

Kompetenzzuschreibungen von Ausbildern – Analysen zur Güte von Ausbilderurteilen

Forschungsstand

In betrieblichen Beurteilungsverfahren kommt den Einschätzungen der Kompetenzen der Auszubildenden durch die Ausbilder erhebliche Bedeutung zu und auch in gängigen Verfahren der Kompetenzdiagnostik (z. B. ERPENBECK/ROSENSTIEL 2003; ERPENBECK 2004) werden Fremdzuschreibungen neben Selbstzuschreibungen und testbasierten Verfahren häufig genutzt. Solche Fremdzuschreibungen beruhen auf mehr oder weniger systematischen Verhaltensbeobachtungen. Relativ aufwändig kommen solche Verhaltensbeobachtungen beispielsweise bei Assessment-Centern zur Anwendung (z. B. LISSMANN 2013), wobei die praktischen Umsetzungen erheblich variieren und häufig keine befriedigende psychometrische Güte erreicht wird. Mit erheblichen Güteproblemen von subjektiven Kompetenzeinschätzungen ist vor allem dann zu rechnen, wenn Ausprägungen relativ komplexer Konstrukte vorgenommen werden und den Urteilenden die (implizite) „Operationalisierung“ überlassen bleibt (NICKOLAUS/SEEGER 2013). Studien zur Güte der Leistungseinschätzungen von Lehrenden liegen vor allem für Lehrer und Lehrerinnen vor (im Überblick SCHRADER 2011), zur Güte von Ausbilderurteilen ist die Forschungslage deutlich eingeschränkter. Vereinzelt sind auch Studien auffindbar, in welchen Lehrer- und Ausbilderurteile vergleichend gegenüber gestellt wurden (AHEIMER 1998; FLEUCHAUS 2004). Dabei zeichnet sich tendenziell ab, dass Ausbilder eher positivere Kompetenzzuschreibungen vornehmen als Lehrer beruflicher Schulen (ebd.). Bei schulischen Lehrkräften ergeben sich Güteprobleme nicht zuletzt durch die üblicherweise genutzten klasseninternen Bezugssysteme (SCHRADER 2011). Bei Ausbildern dürfte das gleiche Problem auftreten, wobei die meist kleineren Gruppengrößen die Referenzproblematik vermutlich eher verschärfen. Schulische Lehrkräfte erreichen eine akzeptable Reliabilität und Validität am ehesten dann, wenn spezifische Leistungsaspekte beurteilt werden (SCHRADER 2011). Werden Testleistungen als Kriterium für die Beurteilung der Lehrerurteile herangezogen, sind deutlich divergierende Korrelationen feststellbar, deren mittlerer Wert bei $r = 0,69$ liegt (ebd.). Lehrerurteile zu Aufgabenschwierigkeiten, die in relativ enger Beziehung zu Kompetenzeinschätzungen stehen, sind gemessen an den tatsächlich erbrachten Lösungsquoten meist systematisch (positiv) verzerrt, d. h. es werden von Seiten der Lehrkräfte höhere Lösungsquoten erwartet, als tatsächlich erzielt werden (SCHRADER 2011; HAOLADER 2010). SCHRADER berichtet für Zusammenhänge zwischen den tatsächlich erreichten und den von Seiten der Lehrenden erwarteten Lösungen ein Korrelationsspektrum von $r = 0,33$ bis $0,56$, HAOLADER ermittelt in einer Untersuchung bei Elektronikern hingegen Korrelationen der Schwierigkeitseinschätzungen von Lehrkräften und den tatsächlich erzielten Lösungen von $r = 0,71$ bis $0,79$ (ebd. S. 134).

Vor allem im Kontext von Assessment Centern werden Anstrengungen unternommen, durch kriteriengeleitete Beobachterschulungen die Beurteilungsgüte zu erhöhen. Probleme bereitet aber selbst bei gut geschulten Personen die Abhängigkeit der Urteile von den Gruppenzusammensetzungen (EISENHARDT 2008) und damit erneut die Referenzproblematik bzw. die mangelnde Standardisierung. Fraglich ist zudem, ob die befriedigenden Beobachterübereinstimmungen, die im Laufe solcher Schulungen häufig erreicht werden, auch valide sind (BECK, 1987). Trotz dieser Probleme sind Ausbilderurteile in vielen Beurteilungssituationen kaum durch psychometrisch hochwertigere Verfahren substituierbar und angesichts der z. T. relativ guten Einschätzungen liegt der Gedanke nahe, der Frage nachzugehen, unter welchen Bedingungen die Urteilsgüte besonders günstig ausfällt und welche Vorkehrungen

geeignet sind, die Urteilsgüte weiter zu optimieren. In einer praktischen Perspektive wäre der Rückgriff auf Ausbilderurteile vor allem bei nur mit großem Aufwand zu generierenden Kompetenzabschätzungen hilfreich. In einer akzeptablen Güte ist das im Anschluss an die Befundlage am ehesten dann zu erwarten, wenn bei der Beobachtung hinreichend klar ist, welche konkreten Realitätsausschnitte und damit assoziierten Kompetenzen in den Blick genommen werden.

Fragestellung

Im Rahmen eines Projekts, in dem die Entwicklung von reliablen und validen Testinstrumenten zur Erfassung berufsfachlicher Kompetenzen von Kfz-Mechatronikern im Mittelpunkt stand, wurde auch der Frage nachgegangen, inwieweit relativ aufwändige, testbasierte Abschätzungen des Handlungswissens, im Bereich der Reparatur und Instandhaltung, mit weniger aufwändigen Ausbildereinschätzungen angenähert werden können. Im Folgenden stellen wir zunächst das videogestützte Testinstrument zur Erfassung des Handlungswissens und das aufwandsärmere Instrument zur Erfassung der einschlägigen Ausbilderurteile vor. Anschließend berichten wir über die Ergebnisse des vorgenommenen Vergleichs.

