Referierte Beiträge

Frank Musekamp / Georg Spöttl / Matthias Becker[1]

Schriftliche Arbeitsaufträge zur Erfassung von Differenzen in der Expertise von Facharbeitern und Auszubildenden

Kurzfassung: Der Versuch berufliche Kompetenzen im großen Maßstab zu erfassen orientiert sich bisher stark am Vorgehen der Leistungsmessungen in allgemeinbildenden Schulen. In diesem Beitrag wird die These entwickelt, dass mit den Forschungsdesigns der momentanen Kompetenzforschung der „Qualität" oder „Natur" von Kompetenzen zu wenig Aufmerksamkeit geschenkt wird. Um diesen Aspekt stärker in den Fokus zu nehmen, wurden im Rahmen des Projekts zur Evaluation des Ausbildungsberufes Kfz-Servicemechaniker schriftliche Arbeitsaufgaben entwickelt. Diese wurden Probanden mit sehr unterschiedlichen Kompetenzniveaus vorgelegt. Die entwickelten Aufgaben orientieren sich inhaltlich sowie in ihrem Erscheinungsbild möglichst nah an den betrieblichen Arbeitsprozessen und zeichnen sich durch eine offene Antwortgestaltung durch die Probanden aus. In diesem Beitrag werden die Ergebnisse dieser Herangehensweise dargestellt und diskutiert. Es wird gezeigt, inwiefern sich berufserfahrene Facharbeiter von Auszubildenden unterscheiden und wie sich beide Gruppen mithilfe der schriftlichen Arbeitsaufgaben differenzieren lassen. Dadurch sind interessante Einsichten über die Qualität und zur Struktur beruflicher Kompetenzen möglich, die in der Zukunft eventuell zur Konzeption empirisch belastbarer Kompetenzmodelle beitragen können.

Abstract: The attempt to record occupational competences on a large scale has so far been strongly oriented towards the practice of performance measurement in general education schools. This article develops the hypothesis that the research designs of current competence research do not adequately take into account the "quality" and "nature" of competences. In order to put this aspect in the centre of interest, the project on the evaluation of the occupation of an automotive service mechanic has developed written work tasks which were presented to subjects with widely varying competence levels. The written work tasks are aligned to corporate work processes both with respect to their contents and their external shape and excel by the option of an open shaping of the answers given by the test person. This article will present the results of this approach. It will be shown how experienced skilled workers differ from trainees and how both groups could be differentiated with the aid of written work tasks. This will result in an interesting glimpse to the nature of occupational competences which could possibly contribute to the future concepts of empirically resilient competence models.

1 Dieser Beitrag beruht auf einer Untersuchung, die im Rahmen des Kfz-Servicemechanikerprojekts (Musekamp/Spöttl/Becker 2007; Musekamp/Becker 2008) realisiert und durch das Ministerium für Arbeit, Gesundheit und Soziales, NRW sowie den ESF finanziert wurde. Für die umfangreichen Auswertungsarbeiten möchten wir Nils Petermann, Hermann Hitz und Klaus Isermann danken. Wertvolle Hinweise zur Auswertung gab Marc Schütte.

1. Problemstellung und Relevanz der Thematik

Spätestens seit den internationalen Leistungsvergleichsstudien im Bereich der allgemeinbildenden Schulen TIMSS[2] und PISA[3] wird darüber diskutiert, ob und wie eine großflächig angelegte Kompetenzmessung („large scale assesment") auch in der beruflichen Bildung realisiert werden könnte. Der Nutzen eines internationalen Vergleiches im Bereich der beruflichen Bildung wird in einer erhöhten Transparenz der höchst verschiedenen europäischen Ausbildungssysteme gesehen, welche die Grundlage für eine hohe Akzeptanz des europaweit gültigen Qualifikationsrahmens (EQF[4]) darstellt. Mit dem EQF soll durch die Gleichstellung von beruflichen Abschlüssen in den europäischen Mitgliedstaaten die grenzüberschreitende berufliche Mobilität der Bürger sowie deren internationale Kooperationsfähigkeit befördert werden (siehe Hanf 2007).

Während auf der einen Seite allenthalben die Notwendigkeit einer standardisierten und wenn möglich internationalen Erfassung beruflicher Kompetenzen gesehen wird, besteht nach wie vor eine unüberschaubar große Bandbreite an wissenschaftlichen und alltäglichen Vorstellungen darüber, was berufliche Kompetenz überhaupt ist. Nickolaus (2008) konstatiert für die Wissenschaft, dass zwar zahlreiche Konzepte von Kompetenz auf theoretischer Ebene entwickelt sind (z. B. der Begriff der beruflichen Handlungskompetenz der KMK 2007), diese jedoch empirisch bisher unzureichend geprüft wurden. Hier steht die Frage nach der Natur von Kompetenzen im Mittelpunkt, also die Frage, was *die* Kompetenz in bestimmten Domänen auszeichnet oder wie sie sich zusammensetzt (vgl. Becker 2009a). In diesem Beitrag wird die Position vertreten, dass sich das „Wesen" von Kompetenz insbesondere im Vergleich von kompetenten und weniger kompetenten Personen erkennen lässt. Es wird deshalb der Frage nachgegangen, worin sich im Kfz-Service erfahrene Facharbeiter (Experten) von jungen Menschen unterscheiden, die sich erst in einem frühen Stadium der Kompetenzentwicklung befinden (Novizen).

Gleichzeitig stellt sich die Frage, ob durch die Wahl bestimmter Methoden zur Erfassung von Kompetenzen nicht schon zu sehr determiniert wird, was der Forscher vom Gegenstandsbereich – also von der Kompetenz an sich – zu Gesicht bekommen wird (Epistemologie). Erhebungsdesigns und Erhebungsinstrumente schaffen Perspektiven, die sich – sind sie einmal eingenommen – nur schwer wieder verändern lassen. Dieser Artikel zeigt auf, dass auch die momentan verstärkt eingesetzten hypothesentestenden Forschungsdesigns zur Kompetenzerfassung, die unter Einsatz probabilistischer Testmodelle realisiert werden, ontologischen Annahmen und epistemologischen Festlegungen unterliegen, die es sich bei bestimmten Fragestellungen zunächst zu verlassen lohnt. Dies gilt insbesondere dann, wenn Aussagen zum Wesen von Kompetenzen getroffen werden sollen, also diejenigen Fähigkeiten erst im Detail zu beschreiben sind, die berufliche Leistungen in der realen Arbeitswelt ermöglichen (explorativer Charakter).

Dazu werden Ansätze zur Kompetenzmessung mithilfe der Item-Response-Theorie (IRT)[5] zunächst in einigen grundlegenden Zügen beschrieben. Im Anschluss

2 Third International Mathematics and Science Study – TIMSS (Baumert/Lehmann/Lehrke 1997; Baumert/Bos/Lehmann 2000)
3 Programme for International Student Assessment (Baumert 2001a, Baumert 2001b)
4 European Qualifications Framework
5 Eine alternative Bezeichnung für die Item-Response-Theorie ist „probabilistische Testtheorie".

wird herausgearbeitet, wie eine explorative Untersuchung im Experten-Novizen Vergleich mithilfe von offenen, schriftlichen Arbeitsaufträgen Einsichten ermöglicht, die auf der Grundlage klassischer Erhebungsdesigns unter Einsatz der IRT so nicht möglich gewesen wären. Abschließend werden Potenzial und Einschränkungen eines derartigen Vorgehens diskutiert.

2. Kompetenzstruktur- und Niveaumodelle

Um eine Vorstellung über die Beschaffenheit beruflicher Kompetenzen zu entwickeln, werden sogenannte Kompetenzmodelle entworfen. Sie sollen (vorläufige) Antworten auf die Frage geben, *was Kompetenz überhaupt ist* (Ontologie).

2.1 Zur Ontologie

In aller Regel werden Kompetenzmodelle in Struktur- und Niveaumodelle unterschieden. Strukturmodelle differenzieren Kompetenzen in unterschiedliche Dimensionen, die den Zweck erfüllen „das Gefüge der Anforderungen, deren Bewältigung von Schülerinnen und Schülern erwartet wird“ (Klieme/Avenarius/Blum/Döbrich/Gruber/Prenzel/Reiss/Riquarts/Rost/Tenorth/Vollmer 2007, S. 74) zu beschreiben. Diese Kompetenzstrukturen lassen sich nach unterschiedlichen Kriterien differenzieren: Prominent ist eine Unterscheidung nach „Lernbereichen, Fächern bzw. ‚Domänen‘“ (ebd. S. 75). Für den kaufmännischen Bereich prüft Seeber (2008) beispielsweise, ob es sich bei dem mit einem multiple-choice-Test erfassten Konstrukt um ein „komplexes, gleichwohl homogenes Fähigkeitsbündel“ der kaufmännischen Fachkompetenz handelt, oder ob das Konstrukt nicht vielmehr anhand inhaltlicher Lernbereiche in volkwirtschaftliche, betriebswirtschaftliche, rechtliche und auf Rechnungswesen bezogene Kompetenzbereiche unterschieden werden kann (S. 79). Eine andere Strukturierungslogik wird von Geschwendtner (2008) im gewerblich-technischen Bereich geprüft. Neben einer inhaltsfachlichen Strukturierung in Fahrzeugtechnik und Elektrotechnik wird eine Differenzierung nach Wissensarten begutachtet (prozedurales vs. deklaratives Wissen), jedoch aufgrund fehlender empirischer Evidenz verworfen. Eine weitere Gliederungsmöglichkeit von Kompetenzstrukturmodellen besteht in der Unterscheidung von Bezugspunkten der Kompetenz, also beispielsweise einer Selbst-, Sach-, Methoden- und Sozialkompetenz, wie sie im Rahmen eines VET-LSA angedacht wird (vgl. Baethge/Achtenhagen/Arends/Babic/Baethge-Kinsky 2006). Die Unterscheidung erfolgt hier in Bezug auf Einheiten, auf die das Handeln ausgerichtet ist, welches durch die entsprechenden (Teil-) Kompetenzen ermöglicht wird. Die Sozialkompetenz ist beispielsweise die kognitive Disposition, die dazu notwendig ist, in Bezug auf andere Menschen erfolgreich zu handeln (vgl. dazu auch Straka/Macke 2008). Eine weitere Möglichkeit besteht darin, Kompetenzdimensionen anhand von Anforderungen abzugrenzen, die Gruppen ähnlicher Situationen an ein Individuum stellen (Hartig/Klieme 2006).

Kompetenzniveaumodelle haben dagegen den Anspruch, zu beschreiben, „welche Abstufungen eine Kompetenz annehmen kann bzw. welche Grade oder Niveaustufen sich bei den einzelnen Schülerinnen und Schülern feststellen lassen“ (Klieme, u.a. 2007, S. 74). Dabei lassen sich je nach Domäne wiederum zahlreiche Strukturierungskriterien differenzieren. Klieme u.a. (2007) schlagen vor, Kompetenz-

stufen einerseits über „kognitive Prozesse und Handlungen von bestimmter Qualität [zu spezifizieren], die Schülerinnen und Schüler auf dieser Stufe bewältigen können, nicht aber auf niedrigeren Stufen“ (S. 76). Andererseits wird auch vorgeschlagen, die Niveaudefinitionen anhand von „Wissensanforderungen“ zu definieren (S. 85). Darüber hinaus sind zahlreiche andere Kriterien zur Formulierung von Kompetenzniveaus denkbar. Die Forschungsgruppe um Lehmann und Seeber (2007) kombinieren Wissensarten wie Faktenwissen, Konzeptwissen und prozedurales Wissen mit kognitiven Verhaltensdimensionen wie Reproduzieren, Verstehen/Anwenden und Kritisieren/Reflektieren und untersuchen, ob sich ein Kompetenzniveaumodell anhand dieser Kategorien hypothesenkonform formulieren lässt (vgl. Seeber 2008, 2005; Hofmeister 2005). Gschwendtner (2008) wiederum präsentiert ein kraftfahrzeugtechnisches Kompetenzniveaumodell as einer Anforderungsperspektive, welches anhand der Bloom'schen Kategorien (vgl. Bloom/Engelhart/Furst/Hill/Krathwohl 1973) und zusätzlich dem Grad der Wissensvernetzung sowie der Vertrautheit von Testinhalten aus der Sekundarstufe formuliert wird.

