebene als auch auf Einzelitemebene untersucht (Linacre, 1994), dabei konnten keine signifikanten Auswirkungen einzelner Rater:innen auf die Itemschwierigkeit festgestellt werden ($\text{Chi-Quadrat} = 3.32$; $\text{df} = 2$; $\text{p-Wert} = .245$).

Zur Durchführung der Analyse wurden die schwierigkeitsbeschreibenden Designkriterien dummy-kodiert. Tabelle 3 zeigt die Regressionsgewichte. Es wird deutlich, dass sich für alle drei schwierigkeitsbeschreibenden Designkriterien ein erwartungsgemäßer (unabhängiger) linearer Einfluss auf die empirische Schwierigkeit nachweisen lässt.

Tabelle 3: Regressionsgewichte als Prädiktoren der Itemschwierigkeit

Adj. $R^2 = 0,802$	Regressionsgewicht	Standardfehler	Betagewicht	p
Konstante	−1.468	.202	–	.000
hohe Wissenskomplexität	2.130	.334	.654	.000
mittlere Wissenskomplexität	1.493	.327	.562	.000
hohe kognitive Aktivierung	1.340	.246	.523	.000
mittlere funktionale Modellierung	.956	.296	.428	.003
mittlere kognitive Aktivierung	.941	.301	.406	.005

Anmerkung: n = 507 (Anzahl der Testitems für beide Ausbildungsberufe)

Insbesondere die schwierigkeitsbeschreibenden Designkriterien „Wissenskomplexität in der Domäne“ und „Kognitive Aktivierung“ sind hoch prädiktiv; sie haben auf allen Stufen prädiktive Vorhersagekraft. Das Designkriterium „funktionale Modellierung“ fällt dahinter leicht zurück. Hierbei dürfte eine Rolle spielen, dass von den 507 gerateten Aufgaben nur 51 Aufgaben bezüglich der funktionalen Modellierung den Wert 3 erreicht haben.

3 Schlussfolgerungen

Vor allem die unterschiedlichen Bildungsbiografien und die verschiedenen Ausbildungserfahrungen erhöhen die Herausforderungen bei der Erstellung kompetenz-orientierter Prüfungen im beruflichen Bereich. Es sind solche Aufgaben zu konstruieren und in Gesamtprüfungen zu binden, die für jede/n Proband:in (unabhängig von einem spezifischen Ausbildungsplatz) zu validen Ergebnissen führen. Eine Grundvoraussetzung hierfür sind Prüfungen, die die psychometrischen Gütekriterien erfüllen. Die Praxis der Prüfungserstellung steht vor der Herausforderung, regelmäßig fachlich aktuelle und authentische Prüfungsangebote zu machen, die als Berufseingangsprüfung das berufliche Handlungsfeld umfangreich abbilden kann und deren Qualität bei den abnehmenden Institutionen anerkannt ist.

Prüfungsaufgaben, die entweder berufsirrelevante Inhalte abfragen oder bei denen wesentliche psychometrische Gütekriterien vernachlässigt werden, gefährden jedoch die Aussagekraft der Prüfungsergebnisse. Die Arbeiten im Verbundprojekt ASPE haben gezeigt, dass ein zentrales Qualitätskriterium die ausgewogene Schwierigkeits-

verteilung der Prüfungsaufgaben ist. Gerade an diesem Punkt weisen die Testinformationskurven vergangener kaufmännischer Prüfungsjahrgänge Konzentrationen auf mittelschwere Aufgaben auf, während die Extreimbereiche – sehr leichte und sehr schwere Aufgaben – kaum repräsentiert sind (mit Folgen für die Reliabilität der Prüfungen). Das Verbundvorhaben ASPE hat sich zum Ziel gesetzt, wissenschaftlich fundierte Schwierigkeitsfaktoren zu definieren und systematisch zu beschreiben, die eine präzise und verlässliche Unterstützung für ehrenamtliche Prüfungsaufgabenerstellende darstellen können. Im Ergebnis erweisen sich drei Designkriterien als prädiktiv: „Wissenskomplexität in der Domäne“, „Kognitive Aktivierung – Verständnistiefe“ und „Funktionale Modellierung – Komplexität“. Diese Faktoren können unter Nutzung einer dreistufigen Systematik 80 Prozent der empirisch erzeugten Aufgabenschwierigkeit erklären. Für die Prüfungsaufgabenerstellenden ist das Vorgehen zugleich effizient und plausibel, da mit den Kriterien die ihnen vertrauten Dimensionen der kognitiven Anforderung und der inhaltlichen Tiefe erfasst werden, die für die Erstellung von Prüfungsaufgaben relevant sind.

Damit ist die Suchrichtung für eine empirisch evidente Schwierigkeitsprognostik durch folgende Logik vorgegeben: Gesucht werden Kriterien, die den Anforderungsgehalt von Prüfungsaufgaben mit der Folge erhöhen, dass die Wahrscheinlichkeiten einer erfolgreichen Aufgabenbewältigung sinken. Es sind folglich theoretische Modelle zu entwickeln, mit denen sich schwierigkeitsbeschreibende Designkriterien – wie beispielsweise die kognitive Aktivierung – auf Verarbeitungs- und Lösungsprozesse beziehen lassen. Um die Charakteristik der Anforderungssituation und die Leistungsfähigkeit der Testperson gemeinsam zu bewerten, sind fachdidaktische Ausarbeitungen sowohl im Hinblick auf die Konstruktion von Testsituationen als auch für die Auswertung der Ergebnisse zwingend notwendig.

Literatur

- Adams, R. J., & Khoo, S. T. (1996). *ACER Quest. interactive test analysis system*. Version 2.1. The Australian Council for Educational Research.
- Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology*, 63, 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>
- Bjørnavold, J. (2000). *Making learning visible*. Luxembourg.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (Eds.) (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. David McKay.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Psychology Press. <https://doi.org/10.4324/9781410600127>
- CEDEFOP (Ed.) (2003). *Terminology of Vocational Training Policy: A Multilingual Glossary for an Enlarged Europe*. Verfügbar unter http://www.europass.cedefop.eu.int/img/dynamic/c-313/cv-1_en_USglos-sary_4030_6k.pdf (Zugriff am 10.07.2025).

- Chang, S.-H., Lin, P.-C., & Lin, Z. C. (2007). Measures of Partial Knowledge and Unexpected Responses in Multiple-Choice Tests. *Educational Technology & Society*, 10(4), 95–109.
- Deutscher, V., Seifried, J., Rausch, A., Thomann, H., & Braunstein, A. (2022). Die LUCA Office Simulation in der Lehrerinnen- und Lehrerbildung – didaktische Design-Empfehlungen und erforderliche Lehrkompetenzen. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach, & M. Stock (Hrsg.), *Digital Literacy in der beruflichen Lehrer:innenbildung* (S. 107–121). wbv. <https://doi.org/10.3278/9783763973019>
- Deutscher, V., & Winther, E. (2018). Instructional sensitivity in vocational education. *Learning and Instruction*, 53, 21–33. <https://doi.org/10.1016/j.learninstruc.2017.07.004>
- Gelman, R., & Greeno, J. G. (1989). On the nature of competence: Principles for understanding in a domain. In L. B. Resnick (Ed.), *Knowing, learning and instruction. Essays in honor of Robert Glaser* (pp. 125–186). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781315044408-5>
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77–89. <https://doi.org/10.1080/19312450709336664>
- Hering, R. von, Rietenberg, A., Heinze, A., & Lindmeier, A. (2021). Nutzen Auszubildende bei der Bearbeitung berufsfeldbezogener Mathematikaufgaben ihr Wissen aus der Schule? Eine qualitative Untersuchung mit angehenden Industriekaufleuten. *Journal für Mathematik-Didaktik*, 42(2), 459–490. <https://doi.org/10.1007/s13138-021-00181-8>
- Hering, R. von, Zingelmann, H., Heinze, A., & Lindmeier, A. (2020). Lerngelegenheiten mit kaufmännischem Kontext im Mathematikunterricht der allgemeinbildenden Schule – Eine Schulbuch- und Aufgabenanalyse. *Zeitschrift für Erziehungswissenschaft*, 23(1), 193–213. <https://doi.org/10.1007/s11618-019-00925-w>
- Kendall, J. S., & Marzano, R. J. (2007). *The new taxonomy of educational objectives*. Corwin Press.
- Klotz, V. K., & Winther, E. (2012). Kompetenzmessung in der kaufmännischen Berufsausbildung: Zwischen Prozessorientierung und Fachbezug. *Eine Analyse der aktuellen Prüfungspraxis. Betriebs- und Wirtschaftspädagogik online (bwp@)*, 22(8), 1–16.
- Krippendorff, K. (2004). Reliability in content analysis. Some common misconceptions and recommendations. *Human Communication Research*, 30(3), 411–433. <https://doi.org/10.1111/j.1468-2958.2004.tb00738.x>
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511815355>
- Linacre, J. M. (1994). Sample size and item calibrations stability. *Rasch Measurement Transactions*, 7(4), 328.
- Marzano, R. J., & Kendall, J. S. (2007). *The new taxonomy of educational objectives* (2nd ed.). Corwin Press.
- Rychen, S., & Salganik, L. H. (Hrsg.) (2003). *Key Competencies for a Successful Life and a Well-Functioning Society*. Hogrefe.
- Sweller, J. (1988). Cognitive Load during Problem Solving: Effects on Learning. *Cognitive Science*, 12, 257–285. https://doi.org/10.1207/s15516709cog1202_4

- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10, 251–296. <https://doi.org/10.1023/A:1022193728205>
- Winther, E. (2010). *Kompetenzmessung in der beruflichen Bildung*. wbv. <https://doi.org/10.3278/6004148w>
- Winther, E., & Klotz, V. (2016). Berufliche Kompetenzniveaumodellierung: Eine Blaupause für die kaufmännische Domäne. *Unterrichtswissenschaft*, 44, 131–148.
- Winther, E., Reimer, J., & Vonarx, A. (2023). Kompetenzorientierte Prüfungserstellung. Eine digitale Workbench zur Konstruktion von Prüfungsaufgaben in der kaufmännischen Berufsbildung. *BWP* 52(3), 16–20.
- Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis*. MESA Press.

Schwierigkeitsbestimmende Aufgabenmerkmale in Prüfungsaufgaben im technischen Bereich am Beispiel des Ausbildungsberufs Mechatroniker:in

ELMAR DAMMANN, PIA SCHÄFER & FELIX WALKER

Zusammenfassung

Prüfungen repräsentieren die letzte und vielfach wesentliche Hürde eines Ausbildungsgangs. Doch was macht Prüfungsaufgaben eigentlich aus? Wodurch kann insbesondere deren Schwierigkeit erklärt werden? Hier zeigt sich für die gewerblich-technische Berufsbildung ein Forschungsdesiderat, zu dessen Auflösung mit den Arbeiten des Projekts TechKom beigetragen werden soll. Untersucht werden Prüfungsaufgaben des Ausbildungsberufs Mechatroniker:in hinsichtlich schwierigkeitsbestimmender Aufgabenmerkmale. Der Fokus liegt dabei auf der Textverständlichkeit und der Bild-Text-Integration. Die statistischen Analysen zeigen kleine bis mittlere Effekte einzelner Merkmale auf die Lösungsquote der untersuchten Prüfungsaufgaben. Erste Empfehlungen zur Gestaltung von Prüfungsaufgaben im gewerblich-technischen Bereich werden abgeleitet.

Schlagerworte: Schwierigkeitsbestimmende Aufgabenmerkmale, Prüfungsaufgaben, multimediales Lernen, Textverständlichkeit, Bild-Text-Integration

Abstract

Examinations represent the final and often essential hurdle in a training program. But what actually constitutes examination tasks? What in particular explains their difficulty? This is a research desideratum for industrial-technical vocational education and training, and the work of the TechKom project is intended to help resolve it. Examination tasks for the training occupation of mechatronics technicians are examined with regard to task characteristics that determine difficulty. The focus is on text comprehensibility and image-text integration. The statistical analyses show small to medium effects of the task characteristics on the solution rate of the tasks. Recommendations for the design of examination tasks in the industrial-technical field are derived.

Keywords: Difficulty-Determining Task Characteristics, Examination Tasks, Multimedia Learning, Text Comprehensibility, Image-Text Integration

1 Problemstellung

Wenn man nach der Ursache für eine erfolgreiche oder auch erfolglose Aufgabenbearbeitung sucht, richtet sich die Aufmerksamkeit zumeist auf die Anstrengungsbereitschaft, die Kompetenz oder Motivation von Personen, die die Aufgabe bearbeiten. Eher selten ist (Forschungskontexte einmal ausgenommen) die Aufgabe selbst Gegenstand der Betrachtungen. Dabei spielen die Anforderungen, welche Aufgaben an die bearbeitende Person stellen, eine entscheidende Rolle für deren Lösung. Welche Inhalte werden mit der Aufgabe angesprochen? Welche kognitiven Prozesse sind für die Lösung erforderlich? Welche sprachlichen Anforderungen stellt die Aufgabe? Diese Fragen sind insbesondere dann von Bedeutung, wenn die für die Bearbeitung der Aufgabe verfügbare Zeit begrenzt ist und die Konsequenz des Lösens oder Nicht-Lösens, wie im Fall von Abschlussprüfungen, erheblich ist (High-Stakes-Kontext). Prüfungen im schulischen Kontext folgen dieser Logik einer Leistungserbringung in definierten Zeiträumen. Dabei sollen Prüfungserstellende gewährleisten, dass die Prüfungsaufgaben von den teilnehmenden Personen grundsätzlich bearbeitet und gelöst werden können und die Lösung der Prüfungsaufgaben primär von den verfügbaren fachlichen Kompetenzen abhängt. Diese Anforderung ist für Prüfungsaufgaben im gewerblich-technischen Bereich weitgehend unerforscht, daher soll das Projekt TechKom¹ zur Bearbeitung des Forschungsdesiderats beitragen.

