

Viola Katharina KLOTZ & Esther WINTHER
(Universität Paderborn)

Kompetenzmessung in der kaufmännischen Berufsausbildung: Zwischen Prozessorientierung und Fachbezug. Eine Analyse der aktuellen Prüfungspraxis

Online unter:

http://www.bwpat.de/ausgabe22/klotz_winther_bwpat22.pdf

in

bwp@ Ausgabe Nr. 22 | Juni 2012

Funktionen und Erträge pädagogischer Diagnostik im wirtschafts- und berufspädagogischen Bereich

Hrsg. v. Ausgabe 22: **Tade Tramm, Susan Seeber & H.-Hugo Kremer**

<http://www.bwpat.de> | ISSN 1618-8543 | *bwp@* 2001–2012

bwp@

www.bwpat.de

Herausgeber von *bwp@* : Karin Büchter, Franz Gramlinger, H.-Hugo Kremer und Tade Tramm

Berufs- und Wirtschaftspädagogik - *online*

Online: www.bwpat.de/ausgabe22/klotz_winther_bwpat22.pdf

Empirische Modellierung und Messbarkeit von Handlungskompetenz bilden zentrale Inhalte des derzeitigen bildungspolitischen Diskurses. In den vergangenen Jahren wurden hierzu in unterschiedlichen beruflichen Domänen Modelle konstruiert und auf empirische Belastbarkeit geprüft. Der Beitrag schließt an diese Arbeiten an und überprüft die Validität und Reliabilität der kaufmännischen Abschlussprüfungen (IHK) anhand von $n = 1768$ Industriekaufleuten. Das IHK-Testinstrument folgt dem curricular formulierten Anspruch, Handlungskompetenz von Auszubildenden geschäftsprozessorientiert durch eine Einteilung der Teststruktur in die prozessualen Dimensionen Planung, Durchführung und Kontrolle (neben der klassischen Einteilung in fachliche Domänen) erfassen zu können. Diskutiert werden in diesem Zusammenhang die folgenden Fragestellungen: (1) Entspricht das dem Test zugrunde liegende Prüfkonstrukt dem Konzept beruflicher Handlungskompetenz? (2) Inwieweit lässt sich das Prüfkonzept empirisch bestätigen? (3) Mit welcher Präzision vermag das Testinstrument Kompetenz zu messen? Die Fragestellungen werden methodisch sowohl anhand Analyseverfahren der klassischen Testtheorie als auch mittels Verfahren der probabilistischen Testtheorie behandelt. Im Ergebnis erweist sich das Modell der Prozessorientierung als faktoriell nicht valide. Stattdessen kann eine 5-Faktor-Lösung generiert werden, die sich eingeschränkt als fachbezogene Domänenlösung und damit als Fachexpertise interpretieren lässt. Bezüglich der Zuverlässigkeit des Testinstrumentes zeigen sich ausreichend hohe Reliabilitäten bei der Erfassung durchschnittlicher Kompetenz bzw. Fachkompetenz, sowie wenig präzise Kompetenzeinschätzungen für tendenziell hohe und niedrige Fachkompetenzausprägungen. Die Befunde werden hinsichtlich ihrer praktischen Relevanz für den nationalen und europäischen Kontext diskutiert.

Measuring competence in vocational training for commercial occupations: Between process-orientation and subject-specialism. An analysis of the current examination practice

Empirical modelling and measurability of action competence are key components of the current educational political discourse. In recent years models were constructed and tested for empirical strength in various different vocational domains. This article follows on from these studies and tests the validity and reliability of the commercial final examinations (IHK) using $n = 1768$ industrial clerks. The IHK test instrument follows the demand formulated in the curriculum that action competence of trainees should be captured in a business process-oriented way through a division of the test structure into the process dimensions of planning, execution and control (alongside the classical divisions into subject domains). In this context, the following questions are discussed: (1) Does the examining construct that is the foundation of the test correspond to the concept of vocational action competence? (2) To what extent is it possible to confirm the examining concept empirically? (3) With what precision is it possible for the test instrument to measure competence? The questions are dealt with using methods including the analytical procedures of classical test theory, as well as

using procedure of probabilistic test theory. The results show that the model of process-orientation is factorially not valid. Instead, a five-factor solution can be generated, which can be interpreted in a restricted way as a subject-specific domain solution and therefore as subject expertise. With regard to the reliability of the test instrument, sufficiently high reliabilities for capturing average competence or subject competence are shown, as well as not very precise competence estimates for subject competence characteristics that tend to be either high or low. The findings are discussed with regard to their practical relevance for the national and European contexts.

Kompetenzmessung in der kaufmännischen Berufsausbildung: Zwischen Prozessorientierung und Fachbezug. Eine Analyse der aktuellen Prüfungspraxis¹

1 Einleitung

Es sind vor allem zwei internationale Leistungsvergleichsstudien im allgemeinbildenden Bereich, 1996 die *TIMS*-Studie (BAUMERT/ LEHMANN 1997) und im Jahr 2000 die *PISA*-Studie (BAUMERT/ KLIEME/ NEUBRAND et al. 2001), die den Impuls für eine breite Diskussion über die Qualität von Schule und Unterricht, auch hinsichtlich der Qualität von Leistungsmessungen und ihrer Aussagekraft gegeben haben. Auf Basis der in diesen Studien gewonnenen Erkenntnisse lässt sich in Deutschland für den allgemeinbildenden Bereich ein Umdenken konstatieren, das als Neuorientierung von einer Input- hin zu einer Outputsteuerung beschrieben werden kann. Ein solcher Paradigmenwechsel hat für den Bereich der beruflichen Bildung in Deutschland allerdings nur bedingt stattgefunden. Zwar wird spätestens seit Veröffentlichung der ECVET-Dokumente von den meisten Autoren auch für die berufliche Bildung eine Hinwendung zur Outputorientierung² im Kontext der Europäisierungsbestrebungen erwartet (BOHLINGER 2006) – beziehungsweise mitunter auch befürchtet (z. B. DREXEL 2005; DREXEL 2008) – belastbare, empirische Befunde zur Qualität des beruflichen Bildungssystems sowie gültige Verfahren der Erfassung von Lernergebnissen sind jedoch defizitär. Hier besteht dringender Forschungsbedarf, bedenkt man die erst kürzlich für Deutschland erzielte Einigung zur Verortung der beruflichen Bildungsabschlüsse innerhalb eines einheitlichen europäischen Qualifikationsrahmens. Für eine solche Verortung müssten ex ante zwei notwendige Voraussetzungen hinsichtlich der Prüfung beruflicher Lernergebnisse erfüllt sein: (1) Klarheit darüber, nach welchen Testschemata Kompetenzen in der Testpraxis der einzelnen Nationen erfasst werden und inwieweit diese auch mit Hinblick auf die empirische Realität die Kompetenzen von Auszubildenden abzubilden vermögen (Validität). Hierbei sei angemerkt, dass nicht notwendigerweise dasselbe Kompetenzverständnis und die gleichen Kompetenzmessverfahren in den einzelnen Nationen bestehen müssten, um eine Vergleichbarkeit der verschiedenen Systeme zu gewährleisten. Gewährleistet sein müsste hingegen zumindest ein gemeinsames Verständnis darüber, hinsichtlich welcher Zielkategorien in den einzelnen Nationen Kompetenzen erfasst werden (vgl. hierzu BAETHGE/ ACHTENHAGEN et al. 2006). Darüber hinaus hat (2) Gewissheit darüber zu bestehen, mit wel-

¹ Der Beitrag ist entstanden im Rahmen des Teilprojekts „Kompetenzorientierte Assessments in der beruflichen Bildung“ (Wi 3597/1-1 sowie Wi 3597/1-2) des DFG-Schwerpunktprogramms „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen (SPP 1293).