Instrumente

Das Instrument für die Erfassung des Handlungswissens (siehe auch SCHMIDT/NICKOLAUS/WEBER 2014) wurde in enger Kooperation mit Experten aus der Bildungspraxis (Prüfungsexperten, Lehrenden (Betrieb/Schule), Lehrbuchautoren) entwickelt und deckt häufig vorkommende Anforderungssituationen ab (Standardservice: Ölservice, Reifenservice, Motor-kühlsystem (zwei Varianten), Bremsenservice (zwei Varianten); Reparatur/Instandhaltung: Beleuchtungsanlage, Bremsanlage, Auspuffanlage (zwei Varianten), Fahrwerk: Spurstange, Instandhaltung: Zahnriemen). Zu diesen Tätigkeiten wurden Videovignetten generiert, die die gezielt in ihrer Güte manipulierten Handlungsvollzüge detailliert dokumentieren. An gezielt platzierten Unterbrechungen der Handlungsvollzüge wurden die Auszubildenden aufgefordert, den weiteren Handlungsstrang (Handlungspläne) zu explizieren (Beispielitem: Welche Arbeitsschritte sind noch notwendig, bevor die Motorabdeckung wieder montiert wird?), die dokumentierten Handlungsvollzüge zu beurteilen (Handlungsbeurteilungen zu Vollzug und Handlungsergebnis, Beispielitem: Wie beurteilen Sie die Vorgehensweise des Mechanikers bei der Reparatur der Leitung?) und Wissensfragen zu den Reparaturhandlungen zu beantworten (handlungsrelevantes Fachwissen, Beispielitem: Was könnte die Ursache für den Wackelkontakt sein? Nennen Sie bitte 3 mögliche Ursachen!). Die folgende Abbildung (Abb. 1) gibt beispielhaft ein Item zur Erfassung von Handlungsplänen aus der Videovignette Beleuchtungsanlage-Reparatur wieder, dem eine Videosequenz vorausging, in der verschiedene Reparaturschritte dargestellt wurden.

Die Reliabilität der beiden Instrumente zur Erfassung des Handlungswissens sind bei einer eindimensionalen Skalierung sowohl im Bereich *Standardservice* (*Ende 2. Ausbildungsjahr*: WLE = 0.84, EAP/PV = 0.85; *Ende 3. Ausbildungsjahr*: WLE = 0.74, EAP/PV = 0.71), als auch im Bereich *Reparatur/Instandhaltung* (*Ende 3. Ausbildungsjahr*: WLE = 0.70, EAP/PV = 0.71) gut bis befriedigend. Bei mehrdimensionalen Skalierungen, die gegen Ende der Ausbildung, d.h. am Ende des *3. Ausbildungsjahrs*, besser auf die Daten passen (SCHMIDT/NICKOLAUS/WEBER 2014), sind die Reliabilitätswerte zwar noch akzeptabel, jedoch optimierungsbedürftig: Bereich *Standardservice* (WLE = 0.55–0.61, EAP/PV = 0.67–0.68); Bereich *Reparatur/Instandhaltung* (WLE = 0.55–0.56, EAP/PV = 0.65–0.68), das gilt insbesondere für die Subdimension des direkt auf die Handlung bezogenen Sachwissens.

Das Instrument zur Erfassung der Ausbildereinschätzungen ist konsequent an den in den Videovignetten abgebildeten beruflichen Handlungen aus dem Bereich Reparatur/

Zahnriemen

Status 12 / 16 35:00

Welche der folgenden Handlungsschritte müssen aus fachmännischer Hinsicht an diesem Fahrzeug noch getätigt werden, bevor es an den Kunden übergeben werden kann?

Beachten Sie: 2 Antworten sind richtig.

- ☐ Abgasuntersuchung durchführen
- ☐ Hochdrehzahltest durchführen
- ☐ Ventilspiel prüfen
- ☐ Probefahrt machen
- ☐ Keilriemenspannung prüfen
- ☐ Zündzeitpunkt prüfen

Fahrzeugschein

Reparaturan-
leitung
Teil 01

Werkstattauftrag

Reparaturan-
leitung
Teil 02

Weiter

angemeldet als:

Abb. 1: Momentaufnahme Videovignette: Zahnriemen-Reparatur

Instandhaltung ausgerichtet. Mit der Wahl häufig vorkommender Handlungen sollte gewährleistet werden, dass die Auszubildenden solche Tätigkeiten schon häufiger ausgeübt haben und die Ausbilder, die mit der Betreuung der Auszubildenden betraut sind, eine fundierte Beurteilungsbasis besitzen. Empirische Befunde sprechen dafür, dass die einbezogenen Handlungen zum Zeitpunkt der Kompetenzmessung, zumindest zu einem gewissen Grad, routinisiert vollzogen werden (SCHMIDT/NICKOLAUS/WEBER 2014). Zur weiteren Absicherung dieser Annahme, wurden auch Ausbildereinschätzungen zur Häufigkeit der jeweiligen Tätigkeiten im Arbeitszuschnitt der Auszubildenden erhoben. Zu vollziehen waren die Beurteilungen bezogen auf die einzelnen Tätigkeiten, wie beispielsweise die Reparatur der Bremsanlage¹, nur dann, wenn diese in einem zur Erhebung nahe liegenden Zeitpunkt zumindest einmal ausgeführt wurde (Erhebung: Frühjahr/Sommer 2013, Beurteilungszeitraum für die Durchführung der Tätigkeit: seit Anfang 2013). Erfasst wurde des Weiteren, (1) der Grad der Selbständigkeit², in dem die Tätigkeit ausgeführt wurde, (2) eine Einschätzung, ob der Auszubildende zum gegenwärtigen Zeitpunkt in der Lage wäre die Tätigkeit vollkommen selbstständig auszuführen³, (3) die Einhaltung der Vorgabezeiten für die Durchführung der Reparatur⁴, (4) die Fähigkeit der Auszubildenden eine Beurteilung der Ausgangslage vorzunehmen⁵, (5) eine Einschätzung zum fachmännischen Vorgehen⁶, (6) eine Aussage zur Häufigkeit von Reklamationen

- 1 Bitte beurteilen Sie die Kompetenzen Ihres Auszubildenden hinsichtlich folgender Arbeiten an der Bremsanlage: Erneuern von Bremsscheiben und Bremsbelägen sowie Erneuern von Bremschläuchen
- 2 In welcher Weise führte der Auszubildende die genannten Arbeiten (oder Ausschnitte davon) seit Jahresbeginn durch? (vierstufige Skala: vollkommen selbstständig bis vollkommen angeleitet)
- 3 Wenn die genannten Arbeiten heute anfallen würden, kann der Auszubildende diese vollständig fachmännisch durchführen? (vierstufige Skala: trifft vollkommen zu bis trifft überhaupt nicht zu)
- 4 Erfüllt die vorgegebenen Richtzeiten vollständig bis erfüllt die Vorgabezeiten nicht
- 5 Am Beispiel der Bremsenreparatur: Wie schätzen Sie die Fähigkeit des Auszubildenden ein, Bremsenverschleiß so zu beurteilen, dass die richtigen Reparaturen folgen? (vierstufige Skala: „Dem Auszubildenden fiel die Beurteilung nie schwer“ bis „Dem Auszubildenden fiel die Beurteilung immer schwer“)
- 6 Wie schätzen Sie die Vorgehensweise des Auszubildenden bei der Durchführung der oben genannten Reparaturen seit Jahresbeginn ein? („Der Auszubildende führte die Reparaturen an der

im Anschluss an die vom Auszubildenden durchgeführte Tätigkeit und abschließend, eine Vergleichseinschätzung zu anderen Auszubildenden im dritten Ausbildungsjahr⁷.

