

Felix Rauner



Berufliche Kompetenzdiagnostik mit COMET

Erfahrungen und Überraschungen
aus der Praxis



Felix Rauner

Berufliche Kompetenzdiagnostik mit COMET

Erfahrungen und Überraschungen
aus der Praxis



© 2018 wbv Publikation
ein Geschäftsbereich der
wbv Media GmbH & Co. KG, Bielefeld

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

Umschlagmotiv: iStock/siraanamwong

Bestellnummer: 6004677
ISBN (Print): 978-3-7639-5994-5
ISBN (E-Book): 978-3-7639-5995-2
Printed in Germany

Das Werk einschließlich seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar. Insbesondere darf kein Teil dieses Werkes ohne vorherige schriftliche Genehmigung des Verlages in irgendeiner Form (unter Verwendung elektronischer Systeme oder als Ausdruck, Fotokopie oder unter Nutzung eines anderen Vervielfältigungsverfahrens) über den persönlichen Gebrauch hinaus verarbeitet, vervielfältigt oder verbreitet werden.

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

Bibliografische Information der Deutschen Nationalbibliothek

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

Inhalt

Vorwort	7
Einleitung: Das COMET-Testverfahren – ein Medium der Qualitätsentwicklung beruflicher Bildung	9
1 Das COMET-Projekt	19
1.1 Gestaltungskompetenz – die Leitidee der modernen beruflichen Bildung	19
1.2 Das COMET-Kompetenzmodell	20
1.3 Bewertung der COMET-Projekte durch die Experten/-innen der Berufsbildungspraxis	30
2 Heterogenität der beruflichen Kompetenzentwicklung – Unterschiede wie Tag und Nacht	33
2.1 Perzentilbänder als Methode der Darstellung der Heterogenität der Kompetenzausprägung von Testgruppen	34
2.2 Die Einordnung und Bewertung der Kompetenzausprägung anhand des Heterogenitätsdiagramms	36
2.3 Beispiele zur Heterogenität von Klassen, Bildungszentren und Regionen	38
2.4 Fazit: Umgang mit Heterogenität	42
3 Die Lernortkooperation: Das Markenzeichen der dualen Berufsausbildung auf dem Prüfstand	45
3.1 Wie Auszubildende und ihre Ausbildenden/Lehrkräfte die Qualität der Ausbildung bewerten	47
3.2 Auszubildende bewerten die Lernortkooperation: Methodisches Vorgehen	48
3.3 Überraschende Ergebnisse	54
3.4 Fazit	63
4 Wie unterschiedlich kompetente Auszubildende die Ausbildungsqualität in ihren Berufen bewerten	65
4.1 Ergebnisse zum Zusammenhang zwischen Kompetenz und Ausbildungsqualität	65
4.2 Erläuterungen anhand der Detailergebnisse	67
4.3 Fazit	72

5	Kaum zu glauben: die Stagnation der Kompetenzentwicklung und wie sie überwunden werden kann	75
5.1	Die Stagnation der Kompetenzentwicklung als ein berufsübergreifendes Phänomen	76
5.2	Eine neue Qualität des Stagnationsphänomens	83
5.3	Die Stagnation der Kompetenzentwicklung kann überwunden werden ..	85
5.4	Zur Aufklärung des Stagnationsphänomens	89
5.5	Die Lehrkräfte als der entscheidende Faktor für die Kompetenzentwicklung: Das Transferphänomen	90
5.6	Fazit	94
6	Lehrer/-innen entdecken ihre Kompetenzen: ein „Aha-Effekt“ und seine Folgen	97
6.1	Entwickeln von Testaufgaben	97
6.2	Das COMET-Messmodell	99
6.3	Durchführen des Pretests	100
6.4	Das veränderte Fachverständnis prägt das didaktische Handeln der Lehrkräfte	107
6.5	Fazit	110
7	Die Testmotivation der Lernenden gibt Rätsel auf	113
7.1	Messen der Testmotivation	113
7.2	Explorative Faktorenanalyse zum Zusammenhang zwischen Testmotivation und Testleistung	114
7.3	Wie sich die Testmotivation auf das Kompetenzniveau und die Kompetenzentwicklung auswirkt	118
7.4	Eine Überraschung: Für die südafrikanische Berufsbildung lassen sich die gewonnenen Erkenntnisse zur Testmotivation nicht anwenden	121
7.5	Fazit	126
8	Was es mit dem beruflichen Wissen auf sich hat und warum es so selten systematisch gefördert wird	127
8.1	Handlungswissen (Arbeitsprozesswissen)	127
8.2	Beispiele: Ein hohes Wissensniveau als eine Voraussetzung für Verantwortung- und Qualitätsbewusstsein	131
8.3	Beispiel: Die Ausbildung von Elektronikern/-innen für Betriebstechnik (NRW) an verschiedenen Standorten	133
8.4	Beispiel Dienstleistungsberufe	135
8.5	Fazit	137
9	Zusammenfassung und Ausblick: Warum die Qualitätsentwicklung in der beruflichen Bildung ohne die Kompetenzdiagnostik nicht auskommt	139