² Mit dem Input eines Bildungssystems sind im Allgemeinen Lehrpläne, Lehrmittel, Infrastruktur etc. gemeint. Der Output umfasst hingegen die Lernergebnisse der Schülerinnen und Schüler in Form von Persönlichkeitsmerkmalen und bezieht sich insbesondere auf Messverfahren. (KLIEME/ AVENARIUS/ BLUM et al. 2003, 12).

cher Präzision bzw. Zuverlässigkeit die Individuen in den einzelnen nationalen Systemen den definierten Kategorien zugeordnet werden (Reliabilität). Nur unter der Voraussetzung einer präzisen Erfassung auf nationaler Ebene kann auch eine zuverlässige Erfassung aller Individuen auf europäischer Ebene gelingen. Unter Berücksichtigung dieser Prämissen ist aktuell zu urteilen, dass der europäische Qualifikationsrahmen sich als Einordnungsrahmen schulischer und beruflicher Leistungsdispositionen sowie der entsprechenden Abschlüsse bzw. Zertifikate zum Zwecke höherer Transparenz und Vergleichbarkeit nicht adäquat eignet. Die Forschungsdesiderate lassen sich vorrangig in der Operationalisierung der Fähigkeits- und Kompetenzfacetten verorten.

Hier schließt der folgende Beitrag an. Untersucht werden auf Basis der kaufmännischen Abschlussprüfungen für Industriekaufleute mögliche Kompetenzmodellierungen, die zu einer präziseren Bewertung der in Bildungsgängen vermittelten und erworbenen Leistungsfähigkeiten beitragen können. Die Datenbasis bilden hierbei $n = 1768$ Abschlussklausuren des Prüfungsbereichs „Geschäftsprozesse“, denen eine besondere Relevanz für die Erfassung und Bewertung von Kompetenzen der Auszubildenden zugemessen werden kann: Zum einen soll über die Abschlussprüfungen das Ausbildungsziel – nämlich der Erwerb beruflicher Handlungsfähigkeit – festgestellt werden; zum anderen sind die Prüfungsanforderungen bundeseinheitlich geregelt und in den Ausbildungsordnungen verankert (vgl. LORIG/ SCHREIBER 2007, 6). Für die Berufsbildung in kaufmännischen Gewerbeberufen ist die Industrie- und Handelskammer (IHK) zuständige Stelle im Sinne des Gesetzes. Sie legt die Ausbildungsinhalte der kaufmännischen Ausbildungsberufe sowie deren Prüfung u.a. durch die AKA (Aufgabenstelle für kaufmännische Abschlussprüfungen) mit Bezug auf das BBiG fest.

2 Das Strukturmodell beruflicher Handlungskompetenz in den kaufmännischen Zwischen- und Abschlussprüfungen

Das Konzept der beruflichen Handlungskompetenz, das die berufs- und wirtschaftspädagogischen Diskussionen bereits seit Mitte der 1980er Jahre (vgl. HACKER 1986; VOLPERT 1983; REETZ 1989a; REETZ 1989b) prägt, muss inzwischen als zentrales Konstituierungsmoment beruflicher Bildung gelten. So griffen Ausbildungsordnungen und Rahmenlehrpläne durch den Beschluss der Kultusministerkonferenz (KMK) berufliche Handlungskompetenz bereits Mitte der 1990er Jahre als oberste Zielkategorie auf. Auch laut BBiG sollen Lerninhalte so vermittelt werden, dass die Auszubildenden im Sinne des § 1 Abs. 2 zum selbständigen Planen, Durchführen und Kontrollieren befähigt werden – hiermit wird auf einen vollständigen Handlungszyklus im betrieblichen Gesamtzusammenhang abgestellt. Da die Ausbildungsordnung gemäß BBiG §5 als Grundlage der Abschlussprüfungen zu dienen hat, steht das primär heuristisch zu sehende Konzept der Handlungskompetenz auch im Zentrum der aktuellen beruflichen Kompetenzmessung. Bei der Aufgabebearbeitung werden entsprechend ein ganzheitliches Denken sowie die Einbeziehung vorausgehender und sich anschließender Prozesse von der Testperson verlangt (vgl. EBBINGHAUS 2005, 2f.). Der Aspekt der *Ganzheitlichkeit* bedingt die Bewältigung berufstypischer Aufgaben in der Breite der sachlichen, humanen und sozialen Anforderungen. Der *Vollständigkeitsanspruch* dagegen bindet

das selbstständige Handeln an den umfassenden Handlungsprozess, der über den Dreischritt Planung, Durchführung und Kontrolle strukturiert werden kann (vgl. BERTELSMANN STIFTUNG 2009, 39). Handlungskompetenz bezieht sich dabei anders als der fachspezifische Kompetenzansatz nicht auf Themen- bzw. Fachbereiche, sondern auf *authentische* Handlungssituationen, die sich durch situative Echtheit bzw. Glaubwürdigkeit auszeichnen. Reale Situationen müssen dementsprechend so aufbereitet werden (Schaffung einer authentischen Lernsituation), dass sie eine intendierte Botschaft/Wirkung (Lernziel) transponieren (vgl. ACHTENHAGEN/ WINTHER 2009, 10; WINTHER 2010). Analog dazu müssen reale Situationen auch für Assessmentsituationen authentisch aufbereitet werden, um das Konstrukt der Handlungskompetenz und seine einzelnen Facetten valide zu erfassen. Die kaufmännischen Abschlussprüfungen werden dem komplexen Gegenstand beruflicher Handlungskompetenz und einer validen Prüfung im betrieblichen Gesamtzusammenhang folglich nur dann gerecht, wenn sie in ihrer Testlogik der Idee authentisch modellierter Arbeitsrealitäten folgen. Während bislang einige – zum Teil divergierende – Forschungsarbeiten hinsichtlich der inhaltlichen Umsetzung und Vermittlung beruflicher Handlungskompetenz in den Curricula und durch die Lehrkräfte vorliegen (z. B. REINISCH 1999; PÄTZHOLD 2000; KREMER/SLOANE 2001; KREMER 2003), existieren hinsichtlich der Umsetzung des Prinzips der Handlungskompetenz im kaufmännischen Prüfungssystem und im Rahmen formativer sowie summativer Assessments der Leistungsstände kaum Befunde. So konstatiert WEIß (2011, 38) ein gänzlich fehlendes an belastbaren Erkenntnissen zur Qualität des beruflichen Prüfungssystems und SEEBER/ NICKOLAUS/ WINTHER et al. (2010, 2) verweisen auf deutliche Lücken in der theoretisch-konzeptionellen Ausdifferenzierung von Kompetenzen in verschiedenen Domänen oder domänenübergreifenden Bereichen. Diese Forschungsdesiderata legen nahe, die Aussagefähigkeit der aktuell eingesetzten Prüfungsinstrumente zunächst zu evaluieren, um anschließend die Prüfungen gegebenenfalls so zu modifizieren, dass sie den gesetzlichen Anforderungen im Sinne einer objektiven, reliablen und validen Messung von beruflicher Handlungskompetenz entsprechen.

Die vorliegende Untersuchung der IHK-Abschlussprüfung des Ausbildungsberufes Industriekaufmann/-frau des Prüfungsjahrgangs 2008 bezieht sich auf den Prüfungsbereich „Geschäftsprozesse“, der mit 180 Minuten Prüfungsdauer und 40 prozentiger Gewichtung der Abschlussnote der umfangreichste Teilprüfungsbereich der Abschlussprüfung ist. Zudem kann vermutet werden, dass durch das offene Antwortformat sowie durch die in betriebliche Prozesse eingebundene Aufgabenstellung dem Konzept der Handlungskompetenz in besonderer Weise entsprochen werden könnte. Hinzukommt, dass sich von den insgesamt 880 Zeitstunden des Rahmenlehrplans 600 Stunden direkt den Aufgaben des Prüfungsbereichs „Geschäftsprozesse“ zuordnen lassen. Der Bereich umfasst insgesamt sechs zusammenhängende, jedoch unabhängig voneinander lösbare, Aufgaben. Mit einer Gesamtpunktzahl von 38 Punkten (zwei Teilaufgaben) stellt das Aufgabengebiet „Marketing und Absatz“ (A1 und A2) den größten Anteil. Die drei anderen Aufgabengebiete „Beschaffung und Bevorratung“ (A3), „Personal“ (A4) sowie „Leistungserstellung“ (A5 und A6) sind gleich gewichtet. Im Zentrum dieses Prüfungsbereichs stehen Prozesse und komplexe Sachverhalte anhand derer die Lernenden Geschäftsprozesse analysieren sowie Problemlösungen ergebnis- und kundenorien-

tiert entwickeln sollen. Die zur Anwendung kommenden, “situativen” Aufgabentypen sollen praxis-, adressaten-, aktivitäts- und entscheidungsorientiert eine berufliche Situation abbilden und für die Berufsausübung wesentliche Kompetenzen abfragen (Aka 2009).

