

# Skalen und Stufen kaufmännischer Kompetenz

**KURZFASSUNG:** Im vorliegenden Beitrag werden die Ergebnisse des BMBF-Projekts „Konstruktvalidität von Simulationsaufgaben: Computergestützte Messung berufsfachlicher Kompetenz (K350600)“ vorgestellt. Das Ziel dieses Projekts war es, in Vorbereitung eines international vergleichenden Large Scale-Assessment in der beruflichen Bildung (VET-LSA) innovative Testformate zu erproben, mit denen es gelingen kann, berufliche Handlungskompetenz in ihren unterschiedlichen Facetten zu erfassen. Hierfür wurde die web-basierte Computersimulation ALUSIM entwickelt und an 264 Auszubildenden im Ausbildungsberuf Industriekaufmann / Industriekauffrau getestet. Kernstück der Computersimulation sind berufliche Anforderungssituationen, die Arbeits- und Geschäftsprozesse von Industrieunternehmen visualisieren und die für die Auszubildenden authentische betriebliche Aufgaben einschließlich der notwendigen Handlungs- und Kommunikationsabläufe darstellen. Die Entwicklung der Aufgaben orientierte sich an dem Kompetenzstrukturmodell für die kaufmännische Bildung (WINTHER & ACHTENHAGEN, 2008). Auf Basis des Modells konnten konkrete kognitionspsychologische **und** fachdidaktische Überlegungen in die Gestaltung der Testaufgaben als interpretierbare Schwierigkeitsparameter einfließen und direkt auf psychometrische Modelle der Item Response Theory bezogen werden. Die Ergebnisse belegen, dass sich berufliche Handlungskompetenz über zwei Skalen umfassend beschreiben lässt: Die Skala der handlungsbasierten Kompetenz zielt auf vorwiegend betriebliche Strukturen und Entscheidungsprozesse innerhalb authentischer betrieblicher Anforderungssituationen ab, während die Skala der verstehensbasierten Kompetenz wissenszentrierte Adaptationen und Transferfähigkeiten vor dem Hintergrund dieser Anforderungssituationen erfasst.

**ABSTRACT:** The paper is dealing with the measurement and scaling approach of vocational competence. In preparing an international comparison of vocational education and training a pre-test was initiated by the German Federal Ministry of Education and Research to clarify the construct validity of authentic business simulations in the field of business and administration. Using modern assessment theory the test development is based on both a competence model and a psychometric IRT model (WINTHER & ACHTENHAGEN, 2008). Combining research from both areas the model is able to adapt results from investigations in the fields of compulsory education (e. g. the TIMS- and PISA-studies) and to define competence in the field of business and administration using two different competence scales: action-based competence is focused on operational sequences; understanding-based competence describes knowledge adaptations and the students' ability to deal with theoretical constructs.

## 1 Projektziele – Zur Messbarkeit beruflicher Handlungskompetenz

Ausgangspunkt des Projekts waren Diskussionen, die sich auf die Feasibility-Studie zum Berufsbildungs-PISA bezogen (vgl. BAETHGE, ACHTENHAGEN, ARENDS, BABIC, BAETHGE-KINSKY & WEBER, 2006). Obwohl dort die Begründungen für die Operationalisierung des Konzepts beruflicher Handlungskompetenz detailliert vorgestellt sind und zugleich gegenüber den Ansätzen eines EQF abgegrenzt werden, wurde von Seiten der Arbeitgeber und der Gewerkschaften die Befürchtung geäußert, dass in einem solchen VET-LSA die typische Leistung des deutschen Systems einer Dualen Berufsausbildung nicht hinreichend zur Geltung käme: die Herausbildung einer beruflichen Handlungskompetenz in den verschiedenen beruflichen Fachgebieten. Weiter: Wenn in einen solchen Vergleich Länder mit einer überwie-

gend schulischen beruflichen Ausbildung einbezogen würden, dann bestünde die Gefahr, dass die Jugendlichen, die ein solches Programm durchlaufen, aufgrund ihres höheren schulischen Ausbildungsanteils bei Papier-und-Bleistift-gestützten Wissensfragen besser abschnitten als die deutschen Auszubildenden mit ihrem geringeren schulischen und höheren betrieblichen Ausbildungsanteil. Von daher müsste – so der Tenor der Schlussfolgerungen – sicher gestellt sein, dass das Spezifikum der deutschen Dualen Ausbildung – nämlich das Befähigen zu selbstständigen und selbstverantworteten Handlungen am Arbeitsplatz, kurz: berufliche Handlungskompetenz – in der Aufgabenstellung und der Aufgabenstruktur eines VET-LSA zentral zur Geltung käme.