Durchführung der Untersuchung

Die Datenerhebung erfolgte im Zeitraum März bis Juli 2013 bei N = 122 Ausbildern. Den Ausbildern lagen zu diesem Zeitpunkt keine Informationen zu den Testergebnissen der Videovignetten ihrer Auszubildenden vor. Zuordnungen der Ausbilderurteile zu den erhobenen Leistungsdaten der Auszubildenden konnten bezogen auf die einzelnen Reparaturaufträge in unterschiedlichem Umfang realisiert werden (Tabelle 1). Diese zuordenbaren Fälle sind im Weiteren auch Basis für die Auswertungen und Ergebnisdarstellungen. Darüber hinaus liegen zu den gleichen Probanden auch Leistungsdaten zum gesamten Fachwissen und zur fachspezifischen Problemlösefähigkeit vor, die die Möglichkeit eröffnen in einer vergleichenden Perspektive die spezifizierten Ausbilderurteile nicht nur zu den inhaltlich direkt assoziierten Leistungen im Handlungswissenstest sondern auch zu anderen Leistungsdaten zuzuordnen.

Tab. 1: Zuordenbare Fälle der Ausbilderbeschreibungen zu den Leistungsdaten der Auszubildenden

Reparaturauftrag	Spurstange	Bremse	Auspuff	Zahnriemen	Beleuchtung
Zuordenbare Ausbilderurteile (N) zu den erhobenen Leistungsdaten der Auszubildenden	111	119	60	21	21

Ergebnisse

Deskriptive Ergebnisse: Wie aus Tabelle 2 zu entnehmen ist, haben die Auszubildenden die einschlägigen Tätigkeiten in den letzten drei Monaten zwischen zwei und viermal ausgeführt, die Standardabweichungen liegen bei allen Reparaturfällen bei ca. 1. Besonders häufig werden Bremsreparaturen durchgeführt, gefolgt von Reparaturen an der Spurstange und dem Auspuff.

Tab. 2: Mittelwerte und Standardabweichungen der Ausbildereinschätzungen zu den einzelnen Reparaturaufträgen

Reparaturauftrag	Spurstange	Bremse	Auspuff	Zahnriemen	Beleuchtung
Häufigkeit der Ausführung	2.98 (1.18)	4.36 (0.92)	2.67 (1.08)	1.89 (0.93)	1.95 (0.97)
Selbständigkeitsgrad	3.23 (0.72)	3.30 (0.77)	3.32 (0.69)	2.39 (0.89)	2.73 (0.88)
Fachmännische Durchführung	3.35 (0.62)	3.56 (0.56)	3.43 (0.61)	2.74 (0.75)	3.27 (0.79)
Einhaltung der Richtzeiten	2.83 (0.69)	2.81 (0.68)	2.84 (0.60)	2.32 (0.83)	2.40 (0.91)

Bremsanlage immer fachmännisch aus“ bis „Der Auszubildende führte die Reparaturen an der Bremsanlage überhaupt nicht fachmännisch aus“).

- 7 Wie schätzen Sie den Auszubildenden bei den genannten Arbeiten im Vergleich zu einem durchschnittlich gut arbeitenden Auszubildenden im 3. Ausbildungsjahr ein? (fünfstufige Skala: überdurchschnittlich bis unterdurchschnittlich)

Reparaturauftrag	Spurstange	Bremse	Auspuff	Zahnriemen	Beleuchtung
Beurteilungsfähigkeit der Ausgangssituation	3.25 (0.69)	3.24 (0.55)	3.42 (0.52)	3.05 (0.72)	3.29 (0.82)
Fachmännische Vorgehensweise	3.39 (0.56)	3.53 (0.53)	3.35 (0.54)	3.32 (0.83)	3.20 (0.77)
Reklamationen (Häufigkeit)	1.07 (0.30)	1.19 (0.45)	1.10 (0.30)	1.05 (0.21)	1.29 (1.06)
Vergleich des Auszubildenden mit einem durchschnittlichen Auszubildenden	3.63 (0.84)	3.60 (0.88)	3.57 (0.80)	3.43 (0.74)	3.53 (0.99)

Legende zu Tabelle 2: Verwendete Skalen: Häufigkeit der Ausführung: 1 = keinmal, 2 = 1–2 mal, 3 = 3–5 mal, 4 = 6–8 mal, 5 = mehr als 8 mal; Selbstständigkeitsgrad: 1 vollkommen angeleitet, 4 vollkommen selbständig; Fachmännische Durchführung: 1 trifft überhaupt nicht zu, 4 trifft vollkommen zu; Einhaltung der Richtzeiten: 1 erfüllt Richtzeiten nicht, 4 erfüllt Richtzeiten vollständig; Beurteilungsfähigkeit der Ausgangssituation: 1 Beurteilung fällt immer schwer, 4 Beurteilung fällt nie schwer; fachmännische Vorgehensweise: 1 überhaupt keine fachmännische Ausführung, 4 immer fachmännische Ausführung; Reklamationen 1 nie, 5 mehr als 8 mal; Vergleich der Auszubildenden: 1 unterdurchschnittlich, 5 überdurchschnittlich.