WINTHER (2011) hat diesen Prüfungsteil erstmalig hinsichtlich der Objektivität und der curricularen Validität untersucht. Dabei konnten systematische Verzerrungen – u. a. verursacht durch ein uneinheitliches Korrekturvorgehen der Lehrkräfte – in den Daten nachgewiesen werden. Zudem wurde auf Basis inhaltsanalytischer Untersuchungen festgehalten, dass der inhaltlichen Dominanz der Leistungserstellung in den Lehrplänen zum Prüfungszeitpunkt 2008 in den Prüfungen nicht adäquat entsprochen worden ist, mit der Folge, dass Fragen der Wertschöpfung und damit des quantifizierbaren Prozesses der Leistungserstellung insgesamt unterrepräsentiert waren. Hinsichtlich des Kriteriums der Konstruktvalidität konnten weder handlungs- noch inhaltspezifische Strukturen identifiziert werden. Diese Befunde machten eine vollständige Neukodierung aller Klausuren notwendig, mit dem Ziel, die Datenbasis einerseits um Verzerrungen durch das uneinheitliche Korrekturvorgehen der Lehrkräfte zu bereinigen und um andererseits eine einheitliche Kodier- und Bewertungsbasis zu schaffen. Hierdurch wurde es im Rahmen des weiteren Forschungsvorgehens möglich auf Grundlage der rekodierten Datenbasis die Konstruktvalidität und zusätzlich die Reliabilität der Abschlussklausuren erneut zu eruieren.

3 Bewertung der Konstruktvalidität kaufmännischer Abschlussprüfungen

Von Konstruktvalidität eines Testinstrumentariums ist dann auszugehen, wenn die postulierten, prozessualen und inhaltlichen Strukturen sinnvoll operationalisiert wurden und sich zudem empirisch in den Testergebnissen widerspiegeln. Im Hinblick auf die kaufmännischen Abschlussprüfungen stellt sich daher zunächst die Frage, inwieweit die verschiedenen Facetten beruflicher Handlungskompetenz (Ganzheitlichkeit, Vollständigkeit und Authentizität) in die Testkonstruktion der Prüfungen Eingang finden. Gemäß der Prüfungsbeschreibung ist in den Abschlussprüfungen die Befähigung zur beruflichen Handlungskompetenz durch das Bearbeiten komplexer, praxisbezogener Aufgaben, die sich auf Ausschnitte betrieblicher Handlungen beziehen, nachzuweisen. Bei genauerer Einsicht in die Klausuraufgaben muss zunächst eine völlige Abwesenheit von Items zur Messung der Sozial- und Humankompetenz konstatiert werden, wodurch der *Aspekt der Ganzheitlichkeit* beruflicher Kompetenz nicht geprüft werden kann. Hinsichtlich des *Authentizitätsaspektes* lässt sich ferner feststellen, dass zwar zu Beginn der Prüfungen eine Rahmensituation im Sinne einer Unternehmensbeschreibung für die Klausurteilnehmer inszeniert wird, an die jedoch bei der Formulierung der eigentlichen Testitems nicht mehr angeknüpft wird. Hierdurch wirkt die Rahmensituation auf die Prüfungsteilnehmer vermutlich eher entrückt und isoliert als authentisch. Dem *Vollständigkeitsanspruch* wird formal entsprochen, indem Handlungskompetenz über die drei Prozessdimensionen Planung, Durchführung und Kontrolle abgebildet werden soll (Aka 2009, 3). Hierzu werden in der Aufgabenbeschreibung jeder einzelnen Prozessdimension entsprechende Items zu deren Messung zugeordnet. Das folgende – zunächst heuristisch auf die Prü-

funktionslogik bezogene – Strukturmodell beschreibt die angenommene Dimensionalität von Kompetenz, d. h. die Struktur, die hinter den Items vermutet wird.

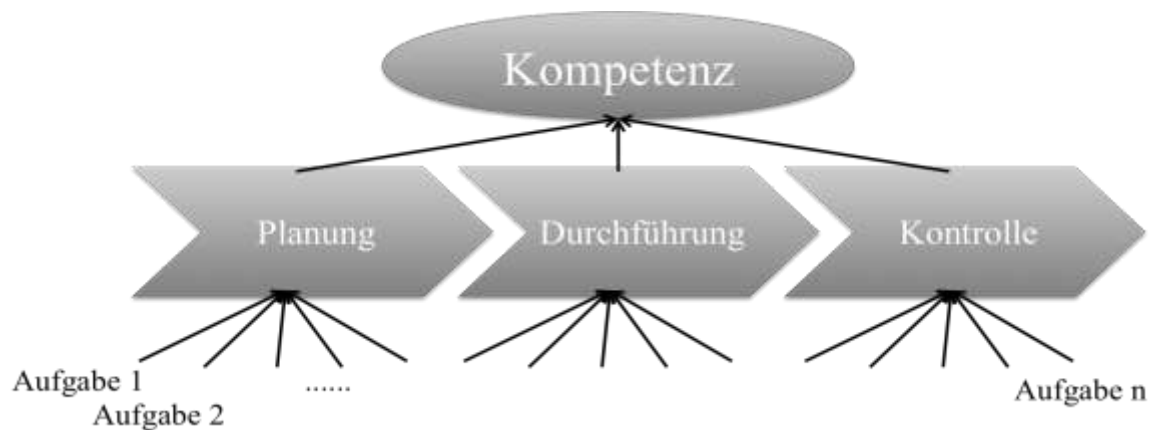


Abb. 1: Prozessorientiertes Kompetenzstrukturmodell der AkA

Während die Facetten Ganzheitlichkeit und Authentizität beruflicher Handlungskompetenz bereits bei inhaltlicher Betrachtung entfallen, könnte sich der Vollständigkeitsanspruch beruflicher Handlungskompetenz empirisch als valide erweisen und damit den in der Prüfungsbeschreibung formulierten Anspruch, berufliche Handlungskompetenz zu messen, partiell rechtfertigen.

Der Datensatz des Testbereiches „Geschäftsprozesse“ besteht aus 1768 Einzeldaten, die an sechs verschiedenen IHK-Standorten³ erhoben wurden. Damit davon ausgegangen werden kann, dass die prozessualen Handlungselemente Planung, Durchführung und Kontrolle einen Kompetenzbereich typischerweise charakterisieren, müssten zur Bewältigung dieser bereichsspezifischen Anforderungssituationen verschiedene Fähigkeitsstrukturen benötigt werden. Statistisch betrifft dies die Frage, ob das Zustandekommen der Korrelationswerte durch einen oder etwa durch mehrere Hintergrundfaktoren bestimmt wird. Oder anders formuliert: Die Frage, ob sich das Antwortverhalten auf ein Testitem (manifeste Indikatorvariablen) auf wenige, dahinterstehende Variablen (latente Kompetenzdimensionen) in Form eines bestimmten Kompetenzstrukturmodells zurückführen lässt. Die methodische Vorgehensweise bei der Überprüfung solcher Messmodelle ist meist faktoranalytisch. Zunehmend werden jedoch auch mehrdimensionale Modelle der Item Response Theory (MIRTs) eingesetzt (HARTIG/HÖHLER 2008), die der Familie probabilistischer Testmodelle entstammen und sich aufgrund ihrer speziellen Eigenschaften zu diesem Zweck besonders eignen.⁴ Für die Analyse

³ Lüneburg, Hannover, Frankfurt, München, Saarland und Nürnberg.