Literatur	141
Abbildungsverzeichnis	149
Tabellenverzeichnis	154
Anhang	155
Anhang 1: Beschreibung der Lernbereiche	155
Anhang 2: Ratingskala	157
Anhang 3: Verzeichnis der COMET-Veröffentlichungen	159

Vorwort

Genauer müsste der Titel dieses Buches lauten: „Die berufliche Kompetenzentwicklung birgt Überraschungen – *wenn man über eine Methode verfügt, berufliche Kompetenz genau zu messen.*“ Damit ist nicht eine Methode gemeint, mit der berufliche Kompetenz als ein genauer Punktwert zum Beispiel auf einer Skala von 0–100 oder einer normierten Skala wie zum Beispiel bei den PISA-Tests gemessen wird. Für das Messen *beruflicher* Kompetenz sind solche Messverfahren ungeeignet, da sie es nicht erlauben, z. B. die Fähigkeit eines/-r Heizungsmonteurs/-in oder von Auszubildenden dieses Gewerbes zu überprüfen, wie gut er einen/-r Kunden/-in bei der Modernisierung der Heizung berät. Bei der Überprüfung des beruflichen Wissens und Könnens kommt es nämlich darauf an, Methoden anzuwenden, die Aussagen darüber ermöglichen, ob zum Beispiel Auszubildende gegen Ende der Ausbildung die für die Berufsfähigkeit charakteristischen Aufgaben in ihrer Komplexität versteht und sie unter Beachtung *aller (!)* relevanten Kriterien lösen kann. Dabei muss die Vielfalt möglicher Lösungen begründet gegeneinander abgewogen werden, um den jeweils gegebenen *Lösungsraum* klug auszuschöpfen. Berufliche Kompetenz zeichnet außerdem aus, bei der Lösung beruflicher Aufgaben die ausgewählte Lösung zu begründen, sodass Auftraggebende in die Aufgabenlösung einbezogen werden können.

Mit dem COMET-Testverfahren ist es möglich, die berufliche Kompetenz in allen Berufen so genau zu messen, dass das Test-/Prüfungsergebnis detailliert Auskunft darüber gibt, ob die Test-/Prüfungsteilnehmenden alle relevanten Lösungsaspekte berücksichtigt haben und auf welchem Wissensniveau sie diese begründen können. Auf der Grundlage guter Test-/Prüfungsergebnisse kann dann zum Beispiel ein/-e Meister/-in einem/-r Gesellen/-in die Beratung von Kunden und die eigenständige Ausführung von Kundenaufträgen übertragen.

Mit dieser neuen Qualität des Messens beruflicher Kompetenz kann nicht nur ein sehr genaues Bild der beruflichen Kompetenz von Test-/Prüfungsteilnehmenden erfasst werden, sondern auch das *Kompetenzprofil* angehender Fachkräfte. Die Kompetenzprofile von Auszubildenden und Fachschulstudierenden erlauben auch Rückschlüsse auf die Problemlösungsmuster ihrer Lehrkräfte, die sie auf die Lernenden übertragen. Mit einer Erweiterung des deutsch-chinesischen COMET-Projektes im Automobil-Servicesektor – auf Initiative von Prof. Zhiqun Zhao – um 79 chinesische Lehrkräfte und Dozenten/-innen (als Testteilnehmende) der beteiligten Ausbildungs- und Studiengänge ist es gelungen, diesen sehr grundlegenden Transfereffekt empirisch aufzuklären. Seither verfügen wir über die Erkenntnis, dass Lehrkräfte, ohne dass es ihnen bewusst ist, ihre fachlichen Problemlösungsmuster – ihr Fachverständnis – nahezu 1 : 1 auf die Lernenden übertragen. Die Aufnahme dieser neuen Erkenntnis der beruflichen Lehr-Lernforschung darf natürlich in einem Buch, in dem anhand auffälliger und überraschender Ergebnisse der Kompetenzdiagnostik

neue Erkenntnisse gewonnen werden, nicht fehlen. Erleben die Lehrkräfte diesen Zusammenhang bewusst anhand der COMET-Testergebnisse, dann löst dies sehr oft einen Aha-Effekt aus und verändert ihr Fachverständnis.

Die neue Form des Erfassens der Kompetenzausprägung hat in den vielen regionalen, nationalen und international vergleichenden Projekten der Kompetenzdiagnostik nicht selten große Überraschungen ausgelöst. Von solchen Überraschungen und worauf sie zurückzuführen sind, berichten wir in diesem Buch. Nicht selten standen auch wir – die Entwickler/-innen dieser Methode – vor einem Rätsel, das erst in einem langwierigen Forschungsprozess aufgeklärt werden konnte.

