

# Konstruktvalidität konstruktiver Problemlösefähigkeit bei Elektronikern für Automatisierungstechnik

**KURZFASSUNG:** Papier-und-Bleistift-Tests sind ökonomisch in Testsituationen einsetzbar. Bislang ist nicht geklärt, ob mit solch einer Repräsentationsform auch die konstruktive Problemlösefähigkeit bei Elektronikern für Automatisierungstechnik valide erfasst werden kann. Dieser Fragestellung geht der Beitrag nach, in dem Testleistungen, die mit Papier-und-Bleistift-Tests erhoben wurden, mit Testleistungen aus einer realen Anforderungssituation in Beziehung gesetzt werden. Acht komplexe Problemstellungen werden deshalb in einem Cross-Over-Design 145 Auszubildenden am Ende ihrer Ausbildung zur Bearbeitung vorgegeben und über einen Vergleich von Strukturgleichungsmodellen geprüft. Im Ergebnis kann gezeigt werden, dass die Testleistung, welche in einem Papier-und-Bleistift-Test erbracht wurde und jene, welche in einer realen Entwicklungsumgebung erbracht wurde, als konvergent valide interpretiert werden kann und der Papier-und-Bleistift-Test zur Erfassung der konstruktiven Problemlösefähigkeit bei Elektronikern für Automatisierungstechnik als geeignet erscheint.

**ABSTRACT:** Paper-pencil-tests are efficient assessment instruments. An uncleared item is nevertheless, if it is possible to measure constructive problem-solving competencies in a valid fashion using paper-pencil-tests within the occupation of electronics technicians for automation technology. To clarify this, this paper examines the relationship between the test performances within a paper-pencil-test on the one hand and within real life workplace test situations on the other. Eight complex problems were developed for both test formats. These complex problems were then administered in a cross-over-design using a sample of 145 electronics technicians for automation technology at the end of their apprenticeship. The relationship between the test formats were then estimated using structural equation models. Based on these analyses our results document that the test performances within a paper-pencil-test are highly comparable to test performances within real life programming environments. Therefore, the paper-pencil-test seems to be convergent valid to the real-life constructive problem-solving situations and shows an efficient way to assess constructive problem-solving competencies.

## 1. Thematische Einordnung

Elektroniker für Automatisierungstechnik<sup>1</sup> arbeiten an komplexen Anlagen der automatisierten Produktion bzw. Fertigung und müssen deren Programmierung, Inbetriebnahme und Instandhaltung beherrschen. Sie werden, neben der Chemie- und Automobilindustrie, hauptsächlich von Betrieben der Elektro- und Maschinenbaubranche beschäftigt (vgl. BUNDESAGENTUR FÜR ARBEIT 2014; VERORDNUNG ÜBER DIE BERUFAUSBILDUNG IN INDUSTRIELLEN ELEKTROBERUFEN, S. 1683).

Ausgelöst u. a. durch die Implementierung des Lernfeldkonzepts und der Abkehr von der Input- und der Hinwendung zu einer stärkeren Outputsteuerung des Berufsbildungssystems wurde die Frage nach entsprechenden Messinstrumenten

1 Der einfacheren Lesbarkeit halber wird nur die männliche Schreibweise verwendet. Grundsätzlich sind jedoch beide Geschlechter gemeint. Aus selbigem Grund wird einheitlich die Berufsbezeichnung Elektroniker für Automatisierungstechnik verwendet. Diese Bezeichnung schließt die entsprechende Berufsbezeichnung des Handwerks mit ein.

zur Erfassung des Outputs, der erworbenen Kompetenzen der Auszubildenden, virulent. Unter dem Anspruch wissenschaftlicher Gütekriterien liegen für die meisten Ausbildungsberufe entsprechende Instrumente nicht vor. Problematisch ist insbesondere die Einlösung des Anspruchs der ganzheitlichen Erfassung des Konstrukts beruflicher Handlungskompetenz mit all ihren Facetten oder Teilkompetenzen (vgl. etwa BADER/MÜLLER 2002; KNÖLL 2007). In den letzten Jahren entstanden Arbeiten zur Kompetenzerfassung, die sich auf einzelne Ausbildungsberufe konzentrieren und die sich auf die i. d. R. empirisch leichter zugänglichen Teilkompetenzen des Bereichs berufsfachlicher Kompetenzen beschränken (vgl. NICKOLAUS u. a. 2011; SEEGER/LEHMANN 2011; WINTHER/ACHTENHAGEN 2013; WITTMANN u. a. 2014), z. T. aber auch emotionale Befindlichkeiten u. ä. mit aufnehmen (vgl. SEMBILL/RAUSCH/KÖGLER 2013; RAUSCH u. a. 2010). Die ASCOT-Initiative des BMBF setzte weitere Impulse zur Entwicklung valider Instrumente der Kompetenzdiagnostik.

Vorarbeiten zur Kompetenzerfassung legen nahe, dass sich auch berufsfachliche Kompetenzen in empirisch trennbare Teilkompetenzen untergliedern lassen und mehrdimensionale Konstrukte vorherrschen (vgl. NICKOLAUS u. a. 2011; GSCHWENDTNER 2011; WINTHER/ACHTENHAGEN 2009; SEEGER 2008; ROSENDAHL/STRAKA 2011). Für die in diesem Beitrag zu verfolgende Fragestellung soll der Erfassung der Problemlösefähigkeit vertiefter nachgegangen werden. Auf Basis von Anforderungsanalysen von Tätigkeiten ist zu vermuten, dass der Kompetenzaspekt fachspezifischer Problemlösefähigkeit ausdifferenziert werden kann in einen konstruktiven und einen analytischen Teil, so dass eine weitere Fokussierung notwendig wird. Damit erfolgt für das hier referierte Forschungsprojekt<sup>2</sup> die Eingrenzung auf die Entwicklung eines arbeitsprozessnahen Messinstruments für die Erfassung konstruktiver fachspezifischer Problemlösefähigkeit. In der gewerblich-technischen Domäne wurde die Einbeziehung konstruktiver Anforderungen in Tests bislang kaum systematisch bearbeitet.

## 2. Begründungslinien der Konstruktentwicklung

### 2.1 Mehrdimensionale Strukturen berufsfachlicher Kompetenz

Der aktuelle Forschungsstand innerhalb der gewerblich-technischen Domäne weist mit dem Fachwissen und mit der Anwendung des Fachwissens (Problemlösefähigkeit) zwei empirisch trennbare Subdimensionen der Fachkompetenz aus (vgl. NICKOLAUS 2011, S. 333f.). Die fachspezifische Problemlösefähigkeit wird hierbei definiert als ein Bündel kognitiver Leistungsdispositionen, welche ein selbstständiges, methodengeleitetes und sachgerechtes Handeln in berufsfachlichen problemhaltigen Situationen ermöglicht (vgl. HARTIG/KLIEME 2007; KMK 2007; ABELE/WALKER/NICKOLAUS 2014).