⁴ Solche Modelle gründen auf der Annahme, dass die Lösungswahrscheinlichkeit für ein bestimmtes Item zum einen von der Personenfähigkeit θ_i (Personenparameter) und gleichzeitig von der Itemschwierigkeit δ_i (Itemparameter) abhängt, wobei diese beiden Parameter in proportional negativer Beziehung zueinander stehen und sich dementsprechend als logische Differenz ($\theta_i - \delta_i$) auf eine gemeinsame Skala transformieren lassen (WRIGHT/STONE 1979, 137). Die Betonung probabilistischer Modelle auf der Latenzität des zu messenden Konstrukts (ROST/SPADA 1978, 60), erweist sich für die Intention einer Dimensionsüberprüfung der AkA-Abschlussklausuren als angemessen. Auch die besondere Anwendbarkeit auf kategoriale Datensätze

der Datenbasis wurde ein mehrdimensionales, ordinales Raschmodell gewählt⁵. Mit diesem konnte innerhalb einer Modellevaluation überprüft werden, wie gut das prozessuale Strukturmodell der IHK-Abschlussklausuren die empirische Wirklichkeit widerspiegelt. Numerisch lässt sich dies anhand von Fit-Indizes bestimmen, die in Form von Kennwerten die Anpassung der empirischen Datenbasis an ein theoretisches Modell beschreiben. Im Ergebnis weisen fast alle Kennwerte der globalen Anpassungsgüte auf eine schlechte bis sehr schlechte Anpassung des Datensatzes an das durch die Abschlussprüfungen postulierte, prozessuale Modell 1 hin (RMSEA: 0,054; CFI: 0,782; TLI: 0,867).⁶

Da sich das postulierte Kompetenzstrukturmodell einer Prozessorientierung als nicht valide erwies, wurde im weiteren Forschungsvorgehen explorativ nach den tatsächlich zu Grunde liegenden Kompetenzstrukturen im Datensatz gesucht.⁷ Empirisch werden dabei diejenigen Kompetenzbereiche ausgewiesen, die hinsichtlich ihrer korrelativen Zusammenhänge hinreichend unabhängig voneinander sind (HARTIG/ KLIEME 2006, 132). Analysiert man die Schülerantworten mittels diesen Verfahrens ergibt sich für einen Großteil der Aufgaben folgendes, fachbezogenes Kompetenzstrukturmodell mit den dazugehörigen Faktorladungen:

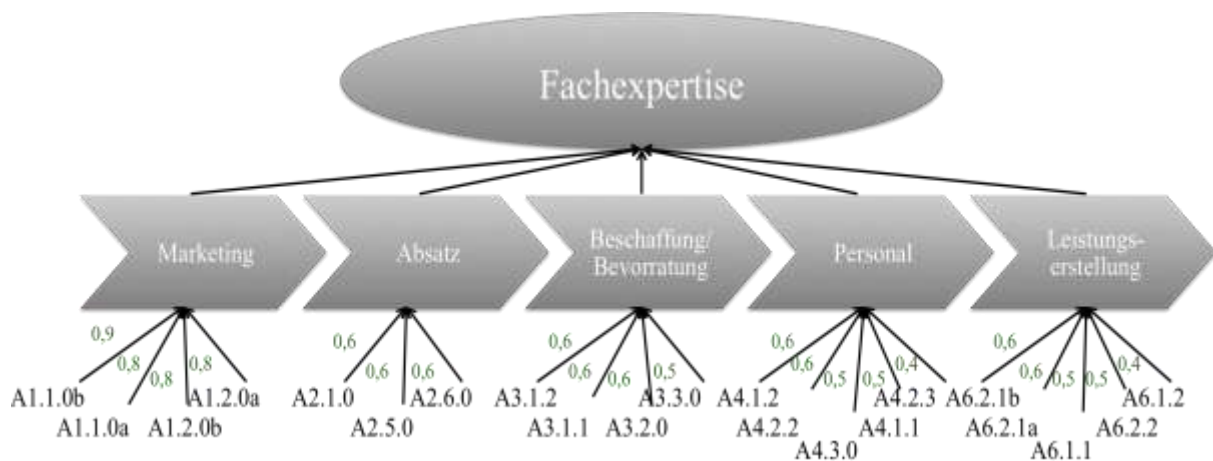


Abb. 2: Fachbezogenes Kompetenzstrukturmodell der Aka

sowie die Robustheit der IRT gegenüber Verletzungen der Normalverteilungsannahme (KIRISCI/ HSU/ YU 2001) prädestiniert den vorliegenden Datensatz geradezu für probabilistische Testmodelle.

⁵ Genauer das Graded Reponse Modell (SAMEJIMA 1969), wobei der Diskriminationsparameter auf 1 fixiert wurde. Dieses bietet sich für die Analyse der Aka-Aufgaben an, da es imstande ist, polytome Items abzubilden, die zudem hinsichtlich der Anzahl ihrer Skalierungsstufen variieren. Für eine vertiefte, formale Herleitung des Graded Response Modells und seiner Parameterschätzung seien an dieser Stelle SAMEJIMA (1969) und LINDEN/ HAMBLEDON (1997) Kapitel 5 (85-97) empfohlen.

⁶ Fitindizes: Root-Mean-Square-Error of Approximation (RMSEA); Comparative-Fit-Index (CFI); Tucker-Lewis-Index (TLI).

⁷ Innerhalb der explorativen Faktorenanalyse wurde eine Hauptkomponentenanalyse mit der obliquen Rotationsmethode Promax verwendet. Zur Bestimmung der Anzahl der zu extrahierenden Faktoren wurde der graphische Scree-Test nach CATTELL (1966) durch eine Gegenüberstellung der Daten mit einer Zufallssimulation ergänzt. Dieses durch HORN (1965) entwickelte Verfahren wurde mittels eines weiteren Programmes (Monte Carlo PA) zur Erstellung der Simulationsdaten für den Datensatz angewandt und vereint die dem Scree-Test eigene Relativbetrachtung mit mehr Objektivität.

Dieses Kompetenzstrukturmodell entspricht weitestgehend einer fachbezogenen Testeinteilung, wobei sich der Bereich „Marketing und Absatz“ empirisch noch einmal in zwei unterschiedliche Inhaltsbereiche („Marketing“ und „Absatz“) aufspaltet. Das aus der explorativen Analyse generierte Modell hielt im Anschluss auch einer Überprüfung mittels des konfirmatorischen IRT-Modells stand, innerhalb derer sich gute Kennwerte für die Modellanpassung ergaben (RMSEA: 0,041; CFI: 0,957; TLI: 0,965). Auf Basis dieser Befunde lässt sich Folgendes zusammenfassen: Für 21 von 35 analysierbaren Items ist ein fachbezogenes Strukturmodell validierbar. Ein Drittel der Items wird damit jedoch nicht empirisch durch die fachbezogene Domänenstruktur repräsentiert. Auffällig ist, dass sich jene Items, welche nicht der Zuordnung entsprechen, durch eine stärkere Transfer- und Fallbezogenheit auszeichnen und häufig in Rechnungen bestehen. Diese Charakteristika scheinen sich so auf die Items auszuwirken, dass sie einer fachlichen Dimension nicht mehr direkt zuordenbar sind, da die angesprochenen, stärker generalisierbaren und situationsbezogenen Fähigkeiten den Fachbezug bei der Aufgabenbearbeitung auflösen. Mit Blick auf die formal über Lernfelder strukturierten Unterrichtsinhalte bleiben vor dem Hintergrund dieser Befunde mehrere Fragen offen: Zum einen ist anzumerken, dass die Anforderungen eines prozessualen Handelns in und Verstehens von betrieblichen Gesamtzusammenhänge scheinbar nicht so in den beruflichen Lehr-Lernprozessen implementiert sind, dass sie in Prüfungskontexten relevant und über verschiedene Fähigkeitsstrukturen repräsentiert werden. Für die Gestaltung von Unterricht und dementsprechend für die Konstruktion von Assessments werden folglich berufliche Handlungsfelder als Metamodelle des Lerngegenstands benötigt (ACHTENHAGEN 2003; WINTHER 2006; WINTHER 2009). In diesem Zusammenhang werden insbesondere Fragen der sequentiellen Gestaltung von Unterricht relevant, die unter Rückgriff auf die aktuelle Kompetenzdiskussion in der beruflichen Bildung Aspekte der Komplexitätsvariation von beruflichen Anforderungssituationen sowie die Integration domänenspezifischer und domänenübergreifender Kompetenzen über berufliche Inhalte mit einschließen (SEEBER 2008; NICKOLAUS/GSCHWENDTNER/ GEISSEL 2008; WINTHER/ ACHTENHAGEN 2008; WINTHER/ ACHTENHAGEN 2009). Oder anders formuliert: Prozessuale Denk- und Handlungsmuster sind im Rahmen des Kompetenzerwerbs zunächst anzulegen und zu fördern, bevor sie zum Gegenstand einer Kompetenzmessung gemacht werden können.