Dass in den ausgewählten Beispielen der Kompetenzdiagnostik in der beruflichen Bildung bei den Beteiligten Überraschungen ausgelöst wurden, zeugt von einer großen Wissenslücke über die Qualität der beruflichen Bildung bei den Verantwortlichen für die Ausbildung beruflicher Fachkräfte. Diese kann erst jetzt mit den Methoden der Kompetenzdiagnostik beruflicher Bildung (COMET) geschlossen werden. Die Beispiele zeigen auch, über welches didaktische Potenzial diese neue Form der Lehr-Lernforschung in der beruflichen Bildung verfügt und dass sich Forschende und Praktiker/-innen bei ihrer Anwendung auf allerlei Überraschungen gefasst machen können. Bei der Auswahl der Beispiele konnte ich mich auf die große Fülle interessanter und für die berufliche Bildung oft wegweisender Ergebnisse der Projekte der Bundesländer Hessen und NRW sowie der internationalen Projekte in der Schweiz, in Norwegen und anderen europäischen Ländern, Chinas und Südafrikas stützen. Den Auswertungsgesprächen mit den Koordinatoren/-innen der beruflichen Projektgruppen verdanke ich ein tieferes Verständnis der Innovationen, die diese COMET-Projekte ausgelöst haben.

Ihre Schlüsselrolle im COMET-Netzwerk zeigen die Berichte und Diskussionsbeiträge z. B. von Karin Gäumann-Felix und Daniel Hofer (Schweiz), Thomas Scholz und Gerald Hubacek (Hessen), Christof Stegemann, Kathrin von Ferde und Markus Großheim (NRW), Helen Brown (Südafrika) sowie Rongxia Zhuang und Li Ji (Peking).

Bei meinem Mitarbeiter Martin Ahrens bedanke ich mich für die professionelle Gestaltung und die redaktionelle Bearbeitung des Buchmanuskriptes mit seiner großen Zahl komplizierter Grafiken.

Bremen, im Oktober 2018
Prof. Dr. Dr. h. c. Felix Rauner

Einleitung: Das COMET-Testverfahren – ein Medium der Qualitätsentwicklung beruflicher Bildung

Ohne eine solide Kompetenzdiagnostik ist eine nachhaltige Qualitätsentwicklung in der beruflichen Bildung kaum möglich. Diese Einsicht wurde immer wieder in den COMET-(COMpetence diagnostics in vocational Education and Training)-Projekten seit 2008 bestätigt. Selbstverständlich ist diese Erkenntnis nicht. Sie stützt sich auf die Testerfahrungen, die beim Messen beruflicher Kompetenz mit offenen komplexen Testaufgaben gewonnen wurden.

Lässt sich berufliche Kompetenz überhaupt messen?

Fachkräfte, die ihren Beruf meisterhaft beherrschen – so wie es die Meisterlehre anstrebt –, sollten über die Fähigkeit verfügen, Probleme und Aufgaben in ihrem Beruf unter Beachtung aller (!) aufgabenrelevanten Kriterien zu lösen. Vergisst ein/-e Kfz-Mechatroniker/-in beim Kfz-Service die Überprüfung eines sicherheitsrelevanten Details – und das können bei einem Kraftfahrzeug sehr viele sein –, dann kann dies schwerwiegende Folgen haben. Ein/-e Prozessleitelektroniker/-in, der/die an der Leitwarte eines Stahlwerkes Prozesse der Stahlproduktion steuert, oder eine Krankenschwester/ein Krankenpfleger, die/der Patienten auf einer Intensivstation betreut, sind weitere Beispiele für die große Zahl der Fachkräfte, die in ihren Berufen große Verantwortung wahrnehmen. Sie müssen sich stets bewusst sein, was beim Bearbeiten und Lösen beruflicher Aufgaben auf dem Spiel steht. Zwischen beruflicher Kompetenz und beruflicher Arbeitsethik besteht daher ein nicht auflösbarer Zusammenhang. Der Überprüfung der Berufsfähigkeit am Ende einer Berufsausbildung wird in allen Berufsbildungssystemen eine große Bedeutung zugemessen. Das Prüfungszeugnis – im Handwerk ist dies zum Beispiel der Gesellenbrief – bescheinigt dem Auszubildenden nach bestandener Prüfung, dass er berechtigt und in der Lage ist, in der ganzen Breite der im jeweiligen Berufsbild ausgewiesenen Tätigkeitsfelder und Kompetenzen in eigener Verantwortung Arbeitsaufträge auszuführen. Dies war einer der Gründe dafür, dass in der dualen Berufsausbildung in Deutschland wieder Formen projektformigen Prüfens eingeführt wurden.

Mit der Einführung projektformiger praktischer Prüfungsformen in der Tradition des „Gesellenstückes“ seit Mitte der 1980er-Jahre (vgl. SCHMIDT 1998) wird eine kontroverse Diskussion über diese praxisnahe Prüfungsform geführt. Unstrittig ist, dass es mit projektformigen Prüfungsaufgaben möglich ist, die berufliche Wirklichkeit authentisch abzubilden. Dies wird von den Auszubildenden positiv bewertet (Abb. 1). Dagegen werden die *Prüfungsergebnisse* von den Prüfungsteilnehmenden als wenig objektiv beurteilt (Abb. 2). Die Kritik an dieser Prüfungsform richtet sich da-

her vor allem darauf, dass kein Bewertungsverfahren entwickelt wurde, mit dem die Prüfungsleistung hinreichend genau erfasst werden kann.