Doch selbst die beiden Dimensionen des Fachwissens und der Problemlösefähigkeit sind psychometrisch betrachtet selbst wiederum keine homogen ausgeprägten Konstrukte. So zeigen z. B. die Leistungsdaten aus spezifisch konstruierten Fachwis-

2 Gefördert wird das Projekt vom Bundesministerium für Bildung und Forschung. Es ist Teil der Forschungsinitiative technologieorientierte Kompetenzmessung in der beruflichen Bildung (Technology-Oriented Assessment of Skills and Competencies in VET-ASCOT), welche im Dezember 2011 startete. Weitere Informationen im Internet unter [www.ascot-vet.net](http://www.ascot-vet.net).

senstests durchgängig mehrdimensionale Strukturen berufsfachlicher Kompetenz an. Diese Struktur gliedert sich am Ende der Ausbildung stärker nach Fach- und Tätigkeitsbereichen aus und nicht nach der ursprünglich angenommenen, psychologisch begründeten dreidimensionalen Ausdifferenzierungen in deklaratives und prozedurales Wissen sowie die Problemlösefähigkeit (siehe z. B. KNÖLL 2007; GEISSEL 2008). So ergeben sich etwa bei Elektronikern für Energie- und Gebäudetechnik Ausdifferenzierungen des Fachwissens in die Bereiche traditionelle Installationstechnik, elektrotechnische Grundlagen und Steuerungs-/moderne Installationstechnik (u. a. Bussysteme). Je nach Modellbildung wird diese Ausdifferenzierung um die fachspezifische Problemlösefähigkeit erweitert (vgl. NICKOLAUS u. a. 2011, S. 86). Auch GSCHWENDTNER konnte bei Kfz-Mechatronikern eine mehrdimensionale Struktur des Fachwissens nachweisen (vgl. ebd. 2011, S. 62f.). Dieser kurze Überblick zur derzeitigen Befundlage in der gewerblich-technischen Domäne zeigt auf, dass mehrdimensionale Strukturen berufsfachlicher Kompetenz eher die Regel als die Ausnahme darstellen und eine vertiefte Beschäftigung mit der validen Erfassung von Teildimensionen notwendig werden lässt.

## 2.2 Diagnostik fachspezifischer analytischer Problemlösefähigkeit

Die testbasierte Erfassung von Problemlösefähigkeit in technischen Systemen wird bislang häufig mit authentischen Computersimulationen umgesetzt, die es erlauben, in der Testsituation komplexe Anforderungen abzubilden und Interaktionen der Probanden mit den (simulierten) technischen Systemen zu initiieren (z. B. Messungen durchführen, Demontage von Komponenten, Anlagenbedienung wie Start-Stop-Funktionen auslösen und das Betätigen von Schaltern sowie das Einstellen von Sollwerten an einer Regelung (vgl. KNÖLL 2007; NITZSCHKE/GEISSEL/NICKOLAUS 2011; GSCHWENDTNER/ABELE/NICKOLAUS 2009; HOCHHOLDINGER/SCHAPER 2013)). GSCHWENDTNER, ABELE und NICKOLAUS (2009) konnten im Beruf des Kfz-Mechatronikers zeigen, dass ihr simulationsbasiertes Testinstrument zur Erfassung analytischer Problemlösefähigkeit als konvergent valide hinsichtlich der erbrachten Testleistungen in einer realen Arbeitsumgebung interpretiert werden kann. Dieser Nachweis ermöglicht es für diesen Beruf und für den Zweck eines Large-Scale-Assessments, diese Teilkompetenz nicht durch reale Arbeitsproben, sondern ökonomischer mittels Computersimulation zu erfassen.

Die fachspezifische Problemlösefähigkeit bei Elektronikern für Automatisierungstechnik kann nicht, wie einleitend bereits vermerkt, in Abhängigkeit der spezifischen Anforderungen als ein eindimensionales Konstrukt aufgefasst werden. Beim Beruf des Kfz-Mechatronikers hingegen dominieren in erster Linie analytische Anforderungssituationen den Berufsalltag, die häufig über die Fehlerdiagnosefähigkeit in einem Kraftfahrzeug operationalisiert wird (vgl. ebd.). Bei Elektronikern für Automatisierungstechnik treten neben ebenfalls analytisch geprägten Anforderungssituationen in der Fehleranalyse in automatisierten Anlagen auch konstruktive Anforderungssituationen auf. Dies ist z. B. beim Entwurf und der Programmierung von Steuerungsprogrammen der Fall, auf welche wir uns im Weiteren beschränken müssen.

### 2.3 Ansätze zur Diagnostik fachspezifischer konstruktiver Problemlösefähigkeit

Erste Analysen zu konstruktiven Anforderungen zeigen, dass für das Erstellen von Steuerungsprogrammen normierte und vordefinierte Symbole und Grafiken verwendet werden, die entsprechend der Problemstellung auszuwählen, zu kombinieren und mit den relevanten Variablen und Operatoren zu versehen sind. Auf der Prozessebene unterscheidet LANGMANN (2004) bei der Programmierung einer speicherprogrammierbaren Steuerung (SPS) drei Phasen. Erstens die Erstellung eines Steuerungsprogrammes für die SPS. Dies erfolgt in der Regel am PC mit einer graphisch gestützten Entwicklungsumgebung. Zweitens die Übertragung der Programmlogik über die Schnittstelle zum Programmiergerät und drittens die zyklische Ermittlung der Eingangszustände und Beschaltung der Ausgänge nach dem Steueralgorithmus. Dies stellt die eigentliche Online-Arbeitsphase der SPS am Automatisierungssystem dar. Es ist ersichtlich, dass die eigentliche Programmierung dabei zeitlich und räumlich getrennt von dem eigentlichen Automatisierungssystem (wie z. B. dem Mensa-Band) erfolgt bzw. erfolgen kann. Erst nach Fertigstellung des Codes in der Entwicklungsumgebung wird das Programm auf die Steuerung geladen.

Im Curriculum und der Ausbildungsordnung des Berufes Elektroniker für Automatisierungstechnik ist das Programmieren von solchen Steuerungen fest verankert und besitzt eine hohe Relevanz: Eine Ausbilderbefragung und Tätigkeitsanalyse des Bundesinstituts für Berufsbildung spricht von einem hohen Stellenwert der SPS-Programmierung gegenwärtig und zukünftig (vgl. ZINKE/SCHENK/KRÖLL 2014).