Der Selbstständigkeitsgrad liegt durchgängig über der Mitte der vierstufigen Skala, bei den häufiger ausgeführten Reparaturen werden Werte von über drei auf der vierstufigen Skala erreicht (vgl. Abb. 2). Bezogen auf die Reklamationshäufigkeiten bei den Tätigkeiten wurden auf der dreistufigen Skala durchweg Werte von unter 2 erreicht (vgl. Abb. 3), was als gegenläufiges Extrem zu der Beurteilung des Selbstständigkeitsgrades anzusehen ist. Die relativ aufwändige und anspruchsvolle Reparatur am Zahnriemen sowie die Reparatur an der Beleuchtung wird ebenfalls noch als eher selbständig ausgeführt eingeschätzt, wenngleich deutliche Unterschiede zu den Brems-, Auspuff- und Spurstangenreparaturen festzustellen sind. Auch bei dieser Skala sind die Standardabweichungen über alle Reparaturfälle hinweg relativ niedrig. Die Einschätzung, ob der zu beurteilende Auszubildende die jeweilige Reparatur zum Befragungszeitpunkt fachmännisch ausführen kann, wird auch durchgängig positiv eingeschätzt, lediglich bei der Zahnriemenreparatur sind die Einschätzungen etwas niedriger, aber immer noch deutlich im positiven Bereich. Etwas kritischer wird die Einhaltung der durch die Hersteller vorgegebenen Richtzeiten für die Reparaturdurchführung beurteilt, wobei sich auch hier das bei den anderen Beurteilungsmerkmalen aufzufindende Befundmuster bestätigt, wonach die Reparaturen am Zahnriemen und der Beleuchtungsanlage etwas skeptischer als die anderen Reparaturfälle eingeschätzt werden. Insgesamt bleibt allerdings zu konstatieren, dass die von den Auszubildenden im dritten Ausbildungsjahr durchgeführten Reparaturen bereits weitgehend im vorgesehenen Zeitrahmen und auch weitgehend ohne Reklamationen ausgeführt werden können, worin auch eine produktive Einsatzfähigkeit der Auszubildenden (im Urteil der Ausbilder) Ausdruck findet. Positiv fallen auch die Zuschreibungen zur eigenständigen Beurteilungsfähigkeit der Ausgangssituation und der fachmännischen Vorgehensweise aus. Insgesamt erweisen sich diese Ausbilderurteile als homogen positiv, wobei die etwas skeptischeren Einschätzungen bei den Reparaturfällen „Zahnriemen“ und „Beleuchtungsanlage“ darauf hindeuten, dass zumindest bei diesen eher anspruchsvollen Reparaturfällen eine gewisse Differenzierung der Urteile auftritt. Insgesamt fallen die Urteile zu den häufiger vorkommenden Tätigkeiten jedoch wenig differenziert aus. Zur Veranschaulichung der Verteilungen geben wir hier beispielhaft jene zum Selbstständigkeitsgrad bei der Bremsenreparatur, die relativ häufig ausgeführt wird und zur Reklamationshäufigkeit bezogen auf den gleichen Anforderungsbereich wieder.

In der abschließenden Beurteilung des jeweiligen Auszubildenden, *im Vergleich zum Durchschnitt der Auszubildenden*, die deutlich über dem Skalenmittel liegt, deutet sich ein

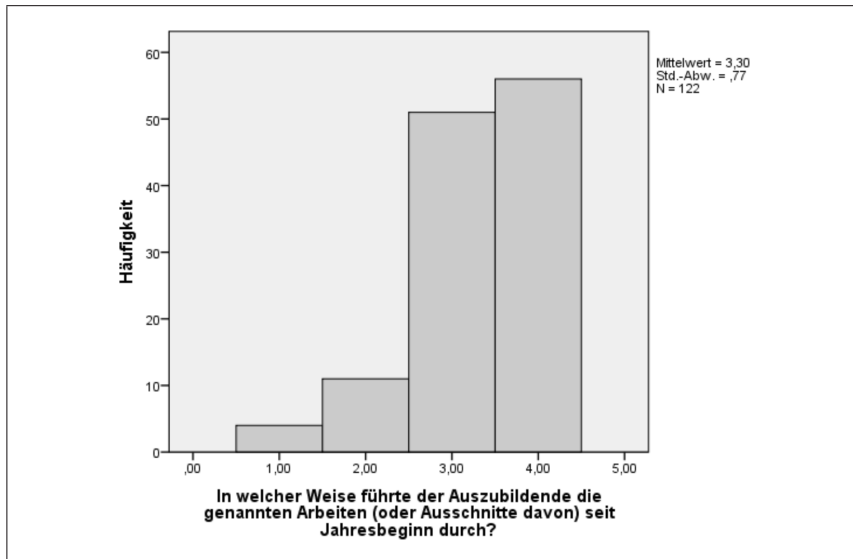


Abb. 2: Selbstständigkeit der Ausführung – am Beispiel: Bremsenreparatur (1 vollkommen angeleitet, 4 vollkommen selbständig)

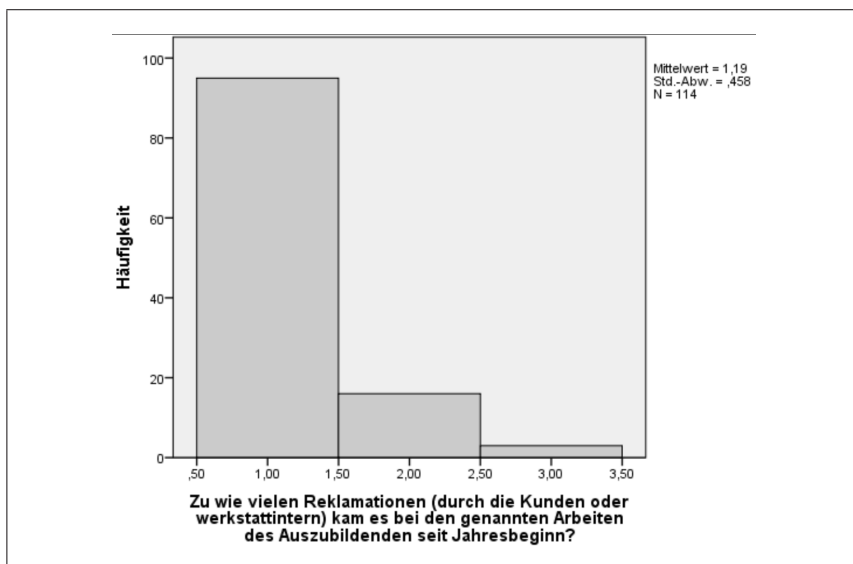


Abb. 3: Reklamationshäufigkeit – am Beispiel: Bremsenreparatur (1: kein einziges Mal, 2: 1–2 Mal, 3: 3–5Mal)

deutlicher Mildeeffekt an⁸ (PRAETORIUS 2014), der auch in den Studien von AHEIMER (1998) und FLEUCHAUS (2004) beobachtet werden konnte. Die Testergebnisse zum Handlungswissen fallen gegenüber den Ausbilderurteilen etwas differenzierter aus (vgl. Tabelle 3). Die Verteilungen liegen zwar partiell immer noch im positiven Bereich, aber weit weniger als im Falle der Ausbildereinschätzungen. Am höchsten sind die Testleistungen bei der Zahnriemenreparatur, am niedrigsten bei der Reparatur der Spurstange und der Beleuchtungsanlage. Die Streuung der Testleistungen ist bei der Beleuchtungsreparatur am größten, gefolgt von der Zahnriemenreparatur. Das sind zugleich jene Reparaturfälle, deren Ausführungshäufigkeit am niedrigsten liegt.