Weder die Auftragnehmer dieses Projekts noch die Autoren der Machbarkeitsstudie sehen hier ein Problem; denn ein selbstständiges und selbstverantwortetes Handeln am Arbeitsplatz ist die Zielsetzung aller Bemühungen zur „transition from school to work“, wie sie in allen Industrienationen oben auf der Bildungsagenda steht. Und die Maßnahmen im Rahmen der deutschen Dualen Berufsausbildung leisten hierzu einen wesentlichen Beitrag. Mit unserem Projekt soll gezeigt werden, wie sich das Konstrukt der „beruflichen Handlungskompetenz“ objektiv, reliabel und valide erfassen lässt und wie dabei das Zusammenspiel von berufsbezogenen Wissens- bzw. Verstehens- und Handlungskomponenten zu berücksichtigen ist<sup>1</sup>.

Eine Vorbemerkung ist notwendig: Bei der Frage einer „angemessenen“ Aufgabenentwicklung für den beruflichen Bereich geht es immer wieder darum, in welchem Maße die Aufgaben in der „Realität“ geprüft würden. Hierzu ist zweierlei zu sagen: Wenn es um eine Prüfung in den Betrieben selbst geht, dann wird der Anspruch erhoben, betriebsspezifische Aufgaben und Handlungen zu prüfen. Das hat zwei Konsequenzen: Zum einen ist ein solches Vorgehen im Rahmen eines Large Scale-Assessment kaum praktikabel; das gilt für den Zeitaufwand, aber auch für Probleme der Aufdeckung vertraulicher Daten, wie z. B. die Preisgestaltung, oder aber für Spezifika der jeweiligen betrieblichen ERP-Systeme. Zum anderen müssten alle Ergebnisse von ihrer Betriebsspezifität in eine Berufstypik rückübersetzt werden, damit sie vergleichbar werden. Diese Feststellung hat Konsequenzen für den Sprachgebrauch: Reserviert man den Terminus „Realität“ für betriebsspezifische Prozesse, dann handelt es sich bei einer authentischen Abbildung berufstypischer Prozesse immer um eine „Simulation“. Berufstypik und Realität wären dann nur über explizite Modelle einer authentischen Aufgabenkonstruktion zuordenbar.

Wir haben unter Vernachlässigung dieser Problematik dennoch versucht, unsere berufstypischen Aufgaben als Arbeitsproben im Hinblick auf eine betriebsspezifische Anwendung testen zu lassen. Die Rückmeldungen aus den Betrieben zeigen, dass diese Art der Kompetenzerhebung nicht als zielführend angesehen wird – was seinen Niederschlag auch darin gefunden hat, dass die Rücksendungen der Aufgaben von der Zahl her zu vernachlässigen sind. Von daher sind Äußerungen im politischen Raum, die die Begriffe „Realität“, „betriebsspezifisch“ und „berufstypisch“ nicht hinreichend differenziert und im Hinblick auf die jeweiligen Konsequenzen begründet verwenden, als unzureichend zurückzuweisen.

1 Ein in der Diskussion um „berufliche Handlungskompetenz“ immer wieder gehörtes Argument bezieht sich auf die „Notwendigkeit“ einer Trennung bzw. Unterscheidung von Wissen und Handlungen. Diese lässt sich aber nicht durchhalten; denn berufliche Handlungen, die sich nicht auf Wissensbestände stützen, scheinen eher kontraproduktiv oder gar widersinnig (vgl. hierzu das Gutachten von ACHTENHAGEN, 2004, für das BMBF).