Zur Erfassung konstruktiver Problemlösefähigkeit stehen prinzipiell die Möglichkeiten von Selbst-/Fremdeinschätzungen, Papier-und-Bleistift-Tests, simulierte Arbeitsproben (Computersimulationen) und reale Arbeitsproben mit ihren jeweiligen Vor- und Nachteilen zur Verfügung (vgl. GSCHWENDTNER/ABELE/NICKOLAUS 2009; NITZSCHKE/GEISSEL/NICKOLAUS 2011). In vorangegangenen Untersuchungen erwiesen sich Papier-und-Bleistift-Tests sowie auf realen Arbeitsprozessen basierende Computersimulationen als geeignete Messinstrumente zur Erfassung fachlicher Kompetenzen im Hinblick auf die Einhaltung wissenschaftlicher Gütekriterien, welche ergänzt werden durch die Kriterien der Praktikabilität (Testzeit, Aufwandsgrenzen) und Skalierbarkeit (vgl. GSCHWENDTNER/ABELE/NICKOLAUS 2009). Vergleichbare Ergebnisse zur Erfassung konstruktiver Problemlösefähigkeit im Beruf des Elektrikers für Automatisierungstechnik liegen bislang nicht vor.

### 3. Fragestellung und Hypothese

Bei der Nutzung einer realen Entwicklungsumgebung treten für Testzwecke jedoch hohe Lizenzkosten<sup>3</sup> auf und erweist sich ferner deren Einsatz in der Kompetenzdiagnostik für Large-Scale-Assessments als nicht praktikabel<sup>4</sup>. Aus diesen Grün-

3 Die Einzelplatzlizenz der SPS-Software Siemens SIMATIC S7 STEP7 V5.4 kostet ca. 1.800 € (die im Funktionsumfang eingeschränkte Schülerversion ca. 40 €).

4 Es werden zur Installation der Software ca. 45 Minuten benötigt (pro Rechner); des Weiteren müssen zwei Gigabyte freier Speicherplatz im lokalen Benutzerprofil zur Verfügung stehen und Administratorrechte vorhanden sein. Zwar ist eine solche Entwicklungsumgebung in Teilen in separaten SPS-Laborräumen an beruflichen Schulen oder in Ausbildungsbetrieben vorhanden, jedoch sind in diesen Räumen meistens ungünstige Erhebungsbedingungen, wie eine geringe

den erscheint der Vergleich der erbrachten Testleistung in der computerbasierten Entwicklungsumgebung mit der Testleistung, welche in einem einfach zu administrierenden Papier-und-Bleistift-Test erbracht wurde, hinsichtlich der Diagnostik konstruktiver Problemlösefähigkeit erfolgsversprechend. Diese Testvariante ist jedoch nur dann zulässig bzw. valide, wenn nachgewiesen werden kann, dass die Papier-und-Bleistift-Testumgebung vergleichbare Testleistungen anzeigt wie Tests, in welchen die reale Entwicklungsumgebung zur Verfügung steht. Als Hypothese formuliert ist in der Untersuchung folgender Zusammenhang zu prüfen:

*Die auf Papier-und-Bleistift-Tests basierenden Testleistungen können in Bezug auf die erbrachten Leistungen in der computerbasierten Entwicklungsumgebung als konvergent valide interpretiert werden.*

#### 4. Forschungsdesign

Zur Hypothesenprüfung müssen Testleistungen von Probanden sowohl von Aufgaben im Format des Papier-und-Bleistift-Tests (PBT) sowie von der realen Entwicklungsumgebung bearbeitet vorliegen. Anschließend können diese beiden Testleistungen mit geeigneten statistischen Verfahren aufeinander bezogen werden und so ein Vergleich bzw. Aussagen zur Interpretation der Validität des PBT ermöglichen. In der Testkonstruktion wurde die Bearbeitungszeit vorab auf 2x60 Minuten fixiert, um einerseits komplexere und realitätsnahe Items entwickeln und einsetzen zu können, die entsprechend eine längere Bearbeitungszeit beanspruchen und andererseits keine Übertestung der Auszubildenden bewirken. Als Ausgangsszenario der Testsituation werden in einem Betrieb Pakete über Förderbänder transportiert, aussortiert und über weitere Förderbänder sowie ein Handhabungsgerät mit Vakuumbreifer in Auffangboxen ausgelagert. Insgesamt wurden acht komplexe konstruktive Problemstellungen in das Szenario eingebettet (insgesamt hat der Test 20 Items). Inhaltlich basieren die in den Problemstellungen abgebildeten Anforderungen auf curricularen Analysen und drei aufeinander aufbauenden Workshops mit Experten der beruflichen Ausbildungspraxis. Die Problemstellungen enthalten immer die Funktionsbeschreibung bzw. den Funktionsablauf, die benötigten Variablenbezeichnungen, das Technologieschema sowie den Arbeitsauftrag.

Folgende Abbildung 1 stellt eine exemplarische Problemstellung in der Entwicklungsumgebung dar. Dabei ist im linken Bild bei den Netzwerken 1 und 2 ein eher geschlossenes Aufgabenformat dargestellt, bei dem die Auszubildenden passende Variablen zuordnen müssen, im rechten Bild bei Netzwerk 3 und 4 ein eher offenes Format dargestellt, bei dem die passenden Bausteine bzw. Operatoren ausgewählt sowie miteinander logisch verknüpft werden müssen<sup>5</sup>.

In beiden Abbildungen sind in der linken Spalte die Bitverknüpfungen, Vergleicher, Zähler etc. systemintern in der Entwicklungsumgebung hinterlegt, die der Programmierer mittels Drag-and-drop-Technik nutzen kann. Die Programmierung erfolgt dann im rechten Bereich und kann mehrere Netzwerke umfassen.

Anzahl an PC-Arbeitsplätzen, Einrichtung der PC-Arbeitsplätze als Partnerarbeitsplätze oder eine für Testsituationen ungünstige fixierte Tischordnung, vorzufinden.