Tab. 3: Fallzahlen und durchschnittliche Testergebnisse zum Handlungswissen-Reparatur

Reparaturauftrag	Spur- stange	Bremse	Auspuff	Zahn- riemen	Beleuch- tung
Fallzahl (N)	251	236	52	62	58
Maximal erreichbare Punktzahl	24	30	30	32	30
Durchschnittliche Punktzahl (Testergebnis)	11,66	18,76	16,87	20,74	16,22
Standardabweichung	3.75	3.59	4.05	4.19	4.86

Im Gegensatz zu den Ausbilderurteilen ergeben sich bei den Testdaten (in der Regel) Normalverteilungen, beispielhaft wiedergegeben sind hier die Verteilungen für die Bremsenreparatur (vgl. Abb.4), die häufig ausgeführt wird und bei der die Testdaten im Mittel über dem Skalenmittelwert liegen.

Ein direkter Vergleich der Daten aus den Ausbildereinschätzungen und den zugehörigen Ausbildereinschätzungen ist insoweit erschwert, als unterschiedliche Skalen zugrunde liegen. Nutzt man als Vergleichsperspektive die Verteilung der Daten bezogen auf die jeweilige Skalenmitte, so wird deutlich, dass die Testergebnisse zum Handlungswissen nicht so günstig ausfallen, wie es aufgrund der sehr positiv ausfallenden Ausbildereinschätzungen zu erwarten wäre. Überraschend ist, dass die Testleistung bei der Zahnriemenreparatur am höchsten ausfällt, da die Ausbilderzuschreibungen in diesem Bereich am schlechtesten ausfielen. Generell deuten die Ergebnisse beider Erhebungsvarianten darauf hin, dass sowohl das eng bezogen auf relativ gut vertraute Tätigkeiten erhobene Handlungswissen, als auch die zugehörigen praktischen Leistungszuschreibungen durch die Ausbilder positiv ausfallen⁹. Dies steht in deutlichem Kontrast zu den vorliegenden Untersuchungsergebnissen im Bereich des Fachwissens (LEHMANN/SEEGER 2007; GEISSEL 2008; GSCHWENDTNER 2008; NICKOLAUS u. a. 2012) und der fachspezifischen Problemlöseleistungen in technischen Domänen, für die durchgängig erhebliche Leistungsdefizite dokumentiert worden sind (z. B. NICKOLAUS u. a. 2012). Mit anderen Worten, die Auszubildenden erreichen, zumindest in jenen Tätigkeitssegmenten, die im Arbeitsalltag dominieren und ein gewisses Routinepotenzial besitzen, ein hinreichendes Leistungsniveau. Gestützt wird diese Befundlage auch durch die Ergebnisse von SCHMIDT u. a. (2014) zu den erbrachten Leistungen in den verschiedenen Subdimensionen des Fachwissens, die (erwartungskonform) deutlich bessere Leistungen

- 8 Es ist hierbei auch auszuschließen, dass es sich bei der Stichprobe um eine Positivauslese handelt. Der T-Test (Levene-Test) unabhängiger Stichproben, auf Mittelwertunterschiede zwischen den Probandenleistungen ohne vorhandene Ausbilderurteile und mit vorhandenen Ausbilderurteilen, zeigt bezogen auf das Fachwissen (ohne Ausbilderurteile: N = 173, MW: 81,81; mit Ausbilderurteilen: N = 97, MW: 81,90) und die Problemlösefähigkeit (ohne Ausbilderurteile: N = 134, MW: 11,28; mit Ausbilderurteilen: N = 88, MW: 10,65) Varianzgleichheit und keine signifikanten Gruppenunterschiede auf (Fachwissen: $p = 0,97$; Problemlösefähigkeit: $p = 0,40$).
- 9 Auch hier nutzen wir als Bezugspunkt die Verteilung der erzielten Leistungen zur Skalenmitte

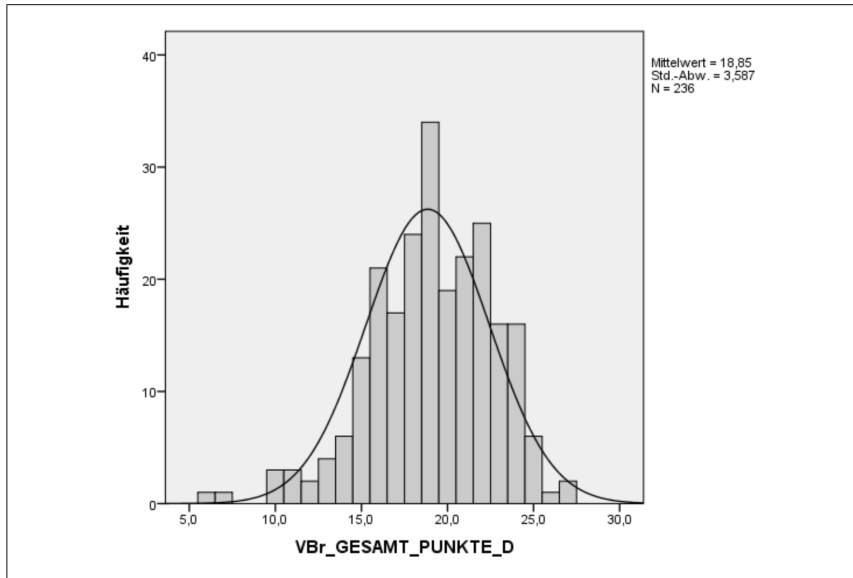


Abb. 4: Häufigkeitsverteilung der Testwerte – am Beispiel: Bremsenreparatur

in jenen Anforderungsbereichen dokumentieren, die auch im betrieblichen Arbeitsalltag der Auszubildenden stark repräsentiert sind¹⁰.