4 Bewertung der Reliabilität kaufmännischer Abschlussprüfungen

Reliabilität kann als Maß der Wahrscheinlichkeit für die richtige Klassifikation eines bestimmten Untersuchungsmerkmals beschrieben werden; Reliabilität beschreibt wie gut ein Schätzer das wahre Personenmerkmal misst (KIPLINGER 2008, 94). Für Kompetenzmessungen lässt sich der Begriff analog als die Wahrscheinlichkeit definieren, mit der die Auszubildenden auf ihr tatsächliches Kompetenzniveau geschätzt werden. Nach KIPLINGER (2008) kann zwischen zwei Formen von Reliabilität unterschieden werden: (1) Reliabilität, die das Design und die Adäquanz des Erfassungssystems betrifft – also insbesondere die Stichprobenziehung (Systemebene) – sowie (2) die Reliabilität der Assessmentergebnisse (Schülerebene). Die *Reliabilität auf Systemebene* misst innerhalb dieser Klassifikation die Konsistenz der Messungen der Schülerleistungen auf Schulebene. Implizit besteht bei inter-

temporalen Vergleichen die Annahme, dass sich andere, die Schülerleistung beeinflussende Faktoren – und hier insbesondere das Fähigkeitsniveau des jeweiligen Jahrgangs – stabil verhalten, d. h. keine Stichprobenfehler vorliegen. Somit können Veränderungen in der Schülerleistung über mehrere Zeitpunkte ausschließlich der Kompetenz zugeordnet werden. Entsprechend kann bei dem umfassenden Datensatz der IHK-Abschlussklausuren weitestgehend von Systemreliabilität ausgegangen werden. Hierauf deuten auch die relativ stabilen Notenverteilungen in der IHK-Prüfst Statistik über mehrere Jahre hinweg. Zudem ermöglicht die Anwendung von IRT-Modellen stichproben- und verteilungsunabhängige Information über die Reliabilität, wodurch die Reliabilitätsproblematik auf Systemebene für die weitere Untersuchung entfällt. Hinsichtlich der *Reliabilität der Assessmentergebnisse* (Schülerebene) bildet der Messfehler das entscheidende Beurteilungskriterium. Dieser entsteht, da ein Test immer nur exemplarisch einige Items eines ganzen Bereiches abbilden kann und darüber hinaus verschiedene andere Faktoren das Testergebnis bei Wiederholung der Messung beeinflussen können. Solche Faktoren bestehen insbesondere in (KIPLINGER 2008, 94-102):

- *Schülerspezifika* wie Motivation, Müdigkeit etc.;
- *Testspezifika* wie das Itemsample oder die Instruktionklarheit;
- *Skalierungsspezifika* wie beispielsweise Zählfehler;
- *Situationsspezifika* wie beispielsweise Klassenraumgröße oder störende Mitschüler.

Ein Messfehler beschreibt folglich die Variabilität, die mit dem Testen einer bestimmten Gruppe innerhalb einer bestimmten Testlokalität einhergeht. Für die Bewertung der Reliabilität der kaufmännischen Abschlussprüfungen auf Schülerebene wurden probabilistische Modelle genutzt. Das besondere bei der Anwendung von IRT-Modellen liegt dabei darin, dass eine probabilistische Modellierung Informationen über die Reliabilität für jede spezifische Ausprägung des Personenmerkmals – und damit für jeden einzelnen Lernenden – ermöglicht. Anders als die klassische Testtheorie, bei der nur die durchschnittliche Reliabilität für den gesamten Test erfasst werden kann.

Nach FISCHER (1974) wird die Reliabilität eines Items als die Information darüber definiert, mit welcher Präzision – statistisch formuliert mit welcher Varianz – ein Proband einer bestimmten Klasse zugeordnet werden kann. Je geringer die Schätzvarianz und damit der Messfehler einer Aufgabe, desto höher fällt die Testinformation und damit auch die Reliabilität eines Items aus (u. a. RAMSAY 1995, 60). Die Informationsmenge des gesamten Tests errechnet sich dann durch Summierung aller einzelnen Iteminformationsfunktionen (BAKER 2001, 104). Die Summierung ist auch intuitiv nachvollziehbar: Je mehr Aufgaben zur Einschätzung eines bestimmten Kompetenzniveaus bearbeitet wurden, desto wahrscheinlicher wird eine richtige Einteilung in eine bestimmte Fähigkeitsklasse in diesem Niveaubereich. Für die Reliabilität der Abschlussprüfungen zeigt sich empirisch die folgende Teststatistik:

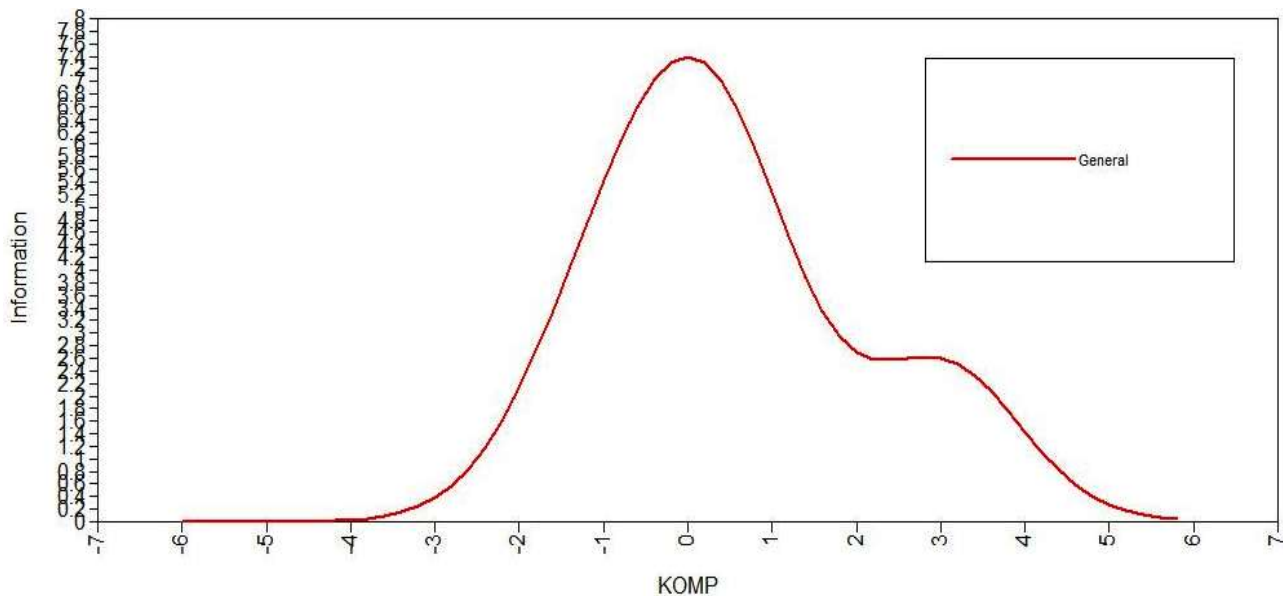


Abb. 3: Fähigkeitsspezifische Reliabilitätssumme über alle Testitems

Ist die Informationsmenge für einen Punkt auf der Kompetenzskala hoch, bedeutet dies, dass ein Proband, dessen wahre Personenkompetenz auf diesem bestimmten Level liegt, mit hoher Präzision auf genau diesem Level geschätzt werden kann. Dabei ergeben sich unterschiedliche Reliabilitäten für unterschiedliche Fähigkeitsniveaus, da die Reliabilität durch IRT-Modellierung als Funktion der Personenkompetenz gebildet wird. Die Funktion für alle Items weist ihren maximalen Wert in etwa bei der durchschnittlichen Personenkompetenz auf. Dies bedeutet, dass die Prüfungsteilnehmer um diesen Bereich am zuverlässigsten auf ihre wahre Kompetenzausprägung geschätzt werden (Reliabilität = 0,76); weiter entfernt von diesem Bereich jedoch mit rapide fallender Messpräzision. Außerdem ist auffällig, dass der rechte Flügel der Funktion auch bei Ausprägungen hoher Personenfähigkeit noch deutlich mehr Information liefert als der linke Flügel für unterdurchschnittliche Ausprägungen der (Fach-)Kompetenz. Dies liegt daran, dass der Test in der Summe mehr schwierige Items beinhaltet als leichte. Die Testinformationsfunktion als Summe aller Iteminformationsfunktionen fällt im Bereich niedriger Kompetenzausprägung folglich geringer aus. Auf Basis der Fähigkeitsspezifischen Reliabilitätssummen lässt sich für die kaufmännischen Abschlussprüfungen urteilen, dass eine Differenzierung zwischen Probanden mit tendenziell niedriger Kompetenzausprägung kaum möglich ist, da hier die Reliabilität des Tests gegen Null läuft.