5 Aus Gründen der Übersichtlichkeit wird in Abbildung 1 nur eine Programmiersprache dargestellt.

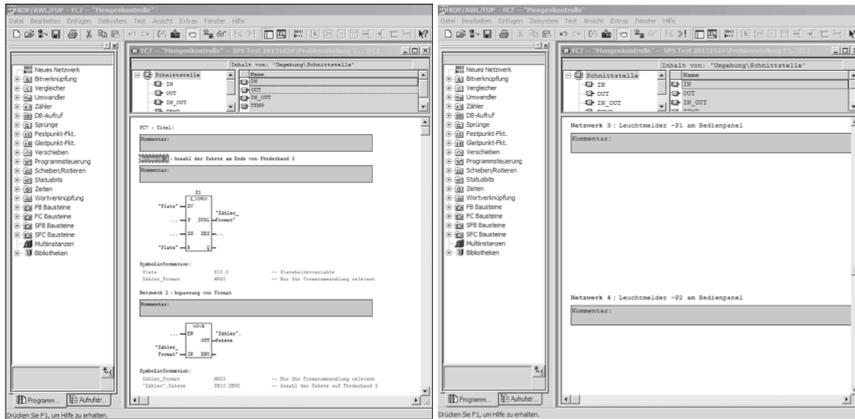


Abb. 1: Darstellung einer exemplarischen Problemstellung in der Entwicklungsumgebung (vgl. LINK 2015, in Vorbereitung)

In Abbildung 2 ist dieselbe Problemstellung im Papier-und-Bleistift-Test in bearbeiteter Form zu sehen. Bei dieser Problemstellung soll die Menge der aussortierten Pakete durch einen Reflexlichttaster „-B7“ erfasst und sobald eine bestimmte Anzahl an Paketen durchlief, der Leuchtmelder „-P1“ am Bedienpanel aktiviert werden. Dazu müssen die Auszubildenden u. a. einen Zählerbaustein (Z1) mit den entsprechenden Variablen versehen. Ferner muss ein Vergleich (CMP  $\geq$ ) realisiert werden, wobei auf das passende Datenformat zu achten ist (Integer).

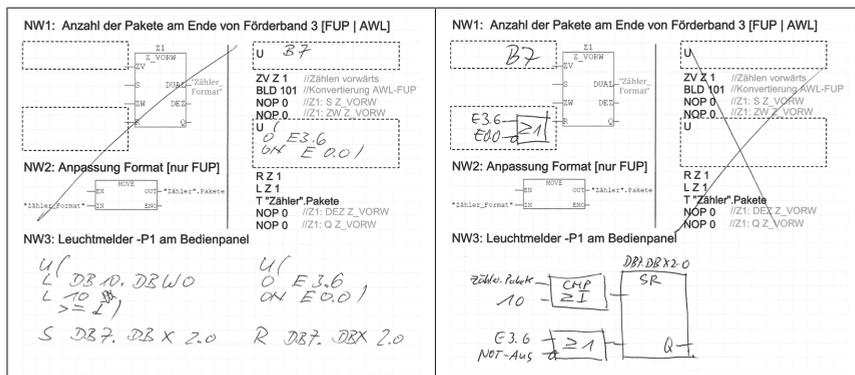


Abb. 2: Ausschnitt einer bearbeiteten Problemstellung von zwei Auszubildenden mit unterschiedlichen Programmiersprachen (vgl. LINK 2015, in Vorbereitung)

Auf der linken Seite wurde die Problemstellung von dem Auszubildenden mit der eher textuellen Programmiersprache AWL (Anweisungsliste) und rechts von einem anderen Auszubildenden mit FUP (Funktionsplan) bearbeitet. Diese beiden Arten zur Programmerstellung werden am häufigsten in der schulischen und betrieblichen Praxis eingesetzt, was ergaben unsere Analysen zu curricularen Schwerpunktsetzungen und Gespräche mit Lehrern und Ausbildern (vgl. LINK 2015, in Vorbereitung).

Die entwickelten acht Problemstellungen, mit mehreren dahinterliegenden Items, wurden in einem Cross-Over-Design in Cluster aufgesplittet und zwei Durchführungsvarianten zugewiesen (siehe Abbildung 3).

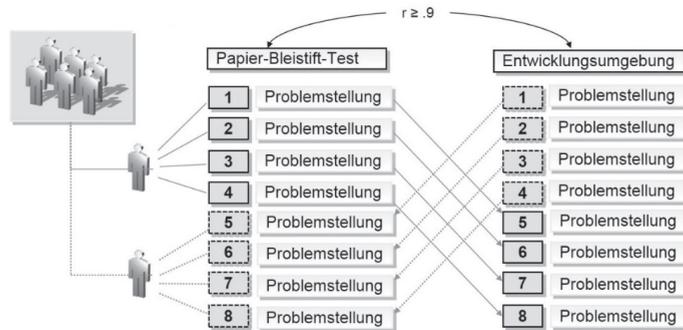


Abb. 3: Untersuchungsanlage der Studie

Gruppe 1 hatte zuerst vier Problemstellungen des PBT zu bearbeiten und anschließend vier Problemstellungen in der realen Entwicklungsumgebung. Bei Gruppe 2 wurde entsprechend umgekehrt gereiht verfahren, d. h. zuerst wurde die Entwicklungsumgebung präsentiert, dann der PBT. Im gewählten Cross-Over-Design muss zur Hypothesenprüfung ex ante ein Kriterium zur Entscheidung über die konvergente Validität definiert werden. Das Kriterium wird, wie bereits in Abbildung 3 angegeben, auf den Betrag von  $r = .9$  der Interkorrelation zwischen den Testleistungen im PBT und der realen Entwicklungsumgebung fixiert.

Es konnte eine Stichprobe von  $N = 145$  Auszubildenden, die sich am Ende ihrer Ausbildung befanden, realisiert werden. Die Auszubildenden waren überwiegend männlich (87,8%), im Mittel 18,9 Jahre alt ( $M = 18,9$ ;  $SD = 1,6$ ). Sie besitzen als höchsten Schulabschluss einen mittleren Bildungsabschluss (84,5%). Die Auswertung beschränkt sich momentan auf die einbezogenen Auszubildenden aus Baden-Württemberg und Bayern.

## 5. Ergebnisse

Die Darstellung der Ergebnisse erfolgt zur differenzierten Nachvollziehbarkeit zunächst getrennt für die beiden Gruppen. Anschließend werden die berechneten Modelle der Gruppe 1 und der Gruppe 2 aufeinander bezogen und eine Prüfung der Hypothese vorgenommen. Für beide Gruppen wurde zunächst ein Modell in Mplus (vgl. MUTHÉN/MUTHÉN 1998-2010) berechnet, indem zunächst die Korrelation zwischen den beiden Testteilen frei geschätzt (Modell A) wurde. Anschließend wurde ein Modell mit einem auf  $.9$  fixierten Betrag des Kriteriums gerechnet (Modell B).

Wie bereits berichtet, bestehen manche Problemstellungen aus mehreren Items. Bei der Modellierung in Mplus mussten von den 20 Items jedoch vier von der weiteren Auswertung ausgeschlossen werden, da diese zu schwer waren und daher zu

geringe Lösungsquoten aufwiesen. Folgende Abbildung 4 zeigt eine gleichmäßige Verteilung der Lösungsquoten der Partial-Credit-Items<sup>6</sup> bei beiden Gruppen.