Trotz dieser vielfältigen Möglichkeiten zur Definition von Strukturen einerseits und von Niveaus andererseits, scheint die Unterscheidung zwischen Kompetenzstruktur- und Niveaumodellen auf den ersten Blick klar. Sie legt nahe, erstere als eine Art zeitlich invariantes Gerüst zu verstehen, dessen Einzelkomponenten (Dimensionen) über die Zeit eine Veränderung über verschiedene Niveaus erfahren können. Damit wären Aussagen denkbar, wie „eine Kompetenz X besteht immer aus den Dimensionen a und b, die jeweils über die Zeit zu- oder abnehmen können“. Jedoch erklären Klieme u.a. (Klieme, u.a. 2007, S. 77), dass weder Struktur- noch Niveaumodelle Aussagen zur zeitlichen Entwicklung von Kompetenzen zulassen. Zwar wäre vorstellbar, dass verschiedene Niveaus von Kompetenzen eine Art aufsteigende Hierarchie aufweisen könnten, die von den Individuen bei zunehmender Kompetenz von niedrigem zu höherem Niveau durchlaufen werden müssten. Dies sei jedoch nicht zwangsläufig der Fall. Dazu seien vielmehr Kompetenzentwicklungsmodelle notwendig. Andererseits schließen beispielsweise Nickolaus u.a. (2008) eine zeitliche Veränderbarkeit von Kompetenzdimensionen (also -strukturen) nicht aus: „Wenn sich am Anfang des ersten Ausbildungsjahres eine Mehrdimensionalität ergibt, kann vermutet werden, dass sich diese im Verlauf der Ausbildung aufgrund paralleler Lehr–Lernprozesse auflöst bzw. die Dimensionen verschmelzen“ (S. 62). Zeitliche Veränderbarkeit ist damit weder ein Kriterium zur Unterscheidung von Struktur- und Niveaumodellen noch spielt sie innerhalb von Kompetenzniveaumodellen bzw. Kompetenzstrukturmodellen eine Rolle. Beide Arten von Modellen geben immer nur einen temporären – quasi statischen – Eindruck einer Kompetenzverteilung innerhalb eine Gruppe von Personen wieder.

Ein weiterer Aspekt von Kompetenzmodellen, der im Rahmen der hier angestellten Überlegungen von Bedeutung ist, wird von Schecker und Parchmann (2006) herausgearbeitet. Sie ergänzen die Unterscheidung von Kompetenzmodellen nach Struktur- vs. Stufenmodellen explizit um eine Differenzierung nach normativen vs. deskriptiven Modellen. Ein normatives Modell ist demnach das „Gefüge einer nach Dimensionen gegliederten Beschreibung der (kognitiven) Voraussetzungen, über die ein Lernender verfügen soll, um Aufgaben und Probleme in einem bestimmten Gegenstands- oder Anforderungsbereich lösen zu können“ (S. 47). Deskriptive Modelle sind dagegen die „typischen Muster [...] (kognitiver) Voraussetzungen“, mit dem man das Lösungsverhalten von Personen bei bestimmten Aufgaben „rekonstruieren bzw. beschreiben“ kann. Damit unterstreichen Schecker und Parchmann den dualen

Charakter von Kompetenzen, welche nicht nur eine psychologisch/biologische Seite aufweisen (Dispositionen), sondern auch eine gesellschaftlich normative (vgl. Spöttl/Musekamp 2009). Normen, die in aller Regel als Lernziele oder neuerdings als Bildungsstandards formuliert werden, müssen nicht mit dem übereinstimmen, was empirisch als Kompetenz erfassbar ist. Dies gilt umso mehr, wenn – wie beim Kompetenzbegriff – ein weiterer schlussfolgernder Schritt unternommen wird. Die Erfüllung von Kompetenzstandards, die in Kompetenzmodellen repräsentiert werden, ist in den meisten Fällen kein Letztkriterium an sich. Angenommen wird, dass Personen, die auf der höchsten Kompetenzstufe stehen, in außerschulischen also lebensweltlichen Zusammenhängen ebenfalls kompetent sind. Insbesondere in der beruflichen Bildung müssen deskriptive Modelle möglichst nahe an diejenigen Kompetenzen angelehnt sein, die zur Erbringung von Leistungen in der Arbeitswelt notwendig sind (vgl. Musekamp 2009). Lernziele in Rahmenlehrplänen sind dagegen immer gesellschaftlich ausgehandelte Normen, zu deren Erfüllung mit einiger Wahrscheinlichkeit andere Kompetenzen notwendig sind, weil im Regelfall Bildungsansprüche mit implementiert sind. Damit ist auch eng verbunden, in welchem Verhältnis Kompetenz als Disposition zur Leistung oder Performanz steht, die ausschließlich unter Einbeziehung lebensweltlicher Zusammenhänge darstellbar ist. Wird dieser Zusammenhang nicht ausreichend und vor allem umfassend genug berücksichtigt, wird mit empirischen Ansätzen die Leistungsfähigkeit einer Person nur eingeschränkt aufgeklärt. „Es ist [dann] schwierig, empirisch begründete Aussagen über das Determinationsgefüge einzelner Komponenten oder Merkmale der Kompetenz zu machen“ (Franke 2005, S. 171).

2.2 Zur Epistemologie

Die Unterscheidung von Kompetenzstruktur- und Niveaumodellen ist eng verknüpft mit der empirischen Herangehensweise zur Konstruktion und Validierung derartiger Modelle und mit den dazugehörigen erkenntnistheoretischen Grundlagen. Diese orientieren sich im beruflichen Bereich zusehends am Vorgehen der Kompetenzmessungen in allgemeinbildenden Schulen. Der Kern jener Erhebungen ist die Messung der Lernergebnisse (Output) mithilfe von standardisierten Testinstrumenten, die meist durch

- eine allgemeine Situationsbeschreibung (z. B. lebensweltliche Situierung),
- einen sehr konkreten Itemstamm (direkte Frage), sowie
- geschlossene Itemformate (multiple–choice)

gekennzeichnet sind. Auch offene Itemformate werden genutzt, die im Anschluss durch Rater geschlossen kodiert werden (z. B. richtig, teilweise richtig, falsch). In neueren Studien werden insbesondere im beruflichen Bereich auch Simulationen zur Datenerhebung eingesetzt (Nickolaus/Gschwendtner/Abele 2009; Achtenhagen/Winther 2009).

Durch die Testleistung der Probanden (beobachtete Variable) wird auf deren Kompetenzausprägung geschlossen (latente Variable). Dies geschieht wie bei PISA und TIMSS mittels Skalierung der erhobenen Daten im Sinne der Item-Response-Theorie (IRT) (vgl. Lehmann/Seeber 2007; Nickolaus/Gschwendtner/Geissel 2008; Winther/Achtenhagen 2008). Dabei wird das Antwortverhalten von Probanden auf

die in Tests gestellten Items wahrscheinlichkeitstheoretisch modelliert, wobei die entsprechenden mathematischen Modelle Eigenschaften aufweisen, die den theoretisch erwartbaren Eigenschaften der zu erfassenden Konstrukte möglichst gut entsprechen (vgl. Musekamp 2009). Entsprechen die in den Tests erhobenen Daten den formulierten Modellen, so lassen sich diese viel weitgehender interpretieren, als dies auf der Grundlage der klassischen Testtheorie möglich wäre. Das am häufigsten eingesetzte probabilistische Modell ist das Raschmodell, auf dessen Grundlage die Schwierigkeit von Items und die Fähigkeit von Probanden geschätzt, und über die Lösungswahrscheinlichkeiten in so genannten Item-Charakteristic-Curves zueinander in Beziehung gesetzt werden (vgl. Rost 2004, Wilson 2005). Neben eindimensionalen Testmodellen, die das latente Konstrukt als kontinuierliche Variable auffassen, gibt es Alternativen, deren latente Variable kategoriale Ausprägungen aufweist und so verschiedene Qualitäten oder Gruppen unterscheidet (z.B. die latente Klassenanalyse LCA, Lazarsfeld/Henry 1968) sowie Mischformen kontinuierlicher und kategorialer Modelle (z.B. das Mixed-Raschmodell, Rost 1990). Langfristig von besonderem Interesse dürften auch solche Modelle sein, die Kompetenzstufen nicht als voneinander unabhängige Lernstände verstehen, sondern als aufeinander aufbauende Entwicklungsstufen (bspw. Verhelst/Glas/de Vries 1997, Wilson 1989)[6]. Die nun folgenden Überlegungen konzentrieren sich jedoch auf das bisher in der Berufsbildungsforschung am häufigsten eingesetzte probabilistischen Testmodell, das eindimensionale Raschmodell.

Das Raschmodell beruht auf der Annahme, dass Kompetenz als eine latente Variable die Ursache der manifesten Leistung im Test ist (vgl. Wilson 2005, S. 13). Für jede Dimension in einem Test wird eine eigene Ursache unterstellt, die dann als eine eigene (Teil-) Kompetenz zu bezeichnen ist (Dimension). Dabei wird allerdings das Konzept der Ursächlichkeit (Kausalität) eng interpretiert. Gemeint ist nicht, dass die Kompetenz einer Einzelperson die Ursache für ein bestimmtes individuelles Handeln darstellen kann[7]. Stattdessen gilt eine Kausalitätsvorstellung auf der Ebene von Gruppen, die an Humes Konzept der Regelmäßigkeit von Ursache und Wirkung angelehnt ist (vgl. Hume 1739, S. 73 ff.). Danach ist eine Kompetenz als ursächlich für ein bestimmtes Testergebnis anzunehmen, wenn

- Kompetenz als Konstrukt bereits vor dem Einsatz eines Tests vorhanden ist (precedence),
- Kompetenz aufseiten der Probanden einer Stichprobe durch die Testsituation mit der Testleistung verbunden ist (contiguity) und
- die Ergebnisse replizierbar sind (regularity).

Replizierbarkeit ist im Rahmen des Raschmodells nicht im Sinne von einem erneuten Teilnehmen einzelner Personen am gleichen Test zu verstehen. Stattdessen sollte ein Rasch-skalierter Test immer zu den gleichen Ergebnissen kommen, solange

6 Insofern ist sicher nicht die mangelnde Verfügbarkeit testtheoretischer Modelle als der Engpass der beruflichen Kompetenzerfassung anzusehen, sondern vielmehr deren empirische Erprobung anhand beruflicher Inhalte.

7 Dass man Handeln auf der Ebene von Individuen nicht im eigentlichen Sinne kausal – also als Ursache Wirkungskette – erklären kann, wurde insbesondere im Rahmen der Holzkamp'schen kritischen Psychologie vertreten. „Handeln" impliziert immer Intentionen, Bedeutungen und Sinn aufseiten des Individuums und Bedarf einer Interpretation im Rahmen gesellschaftlicher Verhältnisse (vgl. sehr pointiert Tolman 1997).

er an Stichproben ein und derselben Population durchgeführt wird (vgl. Borsboom 2005, S. 55). Derartige Tests sind damit stichprobenunabhängig, jedoch nicht unabhängig von der Population, zu der sie Aussagen treffen wollen[8]. Eine Population definiert sich dabei über die Geltung des angenommenen Modells. Nur wenn das Raschmodell in der Population gilt, sind die Testergebnisse stichprobenunabhängig.