Prüfungsaufgaben sind Elemente diagnostischer Instrumente und dienen der Beurteilung individueller Leistungsfähigkeiten (Severing, 2011). Aus testtheoretischer Sicht müssen sie daher Anforderungen an Objektivität, Reliabilität und Validität erfüllen (Hewlett & Kahl-Andresen, 2014). Die Überprüfung dieser Gütekriterien ist ein zentraler Bestandteil der wissenschaftlichen Testentwicklung. Bislang werden Prüfungsaufgaben nur partiell hinsichtlich der Gütekriterien betrachtet, was in der Berufsbildungsforschung kritisiert und die Umsetzung als dringend geboten bewertet wird (Hewlett & Kahl-Andresen, 2014; Weiß, 2011). Für die Absicherung der Validität ist es dabei z. B. entscheidend zu wissen, wie die ermittelten Testwerte interpretiert werden und mit welchen Konstrukten diese Testwertinterpretationen abgesichert werden können (z. B. Hartig et al., 2020).

Untersuchungen aus dem Bereich der Kompetenzforschung zeigen, dass eine Beschreibung von Aufgaben durch unterschiedliche schwierigkeitsbestimmende Aufgabenmerkmale möglich und auch sinnvoll ist (Just et al., 2021; Kauertz, 2014; Nickolaus, 2014; Petsch et al., 2015; Schumann & Eberle, 2011), um Leistungen in Tests kriterienorientiert zu interpretieren (z. B. Hartig, 2007). Die Schwierigkeit einer Aufgabe, die Fachkenntnisse erfassen soll, kann z. B. durch den Textumfang oder die Anzahl zusätzlich benötigter Informationsquellen beeinflusst werden (Brünken et al., 2005; Chandler & Sweller, 1992). In Untersuchungen zu schwierigkeitsbestimmenden Aufgabenmerkmalen wurden bislang primär Aufgabenmerkmale fokussiert, die fachlich-kognitive

¹ TechKom: Technologiebasierte Kompetenzmessung und -förderung in der elektrotechnischen und metalltechnischen Erstausbildung, Förderkennzeichen: 21AP011. Gefördert im Rahmen der BMBF-Förderlinie ASCOT+: Technology-based Assessment of Skills and Competencies in VET (URL: www.ascot-vet.net/)

Anforderungen bzw. die für die Lösung der Aufgabe erforderliche Wissensbasis betreffen (siehe Kapitel 4). Dabei steht eine theoretische Fundierung vieler Aufgabenmerkmale noch aus.

Prüfungsaufgaben zeichnen sich in der Regel u. a. durch textliche Darstellungen aus, für die Lesefähigkeiten von großer Bedeutung sind (Haladyna & Rodriguez, 2013). Die Relevanz der Erforschung von Aufgabenmerkmalen, die die Gestaltung von Prüfungsaufgaben betreffen, insbesondere durch Befunde zu geringen Lesefähigkeiten bei Auszubildenden im gewerblich-technischen Bereich gestärkt (Bals & Noack, 2013; Lehmann & Seeber, 2007; Roche & Terrasi-Haufe, 2019; Seeber, 2005).

Das Projekt TechKom hatte zum Ziel, Prüfungsaufgaben von elektro- und metall-technischen Ausbildungsberufen hinsichtlich der Aufgabenschwierigkeit empirisch zu überprüfen. Dabei wurden gebundene Prüfungsaufgaben der Abschlussprüfungen Teil 2 der Ausbildungsberufe Elektroniker:in für Automatisierungstechnik (EAT), Konstruktionsmechaniker:in (KM) und Mechatroniker:in (ME) fokussiert². Diese Aufgaben wurden von der Prüfungsaufgaben- und Lehrmittelentwicklungsstelle der IHK, Region Stuttgart (PAL) entwickelt und eingesetzt. Die Durchführung von TechKom erfolgte in drei Phasen: (1) Analyse von schriftlichen PAL-Prüfungsaufgaben, (2) Variation von Prüfungsaufgaben bzgl. in (1) identifizierter Aufgabenmerkmale sowie (3) empirische Überprüfung der (variierten) Prüfungsaufgaben.

In diesem Beitrag wird die erste Phase des Projekts vorgestellt, die drei Schritte umfasste: (a) Analyse der bestehenden PAL-Prüfungsaufgaben und Identifikation relevanter schwierigkeitsbestimmender Aufgabenmerkmale, (b) Operationalisierung der identifizierten schwierigkeitsbestimmenden Aufgabenmerkmale und (c) Variation bestehender PAL-Prüfungsaufgaben. Die im Rahmen dieser Untersuchung erstellten Analysen werden exemplarisch für den Ausbildungsberuf Mechatroniker:in (ME) vorgestellt. Mechatroniker:innen erlernen in ihrer Ausbildung den Bau komplexer mechatronischer Systeme aus mechanischen, elektrischen und elektronischen Bauteilen (Bundesagentur für Arbeit, 2025). „Sie prüfen die einzelnen Bauteile und montieren sie zu Systemen und Anlagen. Die fertigen Anlagen nehmen sie in Betrieb, programmieren sie oder installieren zugehörige Software“ (ebd.).

2 Prüfungen und Prüfungsaufgaben der PAL

Die Abschlussprüfungen der dualen Ausbildung wurden in den vergangenen beiden Jahrzehnten umfassend reformiert. Ein Ergebnis dieser Reformbemühungen stellen z. B. die „Empfehlungen des Hauptausschusses für Berufsbildung (BIBB) zur Struktur und Gestaltung von Ausbildungsordnungen – Prüfungsanforderungen –“ (Bundesinstitut für Berufsbildung, 2013; siehe auch Stöhr, 2017) dar. Gründe dieser Reform waren u. a. die große Anzahl unterschiedlicher Prüfungsmethoden (Stöhr, 2017) sowie die Anforderung, das didaktische Leitkonzept der Handlungsorientierung in den Abschlussprüfungen umzusetzen. Eine Neuerung stellt die „gestreckte“ Abschlussprü-

2 <https://www.ascot-vet.net/ascot/de/ascot-projekte/techkom/techkom.html>

fung (GAP) dar, die sich von der „klassischen“ Abschlussprüfung dadurch unterscheidet, dass sie während der Ausbildungszeit (Teil 1 – ehemals Zwischenprüfung) bereits berufliche Handlungskompetenzen auf Facharbeiterniveau prüft (Stöhr, 2017, S.9).

In beiden Teilen der Abschlussprüfung wird zwischen theoretischen und praktischen Prüfungsformen unterschieden. Praktische Prüfungen finden im Betrieb statt und bestehen unter anderem aus Fachgesprächen (Teil 1) und praktischen Aufgaben (Teil 2). Theoretische Prüfungsformen umfassen schriftliche Prüfungsaufgaben im ungebundenen (offenen) und gebundenen (geschlossenen) Aufgabenformat. Teil 1 wird in der Regel im zweiten Ausbildungsjahr durchgeführt, während Teil 2 am Ende der Ausbildung stattfindet.

Die Prüfungsaufgaben- und Lehrmittelentwicklungsstelle der IHK, Region Stuttgart (PAL), ist für die Entwicklung, Erstellung und Umsetzung der bundesweiten Abschlussprüfungen in den gewerblich-technischen Ausbildungsberufen zuständig. Nach eigenen Angaben liefert die PAL viermal jährlich schriftliche, praktische und integrierte Zwischen- und Abschlussprüfungen für rund 136 Berufe, Fachrichtungen und Einsatzgebiete (PAL, 2018). Dabei werden die hauptamtlichen Mitarbeiter:innen durch rund 1.000 ehrenamtliche Mitarbeiter:innen aus Unternehmen und Schulen (Ausbilder:innen und Lehrpersonen) unterstützt. Die Entscheidungsstruktur der PAL hinsichtlich Prüfungsangelegenheiten besteht im Wesentlichen aus drei Gremien, einem Beirat, einem technischen Ausschuss und dem Hauptausschuss. Darüber hinaus gibt es PAL-Fachausschüsse (vgl. ebd.).

Schriftliche Prüfungsaufgaben bestehen nach der QM-Handreichung der PAL zur Erstellung von Prüfungsaufgaben aus einem Informations- und Frageteil (Aufgabenstamm) sowie einem Antwortteil (Prüfungsaufgaben- und Lehrmittelentwicklungsstelle der IHK Region Stuttgart (PAL), 2020): Der Informationsteil besteht aus einem einleitenden Satz, der handlungsorientiert in die Thematik der Aufgabe einführt und zum Frageteil überleitet. Der Frageteil wird in der Regel durch W-Fragen gestaltet und verdeutlicht durch den Einsatz von sogenannten Operatoren, was von den Auszubildenden erwartet wird. Gebundene PAL-Prüfungsaufgaben bieten fünf Antwortmöglichkeiten, von denen eine Antwort richtig ist. Ein Strukturelement der schriftlichen PAL-Prüfung stellen die Tätigkeitsdimensionen dar. Diese unterscheiden sich je nach Beruf und gliedern die PAL-Prüfungsaufgaben nach übergeordneten thematischen Klammern. Für den Ausbildungsberuf ME sind diese Tätigkeitsdimensionen (1) Arbeitsplanung und (2) Funktionsanalyse. Die schriftliche Abschlussprüfung Teil 2 besteht für jede Tätigkeitsdimension aus 28 gebundenen sowie acht ungebundenen Aufgaben. Von den 28 gebundenen Aufgaben sind 25 verpflichtend zu bearbeiten, drei können abgewählt werden. Bei den ungebundenen Aufgaben ist keine Abwahl möglich. Die Möglichkeit des Abwählens von Prüfungsaufgaben stellt ein Selektionskriterium dar, das für die Bewertung und den Vergleich der Prüfungsaufgaben zu berücksichtigen ist.

Um den Forschungsgegenstand dieser Untersuchung besser einordnen zu können, werden im Folgenden zwei Beispielaufgaben der ME-Abschlussprüfung, Teil 2 vorgestellt. Jede der zwei Aufgaben enthält einen **Aufgabenstamm (AS)** sowie fünf **Antwortmöglichkeiten (AM)**. Die erste Beispielaufgabe (Abbildung 1) ist der Tätigkeits-

dimension „Arbeitsplanung“ zugeordnet und gehört zu den abwählbaren Prüfungsaufgaben. Sie wurde im Sommer 2016 eingesetzt und hat eine Lösungsquote von 86,1 % (N = 1.474 Auszubildende).

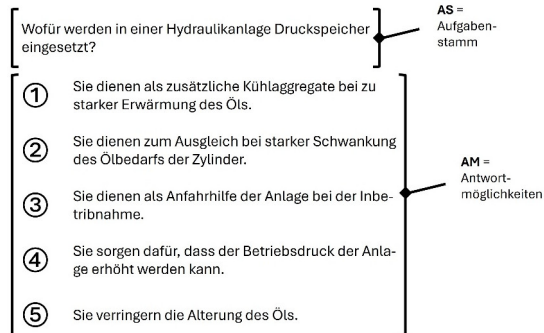


Abbildung 1: Erste Beispielaufgabe der ME-Prüfungen; eigene Darstellung auf Basis der originalen PAL-Prüfungsaufgabe

Der Aufgabenstamm der zweiten Beispielaufgabe (Abbildung 2) besteht ebenfalls nur aus dem Frageteil im AS. Im Unterschied zur ersten Beispielaufgabe ist hier eine **Abbildung (AB)** enthalten (Abbildung 2). Zur Lösung dieser Prüfungsaufgabe können die Auszubildenden Nebenrechnungen in einem bereitgestellten Feld durchführen. Diese Beispielaufgabe ist der Tätigkeitsdimension „Arbeitsplanung“ zugeordnet und gehört zu den nicht abwählbaren Prüfungsaufgaben. Sie wurde im Sommer 2016 eingesetzt und hat eine Lösungsquote von 67,5 % (N = 1.474 Auszubildende). Diese Prüfungsaufgabe erfordert für die Bearbeitung das Hinzuziehen eines **weiteren Dokuments (WD)**, in diesem Fall einer Formelsammlung (Abbildung 2, links).

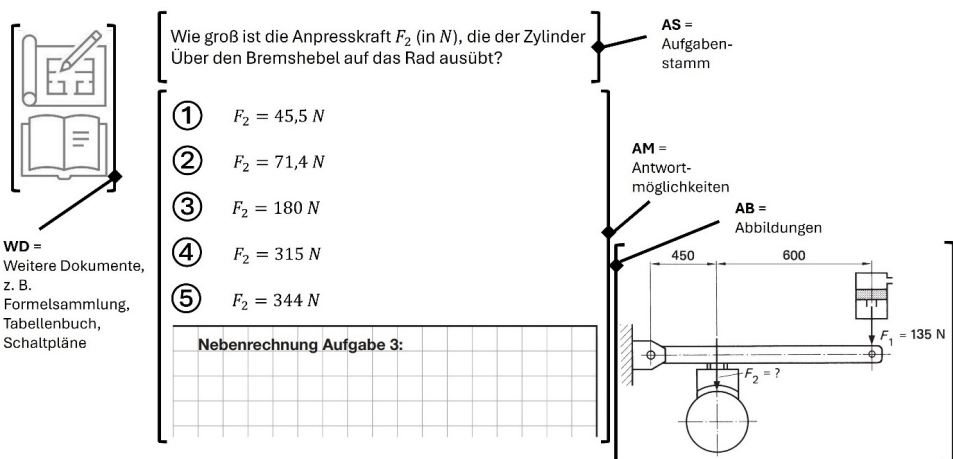


Abbildung 2: Zweite Beispielaufgabe der ME-Prüfungen; eigene Darstellung auf Basis der originalen PAL-Prüfungsaufgabe

Dieser Untersuchung liegen Annahmen über den Einfluss der Aufgabengestaltung auf die Aufgabenschwierigkeit zugrunde. Diese Annahmen werden im folgenden Kapitel durch theoretische Modelle und Forschungsbefunde untermauert.