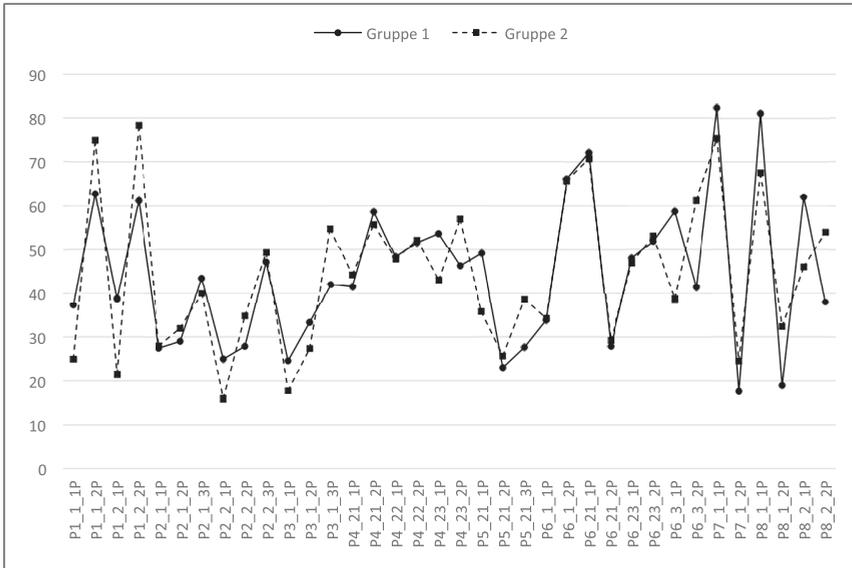


Abb. 4: Diagramm zur Lösungsquote je Partial-Credit-Stufe (in Prozent)

Um die kategorialen Variablen zu modellieren wurde ein WLSMV-Schätzer verwendet. Die SEM-Reliabilität<sup>7</sup> erreicht mit .83 (Gruppe 1) und .88 (Gruppe 2) einen guten Wert.

Für die Gruppe 1, also jene Gruppe der Auszubildenden, die zuerst den PBT und dann den Test in der Entwicklungsumgebung bearbeiteten, ergaben sich für Modell A Parameter gemäß Abbildung 5.

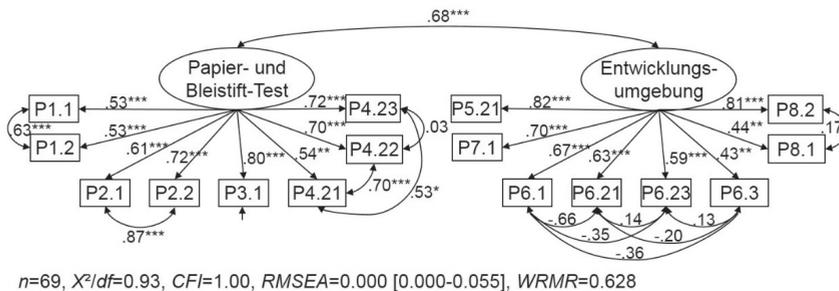


Abb. 5: Modell A mit frei geschätzter Korrelation (Gruppe 1)

- 6 Lesehilfe zu den Items: P1\_1\_1P ist die erste Stufe des Items P1\_1 und ergibt einen Punkt, P1\_1\_2P ist die vollständig richtige Lösung des Items, der Auszubildende erhält hier zwei Punkte.
- 7 Die SEM-Reliabilität wurde im Rückgriff auf GREEN/YANG (2009) berechnet.

Der Chi-Quadrat/df-Wert beträgt .93 und liegt damit zwischen 0 und 2, der Comparative Fit Index (CFI) sollte über .95 bzw. besser .97 für einen guten Modellfit betragen und hat hier den Wert 1.00 (vgl. SCHERMELLEH-ENGEL/MOOSBRUGGER/MÜLLER 2003). Der Root Mean Square Error of Approximation (RMSEA) fällt mit .00 kleiner aus als .06 (vgl. HU/BENTLER 1999) und der Weighted Root Mean Square Residual (WRMR) ist mit .628 kleiner als der empfohlene Cut-Off-Wert von .90 (vgl. YU/MUTHÉN 2001). Wir haben somit bei Modell A einen guten bis sehr guten Modellfit.

Die Bezeichnung der manifesten Variablen (z. B. P1.1) ist so aufgebaut, dass die erste Zahl die Nummer der Problemstellung angibt, die zweite Zahl die Nummer des Items. Bei Problemstellungen, die mehrere Items umfassen, werden noch die Residualkorrelationen zwischen den Items angegeben. Die standardisierten Faktorladungen können hier als Korrelationen zwischen den manifesten Variablen und dem Faktor, auf den sie laden, interpretiert werden. Die Faktorladungen sind zwischen .43 und .82 und mindestens signifikant, viele hochsignifikant.

Neben den Gütekriterien der Modellpassung ist für die Hypothesenprüfung bedeutsam, dass sich bei Modell A zwischen den beiden Testteilen eine hochsignifikante latente Korrelation von  $r = .68$  ergibt.

Bei Modell B ist die Korrelation zwischen den beiden latenten Variablen auf .90 fixiert und es wird geprüft, wie sich die Parameter des Modells verändern (vgl. Abbildung 6).

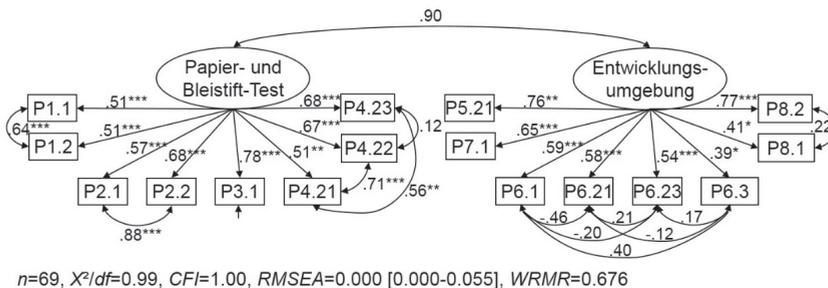


Abb. 6: Modell B mit fixierter Korrelation (Gruppe 1)

Die ausgegebenen Gütekriterien zeigen für den Chi-Quadrat/df-Wert .99, der CFI liegt wieder bei 1.00. Es ergibt sich also auch für das Modell B ein guter Modellfit. Die Faktorladungen liegen zwischen .39 und .78 und sind damit etwas niedriger als bei Modell A. Teile der Modellparameter sind nur noch knapp signifikant.

Der statistische Modellvergleich des Modells A und B für die Gruppe 1 ergibt  $\Delta\chi^2_{(1)} = 3.329$  ( $p = 0.068$ ). Der p-Wert fällt nicht signifikant aus,  $\Delta CFI$  beträgt .00 und ist damit kleiner als .01. Das heißt, das strengere Modell mit der fixierten Korrelation weicht nicht statistisch signifikant von den Daten ab.