Auf Grundlage dieser epistemologischen Annahmen sind mithilfe des Raschmodells statistisch Kompetenzstruktur- und Niveaumodelle konstruierbar. *Kompetenzstrukturmodelle* konstituieren sich über Daten, denen ein bis mehrere latente Variablen als ursächlich – im eben beschriebenen Sinne – zugrunde gelegt werden können. Diese Strukturen lassen sich unter anderem über den Vergleich verschieden dimensionierter Raschmodelle erfassen[9]. Dazu wird geprüft, ob sich die Daten am ehesten über das eindimensionale oder aber über mehrdimensionale Raschmodelle beschreiben lassen. Die Frage ist dabei, welches dieser Modelle relativ am besten auf die erhobenen Daten passt. Der Charakter dieser Dimensionen ist wiederum nicht als (Teil-) Fähigkeit auf der Ebene von Einzelpersonen zu interpretieren, die man sich als physisch oder psychologisch vorstellen könnte. Tests messen in aller Regel willkürlich zusammengesetzte Fähigkeitsbündel, selbst dann, wenn man statistisch-mathematisch eine Eindimensionalität unterstellt. Dies ergibt sich aus der Tatsache, dass zur Lösung von Testaufgaben immer viele Fähigkeiten zugleich vonnöten sind, auch solche, die nicht im eigentlichen Interesse des Forschenden sind. Ein Test zum beruflichen Fachwissen erfordert z. B. auch die Notwendigkeit, die Aufgabenstellung zu lesen. Solange die Fähigkeit zu Lesen jedoch in jedem Item eine vergleichbar große Rolle spielt, während die fachliche Anforderung systematisch variiert, kann der Test als eindimensional betrachtet werden (vgl. Wu/Adams 2007).

Kompetenzniveaumodelle lassen sich entwickeln, weil die IRT eine kriteriumsorientierte Interpretation der Testergebnisse von Probanden erlaubt (vgl. Glaser 1994). Das bedeutet, dass man die Leistung von Probanden anhand derjenigen Items inhaltlich beschreiben kann, die sie zu lösen in der Lage sind. Gleichzeitig sind die Rasch-skalierten Items in eine für alle Probanden gültige schwierigkeitsbezogene Reihenfolge zu bringen. Diese Tatsache macht man sich zunutze, um zu „einer kriteriumsorientierten Interpretation der quantitativen Leistungswerte“ der Probanden zu gelangen, indem die „numerischen Werte auf der Kompetenzskala […] zu konkreten, fachbezogenen Kompetenzen in Bezug“ gesetzt werden (vgl. Hartig 2007, S. 82). Abbildung 1 gibt eine Grafik Hartigs wieder, in der nachzuvollziehen ist, wie die numerisch kontinuierlichen Aufgabenschwierigkeiten auf einer nachträglich in Abschnitte unterteilte Kompetenzskala angeordnet werden. Die einzelnen Abschnitte sind durch diejenigen Itemmerkmale zu beschreiben, die dort wegen ihres Schwierigkeitsparameters verortet sind.

Zur Charakterisierung von Kompetenzniveaus mithilfe der Rasch-skalierten Items können mindestens zwei Wege beschritten werden. Entweder werden sie in erster Linie inhaltlich über sogenannte Markier-Items definiert (Beaton/Allen 1992, im Überblick bei Seeber 2008). Oder die Items werden vor der empirischen Erhebung jeweils anhand verschiedener theoretischer Aspekte kategorisiert, um

8 Dies ist eine der größten Stärken der Item-Response-Theorie im Vergleich zur Klassischen Testtheorie.

9 Es gibt auch probabilistische Testmodelle, die als Konstrukt (latente Variable) kategoriale Ausprägungen aufweisen, also verschiedene Qualitäten unterscheiden, statt eines quantitativen Kontinuums wie beim klassischen Raschmodell. Diese haben aber in der beruflichen Kompetenzforschung bisher eine geringe Bedeutung.

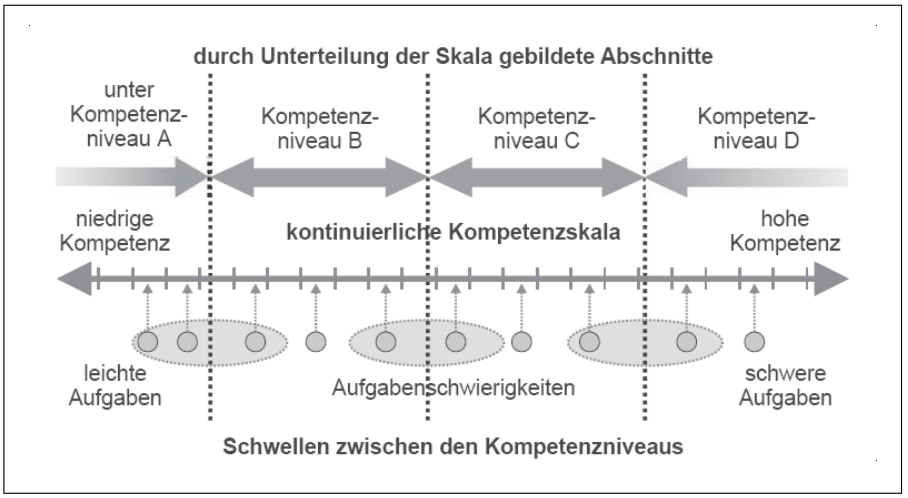

Abb. 1: Veranschaulichung der Unterteilung einer kontinuierlichen Kompetenzskala mit darauf verorteten Aufgabenschwierigkeiten in Kompetenzniveaus. Die Aufgabenschwierigkeiten in Nachbarschaft der Schwellen zwischen den Niveaus sind unterlegt, Quelle: Hartig (2007, S. 83).

anschließend regressionsanalytisch zu prüfen, welche dieser Aspekte die Varianz in den empirischen Aufgabenschwierigkeiten am ehesten aufklären (Hartig 2007). Die jeweils erklärungsstärksten Kategorien können dann genutzt werden, um die Niveaustufen auf abstraktere Weise zu beschreiben, als es auf der Grundlage der tatsächlich eingesetzten Einzelitems möglich wäre.
Während der erste Ansatz vor allem dazu dient, Ergebnisse aus large- scale- Erhebungen den beteiligten und interessierten Gruppen und Institutionen zurückzumelden (vgl. Beaton/Allen 1992, S. 191), erhebt das Verfahren auf regressionsanalytischer Grundlage auch den Anspruch, Erkenntnisse zum Charakter von Kompetenzen zu gewinnen: „Werden Annahmen über die schwierigkeitsbestimmenden Merkmale einer Aufgabe vor der Erhebung empirischer Daten formuliert, gewinnen diese den Status empirisch prüfbarer Hypothesen über Charakteristika der zu erfassenden Kompetenz“ (Hartig 2007, S. 84). Auf diese Weise lassen sich die schwierigkeitsbestimmenden Merkmale der Items im Test auf andere mögliche Items verallgemeinern. Diese verallgemeinerbaren Merkmale beschreiben dann den Charakter der erfassten Kompetenz ganz unabhängig davon, welche Items in einem konkreten Test eingesetzt wurden.

3. Zwischenfazit und Forschungsbedarf

Mithilfe des probabilistischen Vorgehens zur Konzeption von Kompetenzstruktur- und Niveaumodellen wird versucht, nicht nur die Kompetenzstände in bestimmten Stichproben zu erfassen und diese Ergebnisse auf verschiedenen Ebenen des Berufsbildungssystems zu berichten. Es wird ebenfalls erwartet, mit diesem Vorgehen Hinweise auf die „Natur“ von Kompetenz zu generieren, indem von den Merkmalen der Aufgaben im Test auf die Charakteristika von Kompetenzen geschlossen wird.

Ziel ist also nicht allein, Testergebnisse für Fragestellungen der Individualdiagnostik, des Institutionenvergleichs oder der Systemvergleiche zu funktionalisieren, sondern es geht auch um das Wesen der Kompetenz an sich (Ontologie). Insofern hat die Konstruktion von Kompetenzniveaumodellen auch einen explorativen Charakter. Dies birgt jedoch zwei grundsätzliche Probleme: Erstens werden die Tests bisher vornehmlich sehr einheitlichen Stichproben vorgelegt, in denen der Unterschied zwischen kompetent und nicht-kompetent nicht zwangsläufig in seinen wichtigen Aspekten zu beobachten sein muss. Zweitens geht als Kriterium für Kompetenz in die Items nur ein, was zuvor von den Autoren derselben als relevant erachtet wurde. Es handelt sich also zunächst um normative Modelle. Dies soll im Folgenden ausgeführt werden, bevor eine alternative Richtung diskutiert wird.

Zu erstens: Es wird theoretisch angenommen, dass sich die (kognitiven) Voraussetzungen, berufliche Aufgaben fachgerecht zu bearbeiten und zu lösen (Kompetenzen), nicht kontinuierlich vom ersten Tag der Ausbildung bis zum Ende des Erwerbslebens entwickeln. Vielmehr ist von Kompetenzsprüngen und qualitativen Veränderungen im Kompetenzkonstrukt auszugehen, welche z. B. durch Kommunikation und Interaktion im Arbeitsprozess oder auch durch Krisen und deren Bewältigung ausgelöst werden (vgl. Minnameier 2000). Vorstellbar wäre z. B., dass ein Auszubildender, der im Rahmen seiner Lehrzeit ein relativ konsistentes Bild seines Fachgebietes aufgezeigt bekommen hat, durch den Versuch der einfachen Anwendung seines Wissens auf die Praxis erfahren muss, dass das theoretische Fachwissen nur in begrenztem Maße ausreicht, berufliche Aufgaben zu lösen. Das Überwinden einer solchen Krise, die zu einer Integration fachsystematischen/theoretischen Wissens mit den Erfahrungen aus der Praxis führen muss, wird sicher nicht ohne Auswirkungen auf die Kompetenzentwicklung bleiben. Unklar ist jedoch, ob eine solche Krise zu einem bloßen „Mehr" der alten Kompetenzform führen wird, oder eher zu einem auch qualitativ veränderten Konstrukt. Es ist nicht auszuschließen, dass sich über die Zeit nicht nur die Höhe der Kompetenzausprägungen innerhalb der Dimensionen verändert, sondern auch die Struktur der Kompetenz (also die Dimensionen selbst), wie es bereits bei Nickolaus u. a. (2008, S. 62) angedeutet wurde. Dies ist in Abbildung 2 durch verschiedene Ebenen der Kompetenzentwicklung skizziert.

Die Annahme diskontinuierlicher Kompetenzentwicklung wird jedoch in den derzeitigen Designs zur Kompetenzfeststellung nur wenig berücksichtigt, denn bisher werden die Erhebungsinstrumente an relativ begrenzten Ausschnitten der Kompetenzentwicklungsphase konzipiert; häufig an Stichproben einer einzigen Jahrgangsstufe (z. B. die Grundstufe bei Nickolaus u. a. 2008 und die Abschlussklasse bei Lehmann/Seeber 2007). Mit derartigen Designs bleibt jedoch verborgen, ob das Erreichen eines bestimmten Kompetenzniveaus in den früheren Phasen eines Kompetenzentwicklungsverlaufes auch funktional für das Erreichen von Kompetenzqualitäten in späteren Phasen der Kompetenzentwicklung ist. Der Blick auf die Kompetenzstrukturen allein über kurze Etappen der Entwicklungszeit kann nicht verlässlich Auskunft darüber geben, was Kompetenz in der höchsten Ausprägung und in ihrem Wesen ausmacht. Man könnte von einer Art „Konstruktunvollständigkeit" sprechen, die aufgrund der ausschnittsweisen Betrachtung eines (möglichst lebens-) langen Prozesses der Kompetenzentwicklung zustande kommt. Diese mögliche Konstruktunvollständigkeit wird auch von Längsschnittdesigns auf der Grundlage des Raschmodells systematisch ausgeblendet. Zwar ist es ein großer Vorteil, dass das Raschmodell eine Messung der Veränderung eines Konstrukts

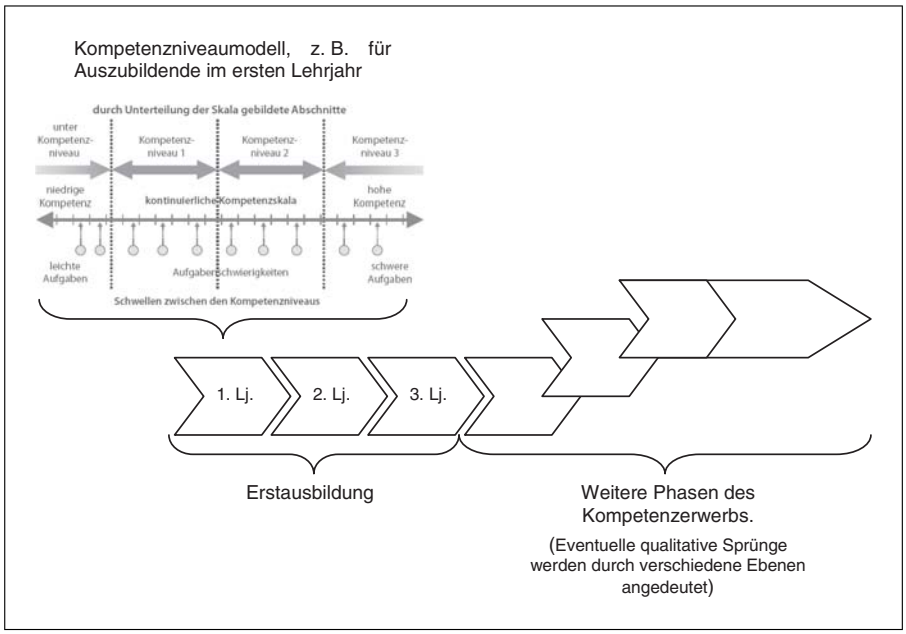

Abb. 2: Kompetenzverlauf unter der Annahme lebenslangen Lernens, Kompetenzniveaumodell entnommen aus Hartig (2007, S. 83, erweitert durch die Autoren)

mithilfe verschiedener Tests (bei nur teilweise identischen Items) erlaubt, jedoch wird Veränderung in erster Linie als quantitative Veränderung bzw. Abnahme eines qualitativ konstanten Konstrukts verstanden. Hinweise auf sich verändernde Konstrukte zu zwei Messzeitpunkten werden ignoriert, indem Items, die zu beiden Messzeitpunkten unterschiedlich funktionieren (DIF), nicht zur Skalenverankerung herangezogen werden (vgl. Wright/Masters 1982).