3 Theoretischer Hintergrund

Aufgabenmerkmale werden dann als schwierigkeitsbestimmend betrachtet, wenn ein Zusammenhang zwischen ihnen und der Schwierigkeit einer Aufgabe nachgewiesen werden kann. Die Schwierigkeit einer Aufgabe kann dabei als Lösungsquote oder Lösungswahrscheinlichkeit (z. B. als Itemparameter im Zuge von Modellierungen der Item-Response-Theory (z. B. Rost, 2004)) operationalisiert werden (z. B. Walpuski & Ropohl, 2014).

Studien zu Aufgabenmerkmalen wurden in verschiedenen fachdidaktischen Forschungsbereichen durchgeführt (Naturwissenschaftsdidaktische Forschung: Kauertz, 2008; Kauertz, 2014; Schmiemann, 2011; Walpuski & Ropohl, 2014; Wirtschaftspädagogik: Deutscher, 2015; Schumann & Eberle, 2011; Winther, 2011; Gewerblich-technische Berufsbildung: Dammann et al., 2016; Geißel, 2008; Gschwendtner, 2008; Just et al., 2021; Nickolaus, 2014; Petsch et al., 2015; Petsch, 2024; Van Waveren & Nickolaus, 2015). Schwierigkeitsbestimmende Aufgabenmerkmale wurden dabei oftmals fachspezifisch untersucht, fachübergreifende Nachweise schwierigkeitsbestimmender Merkmale sind eher selten (Nickolaus, 2014). Darüber hinaus sind zwei Verfahren der Nachweisführung gängig, nämlich die nachträgliche Zuordnung von Aufgabenmerkmalen (post hoc) zu getesteten Aufgaben (Beaton & Allen, 1992) sowie die merkmalsbezogene Aufgabenentwicklung (a priori) und anschließende Überprüfung der Merkmale (Hartig, 2007). Arbeiten mit experimentellen Forschungsdesigns zum gezielten empirischen Nachweis schwierigkeitsbestimmender Aufgabenmerkmale liegen nach aktuellem Kenntnisstand bisher nicht vor.

Aufgabenmerkmale werden in Analysekategorien strukturiert. Prenzel et al. (2002) schlagen z. B. die Kategorien (1) formale Aufgabenmerkmale, (2) kognitive Anforderungen beim Lösen der Aufgaben und (3) Merkmale der für das Lösen der Aufgaben erforderlichen Wissensbasis vor (Prenzel et al., 2002, S. 125). Unter (1) werden Aufgabenmerkmale gefasst, die sich auf „äußere Merkmale der Itemformulierung und des betreffenden Antwortformats“ (ebd.) beziehen. Zu (2) gehören Aufgabenmerkmale, die „kognitive Prozesse bei der Aufgabenbearbeitung“ (Prenzel et al., 2002, S. 128) betreffen. Schließlich werden unter (3) Aufgabenmerkmale gesammelt, bei denen eine Wissensbasis für die Aufgabenlösung erforderlich ist (ebd.).

3.1 Befunde zu schwierigkeitsbestimmenden Aufgabenmerkmalen

In Untersuchungen zu schwierigkeitsbestimmenden Aufgabenmerkmalen konnten Einflüsse auf die Aufgabenschwierigkeit nachgewiesen werden. Prenzel et al. (2002) analysierten den PISA-Naturwissenschaftstest aus dem Jahr 2000 und konnten in ihrem finalen Modell, mit den berücksichtigten Aufgabenmerkmalen, etwa 45 % der Varianz der Aufgabenschwierigkeit erklären. Diese Aufgabenmerkmale sind (ange-

lehnt an die o. g. Kategorien von Prenzel et al. (2002)): (1) die Länge des Aufgabentextes, die Vorgabe einer Grafik oder eines „stilisierten“ Bildes sowie das offene Antwortformat; (2) die Verarbeitung der Textinformationen, die logische Verknüpfung der Textinformationen, die Vorgabe, etwas auszurechnen, räumliches Vorstellungsvermögen und divergentes Denken; (3) terminologisches Wissen, Faktenwissen, kontraintuitives Wissen sowie funktionale Zusammenhänge.

Für den gewerblich-technischen Bereich beschreibt Nickolaus (2014) insgesamt 15 schwierigkeitsbestimmende Aufgabenmerkmale der Analysekategorien (nach ebd.) „Fachwissen“ und „fachspezifisches Problemlösen“, die in unterschiedlichen Untersuchungen im gewerblich-technischen Bereich nachgewiesen werden konnten. Häufig beobachtete Merkmale zur Analysekategorie „Fachwissen“ seien kognitionspsychologisch-taxonomische Merkmale (z. B. die Lernzieltaxonomie nach Bloom (1968) bzw. die überarbeitete Fassung von Anderson (2001)), Modellierungsanforderungen, mathematische Anforderungen, die Komplexität einer Anforderung, die Vertrautheit mit dem Fachinhalt, Aufgabenformate, aufgabenspezifische Verweisqualitäten, domänen-spezifische Merkmale sowie der Umgang mit Texten. Zur Analysekategorie „fachspezifisches Problemlösen“ werden Komplexität, Diagnoseart, Systemdynamik, Erschließung von Informationen, messtechnische Anforderungen sowie technologische Merkmale zugeordnet (Nickolaus, 2014).

Diese bisherigen Untersuchungen fokussierten primär Aufgabenmerkmale der von Prenzel et al. (2002) formulierten Kategorien (2) und (3). Wenig erforscht ist der Einfluss der Aufgabengestaltung auf die Aufgabenschwierigkeit (formale Aufgabenmerkmale nach Prenzel et al. (2002)). Dabei sind Anforderungen des Erfassens und Verarbeitens von medial präsentierten Informationen im Bereich der Psychologie, der kognitiven Linguistik und im Bereich anderer fachdidaktischer Fachbereiche gut erforscht (z. B. Friedrich, 2017; Mayer & Fiorella, 2022; Schnotz, 2023; im Überblick Haladyna & Rodriguez, 2013, S. 25; Downing & Haladyna, 2006, S. 377). Im Bereich der Naturwissenschaftsdidaktiken konnte Stawitz (2010) den Einfluss der Fachsprache, der Sprachkomplexität sowie von Tabellen und Grafiken nachweisen. Die Bedeutung der im Aufgabentext gegebenen Fachinformationen wurde von Ropohl et al. (2015) untersucht. Kleinknecht et al. (2013) identifizieren die sprachlogische Komplexität von Aufgaben als eine von sieben Kategorien der Aufgabenanalyse. Die sprachlogische Komplexität wird u. a. durch zusätzliche Informationen oder komplexe Satzgefüge definiert (Kleinknecht et al., 2013). Aus Perspektive des Lernens mit Multimedia stellt die Gestaltung des Lernmaterials, neben den Merkmalen der Lernenden, einen wesentlichen Einfluss auf den Lernprozess dar (z. B. Mayer & Fiorella, 2022; Schnotz, 2023). Die vielfältigen Erkenntnisse dieser Forschungsbereiche wurden dabei bislang nicht auf Prüfungsaufgaben im Bereich gewerblich-technischer Berufsbildung angewendet.

Ziel dieser Untersuchung ist es, Prüfungsaufgaben des schriftlichen Teils der Abschlussprüfung – Teil 2 des Ausbildungsberufs Mechatroniker:in – hinsichtlich formaler Aufgabenmerkmale (Prenzel et al., 2002) zu untersuchen. Dabei werden die Merkmale „Textverständlichkeit“ und „Bild-Text-Integration“ fokussiert, für die umfangreiche theoretische Vorarbeiten vorliegen (z. B. Friedrich, 2017; Lenhard, 2019; Schnotz, 2023). Diese Vorarbeiten werden im folgenden Kapitel vorgestellt.

3.2 Grundlagen zu Textverständlichkeit und Bild-Text-Integration

Das Verstehen von Texten basiert auf mehreren, ineinandergreifenden Teilprozessen (vgl. Lenhard, 2019, S. 14), die vom Erfassen der „Wortoberfläche eines Textes“ (Lenhard, 2019, S. 14) bis zur Erstellung eines geistigen Abbildes (eines mentalen Modells) des Inhalts reichen (vgl. ebd.). Lesekompetenz ergibt sich nach dem Verständnis der kognitiven Linguistik aus dem Zusammenwirken von „leserbezogenen Aspekten“ (Lenhard, 2019, S. 28) und „textbezogenen Aspekten“ (ebd.; Friedrich, 2017). Zu den leser:innenbezogenen Aspekten werden Merkmale sowie Aktivitäten der Lesenden gezählt (vgl. Lenhard, 2019). Die textbezogenen Aspekte bestehen aus Leseanforderungen und Merkmalen des Textes wie Aufbau und Struktur, Textgattung oder Kohärenz (vgl. ebd.). Das von Friedrich (2017, S. 70 ff.) auf Basis zahlreicher Vorarbeiten entwickelte Konzept der Textverständlichkeit umfasst zehn Merkmale. Diese sind „Wortschwierigkeit, Satzschwierigkeit, Argumentdichte, Propositionsdichte, Aufwand der Inferenzbildung, Aufwand für Reinstatements, Aufwand für Reorganisationen, Hervorhebungen, Anschaulichkeit und Variation der Sprache“ (Friedrich 2017, S. 129). Alle Merkmale bis auf die Merkmale „Argumentdichte“ und „Propositionsdichte“ werden sowohl durch Textmerkmale als auch durch Personenmerkmale beeinflusst (vgl. Friedrich, 2017, S. 137). In den projektimmanenten Erstbetrachtungen der fokussierten ME-Prüfungsaufgaben hat sich gezeigt, dass insbesondere die Merkmale „Wortschwierigkeit“, „Satzschwierigkeit“ und „Hervorhebungen“ in der Perspektive der Textmerkmale bedeutsam sind. Einflussreiche Faktoren seien dabei u. a. die verwendeten Wörter (Wortschwierigkeit), die Satzlänge, Einfachheit oder Satzkomplexität (Satzschwierigkeit) sowie Kohärenzbildungshilfen (Hervorhebungen) (vgl. Friedrich, 2017, 128 ff.). Kohärenzbildungshilfen werden auch im Bereich des Lernens mit Multimedia als Merkmal für das Verstehen von Texten betrachtet, wobei diese im Unterschied zum Verständnis von Friedrich im Kontext multimedialer Darbietungen betrachtet werden, z. B. bei Texten, die auch Abbildungen enthalten (z. B. Brünken et al., 2005). Die Forschung zum Lernen mit Multimedia beschäftigt sich mit der Darbietung und Verarbeitung von multiplen Repräsentationen in Lernmaterialien (z. B. Mayer & Fiorella, 2022). Wichtige Befunde des Forschungsbereichs sind die Begrenztheit des Arbeitsgedächtnisses (Chandler & Sweller, 1992) sowie die Kodierung der aufgenommenen und verarbeiteten Informationen auf mindestens zwei unterschiedliche Arten (Clark & Paivio, 1991). Aus diesen Überlegungen wurden Gestaltungsempfehlungen für Lernmaterialien abgeleitet. Diese betreffen u. a. die Lenkung der Aufmerksamkeit auf wesentliche Aspekte des Lernmaterials, z. B. den Split-Attention-Effekt (Brünken & Leutner, 2001; Chandler & Sweller, 1992), oder die Bezugnahme innerhalb eines Textes sowie zwischen einem Text und weiteren Informationsquellen wie Abbildungen, Tabellen oder Formeln, z. B. Kohärenzbildungshilfe (Brünken et al., 2005; Rabe & Mikelskis, 2007; Schnotz, 2006) und Textkohärenz (Friedrich, 2017; Stawitz, 2010). Der Split-Attention-Effekt ist insbesondere dann gegeben, wenn aufeinander bezogene Informationsquellen getrennt voneinander dargeboten werden (Rey, 2009). „Multiple sources of information that are unintelligible in isolation result in less learning when they are presented in split-attention as opposed to integrated format“ (Sweller, 2010, S. 30).

Unterstützungsmöglichkeiten zur mentalen Kohärenzbildung ergeben sich u. a. in der Reduzierung des Extraneous Cognitive Load (Chandler & Sweller, 1992), indem z. B. „Informationen, die aufeinander bezogen werden sollen, deutlich in Zusammenhang gebracht werden“ (Brünken et al., 2005, S. 63).

Zur Schwierigkeit von Prüfungsaufgaben liegen im Bereich linguistischer Forschung einige wenige Untersuchungen vor. Demnach seien Prüfungsaufgaben aus Platzgründen komprimiert (vgl. Wagner & Schlenker-Schulte, 2015). „Oft kombinieren sie Fachwörter in hoher Zahl mit speziellen sprachlichen Konstruktionen aus dem Bereich der Standardsprache, wie z. B. Substantivketten“ (Wagner & Schlenker-Schulte, 2015, S. 3). Verfahren der Textoptimierung sind dabei wirksame Maßnahmen, um rein sprachliche Barrieren beim Verstehen und Bearbeiten von Prüfungsaufgaben zu vermeiden (Wagner & Schlenker-Schulte, 2006). Textoptimierungen können auf der Ebene der Wörter (z. B. durch geläufige und eindeutige Wörter), der Sätze (z. B. Vermeidung langer Sätze und Negationen) oder des gesamten Textes (z. B. durch Gliederung, Klarheit der Aufgabenstellung) vorgenommen werden (Wagner & Schlenker-Schulte, 2015). Textbezogene Merkmale wie die Textlänge (z. B. Kauertz, 2014; Prenzel et al., 2002) oder die Anzahl an Fachbegriffen (Petsch, 2024) konnten auch in der fachdidaktischen Forschung als schwierigkeitsbestimmende Aufgabenmerkmale identifiziert werden.