Auf Basis des Reliabilitätsverlaufs lässt sich jedoch die Nützlichkeit einer Messung nicht umfassend einschätzen. Ergänzend ist hierzu die Zielsetzung eines Assessments zu berücksichtigen. So kann beispielsweise eine zunehmende Informationsmenge in einem bestimmten Bereich bei gleichzeitigem Abflachen in den Randbereichen explizit erwünscht sein und einer vorgegebenen Testlogik entsprechen. Dies ist immer dann der Fall, wenn Probanden an einem bestimmten Punkt mit besonders hoher Präzision separiert werden sollen. Allerdings gilt hier für die kaufmännischen Abschlussprüfungen folgende Bemerkung: Der entscheidende Separationspunkt ist nicht bei der durchschnittlichen Kompetenzausprägung zu suchen, sondern bei besonders niedrigem Fähigkeitsniveau. Da jährlich ca. 95 Prozent der an den Abschluss-

prüfungen teilnehmenden Auszubildenden diese Prüfung bestehen⁸ verläuft die kritische Einteilungsgrenze in „bestanden“ und „nicht bestanden“ – also weit unter dem durchschnittlichen Kompetenzniveau im linken Extrembereich niedriger Fähigkeit. Gerade hier werden die Prüfungsteilnehmer jedoch mit einer gegen Null verlaufenden Treffsicherheit – und somit praktisch zufällig – in die beiden Kategorien „bestanden“ und „nicht bestanden“ eingeordnet. Hier zeigt sich, dass das Gütekriterium der Reliabilität keineswegs einer statistischen „Testästhetik“ dient, sondern die Konsequenzen einer geringen Reliabilität und damit einer falschen Klassifizierung substantielle – mitunter negative – Auswirkungen auf den beruflichen Werdegang und den weiteren Lebenslauf der Schüler und Schülerinnen haben können.

5 Diskussion und Ausblick

Die Untersuchung der kaufmännischen Abschlussprüfungen im Ausbildungsberuf Industriekaufmann/-frau deutet für den Prüfungsbereich „Geschäftsprozesse“ sowohl auf konzeptionelle als auch auf empirische Limitationen hin. Hiervon sind sowohl die Konstruktvalidität als auch die Reliabilität betroffen. Im Folgenden sollen die Befunde hinsichtlich ihrer praktischen Relevanz – insbesondere mit Hinblick auf die Europäisierungsbestrebungen innerhalb der beruflichen Bildung – diskutiert werden.

Kompetenz bezeichnet im Deutschen Qualifikationsrahmen die Fähigkeit und Bereitschaft des Einzelnen, Kenntnisse und Fertigkeiten sowie persönliche, soziale und methodische Fähigkeiten zu nutzen und sich durchdacht sowie individuell und sozial verantwortlich zu verhalten und wird in diesem Sinne als „umfassende Handlungskompetenz“ verstanden (DQR 2011, 8). Mit Blick auf die aktuellen Prüfungsstrukturen ist festzuhalten, dass das Konzept der Handlungskompetenz primär über eine prozesslogische Struktur – Planung, Durchführung, Kontrolle – operationalisiert wird. Die Annahme, dass sich diese Prozessorientierung als Facette beruflicher Handlungsorientierung auch empirisch in den Testergebnissen der Abschlussprüfungen widerspiegelt, kann nach Überprüfung der globalen Anpassungsgüte des postulierten, prozessualen Kompetenzstrukturmodells nicht verifiziert werden. Statt eines Prozessmodells verweisen die empirischen Befunde hingegen auf ein fachbezogenes Kompetenzstrukturmodell. Hierfür sprechen sowohl die Befunde einer explorativen Faktorenanalyse, innerhalb derer die Daten praktisch „selbstständig“ die domänenspezifische Einteilung in einer 5-Faktoren-Lösung replizieren, als auch die guten Modellwerte des fachbezogenen Modells bei konfirmatorischer Überprüfung. Dies bedeutet, dass sich die Zugehörigkeit einer Aufgabe zu einer bestimmten Prozessdimension nicht als Kompetenzmerkmal für eine Testperson interpretieren lässt und eine dementsprechende Aufgabeneinteilung zu Zwecken der Kompetenzmessung folglich redundant wäre. Dieser Befund hat im Hinblick auf die Evaluation beruflicher Bildungsprogramme vor dem Hintergrund der Lernfelddebatte als normative Setzung der Curricula eine besondere Dynamik und ist an anderer Stelle vorrangig aus einer fachdidaktischen Perspektive zu diskutieren. Überraschend sind diese Befunde allerdings nicht. Sie decken sich mit anderen empirischen Studien aus dem allgemeinbildenden Bereich, in denen ebenfalls eine fach- oder inhaltsbezogene Struktur von Kompetenz modelliert und

⁸ Entnommen der Prüfstatistik der Industrie- und Handelskammer für München und Oberbayern.

empirisch validiert wurde – z. B. PISA 1996 (BAUMERT/ KLIEME/ NEUBRAND et al. 2001) oder TIMSS 2000 (BAUMERT/ LEHMANN 1997) – sowie mit Untersuchungen aus dem beruflichen Bereich, die eine hohe Relevanz betrieblicher Inhalte und deren charakteristischen Einfluss auf Anforderungssituationen und damit letztlich auf die Struktur beruflicher Kompetenz suggerieren (z. B. SEEBER 2008; NICKOLAUS/ GSCHWENDTNER/ AESE 2009; WINTHER 2010). Auch stimmt dieser Befund weitgehend mit den Ergebnissen einer Befragung von Ausbildungsbetrieben im Jahr 2000 durch das BIBB überein, wonach Betriebe sich zwar eine hohe Relevanz von Prüfungsergebnissen wünschen, 77 Prozent der Befragten jedoch bezweifeln, dass das Prüfungszeugnis ausreichend *verlässliche* Hinweise auf die *berufliche Handlungsfähigkeit* liefert. Vielmehr werden die Prüfungen als Indikator für den Erwerb *fachlichen* Wissens angesehen (SCHMIDT 2000; WEIß 2011, 39). Für die berufliche Bildung, innerhalb derer der Begriff beruflicher Handlungskompetenz ein zentrales Konstituierungsmoment bildet, ist ein solches Messkonzept jedoch nicht zufriedenstellend. Denn *was* mit den Abschlussprüfungen gemessen wird, entspricht sowohl inhaltlich als auch empirisch eher der fachlichen Personenfähigkeit als dem Konzept beruflicher Handlungskompetenz. In Konsequenz muss gewissermaßen von einer Entkopplung der Abschlussprüfungen von dem in den nationalen und internationalen Ordnungsgrundlagen festgelegten Prinzip der Handlungsorientierung ausgegangen werden.

Bezüglich der Reliabilitätswerte weisen die Abschlussprüfungen eine hohe Messgenauigkeit für den Bereich mittlerer Kompetenzausprägung aus. In den Randbereichen fallen Testinformation und Reliabilität jedoch rapide ab, was damit zusammenhängt, dass sehr schwierige Items selten, sehr leichte Items noch seltener sind. Entsprechend können über Kompetenzausprägungen in den Randbereichen nur wenig präzise Aussagen getroffen werden. Gerade an der kritischen Bestehens- bzw. Nichtbestehensgrenze werden Schüler und Schülerinnen damit praktisch zufällig und weitgehend unabhängig von ihrer wahren Fähigkeitsausprägung klassifiziert. Dies erscheint umso gravierender, da sich die Relevanz von Testergebnissen und ihr Einfluss auf den weiteren Bildungs- und Lebensweg für die Schülerinnen und Schüler gerade vor dem Hintergrund einer wachsenden Testmentalität im (inter-)nationalen Bildungswesen zunehmend erhöht. Als Beleg hierfür sei darauf verwiesen, dass noch innerhalb des laufenden Jahres alle neuen Qualifikationsbescheinigungen der nationalen Qualifikationssysteme einen *klaren* Verweis auf das jeweilig zutreffende Niveau des Europäischen Qualifikationsrahmens enthalten sollen (DQR 2011, 3). Die hohe Wahrscheinlichkeit eines Klassifikationsfehlers auf nationaler Ebene wird sich jedoch unweigerlich auch durch einen internationalen Metarahmen ziehen.