Werden Kompetenzerhebungsinstrumente jedoch allein an kleinen Ausschnitten der gesamten Kompetenzentwicklung konzipiert, besteht die Gefahr, dass sie geradezu an dem vorbei messen, was Kompetenz ausmacht. Ein illustrierendes Beispiel findet sich bei Hagemeister (1999), der das TIMSS-Item D2 kritisiert, in dem es um ein physikalisches Experiment zum Magnetismus geht (siehe Abbildung 3).

Hagemeister räumt ein, dass Stoff C sicherlich Kaffee sein könnte, fügt aber hinzu: „Um reines Eisenpulver handelt es sich nicht, denn ein solches Häufchen Eisenpulver würde vollständig an einem Magneten hängen bleiben, wie jeder weiß, der schon einmal mit Eisenpulver und Magneten experimentiert hat“ (S. 15). Mit diesem Hinweis auf die Praxiserfahrung wird deutlich, dass Personen aus ihren Erfahrungen heraus häufig nicht nur die eindeutigen Situationen (hier: die reinen Stoffe) kennen, sondern auch wissen, dass die Realität oft mehrdeutig, intransparent oder komplex sein kann. Wer häufig komplexe Probleme löst, weiß, dass die zunächst einleuchtend scheinende Lösung oft nicht die richtige ist. Wer allein die fachlichen Regeln in ihrer didaktischen Reduktion auf das Wesentliche kennt, ist wahrscheinlich eher bereit, die eindeutige Lösung auszuwählen, während der

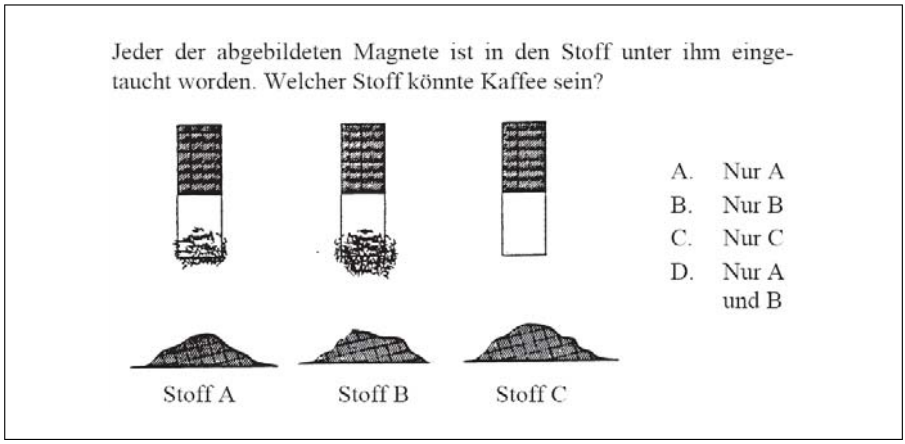

Abb. 3: TIMSS-Item D2, zitiert nach (Hagemeister 1999, S. 15).

Erfahrene zumindest stutzt, weitere Optionen für möglich hält und dann eventuell das falsche Kreuzchen setzt[10].

Zu zweitens: Forscher reproduzieren immer zu gewissen Teilen, was sie zuvor durch je eigene Blickwinkel, implizite Alltagshypothesen oder spezifische Methoden in den Forschungsprozess eingebracht haben. Insofern stellt die Aussage, Testentwickler würden nur erfahren, was sie als Testmaterial in den Forschungsprozess einbringen, kein testspezifisches Problem der Forschung dar. Wenn jedoch Aussagen zur Ontologie von beruflichen Kompetenzen generiert werden sollen, ist es nicht angebracht Items auf der Grundlage von sozialpartnerschaftlich ausgehandelten Berufsbildern, Rahmenlehrplänen oder von fachsystematischen Lehrbüchern zu erstellen. Auf dieser Grundlage lassen sich zwar Kompetenzstruktur- und Niveaumodelle generieren, die eine Beschreibung unterschiedlicher Lernstände ermöglichen. Ob die erreichten Lernstände die Auszubildenden jedoch wirklich in die Lage versetzen, Facharbeit im betrieblichen Kontext mit mehrdimensionalen Anforderungsstrukturen zu leisten, bleibt zunächst eine offene Frage (vgl. Spöttl 2009) .

Zur Lösung des zweiten Aspekts besteht die Möglichkeit, Testleistungen systematisch mit Leistungen in realitätsnahen Simulationen oder Arbeitsproben zu vergleichen, wie Nickolaus u. a. (2009) es erfolgreich zur Validierung eines Messinstruments von Diagnosefähigkeiten am Fahrzeug realisieren (vgl. für den kaufmännischen Bereich Achtenhagen u. a. 2009). Eine alternative Strategie besteht darin, sich dem Kompetenzkonstrukt von einem möglichst weit fortgeschrittenen *empirischen* Niveau aus zu nähern anstatt von normativ formulierten Lernzielen oder Bildungsstandards. Dies lässt sich in der zeitgleichen Einbeziehung von Gruppen erreichen, die sich in einem frühen Stadium der Kompetenz befinden (beispielsweise Auszubildende) und Personen, die schon weiter fortgeschritten sind (wie erfahrene Fachkräfte).

10 Dass daraufhin für den Probanden die Frage in den Vordergrund rückt, was der Tester vom Probanden erwartet, und die eigentliche Frage nach dem Fachinhalt in den Hintergrund gerät, ist eines von weiteren Problemen dieser Testaufgabe, welche nicht weiter bearbeitet werden sollen (vgl. dazu Meyerhöfer 2005).

Damit wird zugleich der unter *erstens* genannte Aspekt der Konstruktunvollständigkeit besser berücksichtigt, da prinzipiell die Möglichkeit gegeben ist, dass die Charakteristika einer hoch ausgeprägten Kompetenz im Unterschied zu Merkmalen einer niedrigen Kompetenzausprägung offensichtlich werden. Ein solches Vorgehen des Experten-Novizen-Vergleichs erinnert an die P-Technique (Cattel 1966) im Rahmen von Faktorenanalysen, bei der nicht Faktoren über den interpersonalen Vergleich mehrerer Variablen modelliert werden (R-Technique), sondern intrapersonale Entwicklung durch die Faktorenanalyse von Testdaten gleicher Probanden zu verschiedenen Zeitpunkten. Weil die Zeitspannen der Kompetenzentwicklung jedoch zu lang sein dürften, kommt es beim Experten-Novizen-Vergleich nicht zu Wiederholungsmessungen bei gleichen Personen, sondern zum Einbezug verschiedener extremer Statusgruppen.

In früheren Stadien des Erkenntnisprozesses ist es zudem angebracht, vor dem Einsatz hypothesenprüfender Verfahren der Kompetenzforschung explorative Verfahren einzusetzen, solange grundlegende Fragen zur Natur von Kompetenzen in beruflichen Domänen nicht geklärt sind (vgl. Becker 2009b, S. 55 ff.). Diese sollten derart offen gestaltet werden, dass die Spezifika einer hohen Kompetenz im Unterschied zu einer weniger hohen Kompetenz auch zum Ausdruck kommen können. Die von den Autoren durchgeführte Untersuchung stellt einen Versuch zur Aufklärung und Einbeziehung dieser Spezifika mithilfe eines Experten-Novizen-Vergleiches dar. Das Vorgehen wird im Folgenden vorgestellt.

4. Experten-Novizen-Vergleich zur Generierung von Kompetenzvorstellungen

In einer eigenen Untersuchung wird versucht, sich der empirischen Beschaffenheit von Kompetenzen explorativ in der methodischen Tradition der Expertiseforschung (Gruber/Ziegler 1996) zu nähern, die im Rahmen der modernen Kompetenzforschung sehr stark in den Hintergrund getreten ist. Dort wurden systematisch Experten mit Novizen verglichen, um so die Ursachen der großen Leistungsunterschiede zwischen Experten und Novizen zu ergründen (Gruber 1994). Im Vokabular der modernen Kompetenzforschung ausgedrückt begibt sich die Expertiseforschung auf die Suche nach den qualitativen Unterschieden in der Kompetenz von Probanden auf verschiedenen (wenngleich extremen) Stufen ihrer Kompetenz. Sie ist damit nicht normativ in dem Sinne, dass – anhand welcher Kriterien auch immer – zunächst Kompetenzmodelle formuliert werden, deren Erreichen im Anschluss getestet wird. Stattdessen ergeben sich Hinweise, anhand welcher Kriterien Kompetenzmodelle strukturiert werden können, die von vornherein als deskriptive bzw. empirische Modelle gelten können. Auf diese Weise ergeben sich Standards (und zugehörige Kompetenzmodelle) aus dem Charakter der empirisch vorhandenen und empirisch notwendigen Kompetenzen zum Lösen beruflicher Aufgaben.

4.1 Forschungsfrage und Erhebungsinstrument

Im Rahmen des Projektes zur Evaluation des Berufes Kfz-Servicemechaniker (Musekamp/Becker 2008, Musekamp 2007) konnte der Zugriff auf die Zielgruppe der Kfz-Auszubildenden genutzt werden, um eine explorative Untersuchung zur Natur

von notwendigen Kompetenzen im Kfz-Service-Sektor zu realisieren. Dazu wurde ein eigenes Instrument, sogenannte schriftliche Arbeitsaufträge entwickelt. Es wurde schnell klar, dass ein derartiges Instrument im Rahmen dieses Projektes wegen eines anders gelagerten Projektschwerpunkts keiner umfassenden methodischen Prüfung unterzogen werden konnte. *Stattdessen wurde die Frage nach den Unterschieden zwischen den Aufgabenlösungen von Personen verschiedener Kompetenzniveaus in den Mittelpunkt gestellt, um so Hinweise zu erhalten, inwieweit sich das Kompetenzniveau bzw. worin sich die Kompetenzstruktur von Novizen und Experten im Kfz-Service unterscheiden.*

Ausgangspunkt der Methodenentwicklung war die zentrale Hypothese, dass sich berufliche Kompetenzen kaum vollständig formal modellieren lassen dürften (siehe auch Nickolaus/Gschwendtner/Knöll 2006, S. 210). Es ist daher notwendig, sich bei der Beschreibung von Kompetenzen auf diejenigen Aspekte zu beziehen, die *bedeutende* Unterschiede in den beruflichen Leistungen hervorrufen. Weiterhin wurden die drei folgenden Annahmen zugrunde gelegt:

1. Berufliche Kompetenzen äußern sich nicht nur darin, dass Personen exakt definierte Teilaufgaben richtig ausführen, sondern vor allem dadurch, dass sie die zur Lösung einer beruflichen Aufgabe notwendigen Anforderungen zunächst erkennen und sich diese dann als *ihre* Aufgaben aneignen. Deshalb wurden offene Antwortformate für die Kompetenzerfassung genutzt, denn wenn in einem Kompetenzfeststellungsverfahren eine geschlossene Frage gestellt wird, kann die Leistung, sich selbst die richtigen Fragen zu stellen, nicht mehr erfasst werden.
2. Offene Erhebungsinstrumente haben überdies den Vorteil, dass Probanden ihre Leistungen in der Form äußern können, wie es ihrem Kompetenzniveau entspricht. Dies verringert die Gefahr, dass durch ex ante festgelegte Formate und Inhalte des Erhebungsverfahrens diejenigen Aspekte von Kompetenz nicht gezeigt werden können, die für außergewöhnliche Leistungen notwendig sind. Dies unterstreicht den explorativen Charakter der hier gewählten Vorgehensweise. Eine vollkommen „theoriefreie" Datenerhebung ist jedoch nicht möglich, da Annahmen natürlich auch in die Gestaltung offener Aufgaben und vor allem in die Wahl der Auswertungsstrategien und –kategorien eingehen.
3. Wenn es richtig ist, dass sich Kompetenzen nicht nur in ihrem Niveau, sondern vor allem in ihrer Qualität über die Zeit ändern, und wenn der erfahrene Facharbeiter den Zielzustand einer entwickelten Kompetenz am ehesten verkörpert, so muss der Facharbeiter bei Untersuchungen zur Natur von Kompetenzen (ontologischer Aspekt) einbezogen werden. Andernfalls besteht die Gefahr, dass ein Verfahren zur Kompetenzerhebung, das nur unter Einbezug von Auszubildenden entwickelt wurde, blind für die Kompetenzqualität von Facharbeitern ist, deren spezifische Ausprägung gerade das Ziel der beruflichen Bildung darstellt.