4 Forschungsfrage der Untersuchung

Auf Basis des im Kapitel 4 vorgestellten Forschungsstandes werden Einflüsse der Aufgabengestaltung auf die Aufgabenschwierigkeit angenommen. Fokussiert werden in der vorliegenden Arbeit die Aufgabenmerkmale „Textverständlichkeit“ (z. B. Friedrich, 2017) und „Bild-Text-Integration“ (z. B. Schnotz, 2023). Aufgabenmerkmale zu kognitiven Anforderungen und zur erforderlichen Wissensbasis werden in der vorliegenden Arbeit nicht berücksichtigt, da hierfür Befunde vorliegen. Da Prüfungsaufgaben in gewerblich-technischen Ausbildungsberufen in ihrer Gestaltung Ähnlichkeiten aufweisen (z. B. durch die Darbietung technischer Abbildungen oder die Art der Textgestaltung), wird angenommen, dass die Aufgabenmerkmale „Textverständlichkeit“ und „Bild-Text-Integration“ in vielen Prüfungsaufgaben des Ausbildungsberufs Mechatroniker:in zu finden sind und diese systematische Effekte auf die Aufgabenschwierigkeit zeigen. Die Forschungsperspektive dieser Untersuchung wird am Beispiel der Prüfungsaufgabe (siehe oben Abbildung 2) in der folgenden Abbildung 3 verdeutlicht. Aufgabenschwierigkeiten können innerhalb der einzelnen Aufgabenteile sowie aus dem Zusammenwirken dieser Aufgabenteile entstehen.

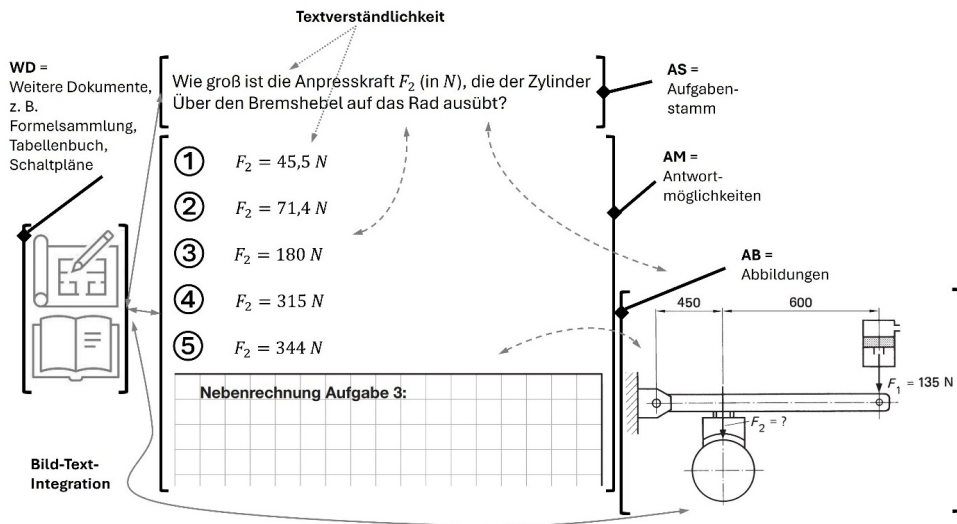


Abbildung 3: Schwierigkeitsbestimmende Merkmale aufgezeigt an der zweiten Beispielaufgabe der ME-Prüfungen; Eigene Darstellung auf Basis der originalen PAL-Prüfungsaufgabe.

Anmerkungen: Durchgezogene Linien = Split-Attention; gestrichelte Linien = Kohärenz; gepunktete Linien = Textverständlichkeit und Fachbegriffe

Aus dem Forschungsstand aus Kapitel 4 sowie den Darlegungen oben ergeben sich für diese Untersuchung die folgenden Forschungsfragen:

FF1: Zeigt sich das Aufgabenmerkmal der Textverständlichkeit bei den analysierten ME-Prüfungsaufgaben der PAL als schwierigkeitsbestimmend?

FF2: Zeigt sich das Aufgabenmerkmal der Bild-Text-Integration bei den analysierten ME-Prüfungsaufgaben der PAL als schwierigkeitsbestimmend?

5 Methodik zur Analyse bestehender Prüfungsaufgaben

Zur Beantwortung der Forschungsfragen wurden Prüfungsaufgaben der PAL herangezogen. Diese stammen aus den Jahrgängen Sommer 2016 bis Sommer 2020. Für den Ausbildungsberuf Mechatroniker:in (ME) wurden neun schriftliche Abschlussprüfungen mit insgesamt 504 Prüfungsaufgaben analysiert.

Die Analysen wurden von einem Projektmitarbeiter durchgeführt, der über Fachkompetenzen im Ausbildungsberuf Mechatroniker:in verfügt. Zur Reflexion des Analysevorgehens sowie für die Entwicklung gemeinsamer Analysekatoren gab es einen regelmäßigen Austausch mit den Projektmitarbeitenden zu den weiteren Ausbildungsberufen Elektroniker:in für Automatisierungstechnik und Konstruktionsmechaniker:in. Im ersten Analyseschritt wurde für jeden Ausbildungsberuf eine Abschlussprüfung gesichtet und die Ergebnisse in der Gruppe diskutiert. Dabei wurden

der grundlegende Aufbau der Prüfungen und Prüfungsaufgaben betrachtet sowie gemeinsame Entscheidungen hinsichtlich der finalen Auswahl zu fokussierender Aufgabenmerkmale und hinsichtlich deren Operationalisierung getroffen. Anschließend folgte die Analyse aller Prüfungsaufgaben bezogen auf diese Operationalisierungen (Ausprägungen siehe nächster Abschnitt und Tabelle 1).

Für die Operationalisierung der beiden Aufgabenmerkmale wurden mehrere der in Kapitel 4 vorgestellten Konstrukte verwendet. Für das Aufgabenmerkmal „Textverständlichkeit“ wurden in dieser Untersuchung die von Friedrich (2017) genannten Aspekte „Wortschwierigkeit“ (erhoben über die Anzahl der Fachbegriffe, siehe Petsch et al., 2015), „Satzschwierigkeit“ (erhoben über den Lesbarkeitsindex LIX, siehe Lenhard & Lenhard, 2014–2022) sowie „Hervorhebungen“ (erhoben über Kohärenzbildungshilfen, z. B. Brünken et al., 2005) berücksichtigt. Das Aufgabenmerkmal „Bild-Text-Integration“ wurde in Anlehnung an Schnotz (2023) über den „Split-Attention-Effekt“ (Ayres & Sweller, 2022; Schnotz, 2023) erfasst. Zusätzlich wurden als Strukturmerkmale der Prüfungsaufgaben die Tätigkeitsdimensionen und die Verbindlichkeit (Wahl- oder Pflichtaufgabe) der Prüfungsaufgaben erfasst.

Jede Operationalisierung umfasst, basierend auf den Erstbetrachtungen der ME-Prüfungsaufgaben, mehrere Ausprägungen (siehe auch Tabelle 1). Für den Split-Attention-Effekt wurden vier Ausprägungen angenommen, bei Ausprägung 1 sind alle notwendigen Informationen in der Aufgabe enthalten. Bei den Ausprägungen 2 bis 4 sind ein oder mehrere weitere Informationsquellen für die Aufgabenbearbeitung erforderlich. Bei der Kohärenzbildungshilfe wurden drei Ausprägungen angenommen. Diese unterscheiden sich zunächst darin, ob aufgrund der Aufgabengestaltung eine Kohärenzbildungshilfe möglich ist (Ausprägungen 1 und 2) oder nicht (Ausprägung 3). Besteht eine Prüfungsaufgabe ausschließlich aus Text, ist die Einbindung einer Kohärenzbildungshilfe nicht möglich. Ist eine Kohärenzbildungshilfe möglich, wird darüber hinaus betrachtet, ob diese in der jeweiligen Aufgabe gegeben ist (Ausprägung 1) oder nicht (Ausprägung 2). Eine Kohärenzbildungshilfe kann dabei z. B. eine Abbildungsbezeichnung oder der Verweis im Text auf eine zugehörige Abbildung sein. Für die Berechnung des Lesbarkeitsindex LIX wird in dieser Untersuchung eine Berechnungsmethode angewendet, wie sie Lenhard und Lenhard (2014–2022) vorgestellt haben. Der Lesbarkeitsindex ergibt sich demnach „aus der Summe der durchschnittlichen Satzlänge eines Textes und des prozentualen Anteils langer Wörter (mehr als sechs Buchstaben)“ (Lenhard & Lenhard, 2014–2022):

$$LIX = SL + \left[\left(\frac{\text{Anzahl langer Wörter}}{\text{Anzahl aller Wörter}} \right) \cdot 100 \right]$$

LIX = Lesbarkeitsindex

SL = durchschnittliche Satzlänge

Die errechneten LIX-Werte werden in eine fünfstufige Skala übertragen, welche die folgenden Stufen enthält: sehr leicht (LIX-Werte kleiner als 30), leicht (LIX-Werte 30 bis 40), mittel (LIX-Werte 40–50), schwer (LIX-Werte 50 bis 60) und sehr schwer (LIX-Werte größer als 60) (angelehnt an Wild & Pissarek, 2021).

6 Ergebnisse

6.1 Deskriptive Befunde

In Tabelle 1 sind die deskriptiven Ergebnisse der Analyse dargestellt. Sie verdeutlichen die Zuordnung der analysierten ME-Prüfungsaufgaben zu den verschiedenen Ausprägungen der Operationalisierungen.

Tabelle 1: Operationalisierungen der Aufgabenanalyse mit Abstufungen und ersten deskriptiven Ergebnissen (OP = Operationalisierung; NA = keine Aufgaben zuordenbar)

OP	Stufe	Erläuterung	Ergebnisse
split	1	alle Informationen sind auf dem Aufgabenblatt	434 (86 %) Aufgaben
	2	ein zusätzliches Dokument ist notwendig	70 (14 %) Aufgaben
	3	zwei zusätzliche Dokumente sind notwendig	NA
	4	> zwei zusätzliche Dokumente sind notwendig	NA
koh	1	eine Kohärenzhilfshilfe ist gegeben	84 (17 %) Aufgaben
	2	eine Kohärenzhilfshilfe ist nicht gegeben	198 (39 %) Aufgaben
	3	eine Kohärenzhilfshilfe ist nicht möglich	222 (44 %) Aufgaben
lix_as/	1	sehr leicht	43 (9 %) / 37 (10 %) Aufgaben
lix_am	2	leicht	127 (62 %) / 74 (20 %) Aufgaben
	3	mittel	174 (35 %) / 103 (28 %) Aufgaben
	4	schwer	116 (23 %) / 59 (16 %) Aufgaben
	5	sehr schwer	44 (9 %) / 100 (27 %) Aufgaben
	999	LIX-Berechnung nicht möglich	NA / 131 Aufgaben
fb_as		Anzahl der Fachbegriffe im Aufgabenstamm	M = 2,8; SD = 2,1
fb_am		Anzahl der Fachbegriffe in den Antwortmöglichkeiten	M = 4,6; SD = 3,3
fb_ab		Anzahl der Fachbegriffe in Abbildungen	M = 6,4; SD = 7,3

Die ersten Analyseergebnisse zeigen spezifische Verteilungen und Schwerpunktsetzungen. Hinsichtlich der Operationalisierung „Split-Attention-Effekt“ (split) sind in den ME-Prüfungsaufgaben die zur Bearbeitung nötigen Informationen überwiegend gegeben, bei 70 ME-Prüfungsaufgaben wird ein weiteres Dokument (z. B. ein Tabellenbuch oder eine Formelsammlung) benötigt. Mehr als ein zusätzliches Dokument wird bei keiner der analysierten ME-Prüfungsaufgaben benötigt (dies stellt sich bei den anderen analysierten Ausbildungsberufen anders dar). Die Operationalisierung „Kohärenz“ (koh) kann grundsätzlich unterschieden werden in Aufgaben, bei denen eine Kohärenzhilfshilfe möglich ist (Stufen 1 und 2, 56 %), und Aufgaben, bei denen keine Kohärenzhilfshilfe möglich ist (Stufe 3, 44 %), da der Aufgabenstamm allein aus Text besteht. Bei Aufgaben, die aus Text und z. B. einer Abbildung bestehen, ist eine Kohärenzhilfshilfe zwar möglich, jedoch überwiegend nicht gegeben (Stufe 2, 39 %). Die Komplexität des Aufgabenstammes (lix_as) liegt bei den ME-Prüfungs-

aufgaben überwiegend auf den Stufen „2“ (leicht) und „3“ (mittel). Die Komplexität der Antwortmöglichkeiten (lix_am) liegt überwiegend auf den Stufen „3“ (mittel) und „5“ (sehr schwer). Bei 131 ME-Prüfungsaufgaben konnte keine Textkomplexität für die Antwortmöglichkeiten ermittelt werden, wenn es sich beispielsweise um das Ergebnis einer Rechenaufgabe handelte (siehe Beispielaufgabe in Abbildung 2). Für die weiteren Analysen werden Aufgaben, für die kein Lesbarkeitsindex der Antwortmöglichkeiten berechnet werden konnte, ausgeschlossen (korrigierte Variable: lix_amk). Die mittlere Anzahl an Fachbegriffen ist in den Abbildungen am größten, in den Aufgabenstämmen am kleinsten. Zu berücksichtigen ist hierbei jedoch, dass nicht bei jeder Aufgabe eine Abbildung enthalten ist.