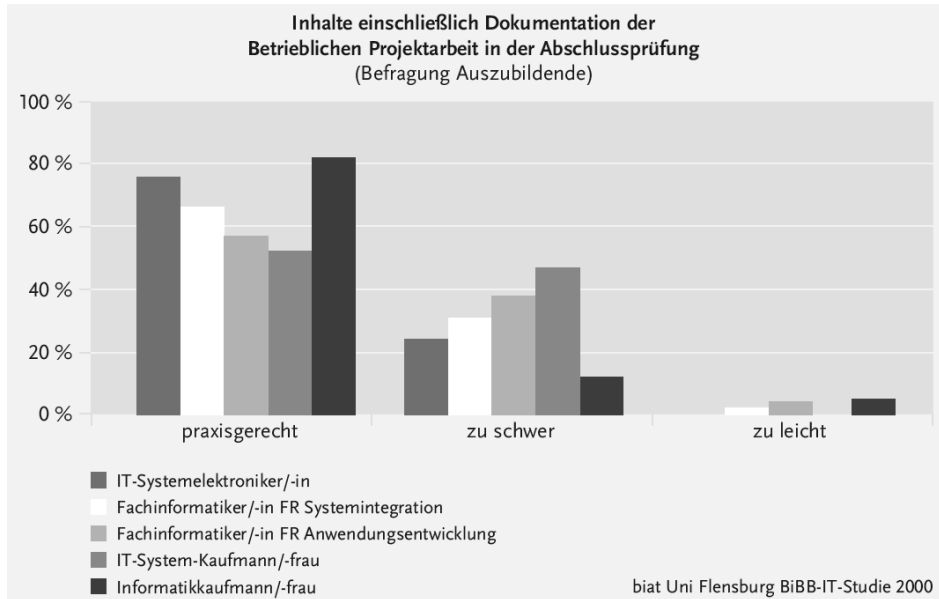


Abbildung 1: Bewertung der betrieblichen Projektarbeit durch die Auszubildenden (PETERSEN, WEHMEYER 2001).

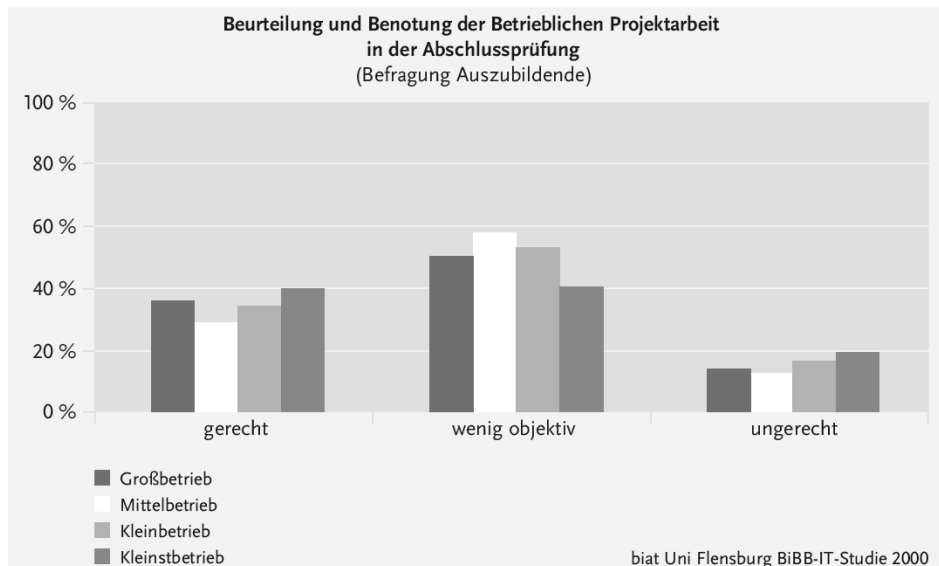


Abbildung 2: Beurteilung der Bewertung der betrieblichen Projektarbeit in der Abschlussprüfung durch die Auszubildenden (ebd., 2001).

Als ein beträchtlicher Nachteil wird von den Prüfungsexperten/-innen auch auf die begrenzte fachliche Breite dieser Prüfungsform verwiesen. Zwar könne nachgewiesen werden, ob Auszubildende *in dem einen oder anderen spezifischen Tätigkeitsfeld* ihren Beruf beherrschten, die Berufsfähigkeit könne mit dieser Form des Prüfens jedoch nicht nachgewiesen werden. Ergänzend werden daher immer noch Multiple-Choice (MC)-Prüfungsaufgaben eingesetzt. Die Konsequenz aus dieser problematischen Praxis sollte daher sein, eine die Ausbildung begleitende Kompetenzdiagnostik einzuführen, die der Qualität der Meisterlehre entspricht. Der/die Meister/-in hatte (hat) nicht nur die Verantwortung für die Ausbildung, sondern auch für die Bewertung der Berufsfähigkeit: die Fähigkeit der Gesellen, ihre Meister kompetent zu vertreten. So verlässt sich – selbstverständlich – in einer Kfz-Werkstatt der/die für die Auftragsabwicklung zuständige Kfz-Meister/-in bei der Fehlerdiagnose auf das Urteil seiner Kfz-Mechatroniker/-innen.