Um Daten unter Berücksichtigung der genannten Annahmen erheben zu können, wurden im Rahmen des Projekts zur Evaluation des Berufes Kfz-Servicemechaniker schriftliche Arbeitsaufträge entwickelt, die sich inhaltlich sowie in ihrem Erscheinungsbild möglichst nah an der betrieblichen Praxis orientieren und keine geschlossenen Fragen enthalten (Abbildung 4). Die Probanden wurden gebeten, diese Aufgabe in drei Schulstunden schriftlich in Aufsatzform zu lösen.

Zusätzlich zur dargestellten Eingangssituation wurden den Probanden ohne weitere Hinweise zahlreiche Informationen zum Fahrzeug zur Verfügung gestellt (z. B.

Vorgehen bei einer Inspektion

Aufgabe:

Ein Kunde möchte an seinem VW Polo eine **Inspektion** durchführen lassen. Wie gehen Sie vor? Die von Ihnen ausgefüllte Auftragskarte liegt bei.

1. Beschreiben Sie mit Skizzen und Worten **genau**, wie Sie die **wichtigsten** Arbeiten durchführen. Worauf achten Sie dabei und warum?

Es handelt sich um folgendes Fahrzeug:

Bild 1

Das Annahmegespräch mit dem Kunden sieht folgendermaßen aus:

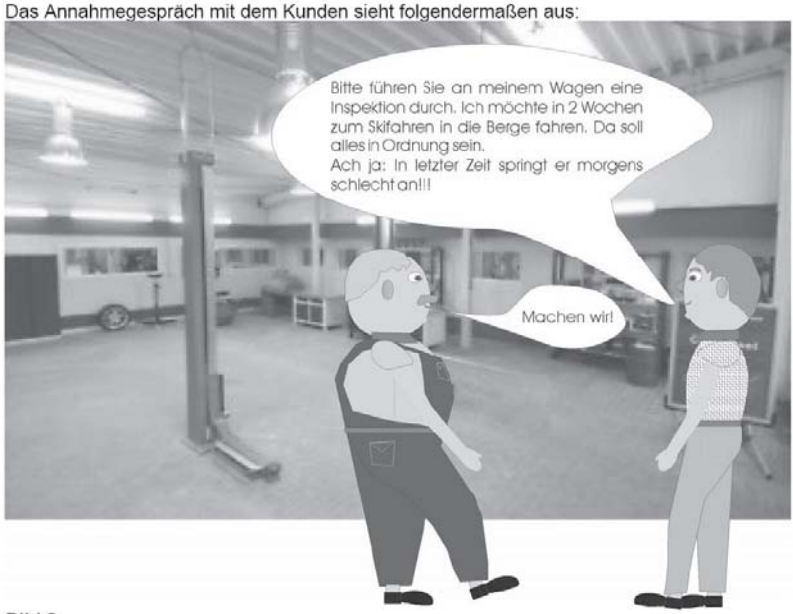

Bild 2

Abb. 4: Offene Aufgabenstellung des schriftlichen Arbeitsauftrags

Fahrzeugpapiere, technische Detailaufnahmen, Auszüge aus Reparaturleitfäden und Ersatzteillisten, Prüfplaketten usw.). Ohne weitere Hinweise deshalb, weil auch in der Werkstatt und an einem Fahrzeug eine nahezu unbegrenzte Menge an Informationen zur Verfügung steht, aus der die jeweils relevante Menge herausgegriffen werden muss. Der erste schriftliche Arbeitsauftrag befasste sich mit einem Inspektionsauftrag in einer Kfz-Werkstatt. Er wurde im Dezember 2005 an sechs nordrhein-westfälischen

Berufskollegs und einer schleswig-holsteinischen Berufsschule vorgelegt. Zielgruppe waren Kfz-Servicemechaniker (Kfz-Sm) der Einstellungskohorte 2005, sowie in vergleichbarer Anzahl Kfz-Mechatroniker (Kfz-M) desselben Einstellungsjahrgangs. Damit lag der Zeitpunkt der Aufgabenbearbeitung für beide Gruppen innerhalb der jeweils ersten fünf Ausbildungsmonate. Zusätzlich wurde der erste Arbeitsauftrag auch mit 21 erfahrenen Facharbeitern (Kfz-Mechanikern) sowie 20 nicht betrieblich eingebundenen Jugendlichen im Berufsgrundbildungsjahr (BGJ) durchgeführt. Der zweite Auftrag wurde an denselben Berufskollegs im März 2007 vorgelegt. Inhaltlich ging es dort um das Finden eines Fehlers am Fahrwerk eines Kundenfahrzeugs, wobei sich die Form des Arbeitsauftrages nicht von dem ersten unterschied. Aus forschungspraktischen Gründen konnten für den zweiten Arbeitsauftrag weder Facharbeiter noch Teilnehmer eines Berufsgrundbildungsjahres gewonnen werden. Eine Übersicht der erreichten Stichprobe gibt Tabelle 1.

Tab. 1: Stichprobenübersicht Arbeitsaufträge I und II

Nummer der EA			Häufigkeit	Prozent
Erster Auftrag (Inspektion)	Gültig	Kfz-Sm	82	48,0
		Kfz-M	48	28,1
		BGJ	20	11,7
		Facharbeiter	21	12,3
		Gesamt	171	100,0
Zweiter Auftrag (Fahrwerk)	Gültig	Kfz-Sm	41	36,6
		Kfz-M	64	57,1
		Gesamt	105	93,8
	Fehlend	System	7	6,3
	Gesamt		112	100,0

Zur einheitlichen Bewertung wurde anhand der heuristischen Dimensionen „Gegenstandsbezug“, „Werkzeuge, Organisation und Methoden der Arbeit“ sowie „Anforderungen von Staat, Gesellschaft und Individuum an Facharbeit“ ein Analyseraster entwickelt[11]. Es enthielt 31 Indikatoren, die bei Beachtung durch einen Werkstattmitarbeiter im Arbeitsprozess zu einem Resultat im Sinne „guter Facharbeit“ führen würden. Ziel war es, die Arbeit – dargestellt in schriftlicher Form – im Sinne objektiver Handlungsresultate zu interpretieren und die Güte des Arbeitsergebnisses anhand der Indikatoren zu bewerten. Demnach handelt es sich nicht um ein Kompetenzmodell in dem „(kognitive) Voraussetzungen“ von Personen normativ und schon gar nicht deskriptiv formuliert werden (vgl. Schecker/Parchmann 2006), sondern um eine Beschreibung der objektiven Anforderungen, die für eine gute Facharbeit erfüllt sein müssen. Straka spricht von externen Bedingungen oder von Sachstruktur (Straka/Macke 2008).

11 Die Bedeutung dieser Dimensionen können bei Spöttl und Gerds (1999) nachgelesen werden, die diese im Rahmen der Entwicklung eines europäischen Berufsbildes „Kfz-Mechatroniker“ ausführlich beschrieben haben.

Die Bewertung der Aufgaben erfolgte durch Berufspädagogen der Fachrichtung Fahrzeugtechnik, die ebenfalls über langjährige Erfahrungen im Werkstattalltag verfügten, sowie durch erfahrene Facharbeiter (Kfz-Meister). Wurden zu bestimmten Indikatoren keine Angaben gemacht, wurde der Indikator als „nicht zutreffend" bewertet.

4.2 Ergebnisse

Zur Auswertung der Daten wurden zwei verschiedene Faktorenanalysen berechnet. Während die erste alle Fälle und alle Erhebungszeitpunkte einschloss, bezog sich die zweite lediglich auf die Fälle, welche die erste Aufgabe bearbeitet hatten (vgl. Tabelle 1). Trotz der großen Heterogenität bezüglich der Probanden und der Aufgaben erbrachte die erste Analyse für den explorativen Zweck ein interessantes Ergebnis, welches sich für die zweite Analyse nur teilweise replizieren ließ. Trotz der offensichtlichen Unzulänglichkeiten der heterogenen Stichprobe soll daher zunächst das Ergebnis der ersten Analyse vorgestellt werden.

Unter Einbezug aller Fälle beider Erhebungszeitpunkte konnten mithilfe einer Faktorenanalyse mit Varimax-Rotation aus den insgesamt 31 Bewertungskriterien des Analyserasters 15 Items selektiert werden, die mit einer Stärke zwischen 0,6 und 0,8 auf drei Faktoren laden (siehe Tabelle 2). Die drei gefundenen Faktoren klären 63 % der Varianz in den Daten auf. Die Interraterreliabilität konnte nur in wenigen Fällen überprüft werden, und nahm dort auch nur mittelmäßige Werte an. Der Weighted-Kappa-Koeffizient mit quadratischen Gewichten lag zwischen 0,41 und 0,52. Hinzu kommt, dass der Vorteil der freien Gestaltbarkeit der Aufgabenlösungen den Nachteil einer erhöhten Bedeutung der schriftlichen Ausdrucksfähigkeit mit sich bringt, die die Ergebnisse in bisher unbekanntem Maße beeinflussen. Diese wenig zufrieden stellenden Parameter müssen bei der Interpretation der Ergebnisse berücksichtigt werden und unterstreichen den explorativen Charakter der Untersuchung.

Die restlichen 16 Items wurden durch die Aufgabenstellung offensichtlich nicht stimuliert. Es gab kaum Äußerungen der Probanden, welche unter den aufgestellten Indikatoren hätten subsumiert werden können. Darunter sind Formulierungen wie „Werkzeuge werden fachgerecht genutzt" oder „Arbeitssicherheit wird gewährleistet". Als Ursachen sind mindestens zwei Erklärungen denkbar. Möglich ist, dass die Aspekte von den Probanden zwar im wirklichen Handlungsvollzug berücksichtigt worden wären, diese jedoch in der Testsituation nicht verbal expliziert werden können (implizites Wissen). Alternativ ist vorstellbar, dass den Probanden, die mit den entsprechenden Indikatoren zu bewertenden Sachverhalte bekannt und auch bewusst waren, ihnen jedoch das detaillierte Aufschreiben als zu aufwendig oder unwichtig erschien. Da diese Indikatoren theoretisch jedoch nicht als unwichtig einzustufen sind, muss davon ausgegangen werden, dass die offene Datenerhebungsmethode „schriftlicher Arbeitsauftrag" in verschiedener Hinsicht blind ist. Dies muss als Schwäche der Erhebungsmethode „offene schriftliche Arbeitsaufträge" gewertet werden.