Zur statistischen Analyse der vorgestellten schwierigkeitsbestimmenden Aufgabenmerkmale wurden vonseiten der PAL für diese Untersuchung Prüfungsergebnisse zu den analysierten 504 ME-Prüfungsaufgaben zur Verfügung gestellt. Bei diesen Prüfungsergebnissen handelt es sich um die aggregierten mittleren Lösungsquoten (PAL-LQ) für jede ME-Prüfungsaufgabe. Die Verteilung dieser Lösungsquoten ist im folgenden Diagramm (Abbildung 4) dargestellt. Personenbezogene Prüfungsdaten standen für diese Untersuchung nicht zur Verfügung. Für die 504 in der Untersuchung berücksichtigten ME-Prüfungsaufgaben ergibt sich eine mittlere Lösungsquote von 72,9 % (SD = 19,4 %).

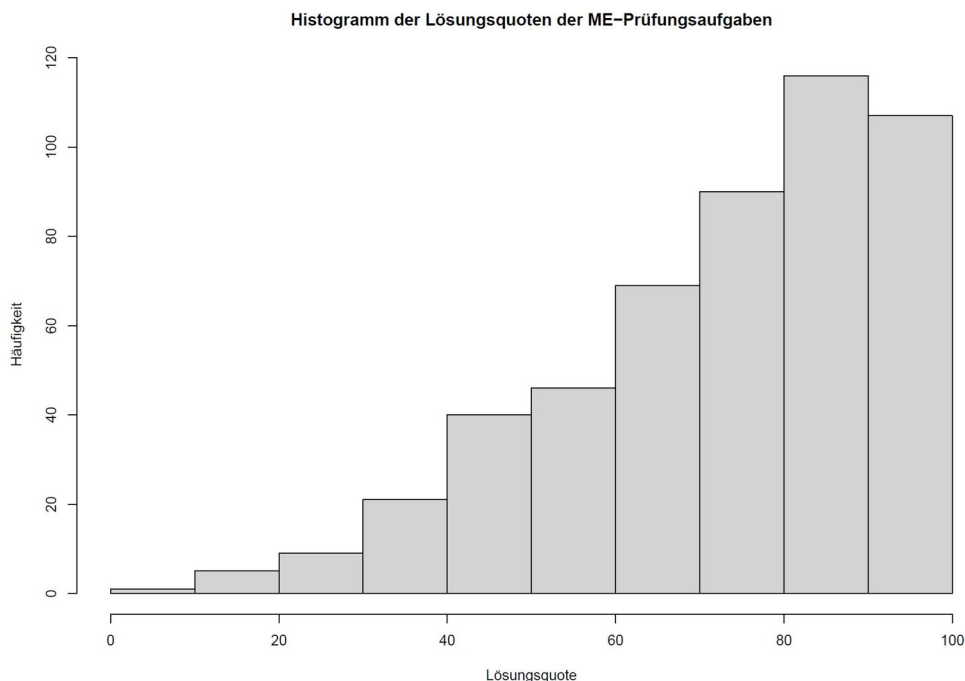


Abbildung 4: Histogramm der Lösungsquoten der analysierten ME-Prüfungsaufgaben

6.2 Identifikation relevanter Aufgabenmerkmale

Bei Untersuchungen zu schwierigkeitsbestimmenden Aufgabenmerkmalen im Rahmen von Kompetenzniveaumodellierungen werden Korrelations- und Regressionsanalysen eingesetzt, um Zusammenhänge zwischen Aufgabenmerkmalen und Lösungshäufigkeiten bzw. Lösungswahrscheinlichkeiten statistisch nachzuweisen (Dammann et al., 2016; Hartig, 2007; Just et al., 2021; Petsch et al., 2015). Für die Analyse der in dieser Untersuchung erhobenen Daten werden für die intervallskalierten Variablen (fb_as, fb_am und fb_ab) Korrelationen berechnet. Da die weiteren Variablen aus mehreren Kategorien bestehen und Vergleiche zwischen diesen Kategorien vorgenommen werden, sind hier Gruppenvergleiche das Mittel der Wahl. Die Ergebnisse der Korrelationsberechnungen werden in Tabelle 2 vorgestellt.

Tabelle 2: Ergebnisse von Pearson-Moment-Korrelationsberechnungen zwischen der Lösungsquote der ME-Prüfungsaufgaben (PAL-LQ) und den intervallskalierten Aufgabenmerkmalen fb_as, fb_am und fb_ab; Signifikanzniveau: * $p < .05$, n. s. = nicht signifikant

	fb_as	fb_am	fb_ab
PAL-LQ	−0,09*	0,04 n. s.	0,07 n. s.

Es zeigt sich lediglich ein kleiner signifikanter Zusammenhang zwischen der Lösungsquote PAL-LQ und der Anzahl der Fachbegriffe im Aufgabenstamm fb_as. ME-Prüfungsaufgaben sind somit etwas leichter, wenn die Anzahl der Fachbegriffe im Aufgabenstamm geringer ist.

Für die Analyse der Zusammenhänge zwischen PAL-LQ und den anderen untersuchten Merkmalen werden Gruppenvergleiche berechnet. Da die Verteilung der Lösungsquoten (PAL-LQ) der ME-Prüfungsaufgaben keiner Normalverteilung folgt, werden für die Gruppenvergleiche nichtparametrische Berechnungsmethoden angewendet. Für Gruppenvergleiche mit zwei unabhängigen Stichproben wird der Mann-Whitney-U-Test (Mann & Whitney, 1947), für Gruppenvergleiche mit mehr als zwei unabhängigen Stichproben der Kruskal-Wallis-Test (Kruskal & Wallis, 1952) eingesetzt. Die Ergebnisse dieser statistischen Tests sind in Tabelle 3 dargestellt.

Tabelle 3: Mittlere Lösungsquoten für die unterschiedlichen Ausprägungen der fokussierten Operationalisierungen (N = Stichprobenumfang; M = Mittelwert; SD = Standardabweichung; Mdn = Median; z = Prüfgröße des Mann-Whitney-U-Tests; H = Prüfgröße des Kruskal-Wallis-Tests; p = Signifikanzniveau)

	Stufe	N	M	SD	Mdn	z/H	p	Effektstärke
split	1	428	74,9	18,2	79,8	z = −4,44	< .001	−0,2
	2	69	66,5	17,2	67,5			
koh	1	84	70,6	19,1	76,2	H(2) = 27,5	< .001	0,23
	2	196	70,2	17,6	71,8			
	3	217	78,1	17,7	81,9			

(Fortsetzung Tabelle 3)

	Stufe	N	M	SD	Mdn	z/H	p	Effektstärke
lix_amk	1	36	72,6	17,6	76,9	H(4) = 4,65	n. s.	–
	2	73	72,5	19,1	75,6			
	3	102	75,5	17,2	79,4			
	4	59	74,9	16,9	79,8			
	5	99	76,4	20,5	79,0			
lix_as	1	43	72,4	18,1	77,1	H(4) = 7,79	n. s.	–
	2	124	73,2	16,8	74,8			
	3	170	73,7	18,6	77,3			
	4	116	76,9	17,8	80,8			
	5	44	68,2	21,3	76,1			

In Bezug auf den Split-Attention-Effekt zeigt sich, dass Prüfungsaufgaben, bei denen alle zur Bearbeitung nötigen Informationen gegeben sind, leichter sind als Prüfungsaufgaben, bei denen ein weiteres Dokument zur Bearbeitung benötigt wird. Der Unterschied ist signifikant und weist einen kleinen bis mittleren Effekt auf (Cohen, 1988; Cohen, 1992). Die Operationalisierung „Kohärenzbildungshilfe“ (koh) weist drei Ausprägungen auf. Signifikant sind die Unterschiede jedoch nur zwischen den Ausprägungen 1 und 3 ($p < .01$) sowie zwischen den Ausprägungen 2 und 3 ($p < .001$). Die Effektstärke liegt bei 0,23 und weist damit einen kleinen bis mittleren Effekt auf (Cohen, 1988; Cohen, 1992). Das bedeutet, dass Prüfungsaufgaben etwas leichter sind, wenn sie ausschließlich aus Text bestehen. Liegen zusätzlich Abbildungen vor, so unterscheidet sich im Mittel die Schwierigkeit der Prüfungsaufgaben abhängig von gegebenen oder nicht gegebenen Kohärenzbildungshilfen nicht. Die Operationalisierung „Lesekomplexität des Aufgabenstammes“ (lix_as) weist fünf Ausprägungen auf. Der Gruppenvergleich wird nicht signifikant. Das gleiche Ergebnis ergibt sich für die Operationalisierung „Lesekomplexität der Antwortmöglichkeiten“ (lix_amk). Das bedeutet, dass sich die Lösungshäufigkeiten der Prüfungsaufgaben bezogen auf die Textkomplexität nicht unterscheiden.

7 Diskussion

Ziel der Untersuchung war die Analyse bestehender Prüfungsaufgaben der PAL zum Ausbildungsberuf Mechatroniker:in hinsichtlich der schwierigkeitsbestimmenden Aufgabenmerkmale „Textverständlichkeit“ und „Bild-Text-Integration“ und deren Einfluss auf die Schwierigkeit der ME-Prüfungsaufgaben. Die Ergebnisse zeigen einen erwartungskonformen, aber kleinen bis mittleren Effekt beim Split-Attention-Effekt

und eher erwartungswidrige und ebenfalls kleine bis mittlere Effekte bei der Kohärenzbildungshilfe. Für die Anzahl an Fachbegriffen im Aufgabenstamm kann ein kleiner Effekt nachgewiesen werden. Keine Effekte zeigen sich bei der Anzahl an Fachbegriffen in den Antwortmöglichkeiten und Abbildungen sowie bei der Lesekomplexität im Aufgabenstamm und in den Antwortmöglichkeiten.

Die Ergebnisse zum Split-Attention-Effekt deuten darauf hin, dass Aufgaben, die die Verwendung zusätzlicher Dokumente erfordern, signifikant schwieriger sind. Dies könnte damit zusammenhängen, dass sich die kognitive Belastung bei Berücksichtigung mehrerer Informationsquellen erhöht. Dieser Befund sollte in einem experimentellen Forschungsdesign überprüft werden, indem Prüfungsaufgaben hinsichtlich der Darbietung relevanter Informationen variiert werden. Einschränkend ist zu berücksichtigen, dass von den vier angenommenen Stufen des „Split-Attention-Effekts“ in den analysierten ME-Prüfungsaufgaben nur die ersten beiden Stufen repräsentiert waren. Das Aufgabenmerkmal „Textverständlichkeit“ weist in der Operationalisierung „Kohärenzbildungshilfe“ (koh) einen erwartungswidrigen Befund auf. Auf Basis der verfügbaren Befunde (z. B. Brünken et al., 2005) wurde erwartet, dass Prüfungsaufgaben schwieriger sind, wenn keine Kohärenzbildungshilfen gegeben sind (z. B. kein Verweis im Aufgabenstamm auf eine zugehörige Abbildung erfolgt). Aufgaben, die nur Text enthalten, sind signifikant leichter als ME-Prüfungsaufgaben, bei denen neben dem Text auch eine Abbildung dargeboten ist. Dieser Befund verlangt nach weiteren Untersuchungen. Zu klären ist dabei u. a., ob Kohärenzbildungshilfen auch bei theoriegeleiteter Aufgabenentwicklung bzw. Veränderung von ME-Prüfungsaufgaben keine Effekte aufweisen.

Für die Operationalisierungen der Lesekomplexität zeigen sich in den Aufgabenanalysen überwiegend keine systematischen Zusammenhänge. Einzig für die korrigierte Operationalisierung „Lesekomplexität der Antwortmöglichkeiten“ (lix_amk) zeigt sich bei der Korrelationsberechnung ein kleiner Effekt, der sich jedoch bei der Analyse der Gruppenvergleiche nicht bestätigt. Der Befund ist hinsichtlich des Forschungsstandes überraschend. Es ist zu überlegen, ob Auszubildende bei der Bearbeitung der ME-Prüfungsaufgaben Lesestrategien anwenden, die es nicht erforderlich machen, die angebotenen Texte vollständig zu lesen. Tuminaro und Redish (2007) berichten für die Bearbeitung von Physikaufgaben von Bearbeitungsstrategien, bei denen Schüler:innen nur die von ihnen als bedeutsam bewerteten Textelemente (z. B. Abkürzungen physikalischer Größen) wahrnehmen. In experimentellen Forschungssettings könnte dieser Befund überprüft werden. Für die Anzahl der Fachbegriffe ergibt sich ein ähnlicher Befund wie beim Merkmal der Lesekomplexität. Ein signifikanter Zusammenhang mit kleinem Effekt kann nur für die Anzahl an Fachbegriffen im Aufgabenstamm festgestellt werden. Dieser Zusammenhang ist negativ, ME-Prüfungsaufgaben sind also leichter, je weniger Fachbegriffe im Aufgabenstamm vorliegen. Wie beim Merkmal „Lesekomplexität“ kann hier gefragt werden, ob die Auszubildenden sämtliche in einer Prüfungsaufgabe dargebotenen Fachbegriffe für ihre Bearbeitung als relevant erachten. Darüber hinaus könnte untersucht werden, ob die identifizierten Fachbegriffe auch aus inhaltlicher Perspektive für die Bearbeitung der

jeweiligen Aufgabe bedeutsam sind. Auch dieser Frage könnte in einem experimentellen Forschungssetting nachgegangen werden.