Für Gruppe 2, also jene Auszubildenden, die zuerst den Test in der Entwicklungsumgebung und dann die PBT-Version bearbeitet haben, wurden auch wie bei Gruppe 1 zwei Modelle gerechnet, siehe Abbildung 7.

Die Gütekriterien geben einen noch akzeptablen bis guten Modellfit an: Der Chi-Quadrat/df-Wert beträgt 1.29, der Comparative Fit Index .96. Der RMSEA hat einen Wert von .062 und der WRMR ist mit 0.822 kleiner als .90. Die Faktorladungen liegen

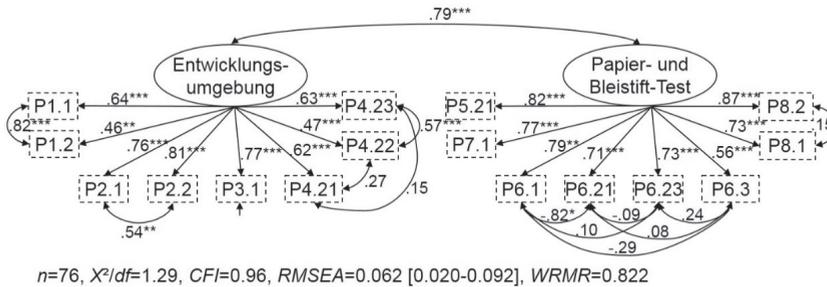


Abb. 7: Modell A mit frei geschätzter Korrelation (Gruppe 2)

zwischen .46 und .87 und sind fast alle hochsignifikant. Bei Modell A zeigt sich im Ergebnis beim 2-Faktormodell eine hochsignifikante latente Korrelation von  $r = .79$ .

Bei Modell B wurde die Korrelation zwischen den beiden latenten Variablen wieder auf  $r = .90$  fixiert (vgl. Abbildung 8).

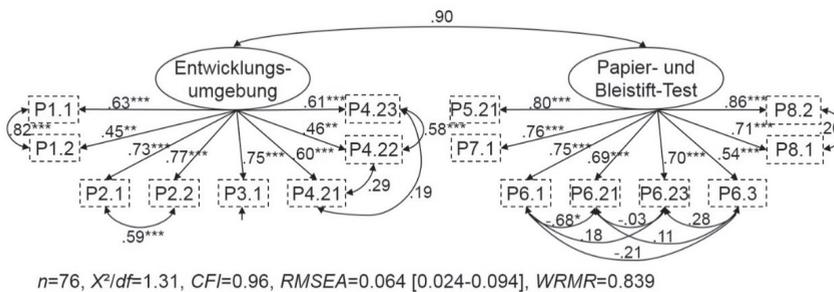


Abb. 8: Modell B mit fixierter Korrelation (Gruppe 2)

Der Chi-Quadrat/df-Wert beträgt 1.31, der CFI ist identisch zu Modell A. Der RMSEA und WRMR sind ähnlich im Vergleich zu Modell A. Folglich ergibt sich auch für Modell B ein akzeptabler bis guter Modellfit. Die Faktorladungen fallen identisch zu Modell A aus. Beim Vergleich der Modelle A und B bei Gruppe 2 ergibt sich beim Chi-Quadrat-Differenztest ein Wert von 2.180 bei einem Freiheitsgrad. Der p-Wert ist nicht signifikant. Delta-CFI ist auch hier kleiner als .01.

Damit weicht auch bei der zweiten Gruppe das strengere Modell mit der fixierten Korrelation nicht signifikant von den Daten ab.

Werden nun die Modelle von Gruppe 1 und Gruppe 2 zur Hypothesenprüfung im Vergleich betrachtet, kann für beide Varianten die PBT-Version mit der Entwicklungsumgebung als konvergent valide interpretiert werden. Die Testleistungen der Auszubildenden weichen unabhängig von der Präsentationsart nicht bedeutsam voneinander ab.

## 6. Diskussion und Ausblick

Durch die Validitätsprüfung konnte bezogen auf die Eingangs formulierte Forschungsfrage gezeigt werden, dass es möglich ist, ohne Verlust in der Diagnosegüte, die konstruktive Problemlösefähigkeit von Auszubildenden im Beruf Elektroniker für Automatisierungstechnik, mit Papier-und-Bleistift-Tests zu erfassen, statt der aufwändigen Testung in der teuren realen Entwicklungsumgebung<sup>8</sup>. Die Modellfit-Werte sind in einem guten bis sehr guten Bereich, einzig der niedrige p-Wert von .068 beim Modelltest der Gruppe 1 fällt weniger deutlich aus als erwartet und kommt nur knapp über die Signifikanzgrenze, sodass durchaus eine gewisse Nähe zur Zweidimensionalität des Modells besteht.

Ein zweidimensionales Modell beim Einsatz von einer authentischen Testumgebung und dem Einsatz von Papier-und-Bleistift-Tests berichtet beispielsweise ABELE (2014a) für die Diagnostik analytischer Problemlösefähigkeit bei Kfz-Mechatronikern. Dort ist es nicht gelungen mittels eines authentischen Papier-und-Bleistift-Tests dasselbe Konstrukt wie durch reale und simulierte Arbeitsproben bei Fehlerdiagnoseszenarien zu erfassen, wenngleich eine hohe Korrelation nachgewiesen werden kann. Ähnliche Befunde liegen von SENKBEIL/IHME (2014) bei der Erfassung von Computer- und Internetanwendungen (ICT Literacy) vor. So korrelieren zwar beide Testformen (Papier-und-Bleistift-Test und computerbasierter Test) miteinander, erfassen aber nicht dasselbe Konstrukt, sondern ICT Literacy auf jeweils eigene Weise.

Wie zu Beginn des Beitrags aufgezeigt hat man sich in den vorangegangenen Untersuchungen vornehmlich mit der analytischen Problemlösefähigkeit beschäftigt, die offensichtlich kaum mit Papier-und-Bleistift-Tests adäquat erfasst werden kann. Ob es im Bereich konstruktiver Problemlösefähigkeit auch in anderen Domänen gelingt das Konstrukt mit Papier-und-Bleistift-Tests zu erfassen, bleibt zu prüfen. Dass es sich bei der analytischen und konstruktiven Problemlösefähigkeit bei Elektronikern für Automatisierungstechnik zumindest um je eigene Konstrukte handelt wird von WALKER/LINK/NICKOLAUS (2015) bestätigt.