Dennoch lassen sich die gefundenen Faktoren sinnvoll interpretieren und ermöglichen damit einen Erkenntnisgewinn: Die Items des ersten Faktors beziehen sich jeweils auf Aufgabenaspekte, die unmittelbar aus der Situationsbeschreibung und den Aufgabenmaterialien hervorgehen, während die Items des zweiten Faktors über die unmittelbare Kernaufgabe hinausgehen, und nicht direkt durch die Materialien

„stimuliert“ werden. Während es aufgrund der Aufgabenbeschreibung mit detaillierten Fotos des Fahrzeugs offensichtlich ist, dass beispielsweise technische Aufgaben zu bearbeiten sind, und diese Aufgaben in einer zeitlichen Reihenfolge günstig zu

Tab. 2: Rotierte Komponentenmatrix (Extraktionsmethode: Hauptkomponentenanalyse, Rotationsmethode: Varimax mit Kaiser-Normalisierung, die Rotation ist in 7 Iterationen konvergiert)[12]

		Komponente		
	Items (Bewertungskriterien)[13]	**1**	**2**	**3**
1	technische Aufgaben werden erkannt	,767	,356	,078
2	Technikbezogene Aufgabenfelder werden benannt	,675	,048	,045
3	Konsequenzen für den Arbeitsprozess werden erkannt	,755	-,084	,190
4	Arbeitsprozess wird organisiert	,720	,399	,227
5	Die richtige Methode zur Bearbeitung wird ausgewählt	,744	,432	,144
6	Fachsprache wird richtig genutzt	,807	,270	,055
7	Aufgabenbearbeitung ist transparent	,802	,391	,169
8	Orientierung an Vorgaben wird beachtet	,479	,625	-,095
9	Umweltverträglichkeit wird sichergestellt	,172	,717	-,017
10	Gesetze usw. werden eingehalten	,297	,541	-,230
11	Sicherheit des Produkts wird gewährleistet	,264	,765	,050
12	Auftragserweiterungen werden bedacht	-,087	,660	,335
13	Kunde wird erfasst	,135	-,138	,775

12 Zahlreiche Kriterien zur Güte der Faktorenanalyse ergeben vielversprechende Werte. Laut Backhaus/Erichson/Plinke/Weiber (2003, S. 275) sind die ausgewählten Variablen dann für eine Faktorenanalyse ungeeignet, wenn mehr als 25 % der nicht diagonalen Elemente in der Anti-Image-Matrix größer 0,09 sind. Das trifft auf 8 der 66 Elemente und damit auf nur rund 12 % zu. Die Daten sind gemäß Anti-Image-Verfahren damit als geeignet anzusehen. Gleiches gilt für die Prüfung anhand des Kaiser-Meyer-Olkin-Kriteriums, welches anzeigt, in welchem Umfang sich die Ausgangsvariablen gruppieren lassen. Das Kriterium, welches als das wichtigste Prüfkriterium für die Güte einer Korrelationsmatrix für eine Faktorenanalyse gilt, ist mit KMK = 0,915 als hervorragend (marvelous) zu bezeichnen. Dies spiegelt sich ebenfalls in guten Werten des „measure of sampling adequacy (MSA)“, das auf Basis der Anti-Image-Korrelationsmatrix für jedes einzelne Item berechnet wird: Zwei der 12 einbezogenen Variablen erreichen einen MSA-Wert größer 0,9 (erstaunlich), acht sind größer 0,8 (verdienstvoll) und zwei immerhin mittelmäßig (> 0,6). Weiterhin sprechen die Kommunalitäten der Faktorenanalyse für deren Aussagekraft. Diese bezeichnen den Umfang an Varianzerklärung, den die Faktoren gemeinsam für eine Ausgangsvariable liefern (ebd. 2003, S. 266). Je nach Variable liegen diese zwischen 46 und 82 %. Damit werden durch die extrahierten Faktoren nach einer Varimax-Rotation knapp 2/3 (62,6 %) der ursprünglichen Gesamtvarianz aufgeklärt.

13 Die Indikatoren wurden mithilfe von Verben in Form von Handlungen formuliert, weil allein das Handlungsresultat in Form eines technischen Artefakts nur wenig Aufschluss darüber gibt, ob gute Facharbeit geleistet wurde oder nicht. Ein funktionierendes Kundenfahrzeug sagt alleine nichts darüber aus, ob z. B. die betrieblichen Interessen eines effizienten Ressourceneinsatzes berücksichtigt wurden. Obwohl Verben wie „erkennen“ oder „benennen“ eine Nähe zur Bloom’schen Taxonomie nahe legen, die eindeutig auf der Ebene der internen Bedingungen angesiedelt sind, ließen sie sich wegen der nötigen Abstraktion nicht vermeiden. Bei der Raterinstruktion wurde jedoch intensiv darauf hingewiesen, dass beobachtbare Handlungsresultate oder Verhalten bewertet werden sollten.

14	Kundenfeedback wird eingeholt	,436	,023	,600
15	Maßnahmen der Kundenpflege werden angewandt	,053	,156	,720

organisieren sind (Items 1 und 4), geht die für eine gute Facharbeit notwendige Berücksichtigung der Sicherheit und Umweltverträglichkeit des Produkts und der zu erbringenden Dienstleistung über die offensichtliche Kernaufgabe hinaus (Items 9 und 11). Diesbezüglich gab es weder in der ersten noch in der zweiten Aufgabe explizite Bearbeitungshinweise. Der dritte Faktor bezieht sich inhaltlich sehr konkret auf Angaben, die hinsichtlich des Kunden gemacht werden. Dementsprechend wurden die drei Faktoren als „unmittelbare Aufgabenbewältigung", „erweiterte Aufgabenbewältigung" und „Kundenorientierung" inhaltlich beschrieben.

Vergleicht man nun die verschiedenen Statusgruppen BGJ-Teilnehmer, Auszubildende und Facharbeiter hinsichtlich der Faktorenwerte[14], so ergeben sich große Unterschiede, die sich entlang der drei Faktoren verschieden stark ausprägen (siehe Abbildung 5).

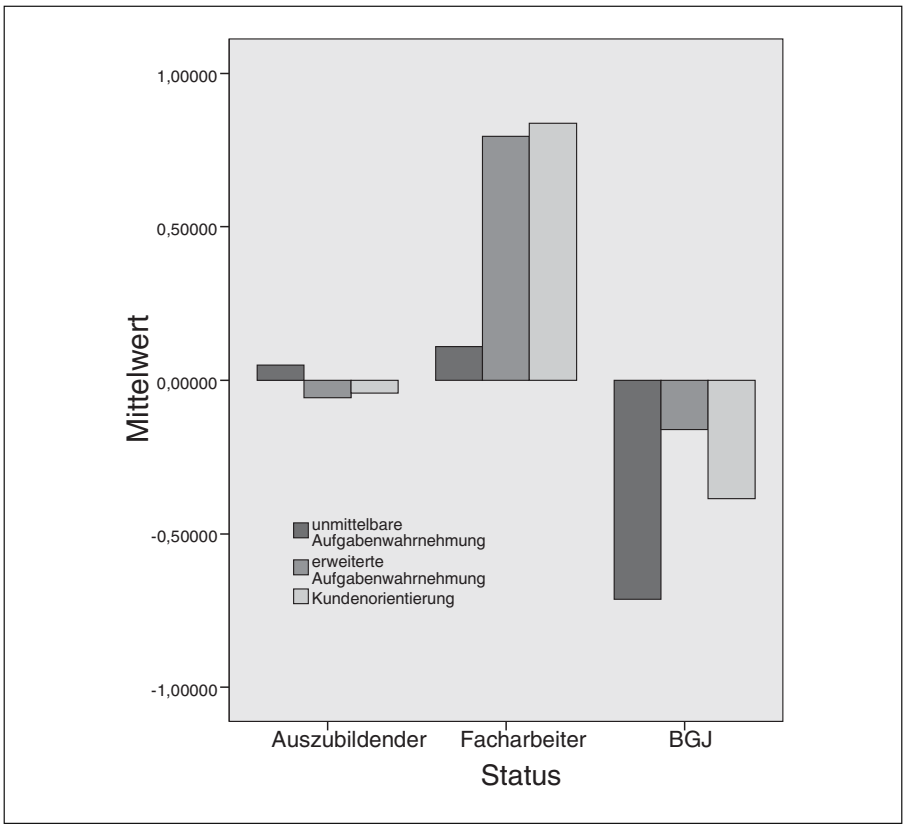

Abb. 5: Vergleich der z-transformierten Faktorwerte nach Statusgruppen (dreifaktorielle Lösung)

14 Faktorwerte werden regressionsanalytisch aus den Antworten der Probanden auf die Einzelitems ermittelt und repräsentieren die Ausprägung der jeweiligen Probanden bezüglich des Faktors.

Erwartungskonform weisen die Facharbeiter tendenziell die höchsten und die BGJ–Teilnehmer die niedrigsten Faktorwerte auf, während die Auszubildenden zum Kfz–Sm und Kfz–M zwischen diesen beiden Gruppen liegen. Deutlich wird, dass der starke Unterschied zwischen dual ausgebildeten und außerbetrieblichen Auszubildenden insbesondere auf den Faktor der unmittelbaren Aufgabenbewältigung zurückgeht, obwohl beide Gruppen eine identische Ausbildungsdauer aufweisen. Andererseits unterscheiden sich die beiden Gruppen mit Erfahrungen im Betrieb (Facharbeiter und Auszubildende) nicht so sehr hinsichtlich dieses ersten Faktors. Langjährige Facharbeiter zeichnen sich gegenüber jungen betrieblichen Auszubildenden jedoch dadurch aus, dass sie über die eigentliche Kernaufgabe hinaus auch die erweiterten Punkte guter Facharbeit mit in den Blick nehmen. Außerdem sind sie sehr viel stärker kundenorientiert als der Nachwuchs, der bisher sein Augenmerk noch sehr stark auf die Technik legt.

Für die zweite Analyse anhand der aus experimenteller Sicht sinnvolleren Stichprobe der drei Statusgruppen in Bezug auf *eine* Aufgabe ließ sich dieses Ergebnis in seinen Grundzügen replizieren.

Aufgrund geringerer Eigenwerte ergab sich jedoch eine zweifaktorielle Lösung, in der sich die zwei ersten Faktoren aus der vorangegangenen Lösung zu einem vereinen, der als „technisch-organisatorische Aufgabenbewältigung" überschrieben werden kann. Der Faktor „Kundenorientierung" blieb jedoch erhalten, wenngleich das Item „Konsequenzen für den Arbeitsprozess werden erkannt" wenig konsistent auch auf diesem Faktor lädt (siehe Tabelle 3).

Tab. 3: Rotierte Komponentenmatrix der zweifaktoriellen Lösung (Extraktionsmethode: Hauptkomponentenanalyse, Rotationsmethode: Varimax mit Kaiser-Normalisierung, Varianzaufklärung: 65 %, die Rotation ist in 3 Iterationen konvergiert)

		Komponente	
		1	**2**
1	Kunde wird erfasst	-,065	,759
2	Auftrag wird geklärt	,156	,723
3	Kundenfeedback wird eingeholt	,266	,734
4	Ergebnis entspricht Kundenanforderungen	,459	,689
5	Konsequenzen für den Arbeitsprozess werden erkannt	,370	,612
6	Aufgaben werden erkannt	,771	,323
7	Arbeitsprozess wird organisiert	,807	,307
8	Orientierung an Vorgaben wird beachtet	,762	,073
9	Die richtige Methode zur Bearbeitung wird ausgewählt	,855	,284
10	Vorgehensweise ist fachlich korrekt	,710	,358
11	Sicherheit des Produkts wird gewährleistet	,718	-,076
12	Fachsprache wird richtig genutzt	,777	,293
13	Aufgabenbearbeitung ist transparent	,870	,317

Damit ist in dieser Zweifaktorenlösung die inhaltliche Beschreibung nicht mehr so klar wie in der ersten Analyse. Dennoch zeigte sich auch hier die prinzipielle Differenz zwischen den drei Gruppen in erwarteter Richtung, und wiederum sind die Unterschiede zwischen allen Gruppen jeweils faktorenspezifisch ausgeprägt (siehe Abbildung 6).