Generell zeigt sich, dass die Analyse von Prüfungsaufgaben hinsichtlich schwierigkeitsbestimmender Aufgabenmerkmale herausfordernd ist. Dies insbesondere deshalb, da Prüfungsaufgaben nicht systematisch vor dem Hintergrund schwierigkeitsbestimmender Aufgabenmerkmale entwickelt wurden und daher große Varianz hinsichtlich der Aufgabenmerkmale aufweisen. Für die im Projekt TechKom durchgeführte Untersuchung der Aufgabenmerkmale „Textverständlichkeit“ und „Bild-Text-Integration“ ergeben sich beispielsweise die folgenden weiteren Forschungsfragen:

Zeigen sich Einflüsse auf die Aufgabenschwierigkeit, wenn...

- 1.) die Lesekomplexität durch Anpassungen bei der Textlänge und eine verständliche Fachsprache (z. B. Fachbegriffe erläutern) optimiert wird?
- 2.) bei Aufgaben mit Text und Abbildungen Kohärenzbildungshilfen (z. B. Beschriftung von Abbildungen, Verweise auf Abbildungen im Text) gegeben werden?
- 3.) relevante Informationen durch zusätzliche Abbildungen untermauert werden, um den Split-Attention-Effekt zu reduzieren?
- 4.) Texte der Antwortmöglichkeiten gekürzt oder in alternativer Weise (z. B. als Tabelle) dargestellt und damit die Lesekomplexität reduziert wird?
- 5.) zu einer Aufgabe relevante Informationen (z. B. Formeln, Zeichnungen oder Angaben auf einem Tabellenbuch) beigelegt werden, um den Split-Attention-Effekt zu reduzieren?

8 Ausblick

Die vorgestellten Forschungsfragen stellen die Grundlage für eine im Projekt TechKom durchgeführte experimentelle Untersuchung dar. Dabei wurde eine Teilmenge der in dieser Untersuchung analysierten ME-Prüfungsaufgaben hinsichtlich der schwierigkeitsbestimmenden Aufgabenmerkmale „Textverständlichkeit“ und „Bild-Text-Integration“ systematisch variiert. Durch diese Veränderung ergaben sich drei Aufgabentypen: 1.) Veränderung der ME-Prüfungsaufgabe hinsichtlich des Merkmals „Textverständlichkeit“, 2.) Veränderung der ME-Prüfungsaufgabe hinsichtlich des Merkmals „Bild-Text-Integration“ und 3.) Veränderung der ME-Prüfungsaufgabe hinsichtlich beider Merkmale. Insgesamt ergab sich aus den einbezogenen originalen und den veränderten ME-Prüfungsaufgaben ein Aufgabenpool von 160 Aufgaben. Diese Aufgaben wurden unter Verwendung eines Booklet-Testheftdesigns rund 400 Auszubildenden des Ausbildungsberufs Mechatroniker:in zur Bearbeitung vorgelegt. Unter Anwendung latenter Schätzmethoden werden die Einflüsse der fokussierten Aufgabenmerkmale in diesem experimentellen Forschungssetting überprüft und die Ergebnisse in weiteren Publikationen veröffentlicht.

Literaturverzeichnis

- Anderson, L. W. (Hrsg.) (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Complete ed.). Longman.
- Ayres, P., & Sweller, J. (2022). The split-attention-principle in multimedia learning. In R. E. Mayer, & L. Fiorella (Hrsg.), *Cambridge handbooks. The Cambridge handbook of multimedia learning* (Third Edition, S. 199–211). Cambridge University Press. <https://doi.org/10.1017/9781108894333.020>
- Bals, T., & Noack, C. (September 2013). *BELkA: Begleitstudie zur Erfassung der Lesekompetenz von Auszubildenden*. Abschlussbericht. Universität Osnabrück. https://archiv.bohnenkamp-stiftung.de/fileadmin/download/BELkA_Abschlussbericht.pdf
- Beaton, A. E., & Allen, N. L. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics*, 17(2), 191–204. <https://doi.org/10.2307/1165169>
- Bloom, B. S. (Hrsg.) (1968). *Taxonomy of educational objectives: The classification of educational goals; handbook* (Repr [d. Ausg.] 1956). McKay.
- Brünken, R., & Leutner, D. (2001). Aufmerksamkeitsverteilung oder Aufmerksamkeitsfokussierung? Empirische Ergebnisse zur „Split-Attention-Hypothese“ beim Lernen mit Multimedia. *Unterrichtswissenschaft*, 29(4), 357–366. <https://doi.org/10.1024/1010-0652.19.12.61>
- Brünken, R., Seufert, T., & Zander, S. (2005). Förderung der Kohärenzbildung beim Lernen mit multiplen Repräsentationen: Fostering Coherence Formation in Learning with Multiple Representations. *Zeitschrift für Pädagogische Psychologie*, 19(1/2), 61–75. <https://doi.org/10.1024/1010-0652.19.12.61>
- Bundesagentur für Arbeit. (2025). *Informationsseite der Bundesagentur für Arbeit zum Ausbildungsberuf Mechatroniker/in*. <https://web.arbeitsagentur.de/berufenet/beruf/2868>
- Bundesinstitut für Berufsbildung. (2013). *Empfehlung des Hauptausschusses des Bundesinstituts für Berufsbildung (BIBB) zur Struktur und Gestaltung von Ausbildungsordnungen – Prüfungsanforderungen –*. <https://www.bibb.de/dokumente/pdf/HA158.pdf>
- Chandler, P., & Sweller, J. (1992). The Split-Attention Effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62(2), 233–246. <https://doi.org/10.1111/j.2044-8279.1992.tb01017.x>
- Clark, J. M., & Paivio, A. (1991). Dual Coding Theory and Education. *Educational Psychology Review*, 3, 149–210. <https://doi.org/10.1007/BF01320076>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second edition). Lawrence Erlbaum Associates Publishers. <https://doi.org/10.4324/9780203771587>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Dammann, E., Behrendt, S., Ștefănică, F., & Nickolaus, R. (2016). Kompetenzniveaus in der ingenieurwissenschaftlichen akademischen Grundbildung: Analysen im Fach Technische Mechanik. *Zeitschrift für Erziehungswissenschaft*, 19(2), 351–374. <https://doi.org/10.1007/s11618-016-0675-5>
- Deutscher, V. (2015). *Diagnostik beruflicher Kompetenzentwicklung*. Springer Gabler. <https://doi.org/10.1007/978-3-658-10681-2>

- Downing, S. M., & Haladyna, T. M. (Hrsg.) (2006). *Handbook of test development*. Erlbaum.
- Friedrich, M. (2017). *Textverständlichkeit und ihre Messung: Entwicklung und Erprobung eines Fragebogens zur Textverständlichkeit*. Waxmann Verlag.
- Geißel, B. (2008). Prädiktoren der Entwicklung zentraler Aspekte von Fachkompetenz in Berufen gewerblich-technischer Erstausbildung. In D. Münk, K. Breuer, & T. Deißinger (Hrsg.), *Schriftenreihe der Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE). Berufs- und Wirtschaftspädagogik – Probleme und Perspektiven aus nationaler und internationaler Sicht: Neuere Forschungserträge aus der Berufs- und Wirtschaftspädagogik* (S. 10–20). Verlag Barbara Budrich. <https://doi.org/10.2307/j.ctvddzs95.4>
- Gschwendtner, T. (2008). Ein Kompetenzmodell für die kraftfahrzeugtechnische Grundbildung. In R. Nickolaus (Hrsg.), *Diskussion Berufsbildung: Bd. 9. Didaktik der gewerblich-technischen Berufsbildung: Konzeptionelle Entwürfe und empirische Befunde* (S. 103–119). Schneider-Verlag.
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge Taylor & Francis Group. <https://doi.org/10.4324/9780203850381>
- Hartig, J. (2007). Skalierung und Definition von Kompetenzniveaus. In E. Klieme, & B. Beck (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)* (S. 83–99). Beltz.
- Hartig, J., Frey, A., & Jude, N. (2020). Validität von Testwertinterpretationen. In H. Moosbrugger (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (3., vollständig neu bearbeitete, erweiterte und aktualisierte Auflage, S. 529–545). Springer. https://doi.org/10.1007/978-3-662-61532-4_21
- Hewlett, C., & Kahl-Andresen, A. (2014). Prüfungsökonomie statt Prüfungsqualität? *Berufsbildung in Wissenschaft und Praxis*, 3, 6–9.
- Just, A., Behrendt, S., Macha, K., & Kögler, K. (2021). Schwierigkeitsbestimmende Aufgabenmerkmale und Grenzen ihres Einsatzes im Hochschulbereich. In H. Binz, & U. Meiser (Hrsg.), *Qualitätspakt Lehre – Individualität und Kooperation im Stuttgarter Studium: Ausgewählte Ergebnisse aus der zweiten Förderphase* (S. 89–179). Universität Stuttgart.
- Kauertz, A. (2008). *Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben* [Zugl.: Duisburg, Essen, Univ., Diss., 2007, Logos-Verl., Berlin]. Deutsche Nationalbibliothek.
- Kauertz, A. (2014). Entwicklung eines Rasch-skalierten Leistungstests. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 341–353). Springer. https://doi.org/10.1007/978-3-642-37827-0_27
- Kleinknecht, M., Bohl, T., Maier, U., & Metz, K. (Hrsg.) (2013). *Lern- und Leistungsaufgaben im Unterricht: Fächerübergreifende Kriterien zur Auswahl und Analyse*. Klinkhardt.
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(160), 583–621. <https://doi.org/10.1080/01621459.1952.10483441>

- Lehmann, R., & Seeber, S. (Hrsg.). (2007). *ULME III: Untersuchung von Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in den Abschlussklassen der Berufsschulen*. Behörde f. Bildung u. Sport.
- Lenhard, W. (2019). *Leseverständnis und Lesekompetenz: Grundlagen – Diagnostik – Förderung* (2., aktualisierte Auflage). Lehren und Lernen. Verlag W. Kohlhammer. <https://doi.org/10.17433/978-3-17-035020-5>
- Lenhard, W., & Lenhard, A. (2014–2022). *Berechnung des Lesbarkeitsindex LIX nach Björnson*. <http://www.psychometrica.de/lix.html> DOI: 10.13140/RG.2.1.1512.3447
- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 50–60. <https://doi.org/10.1214/aoms/1177730491>
- Mayer, R. E., & Fiorella, L. (Hrsg.) (2022). *Cambridge Handbooks. The Cambridge Handbook of Multimedia Learning* (Third Edition). Cambridge University Press. <https://doi.org/10.1017/9781108894333>
- 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.
- Petsch, C. (2024). *Förderung berufsfachlicher Kompetenzen in der Ausbildung der Bauwirtschaft*. Universität Stuttgart.
- Petsch, C., Norwig, K., & Nickolaus, R. (2015). Berufsfachliche Kompetenzen in der Grundstufe Bautechnik – Strukturen, erreichte Niveaus und relevante Einflussfaktoren. In A. Rausch, J. Warwas, J. Seifried, & E. Wuttke (Hrsg.), *Konzepte und Ergebnisse ausgewählter Forschungsfelder der beruflichen Bildung: Festschrift für Detlef Sembill* (S. 59–88). Schneider.
- Prenzel, M., Häußler, P., Rost, J., & Senkbeil, M. (2002). Der PISA-Naturwissenschaftstest: Lassen sich die Aufgabenschwierigkeiten voraussagen? *Unterrichtswissenschaft*, 30(2), 120–135.
- Prüfungsaufgaben- und Lehrmittelentwicklungsstelle der IHK Region Stuttgart (PAL). (2018). *Informationsflyer der PAL*. <https://www.ihk.de/blueprint/servlet/resource/blob/5281904/7b46e6c8c1f63d62783d95c7d95ad63e/pal-flyer-data.pdf>
- Prüfungsaufgaben- und Lehrmittelentwicklungsstelle der IHK Region Stuttgart (PAL). (2020). *QM-Handreichung für Prüfungsaufgabenersteller/-innen, Ratgeber für die Entwicklung von Prüfungsaufgaben*.
- Rabe, T., & Mikelskis, H. F. (2007). Kohärenzbildungshilfe und Selbsterklärungen: Fördern sie das Physiklernen? *Zeitschrift für Didaktik der Naturwissenschaften: ZfDN*, 13, 33–52.
- Rey, G. D. (2009). *E-Learning: Theorien, Gestaltungsempfehlungen und Forschung*. Verlag Hans Huber.
- Roche, J., & Terrasi-Haufe, E. (2019). Sprachkompetenzen fördern an beruflichen Schulen – Unterrichtsgestaltung im Spannungsfeld der Förderung sprachlicher Basiskompetenzen und Berufssprache. In K. Heinrichs, & H. Reinke (Hrsg.), *Wirtschaft – Beruf – Ethik: Bd. 36. Heterogenität in der beruflichen Bildung: Im Spannungsfeld von Erziehung, Förderung und Fachausbildung* (S. 167–179). wbv.