Verwertungsperspektiven der gewonnenen Erkenntnisse bieten sich in mehrerlei Hinsicht an:

a) Eher aus einer diagnostischen Perspektive betrachtet, könnten die hier entwickelten und validierten Instrumente mit den gewonnenen Erkenntnissen nach einem curricularen Abgleich (vgl. NICKOLAUS/GEISSEL 2009) in eine internationale Vergleichsstudie eines Large-Scale-Assessments (VET-LSA) einmünden. Konzeptionelle Vorarbeiten hierzu wurden in einer Machbarkeitsstudie bereitgestellt (vgl. BAETHGE/ARENDS 2009; NICKOLAUS/GEISSEL 2009). Darüber hinaus scheint der hier entwickelte PBT auch für die Verwendung in Prüfungskontexten (z. B. anknüpfend an die Befunde von ABELE 2014b) oder zur standardisierten Evaluation von spezifischen Fördermaßnahmen geeignet.

b) Eher aus einer didaktischen Perspektive betrachtet, eröffnen sich mit dem Papier-und-Bleistift-Test neue Möglichkeiten zur Förderung der konstruktiven Problemlösefähigkeit bei schwächeren Auszubildenden. So konnte im Baubereich bereits mehrfach eine geeignete Fördermaßnahme erfolgreich durchgeführt werden (vgl. NORWIG/PETSCH/NICKOLAUS 2010; PETSCH/NORWIG/NICKOLAUS 2014). Denkbar

8 Zur Absicherung der Ergebnisse werden gegenwärtig die Daten weiterer 150 Probanden aufbereitet und in die weiteren Auswertungen einbezogen.

wäre es, dass gerade die schwächeren Auszubildenden in der komplexen Entwicklungsumgebung anfangs überfordert sind und dass eine Programmierung in einer Papier-und-Bleistift-Umgebung zunächst erfolgsversprechender scheint, bevor im Anschluss am realen System programmiert wird. Ob sich dieser Sachverhalt auch empirisch darstellen lässt, müssen weitere Untersuchungen klären.

## Literatur

- ABELE, S. (2014a): Können Kfz-Fehlerdiagnosekompetenzen mit einem authentischen Papier-und-Bleistifttest valide erfasst werden? 6. ASCOT-Projekttreffen in Bamberg. 30. September 2014.
- ABELE, S. (2014b): Modellierung und Entwicklung berufsfachlicher Kompetenz in der gewerblich-technischen Ausbildung. Stuttgart: Steiner.
- ABELE, S./WALKER, F./NICKOLAUS, R. (2014): Zeitökonomische und reliable Diagnostik beruflicher Problemlösekompetenzen bei Auszubildenden zum Kfz-Mechatroniker. In: Zeitschrift für Pädagogische Psychologie 28, H. 4, S. 167–179.
- BADER, R./MÜLLER, M. (2002): Leitziel der Berufsbildung: Handlungskompetenz. In: Die Berufsbildende Schule 54, H. 6, S. 176–182.
- BAETHGE, M./ARENDTS, L. (Hrsg.) (2009): Feasibility Study VET-LSA. A comparative analysis of occupational profiles and VET programmes in 8 European countries – International report. Bonn, Berlin: bmbf.
- BUNDESAGENTUR FÜR ARBEIT (2014): Elektroniker für Automatisierungstechnik. <http://berufenet.arbeitsagentur.de/berufe/start?dest=profession&prof-id=15630>. 01.11.2014.
- GEISSEL, B. (2008): Ein Kompetenzmodell für die elektrotechnische Grundbildung: Kriteriumsorientierte Interpretation von Leistungsdaten. In: NICKOLAUS, R./SCHANZ, H. (Hrsg.): Didaktik der gewerblich-technischen Berufsbildung. Konzeptionelle Entwürfe und empirische Befunde. Baltmannsweiler: Schneider Verlag Hohengehren, S. 121–141.
- GREEN, S. B./YANG, Y. (2009): Reliability of Summed Item Scores Using Structural Equation Modeling: An Alternative to Coefficient Alpha. In: Psychometrika 74, H. 1, S. 155–167.
- GSCHWENDTNER, T. (2011): Die Ausbildung zum Kraftfahrzeugmechatroniker im Längsschnitt. Analyse zur Struktur von Fachkompetenz am Ende der Ausbildung und Erklärung von Fachkompetenzentwicklung über die Ausbildungszeit. In: Zeitschrift für Berufs- und Wirtschaftspädagogik – Beihefte (ZBW-B), H. 25, S. 55–76.
- GSCHWENDTNER, T./ABELE, S./NICKOLAUS, R. (2009): Computersimulierte Arbeitsproben: Eine Validierungsstudie am Beispiel der Fehlerdiagnoseleistung von Kfz-Mechatronikern. In: Zeitschrift für Berufs- und Wirtschaftspädagogik 105, H. 4, S. 556–578.
- HARTIG, J./KLIEME, E. (Hrsg.) (2007): Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzmodelle. Eine Expertise im Auftrag des Bundesministeriums für Bildung und Forschung. Bonn.
- HOCHHOLDINGER, S./SCHAPER, N. (2013): Gestaltung eines E-Learning-basierten Störungsdiagnosetrainings für die berufliche Ausbildung. In: ROWOLD, J./HOCHHOLDINGER, S./SCHAPER, N. (Hrsg.): Evaluation und Transfersicherung betrieblicher Trainings. Modelle, Methoden und Befunde. Göttingen [u. a.]: Hogrefe, S. 111–130.
- HU, L./BENTLER, P. M. (1999): Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. In: Structural Equation Modeling: A Multidisciplinary Journal 6, H. 1, S. 1–55.
- KMK (2007): Handreichung für die Erarbeitung von Rahmenlehrplänen der Kultusministerkonferenz (KMK) für den berufsbezogenen Unterricht in der Berufsschule und ihre Abstimmung mit Ausbildungsordnungen des Bundes für anerkannte Ausbildungsberufe. Bonn.
- KNÖLL, B. (2007): Differenzielle Effekte von methodischen Entscheidungen und Organisationsformen beruflicher Grundbildung auf die Kompetenz- und Motivationsentwicklung