Zusammenhänge zwischen den Testdaten zum Handlungswissen und den Leistungszuschreibungen der Ausbilder

Im Anschluss an die oben skizzierte Befundlage wäre zu erwarten, dass sich zwischen den Leistungszuschreibungen der Ausbilder (Fremdzuschreibungen) und den Testdaten zum Handlungswissen mittlere Korrelationen beobachten lassen. Mit den auf konkrete und klar abgrenzbare Tätigkeiten bezogenen Fremdeinschätzungen ist auch eine Voraussetzung gegeben, die eher relativ hohe Zusammenhänge erwarten lässt. Die relativ geringen Varianzen in den Leistungszuschreibungen der Ausbilder legen allerdings den Gedanken nahe, dass eher schwache Zusammenhänge gefunden werden (ABELE, 2014, S. 129). Eine etwas höhere Fallzahl, in der wir die Ausbilderurteile und die Testdaten relationieren können, liegt lediglich für die Bremsenreparatur vor ($N = 123$). In diesem Fall ergeben sich signifikante, jedoch schwache Korrelationen der Testdaten mit den Ausbilderereinschätzungen bezogen auf das Beurteilungskriterium „Selbständigkeit der Durchführung“ (vgl. Tabelle 4). Die Zusammenhänge zwischen den Ausbilderzuschreibungen zu den anderen Beurteilungskriterien (Einhaltung der Vorgabezeiten, Anzahl der Reklamationen etc.) und den Testdaten zum Handlungswissen fallen geringer und in der Regel nicht signifikant aus. Die höchste Korrelation¹¹ mit den Testdaten zum Handlungswissen erreicht das vergleichende Ausbilderurteil zu einem imaginären Durchschnittsauszubildenden ($\rho = 0.37$). Dieser Sachverhalt könnte

¹⁰ Das gilt insbesondere für das Fachwissen im Servicebereich.

¹¹ Gerechnet wurden hier Rangkorrelationen, da für die Skalen zu den Ausbilderzuschreibungen keine Metrik unterstellt werden kann.

so gedeutet werden, dass die Ausbilderurteile trotz der vorgegebenen Kriterien eher globale Einschätzungen darstellen. Beurteilt wird von Ausbilderseite nicht auf der Basis einer kriteriengeleiteten Beobachtung am konkreten Ausführungsfall, sondern (retrospektiv) auf Basis von unsystematisch gesammelten Erfahrungen. Dies dürfte auch der „Normalfall“ bei Einschätzungen des Leistungsvermögens von Auszubildenden sein.

Neben den erzielten Testleistungen im Bereich des auf die Bremsenreparatur bezogenen Handlungswissens sind in Tabelle 3 auch die Zusammenhänge der Ausbilderurteile mit den Testleistungen im Fachwissenstest und den Fehlerdiagnoseleistungen wiedergegeben, womit die Möglichkeit eröffnet werden soll, die Leistungszuschreibungen der Ausbilder in diesem konkreten Reparaturfall auch mit anderen Leistungsdaten als dem einschlägigen Handlungswissen in Beziehung zu setzen und damit eine Referenz für die Korrelationen zwischen den Ausbilderzuschreibungen und den Testdaten zum Handlungswissen zu erhalten.

Tab. 4: Zusammenhänge zwischen Testdaten der Auszubildenden und den Leistungszuschreibungen der Ausbilder

Testleistungen/Ausbildereinschätzungen zur Bremsenreparatur	Handlungswissen (Bremsen)	Fachwissen	Fehleranalyseleistung
Selbständigkeit der Durchführung (p-Wert)	0.28 (0.04)	0.37 (0.00)	0.27 (0.00)
Anzahl der Reklamationen (p-Wert)	-0.12 (0.22)	-0.19 (0.05)	-0.16 (0.07)
Vergleich des Auszubildenden mit einem durchschnittlichen Auszubildenden (p-Wert)	0.37 (0.00)	0.16 (0.10)	0.23 (0.02)
Fachmännische Durchführung (p-Wert)	0.12 (0.20)	0.23 (0.01)	0.16 (0.10)
Beurteilungsfähigk. der Ausgangssituation (p-Wert)	0.12 (0.22)	0.16 (0.09)	0.01 (0.85)
Fachmännische Vorgehensweise (p-Wert)	0.18 (0.06)	0.14 (0.14)	0.15 (0.14)
Einhaltung von Richtzeiten (p-Wert)	0.13 (0.19)	0.12 (0.20)	0.03 (0.70)
Häufigkeit der Ausführung (p-Wert)	0.11 (0.25)	0.15 (0.12)	0.11 (0.28)

Legende zu Tabelle 4: Wiedergegeben sind jeweils der Rangkorrelationskoeffizient und die Signifikanz (zweiseitig). Der Vergleich zu einem durchschnittlichen Auszubildenden war explizit bezogen auf die Bremsenreparatur eingefordert.

Die in gleicher Höhe (Handlungswissen/Fehleranalyseleistung) bzw. im Falle des Fachwissens sogar höher ausfallende Korrelation der Testdaten zum Ausbilderurteil zur Selbständigkeit der Durchführung macht deutlich, dass die Selbstständigkeitszuschreibung zur Durchführung der Bremsenreparatur vermutlich eher eine Globaleinschätzung zum Ausdruck bringt. In diesem Sinne könnte man auch den relativ hohen Zusammenhang zwischen der Vergleichseinschätzung und den Testdaten zum Handlungswissen deuten. Dass die Korrelationen der Testdaten zur Anzahl der Reklamationen durchgängig gering und nicht signifikant ausfallen, wird durch die fehlende Varianz der Reklamationsanzahl begünstigt. Reklamationen kommen bei allen Fehlerfällen selten vor, was vermutlich auch darauf zurückzuführen ist, dass Vorkehrungen getroffen werden, die die Auftrittswahrscheinlichkeit von Reklamationen reduzieren. Dafür spricht auch, dass selbst bei jenen Reparaturfällen (Zahnriemen, Beleuchtungsanlage), die relativ selten durchgeführt werden und für die von Ausbilderseite der geringste Selbstständigkeitsgrad unterstellt wird, nahezu keine Reklamationen registriert werden.