- Ropohl, M., Walpuski, M., & Sumfleth, E. (2015). Welches Aufgabenformat ist das richtige? Empirischer Vergleich zweier Aufgabenformate zur standardisierten Kompetenzmessung. *Zeitschrift für Didaktik der Naturwissenschaften*, 21, 1–15. <https://doi.org/10.1007/s40573-014-0020-6>
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2., vollst. überarb. und erw. Aufl.). Huber.
- Schmiemann, P. (2011). Fachsprache in biologischen Testaufgaben: Terminology in Biology Test Items. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 115–136.
- Schnotz, W. (2006). Was geschieht im Kopf des Lesers? Mentale Konstruktionsprozesse beim Textverstehen aus der Sicht der Psychologie und der kognitiven Linguistik. In H. Blühdorn, E. Breindl, & U. H. Waßner (Hrsg.), *Text – Verstehen: Grammatik und darüber hinaus* (S. 222–238). de Gruyter.
- Schnotz, W. (2023). *Multimedia Comprehension*. Cambridge University Press. <https://doi.org/10.1017/9781009303255>
- Schumann, S., & Eberle, F. (2011). Bedeutung und Verwendung schwierigkeitsbestimmender Aufgabenmerkmale für die Erfassung ökonomischer und beruflicher Kompetenzen. In U. Faßhauer, B. Fürstenau, & E. Wuttke (Hrsg.), *Schriftenreihe der Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft. Grundlagenforschung zum Dualen System und Kompetenzentwicklung in der Lehrerbildung* (S. 77–89). Barbara Budrich. <https://doi.org/10.5167/uzh-54458>
- Seeber, S. (2005). Input-Controlling in der beruflichen Bildung. Aspekte der Bestimmung von Lernausgangslagen. *Unterrichtswissenschaft*, 33(4), 314–333.
- Severing, E. (2011). Prüfungen und Zertifikate in der beruflichen Bildung: eine Einführung. In E. Severing, & R. Weiß (Hrsg.), *Schriftenreihe des Bundesinstituts für Berufsbildung, Bonn: Bd. 10. Prüfungen und Zertifizierungen in der beruflichen Bildung: Anforderungen, Instrumente, Forschungsbedarf* (S. 15–36). wbv.
- Stawitz, H. (2010). *Auswirkung unterschiedlicher Aufgabenprofile auf die Schülerleistung* [Zugl.: Duisburg, Essen, Univ., Diss., Logos-Verl., Berlin]. Deutsche Nationalbibliothek.
- Stöhr, A. (2017). *Prüfungen in der dualen Berufsausbildung*. Bundesinstitut für Berufsbildung.
- Sweller, J. (2010). Cognitive Load Theory: Recent Theoretical Advances. In J. L. Plass, R. Moreno, & R. Brünken (Hrsg.), *Cognitive Load Theory* (S. 29–47). Cambridge University Press. <https://doi.org/10.1017/CBO9780511844744.004>
- Tuminaro, J., & Redish, E. F. (2007). Elements of a cognitive model of physics problem solving: Epistemic games. *Physical Review Physics Education Research*, 3(2), 020101.
- Van Waveren, L., & Nickolaus, R. (2015). Struktur- und Niveaumodell des Fachwissens bei Elektronikern für Automatisierungstechnik. *Journal of Technical Education (JOTED)*, 3(2), 61–91. <https://www.journal-of-technical-education.de/index.php/joted/article/view/59>
- Wagner, S., & Schlenker-Schulte, C. (2006). Textoptimierte Prüfungsaufgaben – ein Weg zu Chancengleichheit bei schriftlichen Prüfungen. *Berufsbildung in Wissenschaft und Praxis (BWP)*, 35, 43–46.

- Wagner, S., & Schlenker-Schulte, C. (2015). *Textoptimierung von Prüfungsaufgaben*. Institut für Textoptimierung GmbH. https://zfamedien.de/downloads/ZFA/TOP_Broschue-re-gesamt.pdf
- Walpuski, M., & Ropohl, M. [Martin]. (2014). Statistische Verfahren für die Analyse des Einflusses von Aufgabenmerkmalen auf die Schwierigkeit. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 385–398). Springer. https://doi.org/10.1007/978-3-642-37827-0_30
- Weiß, R. (2011). Prüfungen in der beruflichen Bildung – ein vernachlässigter Forschungsgegenstand. In E. Severing, & R. Weiß (Hrsg.), *Prüfungen und Zertifikate in der beruflichen Bildung: eine Einführung* (S. 37–52). wbv. https://www.pedocs.de/volltexte/2011/4536/pdf/Severing_Pruefungen_und_Zertifikate_2011_D_A.pdf
- Wild, J., & Pissarek, M. (2021). *RATTE 2. Ravensburger Analysetool für Texte*. <https://www.uni-regensburg.de/sprache-literatur-kultur/germanistik-did/downloads/ratte/index.html>
- Winther, E. (2011). Kompetenzorientierte Assessments in der beruflichen Bildung – Am Beispiel der Ausbildung von Industriekaufleuten. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 107(1), 33–54. <https://doi.org/10.25162/zbw-2011-0002>

Kompetenzmessung bei Prüfungen in der beruflichen Bildung – Kommentar zu Schwerpunkt 3

ANDREAS FREY

Zusammenfassung

Es werden die in diesem Band dargestellten Befunde aus den ASCOT+-Projekten TechKom und ASPE kommentiert. Beide Projekte beschäftigen sich mit Abschlussprüfungen im beruflichen Bereich. Sie sind gute Beispiele dafür, wie Grundlagenerkenntnisse und Methoden der empirischen Bildungsforschung zielführend und erfolgreich für eine Qualitätssteigerung in einem gesellschaftlich relevanten Anwendungsbereich der Kompetenzdiagnostik genutzt werden können. Aufbauend auf der Besprechung der beiden Kapitel wird aktueller Forschungs- und Entwicklungsbedarf für die Kompetenzdiagnostik diskutiert, der aus den neuen Möglichkeiten der künstlichen Intelligenz resultiert. Elementar sind dabei konzeptuelle Klarheit bezüglich der zu messenden Kompetenzkonstrukte und eine Nutzung psychometrischer Messmodelle.

Schlagworte: Abschlussprüfungen, empirische Bildungsforschung, Kompetenzdiagnostik, Psychometrie, Testen

Abstract

The findings from the ASCOT+ projects TechKom and ASPE presented in this volume are commented on. Both projects deal with final examinations in the vocational sector. They are good examples of how basic findings and methods of empirical educational research can be used in a purposeful and successful way to improve quality in a socially relevant area of application of educational measurement. Building on the discussion of the two chapters, the current need for research and development in educational measurement resulting from the new possibilities of artificial intelligence is discussed. Conceptual clarity regarding the competence constructs to be measured and the use of psychometric models are elementary.

Keywords: educational measurement, empirical educational research, final examinations, psychometrics, testing

Kompetenzmessung bei Prüfungen in der beruflichen Bildung

Eine der zentralen Entwicklungen der letzten Jahrzehnte in den mit Bildungsprozessen befassten wissenschaftlichen Disziplinen war sicherlich die Herausbildung der empirischen Bildungsforschung. Diese formierte sich ab den 2000er-Jahren als ein interdisziplinärer Forschungsbereich im Schnittbereich von insbesondere Psychologie, Erziehungswissenschaft, Fachdidaktiken und Soziologie. Innerhalb der empirischen Bildungsforschung kommt der Kompetenzdiagnostik eine zentrale Position zu. Durch die Anreicherung der in der Psychologie häufig betrachteten kontextfreien und nicht erlernbaren Konstrukte mit kontextspezifischen, erlernbaren Konstrukten im Sinne einer Verbindung von Wissen und Können zur Bewältigung von Handlungsanforderungen, wurden zahlreiche neue Möglichkeiten zum Verständnis formeller und informeller Bildungsprozesse erschlossen. In mehreren großen Förderinitiativen der Deutschen Forschungsgemeinschaft (DFG) und des Bundesministeriums für Bildung und Forschung (BMBF) konnten wichtige Grundlagenerkenntnisse gewonnen und international sichtbare Expertise zur Kompetenzdiagnostik an Universitäten und Forschungseinrichtungen in Deutschland etabliert werden. Die berufliche Bildung wurde dabei spezifisch in den Förderprogrammen ASCOT und ASCOT+ fokussiert.

Es ist erfreulich zu sehen, wie aufbauend auf den erlangten Grundlagenerkenntnissen in ASCOT+ nun gesellschaftlich relevante Handlungsfelder praxisnah adressiert werden. Exemplarische Beispiele dafür sind die Beiträge zu den Projekten *TechKom* (Dammann et al., in diesem Band) und *ASPE* (Winther et al., in diesem Band). Beide Projekte beschäftigen sich mit Abschlussprüfungen in der beruflichen Bildung. Abschlussprüfungen kompetenzorientiert auszurichten ist direkt eingängig. Hiermit kann erreicht werden, dass die Testergebnisse verlässliche Vorhersagen für die Bewältigung beruflicher Anforderungen erlauben. Dies setzt jedoch voraus, dass hochwertige Instrumente zur Messung der berufsspezifischen Kompetenzen genutzt werden können. Im Folgenden werde ich die Beiträge zu den beiden Projekten kurz diskutieren und dann mit Erörterungen auf übergeordneter Ebene schließen.

Im Projekt *TechKom* wurde unter anderem der Frage nachgegangen, anhand welcher Merkmale die Schwierigkeit von Prüfungsaufgaben für den Ausbildungsberuf Mechatroniker:in gut vorhergesagt werden kann. Dabei wird in konstruktiver Weise eine Brücke zu Forschung zur Textverständlichkeit und zur Text-Bild-Integration beim Lernen geschlagen. Es ist hervorragend, dass dabei im Rahmen des Projekts im Einsatz befindliche Prüfungsaufgaben analysiert werden konnten. Dies eröffnet die Möglichkeit, dass die Ergebnisse der Bildungsforschung direkt auf reales Handeln in einem wichtigen pädagogischen Handlungsfeld Einfluss nehmen. Der Transfer von Forschungsergebnissen in die Praxis ist in der Bildungsforschung – wie in zahlreichen anderen wissenschaftlichen Disziplinen – eine Herausforderung, die oft nicht zufriedenstellend gemeistert wird. Umso erfreulicher ist es, dass dies hier anders ist. In Ergänzung der Ergebnisdiskussion in dem Kapitel von Dammann et al. (in diesem Band) möchte ich anmerken, dass die gefundenen kleinen bis mittleren Effekte von Aufgabenmerkmalen auf die Aufgabenschwierigkeit durchaus auch ein erfreuliches Ergeb-

nis darstellen. Offensichtlich wird die Schwierigkeit der Prüfungsaufgaben eben nicht vornehmlich von Aspekten determiniert, die wenig mit den zu prüfenden berufsspezifischen Kompetenzen zu tun haben. Es wäre doch eher bedenklich gewesen, wenn die Aufgabenschwierigkeiten maßgeblich durch die konstruktfernen Merkmale Textverständlichkeit und Bild-Text-Integration determiniert gewesen wären. Dass sich dies nicht zeigte, wirft meines Erachtens ein positives Licht auf den Itementwicklungsprozess und kann durchaus als Evidenz diskriminanter Validität der Prüfungen angesehen werden.

Möchte man zielsicher erreichen, dass qualitativ hochwertige Kompetenztests bei Abschlussprüfungen im berufsbildenden Bereich im turnusmäßigen Betrieb eingesetzt werden, sollte man dafür Sorge tragen, dass sie im Rahmen etablierter Strukturen entwickelt, vorgegeben, skaliert und gepflegt werden. Wie man Personen ohne tiefergehende Kenntnisse der Kompetenzdiagnostik bei der Aufgabenentwicklung unter die Arme greifen kann, um bildungswissenschaftliche Erkenntnisse zu schwierigkeitsgenerierenden Merkmalen in professioneller, niedrigschwelliger Weise zu nutzen, zeigt das zweite Projekt *ASPE*. In diesem Verbundvorhaben wurde die *ASPE-Workbench* konzipiert, mit der kaufmännische Abschlussprüfungen auf Basis von Designstandards konstruiert werden können. Hierbei wird unter anderem die Schwierigkeit konzipierter Prüfungsaufgaben auf Basis schwierigkeitsgenerierender Merkmale vorhergesagt. Die Facettierung der schwierigkeitsgenerierenden Merkmale hat einen gut gewählten Auflösungsgrad. Mit drei Facetten mit jeweils drei Stufen kann differenziert vorgegangen werden. Das Ganze ist aber noch überschaubar und hat ein hohes Potential, bei den mit der Aufgabenkonstruktion betrauten Personen auf Akzeptanz zu stoßen. Unabhängig davon, dass die Bezeichnung der ersten Facette als „kognitive Aktivierung“ nicht optimal erscheint, da eine Verwechslung mit der gleichnamigen Tiefenstruktur von Unterricht (Kunter & Trautwein, 2013) möglich ist, zeigen die in dem Beitrag von Winther et al. (in diesem Band) berichteten Ergebnisse, dass die Vorhersage der Aufgabenschwierigkeiten gut funktioniert. Sollte es im Projekt nicht schon umgesetzt worden sein, wäre ein sinnvoller nächster Schritt die Integration auch von Laien nutzbarer Funktionen bzw. eine Schnittstelle zu entsprechender Software, mit der Testergebnisse mit einem Modell der Item-Response-Theory (z. B. van der Linden, 2016) auch von psychometrischen Laien skaliert werden können. Hierdurch kann das volle Potential der Kompetenzdiagnostik erschlossen werden, um beispielsweise kriteriumsorientierte Bestehensgrenzen zu definieren. Ein gut auf Prüfungen im beruflichen Bereich übertragbares Vorgehen wurde für den Anwendungskontext kompetenzorientierter Hochschulklausuren von Frey (2021) beschrieben und in der Software *KAT-HS-App* (Fink et al., 2021) implementiert.