Die derzeitige Prüfungspraxis lässt zumindest Zweifel zu, ob die durch den Qualifikationsrahmen angestrebte höhere Transparenz glaubwürdig erreicht werden kann – die Autoren vertreten die Meinung, dass ohne ausreichend gesicherte empirische Evidenzen der Leistungsfähigkeiten einzelner Bildungsprogramme eine systematische Einordnung nicht erfolgen kann und die derzeit vorgenommenen Zuordnungen in den Bereich der Kompetenzlyrik gehören. Eine systematische Einordnung zu Zwecken der Vergleichbarkeit wird erst dann sinnvoll, wenn die Kompetenzen der einzelnen Individuen innerhalb der jeweiligen nationalen Systematisierung mit ausreichend hoher Sicherheit diagnostiziert werden können und wenn inner-

halb der nationalen Tests die tatsächlich zu erfassenden Merkmale vergleichbar werden. Es ist daher zu prüfen, in welcher Form individuelle Kompetenzen sowie Systemleistungen bestmöglich zu erfassen wären. Möglichkeiten ergeben sich hier insbesondere bei der Aufgabenkonzeption. Aspekte der Vollständigkeit und der Authentizität könnten durch eine glaubwürdig an Unternehmensprozessen orientierte Assessmentkonstruktion Berücksichtigung finden. Hierbei sollte auch der Umgang mit Informations- und Kommunikationstechniken, der insbesondere für den kaufmännischen Bereich unverzichtbar ist, in den Abschlussprüfungen eine zentralere Rolle spielen. Erste Arbeiten für den kaufmännischen Bereich bestehen hier bereits durch WINTHER/ ACHTENHAGEN (2009) sowie WINTHER (2010) mit der technologiebasierten Testumgebung *ALUSIM*. Die Messgenauigkeit ließe sich im Zuge der Testkonstruktion verbessern indem mehr schwierige und leichte Items konzipiert würden, womit die Prüfergebnisse zuverlässigere Rückschlüsse auf die Kompetenzen der Auszubildenden zulassen würden. Soll die Testlänge allerdings nicht weiter ausgedehnt werden, bietet sich eine weitere Möglichkeit im Einsatz computerbasierter, adaptiver Testverfahren, bei denen die Schwierigkeit der Aufgaben dynamisch dem individuellen Leistungsniveau angepasst werden kann.

Literatur

ACHTENHAGEN, F./ WINTHER, E. (2009): Konstruktvalidität Von Simulationsaufgaben: Computergestützte Messung berufsfachlicher Kompetenz am Beispiel der Ausbildung für Industriekaufleute. Abschlussbericht zur Einreichung beim BMBF. Professur für Wirtschaftspädagogik der Georg-August Universität Göttingen.

ACHTENHAGEN, F. (2003): Konstruktionsbedingungen für komplexe Lehr-Lern-Arrangements und deren Stellenwert für eine zeitgemäße Wirtschaftsdidaktik. In: KAISER/ F.-J. KAMINSKI, H. (Hrsg.): Wirtschaftsdidaktik. Bad Heilbrunn, 77-97.

AUFGABENSTELLE FÜR KAUFMÄNNISCHE ABSCHLUSS- UND ZWISCHENPRÜFUNGEN (AKA) (Hrsg.) (2009): Prüfungskatalog für die IHK-Abschlussprüfungen. Nürnberg.

BAETHGE, M./ ACHTENHAGEN, F./ ARENDS, L. / BABIC, E./ BAETHGE-KINSKY, V./ WEBER, S. (2006): Berufsbildungs-PISA-Machbarkeitsstudie. Stuttgart.

BAKER, F. (2001): The Basics Of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park.

BAUMERT, J./ LEHMANN, R. (Eds.) (1997): TIMSS - Third International Mathematics And Science Study: Dritte Internationale Mathematik- und Naturwissenschaftsstudie: Anlage, Fragestellungen und Durchführung der TIMSS-Studie in der Bundesrepublik Deutschland. Opladen.

BAUMERT, J./ KLIEME, E./ NEUBRAND, M./ PRENZEL, M./ SCHIEFELE, U./ SCHNEIDER, W./ STANAT, P./ TILLMANN, K.-J./ WEIß, M. (Hrsg.) (2001): PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen.

BERTELSMANN STIFTUNG (Hrsg.) (2010): Berufsausbildung 2015. Eine Entwicklungsperspektive für das duale System. Gütersloh.

BERUFSBILDUNGSGESETZ (BBiG) vom 23. März 2005.

BMBF (2012): Wichtiger Schritt hin zum Bildungsraum Europa - Einigung bei der Einführung des Deutschen Qualifikationsrahmens. Pressemitteilung von 31.01.2012. Online: www.bmbf.de/press/3230.php (11-04-2012).

BOHLINGER, S. (2006): Lernergebnisorientierung als Ziel beruflicher Qualifizierung? Absehbare und nicht absehbare Folgen der Einführung des Europäischen Qualifikationsrahmens. In: *bwp@ Berufs- und Wirtschaftspädagogik – online*, Ausgabe 11. Online: www.bwpat.de/ausgabe11/bohlinger_bwpat11.pdf (11-04-2012).

CATTELL, R. B. (1966): The Scree Test For The Number Of Factors. In: *Multivariate Behavioral Research*, 1, 245-276.

COMMISSION OF THE EUROPEAN COMMUNITIES (2005): Towards A European Qualifications Framework For Lifelong Learning. Brüssel.

ARBEITSKREIS DEUTSCHER QUALIFIKATIONSRAHMEN (2011): Deutsche Qualifikationsrahmen für lebenslanges Lernen, verabschiedet vom am 22. März 2011. Online: www.deutscherqualifikationsrahmen.de/de/aktuelles/deutscher-qualifikationsrahmen-f%C3%BCr-lebenslanges-le_gh3psgo.html (11-04-2012).

DREXEL, I. (2008): Berufsprinzip oder Modulprinzip? Zur künftigen Struktur beruflicher Bildung in Deutschland. In: URBANSKI, D. (Hrsg.): Die Berufskollegs stärken heißt die berufliche Bildung zu stärken: 10 Jahre Berufskolleg in NRW - ein Erfolgsmodell auf dem Weg zur Eigenständigkeit. Krefeld.

EBBINGHAUS, M. (2005): Stand und Perspektiven bei beruflichen Prüfungen – Ansätze zur Reform des Prüfungswesens in der dualen Ausbildung. In: *bwp@ Berufs- und Wirtschaftspädagogik – online*, Ausgabe 8. Online: www.bwpat.de/ausgabe8/ebbinghaus_bwpat8.pdf (11-04-2012).

FISCHER, G. H. (1974): Einführung in die Theorie psychologischer Tests. Bern.

HACKER, W. (1986): Arbeitspsychologie. Psychische Regulation von Arbeitstätigkeiten. Bern.

HARTIG, J./ KLIEME, E. (2006): Kompetenz und Kompetenzdiagnostik. In: SCHWEIZER, K. (Hrsg.): *Leistung und Leistungsdiagnostik*. Heidelberg, 127-143.

HARTIG, J./ HÖHLER, J. (2008): Representation Of Competencies In Multidimensional IRT-Models With Within-Item And Between-Item Multidimensionality. In: *Zeitschrift für Psychologie*, 216, H. 2, 89-101.

HORN, J. L. (1965): A Rationale And Test For The Number Of Factors In Factor Analysis. In: *Psychometrika*, 30, 179-185.

KIPLINGER, L. (2008): Reliability Of Large Scale Assessment And Accountability Systems, In: RYAN, K. E./ SHEPARD, L. A. (Eds.): The Future Of Test-Based Educational Accountability. New York, 93-114.

KIRISCI, L./ HSU, T.-C./ YU, L. (2001): Robustness Of Item Parameter Estimation Programs To Assumptions Of Unidimensionality And Normality. In: Applied Psychological Measurement, 25, 146–162.

KLIEME, E./ AVENARIUS, H./ BLUM, W. et al. (2003): Zur Entwicklung nationaler Bildungsstandards. Eine Expertise. Online:
www.dipf.de/publikationen/zur_entwicklung_nationaler_bildungsstandards.pdf (11-04-2012).