Ist ein „Berufsbildungs-PISA“ eine zeitgemäße Form des Prüfens und der Kompetenzdiagnostik in der beruflichen Bildung?

Das PISA-Projekt und seine internationale Reputation löste auch in der Berufsbildungsforschung und -politik eine Diskussion über ein „Berufsbildungs-PISA“ aus. Drei Fragen bestimmen seither diese Diskussion:

1. Lässt sich die PISA-Methode der Kompetenzdiagnostik auf die berufliche Bildung übertragen?
2. Ist es möglich, ein PISA-ähnliches Testverfahren zu entwickeln, mit dem berufliche Kompetenz länderübergreifend gemessen werden kann (zum Beispiel auf EU- oder OECD-Ebene)?
3. Steht damit schließlich auch ein Verfahren zur Verfügung für eine psychometrisch hochwertige Prüfungsmethode?

In einer vom BMWA (Bundesministerium für Wirtschaft und Arbeit) vergebenen Machbarkeitsstudie wurde auf die Schwierigkeiten hingewiesen, die mit der Realisierung eines Berufsbildungs-PISA verbunden sind. Dies kommt in einer zusammenfassenden Aussage der Studie zum Ausdruck: *„Ein internationaler Vergleich der Leistungsfähigkeit beruflicher Bildung via Kompetenzmessung bedarf eines gemeinsamen Verständnisses von den Zielen der Berufsbildung. Ein solches gemeinsames Verständnis kann nicht unterstellt, es muss konsensuell wissenschaftlich, gegebenenfalls auch politisch, hergestellt werden“* (BAETHGE u. a. 2006, 12). Wie ein gemeinsames Verständnis über die Ziele beruflicher Bildung in einem internationalen Vergleichsprojekt gegebenenfalls *„politisch hergestellt“* werden soll, lassen die Autoren aus verständlichen Gründen offen, da weder die EU noch die OECD über die Legitimation noch die Instrumente für ein solches Verfahren verfügen.

Auf der Grundlage der Machbarkeitsstudie gelang es zwar nicht, ein EU- oder OECD-Projekt zu initiieren, das BMBF stellte jedoch Mittel für ein nationales Forschungsprogramm zur Verfügung, um in Anlehnung an die PISA-Methode die Grundlagen für eine Kompetenzdiagnostik beruflicher Bildung zu untersuchen.

2014 bilanzierte das BMBF in seinem Forschungsbericht 2014 die Ergebnisse dieses Programms (ASCOT):

„Laut dem Berufsbildungsbericht gibt es Bestrebungen, ein europäisches Berufsbildungs-PISA auf der Grundlage von ASCOT einzuführen. Solche Überlegungen wären jedoch überhaupt nur vor dem Hintergrund eines gemeinsamen Verständnisses von Berufsbildung denkbar. Davon sind die EU bzw. die Staaten der OECD noch meilenweit entfernt. Zudem hat sich der Hauptausschuss des Bundesinstituts für Berufsbildung bereits zweimal gegen diese Initiative ausgesprochen. Die Kritik des Hauptausschusses richtet sich vor allem gegen den Anspruch, eine neue Form des Prüfens zu entwickeln. Dieser Ansatz wird als nicht praxistauglich angesehen. Darüber hinaus birgt dieser Ansatz das Risiko der Verbreitung eines verkürzten Konzeptes beruflicher Bildung. Die Gruppe der Beauftragten der Arbeitnehmer erinnert an diese Einschätzung des BIBB-Hauptausschusses und spricht sich gegen die Durchführung eines Berufsbildungs-PISA auf der Grundlage von ASCOT aus“ (BMBF 2014, 165 f.).

Auf die entscheidende Voraussetzung für das Gelingen einer Kompetenzdiagnostik in der beruflichen Bildung haben Thomas MARTENS und Jürgen ROST bereits 2009 hingewiesen: *„Grundsätzlich können [in der Kompetenzdiagnostik] Schwierigkeitsmodelle und Fähigkeitsmodelle unterschieden werden [...]. Im COMET-Projekt wird ein Fähigkeitsmodell überprüft. Es geht darum, modellhaft abzubilden, wie Probanden, deren Lösungen unterschiedliche Ausprägungsgrade aufweisen, offene berufliche Aufgabenstellungen bewältigen“* (MARTENS, ROST 2009, 98). In einem Gutachten der BMWA-Machbarkeitsstudie hatte Jürgen ROST auf dieses Problem des von den Autoren präferierten Kompetenzmessverfahrens aufmerksam gemacht: Es bestehe Einigkeit darüber,

„dass Arbeitsproben in vielen Bereichen die valideren Ergebnisse versprechen. Es geht also darum, Itemformate zu entwickeln, die aussagekräftig genug sind, um die Ergebnisse von Arbeitsproben ersetzen oder gar simulieren zu können. Vorerfahrungen bestehen mit Szenarien als Antwortformat, die sich nach verschiedenen Aspekten kodieren lassen“ (ROST 2006, XXXIV).