- in der gewerblich-technischen Erstausbildung. Eine empirische Untersuchung in der Grundausbildung von Elektroinstallateuren. Aachen: Shaker.
- LANGMANN, R. (2004): Modelle. In: LANGMANN, R. (Hrsg.): Taschenbuch der Automatisierung. München [u. a.]: Fachbuchverl. Leipzig im Carl Hanser Verl., S. 226–278.
- LINK, N. (2015, in Vorbereitung): Problemlösen bei der Programmierung von speicherprogrammierbaren Steuerungen in komplexen automatisierten Systemen. Dissertation. Ludwigsburg.
- MUTHÉN, L. K./MUTHÉN, B. O. (1998–2010): Mplus User's Guide. Los Angeles.
- NICKOLAUS, R. (2011): Die Erfassung fachlicher Kompetenzen und ihrer Entwicklungen in der beruflichen Bildung – Forschungsstand und Perspektiven. In: ZLATKIN-TROITSCHANSKAIA, O. (Hrsg.): Stationen Empirischer Bildungsforschung. Traditionslinien und Perspektiven. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 331–351.
- NICKOLAUS, R./GEISSEL, B. (2009): Electricians. In: BAETHGE, M./ARENDTS, L. (Hrsg.): Feasibility Study VET-LSA. A comparative analysis of occupational profiles and VET programmes in 8 European countries – International report. Bonn, Berlin: bmbf, S. 48–70.
- NICKOLAUS, R./GEISSEL, B./ABELE, S./NITZSCHKE, A. (2011): Fachkompetenzmodellierung und Fachkompetenzentwicklung bei Elektronikern für Energie- und Gebäudetechnik im Verlauf der Ausbildung – Ausgewählte Ergebnisse einer Längsschnittstudie. In: NICKOLAUS, R./PÄTZOLD, G. (Hrsg.): Lehr-Lernforschung in der gewerblich-technischen Berufsbildung. Stuttgart: Steiner, S. 77–94.
- NITZSCHKE, A./GEISSEL, B./NICKOLAUS, R. (2011): Erfassung von Fachkompetenz bei Elektronikern und deren Prädiktoren über den Verlauf der Ausbildung. In: FISCHER, M./BECKER, M./SPÖTTL, G. (Hrsg.): Kompetenzdiagnostik in der beruflichen Bildung. Probleme und Perspektiven. Frankfurt am Main: Peter Lang, S. 111–128.
- NORWIG, K./PETSCH, C./NICKOLAUS, R. (2010): Förderung lernschwacher Auszubildender – Effekte des berufsbezogenen Strategietrainings (BEST) auf die Entwicklung der bautechnischen Fachkompetenz. In: Zeitschrift für Berufs- und Wirtschaftspädagogik (ZBW) 106(2), S. 220–239.
- PETSCH, C./NORWIG, K./NICKOLAUS, R. (2014): Kompetenzförderung leistungsschwächerer Jugendlicher in der beruflichen Bildung – Förderansätze und ihre Effekte. In: Zeitschrift für Erziehungswissenschaft 17, H. 1, S. 81–101.
- RAUSCH, A./SCHEJA, S./DREYER, K./WARWAS, J./EGLOFFSTEIN, M. (2010): Emotionale Befindlichkeit in Lehr-Lern-Prozessen – Konstruktverständnis und empirische Zugänge. In: SEIFRIED, J./WUTTKE, E./NICKOLAUS, R./SLOANE, P. (Hrsg.): Lehr-Lern-Forschung in der kaufmännischen Berufsbildung – Ergebnisse und Gestaltungsaufgaben. Stuttgart: Steiner, S. 193–215.
- ROSENDAHL, J./STRAKA, G. A. (2011): Analysen zur wirtschaftlichen Fachkompetenz angehender Bankkaufleute. In: Zeitschrift für Berufs- und Wirtschaftspädagogik 107, H. 2, S. 190–217.
- SCHERMELLEH-ENGEL, K./MOOSBRUGGER, H./MÜLLER, H. (2003): Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. In: Methods of Psychological Research Online 8, H. 2, S. 23–74.
- SEEBER, S. (2008): Ansätze zur Modellierung beruflicher Fachkompetenz in kaufmännischen Ausbildungsberufen. In: Zeitschrift für Berufs- und Wirtschaftspädagogik 104, H. 1, S. 74–97.
- SEEBER, S./LEHMANN, R. H. (2011): Determinanten der Fachkompetenz in ausgewählten gewerblich-technischen Berufen. In: NICKOLAUS, R./PÄTZOLD, G. (Hrsg.): Lehr-Lernforschung in der gewerblich-technischen Berufsbildung. Stuttgart: Steiner, S. 95–112.
- SEMBILL, D./RAUSCH, A./KÖGLER, K. (2013): Non-cognitive facets of competence – Theoretical foundations and implications for measurement. In: BECK, K./ZLATKIN-TROITSCHANSKAIA, O. (Hrsg.): From diagnostics to learning success. Proceedings in vocational education and training. Rotterdam, Boston: Sense, S. 199–212.
- SENKBEIL, M./IHME, J. M. (2014): Wie valide sind Papier-und-Bleistift-Tests zur Erfassung computerbezogener Kompetenzen? In: Diagnostica 60, H. 1, S. 22–34.

- SONNTAG, K./SCHAPER, N. (Hrsg.) (1997): Störungsmanagement und Diagnosekompetenz. Leistungskritisches Denken und Handeln in komplexen technischen Systemen. Zürich: Vdf, Hochschulverl. an der ETH Zürich.
- WALKER, F./LINK, N./NICKOLAUS, R. (2015): Berufsfachliche Kompetenzstrukturen bei Elektronikern für Automatisierungstechnik am Ende der Berufsausbildung. In: Zeitschrift für Berufs- und Wirtschaftspädagogik (im Druck).
- WINTHER, E./ACHTENHAGEN, F. (2009): Skalen und Stufen kaufmännischer Kompetenz. In: Zeitschrift für Berufs – und Wirtschaftspädagogik 105, H. 4, S. 521–556.
- WINTHER, E./ACHTENHAGEN, F. (2013): Kompetenzdiagnostik im kaufmännischen Bereich – Ein Umsetzungsbeispiel. In: NICKOLAUS, R./RETELSDORF, J./WINTHER, E./KÖLLER, O. (Hrsg.): Mathematisch-naturwissenschaftliche Kompetenzen in der beruflichen Erstausbildung. Stand der Forschung und Desiderata. Stuttgart: Steiner, S. 203–226.
- WITTMANN, E./WEYLAND, U./NAUERTH, A./DÖRING, O./RECHENBACH, S./SIMON, J./WOROFKA, I. (2014): Kompetenzerfassung in der Pflege älterer Menschen – theoretische und domänenspezifische Anforderungen der Aufgabenmodellierung. In: SEIFRIED, J./FASSHAUER, U./SEEGER, S. (Hrsg.): Jahrbuch der berufs- und wirtschaftspädagogischen Forschung 2014. Leverkusen: Budrich, Barbara, S. 53–66.
- YU, C.-Y./MUTHÉN, B. O. (2001): Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. Technical report in preparation. Los Angeles.
- ZINKE, G./SCHENK, H./KRÖLL, J. (2014): Ergebnisse einer Online Befragung zur Berufsfeldanalyse der industriellen Elektroberufe. Bonn.

Anschrift der Autoren: Dipl.-Gwl. Nico Link, link@ph-ludwigsburg.de, Prof. Dr. phil. Bernd Geißel, geissel@ph-ludwigsburg.de, Pädagogische Hochschule Ludwigsburg, Institut für Naturwissenschaften und Technik, Abteilung Technik, Reuteallee 46, D-71634 Ludwigsburg