Die Entwicklung zu Abschlussprüfungen in der beruflichen Bildung im Speziellen und zur Kompetenzdiagnostik im Allgemeinen ist mit dem erreichten Stand keineswegs abgeschlossen. Weiterer Forschungs- und Entwicklungsbedarf resultiert insbesondere aus den neuen Möglichkeiten der künstlichen Intelligenz (KI), die seit Ende 2022 durch die Veröffentlichung von ChatGPT einfach zugänglich geworden sind. Diesbezüglich sind zwei zentrale Änderungen zu verzeichnen. Erstens sind die beruf-

lichen Anforderungen am Arbeitsplatz dabei, sich deutlich zu verändern. Nicht nur in techniknahen Gebieten wird der Umgang mit KI-Systemen voraussichtlich deutlich an Wichtigkeit gewinnen. Dies wird sich in der Folge in geänderten Ausbildungsinhalten und Abschlussprüfungen niederschlagen. Zweitens ergeben sich neue Möglichkeiten der Konstruktion von Abschlussprüfungen mit Hilfe generativer KI. Beispiele großer Systeme, wie sie beispielsweise für den Duolingo-English-Test genutzt werden (Burstein et al., 2024), existieren bereits. Dutzende wissenschaftliche Publikationen beschreiben die besten Strategien zur KI-basierten Generierung und Optimierung von Testverfahren. Dies macht auch vor der KI-basierten Kalibrierung von Verfahren nicht halt, wozu ebenfalls erste Publikationen vorliegen.

Es wird in Zukunft darauf ankommen, die neuen KI-basierten Möglichkeiten nutzbar zu machen und gleichzeitig die elementaren Ecksteine der Kompetenzdiagnostik zu fixieren. Diese werden beispielsweise durch die Schritte der Konstruktion von Kompetenztests (Frey & Hartig, 2022) beschrieben. Elementar ist dabei, dass eine konzeptuelle Klarheit bezüglich der zu messenden Kompetenzen besteht. Hierbei wird es auch weiterhin darauf ankommen, berufsspezifische und generische kognitive Anforderungen zu konkretisieren, die für die Bewältigung beruflicher Anforderungen nötig sind. Weiterhin unverzichtbar ist die Quantifizierung von Kompetenzständen auf Basis psychometrischer Modelle. Erst deren Nutzung erlaubt die eindeutige Ableitung kriteriumsorientierter Testwertinterpretationen (Herzberg & Frey, 2011) als belastbare Grundlage für das Erteilen einer Qualifikation. Diese beiden Aspekte – detaillierte Definition von Konstrukten und psychometrische Modellierung – sind bei aller Aufbruchsstimmung und zuweilen auch Euphorie aufgrund der Möglichkeiten durch KI stets im Mittelpunkt der Kompetenzdiagnostik zu behalten.

Literatur

- Burstein, J., LaFlair, G. T., Yancey, K., von Davier, A. A., & Dotan, R. (2024). *Responsible AI for test equity and quality: The Duolingo English Test as a case study*. arXiv. <https://arxiv.org/abs/2409.07476>
- Fink, A., Spoden, C., Frey, A., & Naumann, P. (2021). Kriteriumsorientiertes adaptives Testen mit der KAT-HS-App. *Diagnostica*, 67, 110–114. <https://doi.org/10.1026/0012-1924/a000268>.
- Frey, A. (2021). Hochschulklausuren als kriteriumsorientierte Testverfahren. In C. Spoden, & A. Frey (Hrsg.), *Psychometrisch fundierte E-Klausuren für die Hochschule* (S. 39–48). Pabst Science Publishers.
- Frey, A., & Hartig, J. (2022). Kompetenzdiagnostik. In M. Harring, C. Rohlfs, & M. Gläser-Zikuda (Hrsg.), *Handbuch Schulpädagogik* (2. Auflage, S. 928–937). Waxmann. <https://doi.org/10.36198/9783838587967>
- Herzberg, P. Y., & Frey, A. (2011). Kriteriumsorientierte Diagnostik. In L. F. Hornke, M. Amelang, & M. Kersting (Hrsg.), *Methoden der psychologischen Diagnostik* (S. 281–324). Hogrefe.

- Kunter, M., & Trautwein, U. (2013). *Psychologie des Unterrichts*. Schoeningh. <https://doi.org/10.36198/9783838538952>
- van der Linden, W. J. (2016). *Handbook of item response theory. Volume one: models*. Chapman & Hall/CRC. <https://doi.org/10.1201/9781315374512>

Abkürzungsverzeichnis

AG BFN	Arbeitsgemeinschaft Berufsbildungsforschungsnetz
AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
ASCOT+	Technology-based Assessment of Skills and Competences in Vocational Education and Training
BBiG	Berufsbildungsgesetz
BIBB	Bundesinstitut für Berufsbildung
BIC	Bayesian Information Criterion
BMBF	Bundesministerium für Bildung und Forschung
BMWA	Bundesministerium für Wirtschaft und Arbeit
BMWT	Bundesministerium für Wirtschaft und Technologie
CFI	Comparative Fit Index
CI	Confidence Interval
COVID-19	coronavirus disease 2019
DF	degrees of freedom
DIF	Differential Item Functioning
EAP	Expected-a-posteriori
ECVET	European Credit System for Vocational Education and Training
EQF	European Qualification Framework
ESNaS	Evaluation der Standards in den Naturwissenschaften
EU	Europäische Union
IHK	Industrie- und Handelskammer
ILIAS	Integriertes Lern-, Informations- und Arbeitskooperations-System
IRT	Item-Response-Theory
KMK	Kultusministerkonferenz
OECD	Organisation for Economic Co-operation and Development

PAL	Prüfungsaufgaben- und Lehrmittelentwicklungsstelle
PISA	P rogramme for I nternational S tudent A ssessment
PISA-VET	P rogramme for I nternational S tudent A ssessment- V ocational Education and T raining
QM	Qualitätsmanagement
RM-ANOVA	Repeated Measures Analysis of Variance
RMSEA	Root mean square error of approximation
SOFI	Soziologisches Forschungsinstitut
SRMR	Standardized Root Mean Residual
TLI	Tucker–Lewis Index
VERA	V ergleichsarbeiten
VET	Vocational Education and Training
VET-LSA	Vocational Education and Training-Large Scale Assessment
WLE	Weighted Likelihood Estimates
wMNSQ	weighted mean squared
4C/ID	Four-Component Instructional Design

Autor:innenverzeichnis¹

Abele, Stephan

Prof. Dr. Stephan Abele; Technische Universität Dresden; Professur für Berufspädagogik; stephan.abele@tu-dresden.de

Beck, Klaus

Prof. Dr. Klaus Beck; Johannes Gutenberg-Universität Mainz; Wirtschaftspädagogik; beck@uni-mainz.de

Dammann, Elmar

Dr. Elmar Dammann; Universität Hamburg; Berufliche Bildung und lebenslanges Lernen; elmar.dammann@uni-hamburg.de

Danek, Amory H.

PD Dr. Amory H. Danek; Technische Universität München; Zentrum für internationale Bildungsvergleichsstudien (ZIB); amory.danek@tum.de

Deutscher, Viola

Prof. Dr. Viola K. Deutscher, Georg-August-Universität Göttingen; Wirtschaftspädagogik und digitales berufliches Lernen; viola.deutscher@uni-goettingen.de

Fortunati, Fabio

Dr. Fabio Fortunati; Universität Duisburg-Essen; Berufliche Aus- und Weiterbildung; fabio.fortunati@uni-due.de

Frey, Andreas

Prof. Dr. Andreas Frey; Goethe-Universität Frankfurt am Main, Pädagogische Psychologie mit Schwerpunkt Beratung, Diagnostik & Evaluation; frey@psych.uni-frankfurt.de

Glogger-Frey, Inga

Prof. Dr. Inga Glogger-Frey; Universität Erfurt; Professur für Bildungspsychologie; inga.frey@uni-erfurt.de

Greiff, Samuel

Prof. Dr. Samuel Greiff, Technische Universität München, Lehrstuhl für Educational Monitoring and Effectiveness; samuel.greiff@tum.de

¹ Stand Juli 2025

Gschwendtner, Tobias

Prof. Dr. Tobias Gschwendtner; Pädagogische Hochschule Ludwigsburg; Abteilung Technik; gschwendtner@ph-ludwigsburg.de

Hartmann, Philipp

Dr. Philipp Hartmann; Georg-August-Universität Göttingen; Professur für Anwendungssysteme und E-Business; philipp.hartmann@uni-goettingen.de

Hartmann, Stefan

Dr. Stefan Hartmann; Pädagogische Hochschule Ludwigsburg; Abteilung Technik; stefan.hartmann@ph-ludwigsburg.de

Hesse, Peter

Dipl. -Ing. Peter Hesse; Technische Universität Dresden; Professur für Berufspädagogik; peter.hesse@tu-dresden.de

Hill, Svenja

M.Ed. Svenja Hill; Technische Universität München; Lehrstuhl für Berufspädagogik; svenja.hill@tum.de

Kaepernick, Jödis

M.Sc. Jödis Kaepernick; Technische Universität Dresden; Professur für Berufspädagogik; joerdis.kaepernick@tu-dresden.de

Kerres, Michael

Prof. Dr. Michael Kerres, Universität Duisburg-Essen; Learning Lab; michael.kerres@uni-duisburg-essen.de

Kohlus, Alexander

M. A. Alexander Kohlus; Universität Duisburg-Essen; Berufliche Aus- und Weiterbildung; alexander.kohlus@uni-due.de

Krebs, Philine

Dr. Philine Krebs; Georg-August-Universität Göttingen; Wirtschaftspädagogik und Personalentwicklung; philine.krebs@uni-goettingen.de

Leutner, Detlev;

Prof. Dr. Dr. h. c. Detlev Leutner; Universität Duisburg-Essen; Lehrstuhl für Lehr-Lernpsychologie; detlev.leutner@uni-due.de

Meiners, Hanna

M.Ed. Hanna Meiners; Georg-August-Universität Göttingen; Wirtschaftspädagogik und Personalentwicklung; hanna.meiners@uni-goettingen.de

Rausch, Andreas

Prof. Dr. Andreas Rausch; Universität Mannheim; Lehrstuhl für Wirtschaftspädagogik, Lernen im Arbeitsprozess; rausch@uni-mannheim.de

Rexhäuser, Dave;

M.Sc. Dave Rexhäuser; Technische Universität Dresden; Professur für Berufspädagogik; dave.rexhaeuser@tu-dresden.de

Seeber, Susan

Prof. Dr. Susan Seeber, Georg-August-Universität Göttingen; Professur für Wirtschaftspädagogik und Personalentwicklung; susan.seeber@wiwi.uni-goettingen.de

Seifried, Jürgen

Prof. Dr. Jürgen Seifried; Universität Mannheim; Lehrstuhl für Wirtschaftspädagogik, Berufliches Lehren und Lernen; juergen.seifried@uni-mannheim.de

Spiel, Christiane

emer. Prof. Dr. Dr. Christiane Spiel; Universität Wien; Bildungspsychologie; christiane.spiel@univie.ac.at

Striković, Aldin

Dr. Aldin Striković; Technische Universität München; Lehrstuhl für Berufspädagogik; aldin.strikovic@tum.de

Schäfer, Pia

M.Ed. Pia Schäfer; Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau; Fachdidaktik in der Technik; Pia.Schaefer@cs.rptu.de

Schlünkes, Sebastian

M.Sc. Sebastian Schlünkes; Universität Duisburg-Essen; Berufliche Aus- und Weiterbildung; sebastian.schlunkes@uni-due.de

Schumann, Matthias

Prof. Dr. Matthias Schumann; Georg-August-Universität Göttingen; Professur für Anwendungssysteme und E-Business; mschuma1@uni-goettingen.de

Velten, Stefanie

Dipl.-Psych. Stefanie Velten, Bundesinstitut für Berufsbildung; Arbeitsbereich „Kompetenzentwicklung“; velten@bibb.de

Vogel, Wolfgang

Dr. Wolfgang Vogel, Aufgabenstelle für kaufmännische Abschluss- und Zwischenprüfungen; wolfgang.vogel@nuernberg.ihk.de

Walker, Felix

Prof. Dr. Felix Walker; Universität Hamburg; Professur für Berufspädagogik insb. technischer Fächer; felix.walker@uni-hamburg.de

Warwas, Julia

Prof. Dr. Julia Warwas; Universität Hohenheim; Professur für Wirtschaftspädagogik, insbesondere Theorie und Didaktik beruflicher Bildung; julia.warwas@uni-hohenheim.de

Weyland, Ulrike

Prof. Dr. Ulrike Weyland; Universität Münster; Professur für Erziehungswissenschaft mit dem Schwerpunkt Berufspädagogik; ulrike.weylend@uni-muenster.de

Wilczek, Larissa

M. A. Larissa Wilczek; Universität Münster; Professur für Erziehungswissenschaft mit dem Schwerpunkt Berufspädagogik; larissa.wilczek@uni-muenster.de

Winther, Esther

Prof. Dr. Esther Winther; Universität Duisburg-Essen; Berufliche Aus- und Weiterbildung; esther.winther@uni-due.de

Wittmann, Eveline

Prof. Dr. Eveline Wittmann; Technische Universität München; Lehrstuhl für Berufspädagogik; eveline.wittmann@tum.de

Wuttke, Eveline

Prof. Dr. Eveline Wuttke; Goethe-Universität Frankfurt am Main; Wirtschaftspädagogik; wuttke@em.uni-frankfurt.de

Wie können Forschungsergebnisse wirksam in Ausbildungspraxis und Prüfungssysteme überführt werden? Der Sammelband präsentiert zentrale Ergebnisse der BMBF-Initiative ASCOT+ (2019–2023). Im Fokus stehen digitale Lehr-Lern- und Prüfungsinstrumente, die berufliche Kompetenzen fördern, Problemlösefähigkeit stärken und prüfungsrelevante Aufgaben kompetenzorientiert gestalten. Beiträge aus sechs Projekten sowie Kommentare namhafter Expert:innen verbinden wissenschaftliche Fundierung mit Praxisnähe. Damit leistet der Band einen Beitrag zur Weiterentwicklung beruflicher Bildung im digitalen Wandel.

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.