Das bedeutet auch, dass mit Testaufgaben, die nach dem Richtig/Falsch-Schema zu lösen sind (wie die Multiple-Choice-Aufgaben), keine beruflichen Kompetenzen gemessen werden können.

Die psychometrische Evaluation des COMET-Testverfahrens hat die Einschätzung von MARTENS und ROST bestätigt (vgl. RAUNER, HEINEMANN 2015, 64). Darauf basiert auch die hohe Akzeptanz des COMET-Testverfahrens in der Berufsbildungspraxis (vgl. die Beiträge von VESPER, SCHOLZ und HUBACEK in: FISCHER, RAUNER, ZHAO 2015) sowie in den international vergleichenden COMET-Projekten in einer großen Breite von Berufen.

Die Testergebnisse dieser Projekte sowie die Kontextanalysen ergeben ein sehr genaues und vor allem inhaltlich valides Bild über die Qualität der Berufsausbildung in den jeweiligen Berufen und Studiengängen der beteiligten Regionen und Länder.

Warum mit Multiple-Choice-Aufgaben keine berufliche Kompetenz gemessen werden kann

Anfang der 1970er-Jahre wurden als eine neue rationelle Form des Prüfens in den Abschlussprüfungen der dualen Berufsausbildung die Multiple-Choice (MC)-Prüfungsaufgaben eingeführt. Zu den auf das Fachwissen ausgerichteten Fragen werden mehrere Antwortmöglichkeiten angeboten, von denen die richtige von der zu prüfenden Person markiert werden muss (Tab. 1).

Tabelle 1: Beispiel einer Multiple-Choice-Aufgabe für Kfz-Mechatroniker (Teil 2 der Gesellenprüfung 2013).

Welche Aufgabe hat ein Partikelfilter?	
Er filtert Stickstoff aus dem Abgasstrom?	
Er filtert Rußpartikel aus dem Abgasstrom?	
Er wandelt Stickstoff in Sauerstoff um?	
Der Diesekraftstoffverbrauch wird gesenkt?	

Der Vorteil dieses Testverfahrens besteht darin, dass es eine sehr rationelle Durchführung und Auswertung erlaubt. Zwei Kritikpunkte werden vor allem von den Ausbildungsbetrieben hervorgehoben:

(1) Im Einzelfall könne der Anteil richtig geratener Antworten eine viel zu hohe Quote erreichen. Damit könnten Prüfungsteilnehmende die Prüfung bestehen, die faktisch nicht die Berufsfähigkeit erreicht haben. Dem wird zu Recht von den Befürwortern/-innen der Multiple-Choice-Aufgaben entgegengehalten, dass die zufallskorrigierte Punktzahl X' leicht zu ermitteln ist.¹

$$X' = X_R - \frac{X - X_R}{m - 1}$$

(2) Der anhand von Auswahlantworten überprüfbare Anteil der beruflichen Fähigkeiten sei sehr begrenzt. Dieser zweite Kritikpunkt wiegt schwerer.

Anhand der Konstruktionskriterien: Schwierigkeitsgrad P , Trennschärfeindex T sowie die Gleichverteilung der Lösungen auf die Distraktoren (falsche Antwortmöglichkeiten) soll bei MC-Aufgaben sichergestellt werden, dass die beruflichen Fähigkeiten valide überprüft werden können. Im Folgenden wird begründet, warum dieses Ziel mit MC-Aufgaben nicht erreicht werden kann.

Der *Schwierigkeitsgrad* P einer Prüfungsaufgabe entspricht dem prozentualen Anteil der Prüfungsteilnehmenden, welche eine Multiple-Choice-Prüfungsaufgabe richtig beantwortet haben.

¹ Die erreichte Punktzahl X_R (Rohwert) wird um einen Faktor reduziert, der sich aus der Differenz zwischen der Gesamtpunktzahl X und X_R , dividiert durch die um 1 reduzierte Zahl der Auswahlantworten m , ergibt.

$$P = 100 \cdot \frac{N_R}{N}$$

Dabei steht N_R für die Anzahl der Prüfungsteilnehmenden, die die Aufgabe richtig gelöst haben, und N für die Gesamtzahl der Teilnehmenden.