KMK (1996): Handreichung für die Erarbeitung von Rahmenlehrplänen der Kultusministerkonferenz (KMK) für den berufsbezogenen Unterricht in der Berufsschule und ihre Abstimmung mit Ausbildungsordnungen des Bundes für anerkannte Ausbildungsberufe. Bonn.

KREMER, H. (Hrsg.) (2003): Handlungs- und Fachsystematik im Lernfeldkonzept. In: *bwp@ Berufs- und Wirtschaftspädagogik – online*, Ausgabe 4. Online:
www.bwpat.de/ausgabe4/kremer_bwpat4.pdf (11-04-2012).

KREMER, H.-H./ SLOANE, P. F. E. (2001): Lernfelder implementieren. Zur Entwicklung und Gestaltung fächer- und lernortübergreifender Lehr-Lernarrangements im Lernfeldkonzept. Paderborn.

LINDEN, W. J./ HAMBLETON, R. K. (1997): Handbook Of Modern Item Response Theory. New York.

LÖRIG, D./ SCHREIBER, B. (2007): Ausgestaltung kompetenzbasierter Ausbildungsordnungen – Grundlage für Kompetenzmessung und Kompetenzbewertung. In: Zeitschrift für Berufsbildung in Wissenschaft und Praxis, 6, 5-9.

NICKOLAUS, R./ GSCHWENDTER, T./ ABELE, S. (2009): Die Validität von Simulationsaufgaben am Beispiel der Diagnosekompetenz von Kfz-Mechatronikern. Vorstudie zur Validität von Simulationsaufgaben im Rahmen eines VET-LSA. Abschlussbericht für das Bundesministerium für Bildung und Forschung. Stuttgart.

NICKOLAUS, A./ GSCHWENDTER, T./ GEIßEL, B. (2008): Modellierung und Entwicklung beruflicher Fachkompetenz in der gewerblich-technischen Erstausbildung. In: Zeitschrift für Berufs- und Wirtschaftspädagogik, 104, 48-73.

PÄTZHOLD, G. (2000): Lernfeldorientierung – Berufliches Lehren und Lernen zwischen Handlungs- und Fachsystematik. In: BADER, R./ SLOANE, P. F. E. (Hrsg.): Lernen in Lernfeldern. Theoretische Analysen und Gestaltungsansätze zum Lernfeldkonzept. Markt Schwaben, 123-140.

RAMSAY, J. O. (1995): TestGraf. A Program For The Graphical Analysis Of Multiple Choice Test And Questionnaire Data. Montreal.

REETZ, L. (1989a): Zum Konzept der Schlüsselqualifikationen in der Berufsbildung. In: Berufsbildung in Wissenschaft und Praxis, 5, 3-10.

REETZ, L. (1989b): Zum Konzept der Schlüsselqualifikationen in der Berufsbildung. In: Berufsbildung in Wissenschaft und Praxis, 6, 24-30.

REINISCH, H. (1999): Zu einigen curriculumtheoretischen Implikationen des Lernfeldansatzes – Überlegungen anlässlich der Beiträge von Clement, Kremer, Sloane und Tramm. In: *bwp@* Berufs- und Wirtschaftspädagogik – online, Ausgabe 4. Online: www.bwpat.de/ausgabe4/reinisch_bwpat4.pdf (11-04-2012).

ROST, J./ SPADA, H. (1978): Probabilistische Testtheorie. In: Klauer, K. J. (Hrsg.): Handbuch für pädagogische Diagnostik. Düsseldorf, 59-97.

SAMEJIMA, F. (1969): Estimation Of Latent Ability Using A Response Pattern Of Graded Scores. In: Psychometrika, 34, H. 4, 100-114.

SCHMIDT, J. U. (2000): Prüfungen auf dem Prüfstand – Betriebe beurteilen die Aussagekraft von Prüfungen. In: Berufsbildung in Wissenschaft und Praxis, 29, H. 5, 27-31.

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

SEEBER, S./ NICKOLAUS, R./ WINTHER, E./ ACHTENHAGEN, F./ BREUER, K. et al. (2010): Kompetenzdiagnostik in der Berufsbildung Begründung und Ausgestaltung eines Forschungsprogramms. In: Berufsbildung in Wissenschaft und Praxis, Beilage zu 1/2010, 1-15.

VOLPERT, W. (1983): Handlungsstrukturanalyse als Beitrag zur Qualifikationsforschung. Köln.

WEIß, R. (2011): Prüfungen in der beruflichen Bildung – ein vernachlässigter Forschungsgegenstand. In: SEVERING, E./ WEIß, R. (Hrsg.): Prüfungen und Zertifizierung in der beruflichen Bildung – Anforderungen – Instrumente – Forschungsbedarf. Bonn, 37-52.

WINTHER, E. (2006): Konstruktion, Implementierung und Evaluation eines Lehrerhandbuchs im Rahmen integrierter Kompetenzentwicklung. In: GONON, P./ KLAUSER, F./ NICKOLAUS, R. (Hrsg.): Kompetenz, Qualifikation und Weiterbildung im Berufsleben. Opladen, 41-54.

WINTHER, E. (2009): The Effectiveness of Learning Environments or About the Difficulties to Compare Apples and Oranges. In: OSER, F./ RENOLD, U./ JOHN, E. G./ WINTHER, E./ WEBER, S. (Hrsg.): VET Boost: Towards a Theory of Professional Competencies. Essays in Honor of Frank Achtenhagen. Rotterdam, 201-218.

WINTHER, E. (2010): Kompetenzmessung in der beruflichen Bildung. Bielefeld.

WINTHER, E. (2011): Kompetenzorientierte Assessments in der beruflichen Bildung – Am Beispiel der Ausbildung von Industriekaufleuten. In: Zeitschrift für Berufs- und Wirtschaftspädagogik, 107, 1, 33-54.

WINTHER, E./ ACHTENHAGEN, F. (2008): Kompetenzstrukturmodell für die kaufmännische Bildung. Adaptierbare Forschungslinien und theoretische Ausgestaltung. In: Zeitschrift für Berufs- und Wirtschaftspädagogik, 104 (4), 511-538.

WINTHER, E./ ACHTENHAGEN, F. (2009): Skalen und Stufen kaufmännischer Kompetenz. In: Zeitschrift für Berufs- und Wirtschaftspädagogik, 105 (4), 521-556.

WINTHER, E./ ACHTENHAGEN, F. (2010): Berufsfachliche Kompetenz: Messinstrumente und empirische Befunde zur Mehrdimensionalität beruflicher Handlungskompetenz. In: Berufsbildung in Wissenschaft und Praxis, 1, 18-21.

WRIGHT, B. D./ STONE, M. H. (1979): Best Test Design. Chicago.

Dieser Beitrag wurde dem *bwp@*-Format: **FORSCHUNGSBEITRÄGE** zugeordnet.

Zitieren dieses Beitrages

KLOTZ, V. K./ WINTHER, E. (2012): Kompetenzmessung in der kaufmännischen Berufsausbildung: Zwischen Prozessorientierung und Fachbezug. Eine Analyse der aktuellen Prüfungspraxis. In: *bwp@* Berufs- und Wirtschaftspädagogik – online, Ausgabe 22, 1-16. Online: http://www.bwpat.de/ausgabe22/klotz_winther_bwpat22.pdf (26-06-2012).

Die Autorinnen



VIOLA KATHARINA KLOTZ

Department 5: Wirtschaftspädagogik, Lehrstuhl für
Wirtschaftspädagogik III, Universität Paderborn
Warburger Str. 100, 33098 Paderborn

E-mail: [viola.klotz \(at\) wiwi.upb.de](mailto:viola.klotz@wiwi.upb.de)

Homepage: <http://wiwi.uni-paderborn.de/department5/lehrstuhl-prof-dr-esther-winter/viola-klotz-m-sc/>



Prof. Dr. ESTHER WINTHER

Department 5: Wirtschaftspädagogik, Lehrstuhl für
Wirtschaftspädagogik III, Universität Paderborn
Warburger Str. 100, 33098 Paderborn

E-mail: [esther.winther \(at\) wiwi.upb.de](mailto:esther.winther@wiwi.upb.de)

Homepage: <http://wiwi.uni-paderborn.de/department5/lehrstuhl-prof-dr-esther-winter/prof-dr-esther-winter/>