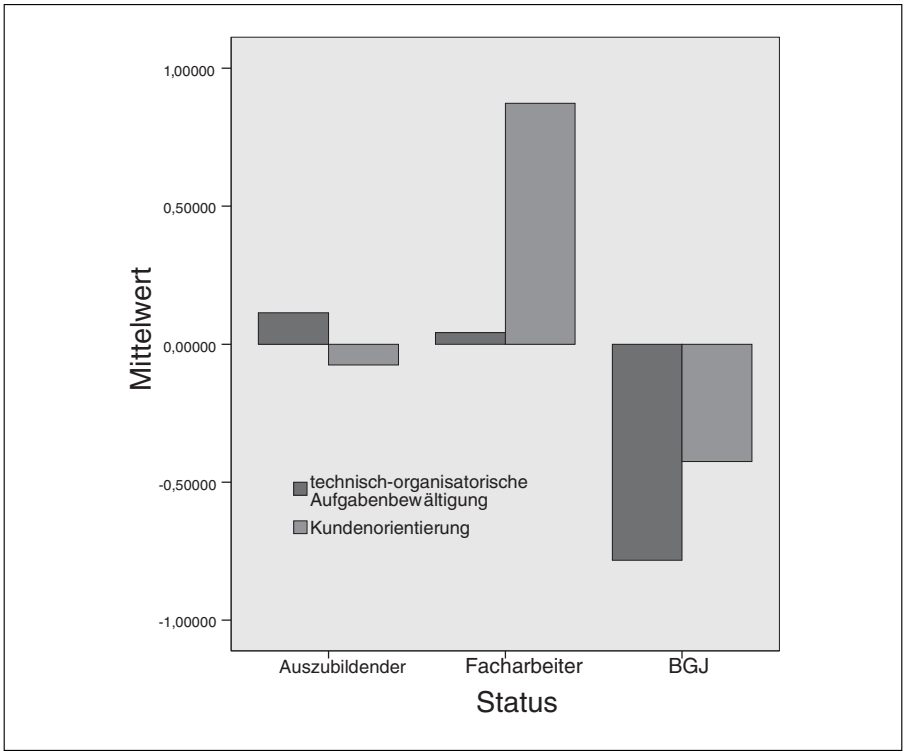

Abb. 6: Vergleich der z-transformierten Faktorwerte nach Statusgruppen (zweifaktorielle Lösung)

Die Ergebnisse beider Faktorenanalysen lassen sich hinsichtlich der eingangs erläuterten Aspekte Konstruktunvollständigkeit und offene Testformate interpretieren, was im folgenden letzten Abschnitt geschehen soll.

5. Schlussfolgerungen und Ausblick

Ausgangspunkt des vorliegenden Artikels war die Feststellung, dass es zurzeit noch wenig elaborierte Vorstellungen zu Kompetenzen und Kompetenzmodellierungen in beruflichen Domänen gibt, die einer empirischen Prüfung standhalten können. Was genau Kompetenzen ausmacht, um in beruflichen Situationen herausragende Leistungen zu ermöglichen, ist bisher nicht geklärt (ontologischer Aspekt). Will man sich in erster Linie ein Bild vom Wesen der Kompetenzen machen und Kompetenzmessung nicht im

Rahmen anderer Fragestellungen instrumentalisieren (z. B. für Systemvergleiche), so sind explorativ qualitative Forschungsmethoden ein vielversprechender Ansatzpunkt. Über einen Experten-Novizen-Vergleich sollte in Erfahrung gebracht werden, worin sich kompetente von weniger kompetenten Personen unterscheiden. Diese empirisch differenzierenden Faktoren sind möglicherweise besser zur Formulierung von deskriptiven Kompetenzstruktur- und Niveaumodellen geeignet als tendenziell normative Kategorien wie fachbezogene Systematiken, kognitive Prozesse oder Wissensarten.

Durch den Einbezug sehr heterogener Probandengruppen war es möglich, die berufliche Kompetenz im Kfz-Service über ein sehr breites Spektrum in den Fokus zu nehmen und die dort relevanten Aspekte zu identifizieren. Hätte man die Faktoren nur anhand einer Probandengruppe mit BGJ-Teilnehmern und Auszubildenden identifiziert, hätte aus dem Ergebnis geschlossen werden können, dass ein eindimensionaler Test auf Grundlage des Faktors „unmittelbare Aufgabenbewältigung" gut geeignet ist, um die Kompetenzentwicklung von Auszubildenden zu erfassen. Weil der Faktor „unmittelbare Aufgabenbewältigung" aber nicht zwischen Auszubildenden und Facharbeitern unterscheidet (siehe Abbildung 6), scheint dies nur für die frühen Phasen der Kompetenzentwicklung zu gelten. Dies sagt jedoch wenig darüber aus, wie weit die Probanden auf ihrem Weg zu einer Kompetenzqualität sind, die eher für Facharbeiter typisch ist.

Die vorgenommene Untersuchung mithilfe schriftlicher Arbeitsaufträge zeigt auf, dass die Unterschiede in den Lösungsprofilen von Facharbeitern, Auszubildenden und BGJ-Teilnehmern nicht nur quantitativer, sondern vor allem qualitativer Natur sind. Facharbeiter achten bei der Aufgabenbearbeitung nicht so sehr auf *mehr*, sondern auf *andere* Aspekte als Auszubildende. Die besondere Leistung der Erfahrenen macht es scheinbar aus, an fachrelevante Aspekte zu denken, die nicht unmittelbar mit dem Auftrag oder einer Testaufgabe zu tun haben (Faktor „erweiterte Aufgabenbewältigung"). Diese Aspekte stellen an sich oft keine besondere fachliche Herausforderung dar. So zeichnet sich z. B. ein guter Facharbeiter dadurch aus, dass er bei der Abarbeitung eines Auftrages versucht, ein Zusatzgeschäft zu realisieren oder den Kunden an den Betrieb zu binden: „*Die Reifen haben die Verschleißgrenze erreicht. Das wurde dem Kunden mitgeteilt und der Auftrag daraufhin erweitert. Der Kunde wurde beraten, sich einen Satz Sommer- und Winterreifen zuzulegen. Daraufhin wollte der Kunde die Winterreifen aufgezogen haben und die Sommerreifen bei mir einlagern*" (Facharbeiter, Fall 143).

Die Entscheidung, ein offenes Erhebungsformat zu wählen hat sich insofern ausgezahlt, als dass die Aufgabenwahrnehmung und -redefinition tatsächlich eine Rolle bei der Bearbeitung gespielt hat. Bei einer geschlossenen Frage nach derartigen Aspekten sind oft auch Laien mit ihrem Alltagsverständnis von Werkstattalltag oder Betriebswirtschaft in der Lage zu erkennen, dass der Verkauf eines zusätzlichen Satzes Reifen einem guten Betriebsergebnis nicht abträglich ist. Diesen Aspekt aus eigenem Antrieb jedoch mitzudenken, obwohl eine technische Kernaufgabe (wie die Durchführung einer Inspektion) im Mittelpunkt steht, macht anscheinend eine Leistung aus, die (neben anderen Aspekten) den Laien vom Experten unterscheidet.

Andererseits besteht ein großer Nachteil der offenen Erhebungsform darin, dass die Probanden immer nur einen Ausschnitt dessen mitteilen, was sie in einer tatsächlichen Situation unternehmen würden. Dabei wird offensichtlich keine besondere Tiefe erreicht, wie sich aufgrund des ersten Faktors der „unmittelbaren Aufgabenbewältigung" vermuten lässt. Auch dort wäre anzunehmen, dass ein Fach-

arbeiter eine bessere Leistung zeigt als ein Auszubildender, jedoch kommt dies in den Aufgabenlösungen nicht zum Ausdruck, weil anscheinend ein Maß an Beschreibungstiefe nicht überschritten wird, welches auch die Auszubildenden erreichen.

An dieser Stelle soll noch einmal der explorative Charakter der vorgenommen Überlegungen hervorgehoben werden, die keinen umfassenden Kompetenzbegriff für den Kfz-Service fundieren können. Es ist nicht anzunehmen, dass sich das umfassende Konstrukt „Kompetenz" entlang der Faktoren „unmittelbare Aufgabenbearbeitung", „erweiterte Aufgabenbewältigung" und „Kundenorientierung" beschreiben lässt. Ebenso wenig ist es angebracht, eher traditionelle Kategorien wie berufliches Fachwissen aus Kompetenzerhebungen auszuklammern. Offenbar lassen sich aber erfahrene von weniger erfahrenen Facharbeitern gut anhand dieser Faktoren unterscheiden, sofern die Aufgabenstellung durch eine offene Gestaltung Raum für das Einbringen von betrieblicher Erfahrung lässt. Als ein nächster Schritt der weiteren Forschung ist eine Analyse der gefundenen Faktoren mithilfe hypothesentestender konfirmatorischer Faktorenanalysen geboten, um die gefundenen Unterschiede zwischen verschieden erfahrenen Kfz-Facharbeitern einer Prüfung zu unterziehen. Eine solche Prüfung bedarf jedoch einer erneuten – im Sinne dieser Fragestellung – systematischeren Datenerhebung sowie einer verstärkten methodischen Kontrolle von Ratereffekten sowie Effekten, die auf die Lesekompetenz der Probanden zurückzuführen sind.

Eine Verstärkung der qualitativen Forschung mithilfe von Experten-Novizen-Vergleichen ist eine vielversprechende Strategie, Hypothesen zur Natur und Genese von Kompetenzen zu entwickeln. Diese könnten dann in die Entwicklung von neuen Items zur large scale Erhebung von Kompetenzen auf der Grundlage kontinuierlicher Item-Response-Modelle eingehen. Gleichzeitig kann ebenso mithilfe Raschskalierter Tests zu beruflichen Kompetenzen oder zu beruflichem Fachwissen der Frage qualitativer Kompetenzveränderungen im Entwicklungsverlauf nachgegangen werden, indem sie Probandengruppen verschiedener Kompetenzniveaus vorgelegt werden (bestimmt z. B. anhand des Merkmals „Berufserfahrung"). Bleiben die Tests trotz heterogener Stichprobe Rasch-skalierbar, spricht dies für eine kontinuierliche Entwicklung des betrachteten Konstrukts. Funktionieren die Items im Test je nach Statusgruppe unterschiedlich (DIF) muss von qualitativen Kompetenzveränderungen im Entwicklungsverlauf ausgegangen werden, die es genauer zu untersuchen gilt.

6. Literaturverzeichnis

Achtenhagen, Frank/Winther, Esther (2009). Konstruktvalidität von Simulationsaufgaben: Computergestützte Messung berufsfachlicher Kompetenz – am Beispiel der Ausbildung von Industriekaufleuten. Göttingen. Online verfügbar: http://www.bmbf.de/pub/Endbericht_BMBF09.pdf [04.11.2009].

Baethge, Martin/Achtenhagen, Frank/Arends, Lena/Babic, Edwin/Baethge-Kinsky, Volker (2006). Berufsbildungs-PISA. Machbarkeitsstudie. Stuttgart: Steiner.

Baumert, Jürgen (2001a). OECD PISA: Programme for International Student Assessment. Schülerleistungen im internationalen Vergleich. Berlin.

Baumert, Jürgen (2001b). PISA – Programme for International Student Assessment. Zielsetzung, theoretische Konzeption und Entwicklung von Messverfahren. In: Weinert, Franz Emanuel (Hrsg.): Leistungsmessungen in Schulen. Weinheim u a.: Beltz-Pädagogik, S. 219–235.

Baumert, Jürgen/Bos, Wilfried/Lehmann, Rainer (2000). TIMSS/III: Dritte internationale Mathematik- und Naturwissenschaftsstudie; mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Opladen: Leske + Budrich.

Baumert, Jürgen/Lehmann, Rainer H./Lehrke, Manfred (1997). TIMSS – mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich: deskriptive Befunde. Opladen: Leske + Budrich.

Beaton, Albert E./Allen, Nancy L. (1992). Interpreting Scales Through Scale Anchoring. In: Journal of Educational Statistics, Jg. 17, H. 2, S. 191–204.