Fazit

Insgesamt legt die Befundlage u.E. folgendes Fazit nahe: Gegen Ende der Ausbildung erreichen die Auszubildenden in der Regel sowohl bei häufiger auftretenden (z. B. Bremsen- und Auspuffreparatur) als auch bei seltener durch die Auszubildenden durchgeführten Reparaturen (Zahnriemen; Beleuchtungsanlage) ein Niveau, das aus Ausbilersicht hinreichend ist, um eine fachmännische und weitgehend selbständige Erledigung der Reparaturtätigkeiten zu sichern.

Die Varianz der Ausbilderurteile ist relativ gering. Besonders gering fällt sie bei der Anzahl der Reklamationen aus und das auch bei jenen Reparaturfällen (Zahnriemen, Beleuchtungsanlage), die relativ selten ausgeführt werden und für die die Ausbilder eine unterdurchschnittliche Selbständigkeit in der Ausführung unterstellen. Denkbar wäre, dass dies auch durch Vorkahrungen begünstigt wird, die darauf zielen, Reklamationen niedrig bzw. die Qualität des Leistungserstellungsprozesses auch bei Ausführung durch die Auszubildenden hoch zu halten. Das positive Leistungsbild bestätigt sich auch in den Testdaten zum reparaturrelevanten Handlungswissen, das im Gegensatz zu Leistungsdaten zum Fachwissen und der Problemlöseleistungen positiv ausfällt, sofern man die Verteilungscharakteristik in Bezug zum Mittelwert der jeweiligen Leistungsskalen als Beurteilungsmaßstab zugrunde legt. Die Varianz des Handlungswissens ist größer als jene der Ausbilderurteile, besonders gering ist sie dabei im Falle der Bremsenreparatur, welche besonders häufig ausgeführt wird und bei der eine deutlich rechtsschiefe Verteilung der Testergebnisse beobachtet werden kann. Die Ausbilderurteile spiegeln die immer noch bestehenden Leistungsunterschiede im Handlungswissen der Auszubildenden nicht wieder. Inwieweit diese Leistungsunterschiede im praktischen Handeln relevant werden, lässt sich allein über die Ausbilderurteile nicht hinreichend abschätzen. Denkbar wäre einerseits, dass die Auszubildenden in routinetragenden Reparaturbereichen, zu welchen wir auch Bremsenreparaturen rechnen, bereits in größerem Umfang implizites Wissen nutzen und die Testleistungen zum Handlungswissen¹² nur einen Teil des zum Einsatz kommenden Wissens spiegeln. Dafür sprechen partiell die Befunde von SCHMIDT u. a. 2014, die bei Kfz-Mechatronikern bereits im Ausbildungsverlauf für Tätigkeiten im Service- und Reparaturbereich ansatzweise Routinebildungsprozesse dokumentieren, jedoch zugleich auch noch am Ende der Ausbildung hohe Assoziationsmaße zwischen dem (reparaturrelevanten) Fachwissen und den Leistungsmaßen der Handlungspläne und Handlungsbeurteilungen dokumentieren. Denkbar ist andererseits aber auch, dass in der Reparaturausführung noch Leistungsunterschiede auftreten, die sich in den so operationalisierten Ausbilderurteilen nicht bzw. nur eingeschränkt niederschlagen.

Die vergleichenden Urteile der Ausbilder zu den Reparaturleistungen des jeweiligen Auszubildenden mit den (fiktiven) Leistungen eines Durchschnittsazubildenden liegen im Mittel deutlich über der Skalenmitte und dokumentieren damit einen deutlichen Mildeeffekt. Dabei ist zu berücksichtigen, dass die Urteile von Ausbildern eingebracht werden, die die Auszubildenden in kleinbetrieblichen Ausbildungskontexten und damit im Leistungserstellungsprozess betreuen. Gleichwohl scheint die Wertschätzung der Arbeit der Auszubildenden hinreichend, um die oben dokumentierte, positive Leistungszuschreibung vorzunehmen. Gegen die Annahme eines Mildeeffekts wäre theoretisch zwar einwendbar, dass es sich bei der Stichprobe um eine Positivselektion handelt, die dadurch zustande kam, dass der Zugang zu den zuordenbaren Ausbilderdaten nur bei Zustimmung der Auszubildenden möglich war. Dagegen spricht allerdings, dass sich diese Gruppe gemessen an den Leistungsdaten zum Fachwissen und zur Problemlösefähigkeit von der Gesamtgruppe nicht positiv unterscheidet.

12 Im Handlungswissen sind sowohl die Fähigkeiten Handlungspläne zu entwerfen, Handlungsvollzüge zu beurteilen als auch das dafür notwendige Sachwissen zusammengefasst. Die von Schmidt u. a. (2014) vorgelegten Ergebnisse deuten darauf hin, dass bei den Beurteilungen und bei den Handlungsplänen bereits gegen Ende der Ausbildung partiell auch implizites Wissen relevant wird.

Vor diesem Hintergrund kann die überdurchschnittliche Einschätzung des je eigenen Auszubildenden u. E. nur im Sinne des Mildeeffekts gedeutet werden. Die vergleichsweise hoch ausfallende Korrelation der vergleichenden Einschätzung der Auszubildenden durch die Ausbilder mit dem testbasiert erhobenen Handlungswissen ($r = 0.37$) deutet zwar an, dass die Ausbilder relational zu den anderen Auszubildenden noch am ehesten differenzierte urteilen können, aber als befriedigen ist dieser Differenzierungsgrad angesichts der absolut doch geringen Korrelation kaum bewertbar.

Zur Validierung der Testdaten des Handlungswissens, im Hinblick auf die tatsächlich erbrachten Leistungen, erweisen sich die Ausbilderurteile trotz einer an Kriterien ausgerichteten und auf klar umrissene Aufgaben bezogenen Beurteilung als weitgehend ungeeignet. Ursächlich könnte dafür sein, dass die Urteile auf Leistungsbereiche bezogen sind, die schon relativ gut beherrscht werden und damit die Trennschärfe der Urteile eingeschränkt ist. Zu prüfen wäre, ob das auch für Urteile gilt, die sich auf eine frühere Phase des Fertigkeitserwerbs bezieht. Vermutlich wären verlässlichere Urteile zu gewinnen, wenn die Ausbilder die Arbeitsprozesse kriterienbezogen direkt beobachten würden, anstatt retrospektiv zu urteilen. Eine systematische kriterienbezogene Beurteilung dürfte allerdings auch in betrieblichen Beurteilungsprozessen eher die Ausnahme sein. Unterstellt man, dass es sich bei solchen häufiger vorkommenden Reparaturarbeiten bereits um Routinetätigkeiten handelt, wäre als Validierungskriterium des Handlungswissens vermutlich eine präzise Zeiterfassungen in Verbindung mit der Ausführungsgüte angezeigt. Die Einhaltung der Richtzeiten scheint in Bezug auf die Ausführungszeit zumindest bei den häufiger ausgeführten Reparaturen bereits ein zu grobes Kriterium.