Der *Trennschärfeindex* T gibt an, wie gut eine Multiple-Choice-Prüfungsaufgabe zwischen „guten“ und „schwachen“ Prüfungsteilnehmenden unterscheidet.

$$T = \frac{R_o - R_u}{N} \cdot 100$$

R_o steht dabei für die Anzahl der Prüfungsteilnehmenden aus der oberen Hälfte der Prüfungsteilnehmenden, die eine Aufgabe richtig gelöst haben, R_u für die Anzahl der Prüfungsteilnehmenden der unteren Hälfte, die diese Aufgabe ebenfalls richtig gelöst haben, und N bezeichnet die Gesamtzahl der Prüfungsteilnehmenden.

Die Ober- und Untergruppe wird gebildet, indem man die Gesamtprüfungsergebnisse nach steigender Punktzahl ordnet und in eine gleich große obere und untere Hälfte teilt. Der Schwierigkeitsgrad und der Trennschärfeindex hängen unmittelbar miteinander zusammen (Abb. 3). Die maximale Trennschärfe wird erreicht bei einem Schwierigkeitsindex von 50 (mittlerer Schwierigkeitsgrad). Dagegen ist für Prüfungsaufgaben der Trennschärfeindex $T=0$, wenn der Schwierigkeitsindex $P=0$ oder 100 beträgt, wenn also entweder von allen Prüfungsteilnehmenden oder von keinem/-r die Aufgabe gelöst wird. Da solche Aufgaben nicht dazu geeignet sind, zwischen „guten“, „weniger guten“ und „schlechten“ Prüfungsteilnehmenden zu

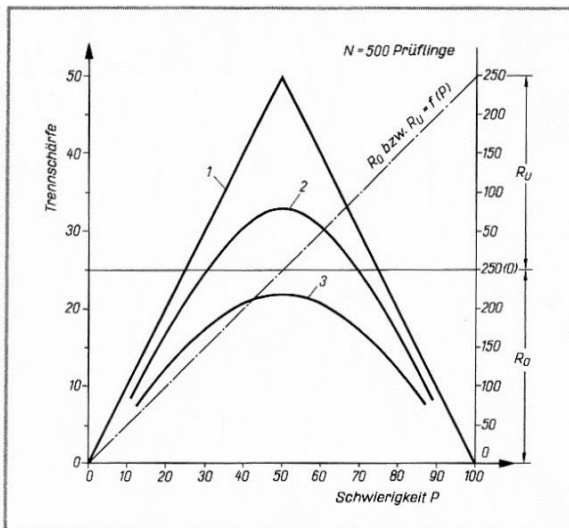


Abbildung 3: Abhängigkeit der Trennschärfe vom Schwierigkeitsindex (LÜDTKE 1974).

unterscheiden, sie also nicht valide im Sinne der *Trennschärfevalidität* sind, gelten sie nach dieser Prüfungskonzeption als ungeeignete Prüfungsaufgaben.

Neben der Ideallinie „1“ zeigen die Kurven „2“ und „3“ einen Verlauf, der empirisch erreicht werden kann. Dies liegt darin begründet, dass es bei der praktischen Anwendung von MC-Aufgaben immer auch Prüfungsteilnehmende der Untergruppe gibt, die auch einzelne schwierige Aufgaben lösen, und umgekehrt Mitglieder der Obergruppe gelegentlich auch leichte Aufgaben nicht lösen können.

Im Standardwerk zu „Forschungsmethoden und Evaluation für Human- und Sozialwissenschaften“ von BORTZ und DÖRING (2003) heißt es dazu: „Bei sehr leichten und sehr schweren Items wird man [...] Trennschärfefeinbußen in Kauf nehmen müssen. Items mit mittleren Schwierigkeiten besitzen die höchste Trennschärfe“ (ebd., 219).

Daraus wird die Schlussfolgerung abgeleitet: „Grundsätzlich sind möglichst hohe Trennschärfen erstrebenswert“ (ebd., 219). Und diese hohen Trennschärfewerte werden erreicht, wenn die Testaufgaben so konstruiert werden, dass sie „idealerweise“ bei einem mittleren Schwierigkeitsgrad ($P = 50$) liegen bzw. einen Schwierigkeitsgrad zwischen 30 und 70 oder auch zwischen 20 und 80 aufweisen (ebd., 218).