Becker, Matthias (2009a). Kompetenzmodell zur Erfassung beruflicher Kompetenz im Berufsfeld Fahrzeugtechnik. In: Fenzl, Claudia; Spöttl, Georg; Howe, Falk; Becker, Matthias (Hrsg.): Berufsarbeit von morgen in gewerblich-technischen Domänen – Forschungsansätze und Ausbildungskonzepte für die berufliche Bildung. Bielefeld: Bertelsmann, S. 239–245.

Becker, Matthias (2009b). Wie lässt sich das in Domänen verborgene „Facharbeiterwissen" erschließen. In: Becker, Matthias; Fischer, Martin; Spöttl, Georg (Hrsg.): Von der Arbeitsanalyse zur Diagnose beruflicher Kompetenzen. Frankfurt am Main [u. a.]: Peter Lang, S. 54–65.

Bloom, Benjamin S./Engelhart, Max D./Furst, Edward J./Hill, Walker H./Krathwohl, David R. (1973). Taxonomie von Lernzielen im kognitiven Bereich. Weinheim: Beltz.

Borsboom, Denny (2005). Measuring the Mind. Conceptual issues in modern psychometrics. Cambridge: Cambridge University Press.

Glaser, Robert (1994). Instructional Technology and the Measurement of Learning Outcomes: Some Questions. In: Educational measurement: issues and practice, Jg. 13, H. 4, S. 6–8.

Gruber, Hans (1994). Expertise. Modelle und empirische Untersuchungen. Opladen: Westdt. Verl.

Gruber, Hans/Ziegler, Albert (1996). Expertiseforschung. Theoretische und methodische Grundlagen. Opladen: Westdt. Verl.

Gschwendtner, Tobias (2008). Ein Kompetenzmodell für die Kraftfahrzeugtechnische Grundbildung. In: Nickolaus, Reinhold; Schanz, Heinrich (Hrsg.): Didaktik der gewerblich-technischen Berufsbildung. Konzeptionelle Entwürfe und empirische Befunde. Baltmannsweiler: Schneider-Verl. Hohengehren, S. 103–119.

Hagemeister, Volker (1999). Was wurde bei TIMSS erhoben? Rückfragen an eine standardisierte Form der Leistungsmessung. In: Die Deutsche Schule, Jg. 91, S. 160–177.

Hanf, Georg (2007). Der Europäische Qualifikationsrahmen – Ziele, Gestalt, Verfahren. In: Grollmann, Philipp; Spöttl, Georg; Rauner, Felix (Hrsg.): Europäisierung beruflicher Bildung – eine Gestaltungsaufgabe. Hamburg: Lit-Verl., S. 53–64.

Hartig, Johannes (2007). Skalierung und Definition von Kompetenzniveaus. In: Beck, Bärbel; Klieme, Eckhard (Hrsg.): Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International). Weinheim: Beltz, S. 79–95.

Hartig, Johannes/Klieme, Eckhard (2006). Kompetenz und Kompetenzdiagnostik. In: Schweizer, K. (Hrsg.): DIPF-Arbeitseinheit: Bildungsqualität und Evaluation. Berlin: Deutsches Institut für Internationale Pädagogische Forschung; Springer, S. 127–143.

Hofmeister, Wiebke (2005). Erläuterung der Klassifikationsmatrix zum ULME-Kompetenzstufenmodell. In: bwp@ (Berufs-Wirtschaftspädagogik – online), H. 1618–8543, S. 1–21. Online verfügbar: http://www.bwpat.de/ausgabe8/hofmeister_bwpat8.pdf [11.11.2008].

Hume, David (1739). A treatise of human nature. Herausgegeben von L.A. Selby-Bigge und P.H. Nidditch. Oxford: Clarendon Press, 1978.

Klieme, Eckhard/Avenarius, Werner/Blum, Werner/Döbrich, Peter/Gruber, Hans/Prenzel, Manfred/Reiss, Kristina/Riquarts, Kurt/Rost, Jürgen/Tenorth, Heinz-Elmar/Vollmer, Helmut J. (2007). Zur Entwicklung nationaler Bildungsstandards. Herausgegeben von Bundesministerium für Bildung und Forschung. Bonn, Berlin. Online verfügbar: http://www.bmbf.de/pub/zur_entwicklung_nationaler_bildungsstandards.pdf [14.11.2008].

KMK (2007). Handreichungen für die Erarbeitung von Rahmenlehrplänen der Kultusministerkonferenz (KMK) für den berufsbezogenen Unterricht in der Berufsschule und ihre Abstimmung mit Ausbildungsordnungen des Bundes für anerkannte Ausbildungsberufe. Herausgegeben von Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Bonn.

Lazarsfeld, P. F./Henry, N. W. (1968). Latent Structure Analysis. Boston: Houghton Mifflin Company.

Lehmann, Rainer/Seeber, Susan (2007). ULME III. Untersuchung von Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in den Abschlussklassen der Berufsschulen. Hamburg.

Meyerhöfer, Wolfram (2005). Tests im Test: das Beispiel PISA. Opladen: Budrich.

Minnameier, Gerhard (2000). Entwicklung und Lernen – kontinuierlich oder diskontinuierlich. Grundlagen einer Theorie der Genese komplexer kognitiver Strukturen. Münster: Waxmann.

Musekamp, Frank (2007). Der Kfz-Servicemechaniker: Verkürzte Ausbildung als Weg aus der Benachteiligung? Hintergrund, Forschungsdesign und erste Ergebnisse einer Evaluation. In: Münk, Dieter; van Buer, Jürgen; Breuer, Klaus; Deißinger, Thomas (Hrsg.): Hundert Jahre kaufmännische Ausbildung in Berlin. Opladen u. a.: Verlag Barbara Budrich, S. 224–234.

Musekamp, Frank (2009). Entwicklung eines standardisierten Instruments zur Kompetenzerhebung im Kfz-Service. In: Fenzl, Claudia; Spöttl, Georg; Howe, Falk; Becker, Matthias (Hrsg.): Berufsarbeit von morgen in gewerblich-technischen Domänen – Forschungsansätze und Ausbildungskonzepte für die berufliche Bildung. Bielefeld: Bertelsmann.

Musekamp, Frank/Becker, Matthias (2008). Ein zweijähriger Ausbildungsberuf zwischen Integrationsanspruch und Qualifikationsbedarf Ergebnisse aus der Evaluation des Ausbildungsberufes Kfz-Servicemechaniker. In: Berufsbildung in Wissenschaft und Praxis, H. 5, S. 43–47.

Musekamp, Frank/Spöttl, Georg/Becker, Matthias (2007). Synthesebericht. Bilanz der Erprobung des Ausbildungsberufes Kfz-Servicemechaniker und Handlungsszenarien. Unter Mitarbeit von Nils Petermann, Hermann Hitz und Klaus Isermann. Institut Technik und Bildung. Bremen.

Nickolaus, Reinhold (2008). Modellierungen zur beruflichen Fachkompetenz und ihre empirische Prüfung. In: Zeitschrift für Berufs- und Wirtschaftspädagogik, 104. Band, H. 1, S. 1–6.

Nickolaus, Reinhold/Gschwendtner, Tobias/Abele, Stephan (2009). Die Validität von Simulationsaufgaben am Beispiel der Diagnosekompetenz von Kfz-Mechatronikern. Vorstudie zur Validität von Simulationsaufgaben im Rahmen eines VET-LSA. Stuttgart. Online verfügbar: http://www.bmbf.de/pub/Abschluss-Bericht_Druckfassung.pdf [03.11.2009].

Nickolaus, Reinhold/Gschwendtner, Tobias/Geissel, Bernd (2008). Entwicklung und Modellierung beruflicher Fachkompetenz in der gewerblichtechnischen Grundbildung. In: Zeitschrift für Berufs- und Wirtschaftspädagogik, 104. Band, H. 1, S. 48–73.

Nickolaus, Reinhold/Gschwendtner, Tobias/Knöll, Bernd (2006). Handlungsorientierte Unterrichtskonzepte als Schlüssel zur Bewältigung problemhafter Aufgaben. In: Minnameier, Gerhard Beck Klaus (Hrsg.): Berufs- und wirtschaftspädagogische Grundlagenforschung: Lehr-Lern-Prozesse und Kompetenzdiagnostik; Festschrift für Klaus Beck. Frankfurt am Main u. a.: Lang, S. 209–224.

Rost, Jürgen (1990). Rasch models in latent classes: An integration of two approaches to item analysis. In: Applied Psychological Measurement, Jg. 14 , H. 3, S. 271–282.

Rost, Jürgen (2004). Lehrbuch Testtheorie – Testkonstruktion. Psychologie Lehrbuch. Bern u. a.: Huber.

Schecker, Horst/Parchmann, Ilka (2006). Modellierung naturwissenschaftlicher Kompetenz. In: Zeitschrift für Didaktik der Naturwissenschaften, H. 12, S. 45–66. Online verfügbar: http://www.ipn.uni-kiel.de/zfdn/pdf/003_12.pdf [14.11.2008].

Seeber, Susan (2005). Zur Erfassung und Vermittlung berufsbezogener Kompetenzen im teilqualifizierenden Bildungsgang „Wirtschaft und Verwaltung" an Hamburger Berufsschulen. In: bwp@ (Berufs-Wirtschaftspädagogik – online), Jg. 8, S. 1–16. Online verfügbar: http://www.bwpat.de/ausgabe8/seeber_%20bwpat8.pdf [14.11.2008].

Seeber, Susan (2008). Ansätze zur Modellierung beruflicher Fachkompetenz in kaufmännischen Ausbildungsberufen. In: Zeitschrift für Berufs- und Wirtschaftspädagogik, 104. Band, H. 1, S. 74–97.

Spöttl, Georg (2009). Kompetenzmodelle in der beruflichen Bildung – Grenzen und Chancen. In: Fenzl, Claudia; Spöttl, Georg; Howe, Falk; Becker, Matthias (Hrsg.): Berufsarbeit von morgen in gewerblich-technischen Domänen – Forschungsansätze und Ausbildungskonzepte für die berufliche Bildung. Bielefeld: Bertelsmann, S. 233–238.

Spöttl, Georg/Musekamp, Frank (2009). Berufsstrukturen und Messen beruflicher Kompetenz. In: berufsbildung, Jg. 63, H. 118, S. 20–23.

Straka, Gerald A./Macke, Gerd (2008). Handlungskompetenz – und wo bleibt die Sachstruktur. In: Zeitschrift für Berufs- und Wirtschaftspädagogik, 104. Band, S. 590–600 [11.11.2008].

Tolman, Charles W. (1997). Erklären, Verstehen und die Kritische Psychologie. In: Forum Kritische Psychologie, Jg. 38, S. 147–157.

Verhelst, N. D./Glas, C. A. W./de Vries, H. H. (1997). A Steps Model to Analyze Partial Credit. In: W. J. d. van Linden (Hrsg.): Handbook of modern item response theory. New York, NY: Springer, S. 123–138.

Wilson, M. R. (1989). Saltus: A Psychometric Model of Discontinuity in Cognitive Development. In: Psychological Bulletin, Jg. 105, H. 2, S. 276–289.

Wilson, Mark R. (2005). Constructing measures. An item response modeling approach. Mahwah, NJ: Erlbaum.

Winther, Esther/Achtenhagen, Frank (2008). Komptenzstrukturmodell für die kaufmännische Bildung. Adaptierbare Forschungslinien und theoretische Ausgestaltung. In: Zeitschrift für Berufs- und Wirtschaftspädagogik, 104. Band, H. 4, S. 511–538.

Wright, Benjamin D./Masters, Geoff. N. (1982). Rating Scale Analysis. Rasch measurement. Chicago, Mesa Pr.

Wu, Margaret/Adams, Raymond L. (2007). Applying the Rasch-Model to Psycho-Social Measurement. A Practical Approach. Melbourne: Educational Measurement Solutions.

Anschrift der Autoren:
Frank Musekamp, Wissenschaftlicher Mitarbeiter am ITB, Universität Bremen, musekamp@uni-bremen.de
Georg Spöttl, Prof. Dr., Direktor des ITB, Universität Bremen, spoettl@uni-bremen.de
Matthias Becker, Prof. Dr., biat, Universität Flensburg, becker@uni-flensburg.de