Bezogen auf den Eingangs skizzierten Forschungsstand deuten wir die Ergebnisse so, dass Ausbilderurteile zu den erbrachten Reparaturleistungen (am Ende der Ausbildung) auch bei Vorgabe von Beurteilungskriterien wenig differenziert sind, obgleich die testbasierten Daten Anlass geben eine deutlich größere Varianz zu unterstellen. Bemerkenswert sind die starken Mildeeffekte, die die Güte der kriterienbezogenen Urteile substantiell reduzieren. Als verlässliche Kompetenzmaße kommen Ausbilderurteile, die auf diese Weise gewonnen werden vor diesem Hintergrund zumindest im analysierten Anwendungsbereich nicht in Frage.

Literatur

- ABELE, S. (2014): Modellierung und Entwicklung der berufsfachlichen Kompetenz in der gewerblich-technischen Ausbildung. Steiner: Stuttgart.
- AHEIMER, R. (1998): Adaption und Test eines Instrumentariums zur Erfassung von ausgewählten Schlüsselqualifikationen in der dualen Ausbildung. Univ. Stuttgart, Diplomarbeit.
- BECK, K. (1987). Die empirischen Grundlagen der Unterrichtsforschung. Eine kritische Analyse der deskriptiven Leistungsfähigkeit von Beobachtungsmethoden. Hogrefe: Göttingen.
- EISENHARDT, N. (2008): Test eines Instruments zur Messung von Sozialkompetenz – eine Empirische Analyse auf Gruppenebene. Diplomarbeit, Universität Stuttgart
- ERPENBECK, J. (2004): KODE® im Tableau quantitativer, qualitativer und komparativer Kompetenzmessverfahren in Deutschland. In: HEYSE, V./ERPENBECK, J./MAX, H. (Hrsg.): Kompetenzen erkennen, bilanzieren und entwickeln. Münster: Waxmann, S. 118–131.
- ERPENBECK, J./ROSENSTIEL, L. VON (HRSG.) (2003): Handbuch Kompetenzmessung. Stuttgart: Schäffer-Poeschel.
- FLEUCHAUS (2004): Kommunikative Kompetenzen von Auszubildenden in der beruflichen Ausbildung: Ausprägungen, Förderung und Relevanz im Urteil von Ausbildern, Lehrern und Auszubildenden. Hamburg: Verlag Dr. Kovac (Studien zur Berufsbildung, Bd.10)
- GEISSEL, B. (2008): Ein Kompetenzmodell für die elektrotechnische Grundbildung: Kriteriumsorientierte Interpretation von Leistungsdaten. In: Nickolaus, R./Schanz, H. (Hrsg.): Didaktik der gewerblichen Berufsbildung. Konzeptionelle Entwürfe und empirische Befunde. Baltmannsweiler: Schneider Verlag Hohengehren (Diskussion Berufsbildung, Bd. 9), S. 121–141.

- GSCHWENDTNER, T. (2008): Ein Kompetenzmodell für die kraftfahrzeugtechnische Grundbildung. In: Nickolaus, R./Schanz, H. (Hrsg.): Didaktik der gewerblichen Berufsbildung. Konzeptionelle Entwürfe und empirische Befunde. Baltmannsweiler: Schneider Verlag Hohengehren (Diskussion Berufsbildung; Bd. 9), S. 103–119.
- HAOLADER FARUQUE AHMED (2010): Technical and Vocational Education and Training – Curricular Reform demand in Bangladesh. Dissertation, Universität Stuttgart
- LEHMANN, R./SEEBER, S. (HRSG.) (2007): ULME 3. Untersuchungen von Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in den Abschlussklassen der Berufsschulen. Hamburg: Behörde für Bildung und Sport.
- LISSMANN, U. (2013): Forschungs- und Erhebungsmethoden. In: Frey, A./LISSMANN, U./SCHWARZ, B. (Hrsg.): Handbuch Berufspädagogischer Diagnostik. Beltz: Weinheim Basel, S. 87–125
- NICKOLAUS, R./ABELE, S. D./GSCHWENDTNER, T./NITZSCHKE, A./GREIFF, S. (2012): Fachspezifische Problemlösefähigkeit als zentrale Kompetenzdimension beruflicher Handlungskompetenz – Modellierung, erreichte Niveaus und relevante Einflussfaktoren in der gewerblich-technischen Berufsausbildung. In: Zeitschrift für Berufs- und Wirtschaftspädagogik, 108, H. 3.
- NICKOLAUS, R./SEEBER, S. (2013): Berufliche Kompetenzen: Modellierungen und diagnostische Verfahren. In Frey, A./Lissmann, U./Schwarz, B. (Hrsg.): Handbuch Berufspädagogische Diagnostik. Weinheim und Basel: Beltz, S. 155–180
- PRAETORIUS, A.-K. (2014). Messung von Unterrichtsqualität durch Ratings. Münster: Waxmann, S. 39
- SCHMIDT, T./NICKOLAUS, R./WEBER, W. (2014): Modellierung und Entwicklung des fachsystematischen und handlungsbezogenen Fachwissens von Kfz-Mechatronikern. In: Zeitschrift für Berufs- und Wirtschaftspädagogik (im Druck)
- SCHRADER, F. (2011): Lehrer als Diagnostiker. In: Terhart, E./Bennewitz, H./Rothland, M. (Hrsg.): Handbuch der Forschung zum Lehrerberuf. Münster: Waxmann, S. 683–689.

Anschrift der Autoren: Dipl.-Päd. Wolfgang Weber, Dipl.-Gwl. Thomas Schmidt, Dr. Stephan Abele, Dipl.-Gwl. Sebastian Heilig, Dr. Anja Sarnitz, Prof. Dr. Reinhold Nickolaus, Institut für Erziehungswissenschaft, Abteilung Berufs-, Wirtschafts- und Technikpädagogik, Geschwister-Scholl-Straße 24D, 70174 Stuttgart, wolfgang.weber@bwt.uni-stuttgart.de, stephan.abele@bwt.uni-stuttgart.de, reinhold.nickolaus@bwt.uni-stuttgart.de