Für Prüfungsaufgaben, die schwieriger oder leichter sind, wäre der Trennschärfeindex zu niedrig, um zwischen „guten“ und „schwachen“ Prüfungsteilnehmenden zu unterscheiden. SCHELLEN kommt daher zu dem Schluss, dass Testaufgaben, die aus dem so definierten Rahmen herausfallen, „gänzlich revidiert oder durch neue ersetzt werden [müssen]“ (SCHELLEN 1994, 135). Es kommt bei dieser Form normorientierter Testaufgaben also nicht darauf an, zu überprüfen, ob ein/-e Testteilnehmer/-in über eine spezifische berufsfachliche Fähigkeit verfügt – dann käme es nämlich auf die *inhaltliche Validität* der Testaufgabe an –, sondern Testaufgaben so zu konstruieren, dass die vorgegebene Bandbreite des Schwierigkeitsgrades und ein entsprechend hoher Trennschärfewert erreicht werden. Diese Werte erreicht man durch das Justieren der Distraktoren (die falschen Antwortvorgaben) bei den MC-Aufgaben. Ist z. B. der Schwierigkeitsgrad einer MC-Prüfungsaufgabe zu niedrig, dann werden durch eine geschickte Formulierung der falschen Antworten weitere Testteilnehmenden dazu verführt, sich für eine falsche Antwort zu entscheiden. Auf diese Weise wird der Schwierigkeitsgrad einer Test- bzw. Prüfungsaufgabe *eingestellt*. In einer Expertise, die Hermann RADEMACKER 1975 im Auftrag des Bundesinstituts für Berufsbildungsforschung (BBF), dem heutigen BIBB, erstellt hat, wurde in aller Deutlichkeit herausgearbeitet, *dass mit dieser Form normorientierter Testaufgaben berufliche Fähigkeiten nicht überprüft werden können*. Er veranschaulichte dies u. a. an einem Beispiel der Ausbildung für Piloten/-innen.



Beim Abschluss der Ausbildung an einer „Pilotenschule“ wurde regelmäßig überprüft, ob die angehenden Piloten/-innen in der Lage sind, die Anzeigen des künstlichen Horizonts richtig zu interpretieren.² Die Testaufgabe lautete: „Geben Sie bitte zur folgenden Darstellung der Anzeige des künstlichen Horizonts an, in welchem Flugzustand sich ihr Flugzeug befindet!“ Die richtige Antwort lautet: „Sinken in einer Linkskurve“ (RADEMACKER 1975, 80). Regelmäßig wurde diese Aufgabe von *allen* Teilnehmenden der Ausbildung richtig gelöst. Dies ist in keiner Weise überraschend, da das Ablesen des künstlichen Horizonts bei einer Vielzahl von Probeflügen sowie im Flugzeugsimulator trainiert wird.

Die Auszubildenden (erfahrene Piloten/-innen) waren mit diesem Testergebnis (stets) sehr zufrieden. Alle Pilotenschüler/-innen hatten einen wesentlichen Aspekt der Berufsfähigkeit (als Pilot/-in) nachgewiesen. Sollte ein/-e Absolvent/-in der Ausbildung nicht über diese Fähigkeit verfügen, dann dürfte ihm die Berufsfähigkeit als Pilot/-in (selbstverständlich) nicht attestiert werden.

Die psychometrische Evaluation des etablierten Testverfahrens durch einen Testpsychologen kam jedoch zu dem Ergebnis, diese Aufgabe aus der Prüfung zu streichen oder sie umzuformulieren, da sie in der vorliegenden Form den Gütekriterien der einschlägigen Testtheorie nicht entspreche. Der Schwierigkeitsgrad und der Trennschärfeindex lägen außerhalb der einzuhaltenden Grenzwerte. Die Aufgabenstellung wurde so geändert, dass ein höherer Schwierigkeitsgrad und damit auch ein hinreichend hoher Trennschärfewert erreicht wurden. Die umformulierte Aufgabe lautete jetzt: „Zeichnen Sie bitte in die Abbildung [ein leerer Kreis, der den künstlichen Horizont symbolisierte] die Stellung des künstlichen Horizonts ein, die anzeigt, wenn Sie mit Ihrem Flugzeug im Steigen eine Linkskurve fliegen“.

Ein hinreichend großer Anteil der angehenden Piloten/-innen löste die Aufgabe nun falsch, obwohl sie alle bei ihren „Schulungsflügen“ und im Flugsimulator den sicheren und fehlerfreien Umgang mit dem künstlichen Horizont nachgewiesen hatten.

Fazit: Dieses Beispiel zeigt, dass normorientierte Tests für die Überprüfung beruflicher Kompetenz ungeeignet sind. Bei der Überprüfung beruflicher Fähigkeiten, vor allem bei solchen, die sicherheitsrelevant sind, ist der Einsatz normorientierter Testaufgaben nicht nur ungeeignet, sondern auch sehr riskant, da die inhaltliche Validität der Test- bzw. Prüfungsaufgaben nicht gegeben ist.³ So ist es z. B. auch unerlässlich, dass die VDE-Sicherheitsvorschriften bei der Installation elektrischer Anlagen von Elektrofachkräften sicher beherrscht werden. Eine Prüfungspraxis, die dies nicht überprüft, birgt unkalkulierbare Risiken, da mit einer bestandenen Prüfung auch die Berechtigung zur Installation elektrischer Anlagen verbunden ist.

Die Überprüfung beruflicher Kompetenz setzt daher notwendigerweise *inhaltlich valide* Test- und Prüfungsformen voraus (vgl. RADEMACKER 1975).

2 Der „künstliche Horizont“ gehört zu den zentralen Navigationshilfen für den Piloten.

3 Es ist bemerkenswert, dass bis heute in der einschlägigen berufspädagogischen Diskussion und Praxis normorientierte Test- und Prüfungsverfahren empfohlen und angewendet werden.