Die Itemkonstruktion zur Erfassung der Verstehens- und Handlungskomponenten berufsfachlicher Kompetenz im Bereich der kaufmännischen Ausbildung orientiert sich an dem Kompetenzstrukturmodell für die kaufmännische Bildung (WINTHER & ACHTENHAGEN, 2008; WINTHER, in Vorbereitung). Vor dem Hintergrund des Modells definieren wir *kaufmännische Kompetenz als Fähigkeit, auf Grundlage eines systemischen Verstehens betrieblicher Teilprozesse und deren Rekonstruktion aus realen Unternehmensdaten in berufsrealen Situationen unternehmerische Entscheidungen treffen und diese validieren zu können, um damit das eigene Wissens- und Handlungspotential vor dem Hintergrund der Entwicklung individueller beruflicher Regulationsfähigkeit auszubauen* (vgl. WINTHER & ACHTENHAGEN, 2008). Diese Definition unterstreicht die Bedeutung der kompetenten Erstellung angemessener Handlungspläne. Ziel der Auswahl von Testinhalten ist es, die handlungsleitende Funktion von Wissensprinzipien einschließlich ihrer einschränkenden Bedingungen herauszustellen (vgl. GELMAN & GREENO, 1989, p. 172). Über die Inhalts- und Zugriffsebene im Kompetenzstrukturmodell gelingt es, berufliche Handlungskompetenz so zu operationalisieren, dass eine Messung valide und reliabel möglich wird und dass sich eine Analyse der Ergebnisse auf spezifische Facetten der Handlungskompetenz beziehen lässt. Das Modell geht dabei im Kern von einer differenzierten Struktur beruflicher Handlungskompetenz aus, die sich sowohl auf die betrieblichen und schulischen Inhalte als auch auf die verschiedenen kognitiven Zugriffe, mit denen typische berufliche Inhalte bewältigt werden, beziehen lässt. Damit wird dem Umstand Rechnung getragen, dass Kompetenz in einer Domäne (1) inhaltsbezogen und (2) keine eindimensionale kognitive Disposition ist und es folglich im Hinblick auf Kompetenzerwerb und -entwicklung nicht genügen wird, nur auf eine spezifische kognitive Struktur zu fokussieren (KILPATRICK, SWAFFORD & FINDELL, 2001; BRAND, HOFMEISTER & TRAMM, 2005). Aus dieser Perspektive werden im Modell drei verschiedene Kompetenzstrukturen betrachtet: die Bearbeitung einer definierten konkreten Lern- und Arbeitsanforderung durch aktivierte deklarative Wissensbestände (conceptual competence) vor dem Hintergrund eines ganzheitlichen und funktionalen Verständnisses zentraler domänenbezogener Ideen, die Selektion und Ausführung domänenbezogener Handlungen und Prozeduren (procedural competence) sowie die kognitive Erfassung der Anforderungssituation und die angemessene Interpretation und Validation der gefundenen Lösung (interpretative competence). Die Kompetenzstruktur bezogen auf die drei inhaltlichen Kompetenzbereiche economic literacy, economic numeracy und Geschäftsvorfall ist damit **prozessual und nicht hierarchisch**. Dies impliziert, dass keine Zunahme an Komplexität im Hinblick auf die Kompetenzstrukturen angenommen wird. Die verschiedenen Kompetenzstrukturen charakterisieren ausschließlich den Zugriff auf spezifische Inhaltsbereiche und stellen damit die Grundlage für die Auswahl von Testinhalten dar.

Innerhalb des Projekts lag der Schwerpunkt der Inhaltsauswahl auf domänenspezifischen Geschäftsvorfällen. Unter Geschäftsvorfall wird in diesem Zusammenhang, eine über konkrete Arbeitsprozesse definierbare Anforderungssituation verstanden, die sich in unternehmensspezifischen Geschäftsprozessen verorten lässt. Ein Geschäftsvorfall stellt damit eine inhaltlich-systematisch ausgestaltete arbeitsplatzspezifische Situation dar, die sich detailliert im Hinblick auf das Anforderungsniveau, die Handlungsspielräume sowie die intendierten Zielsetzungen beschreiben lässt (WINTHER, in Vorbereitung).

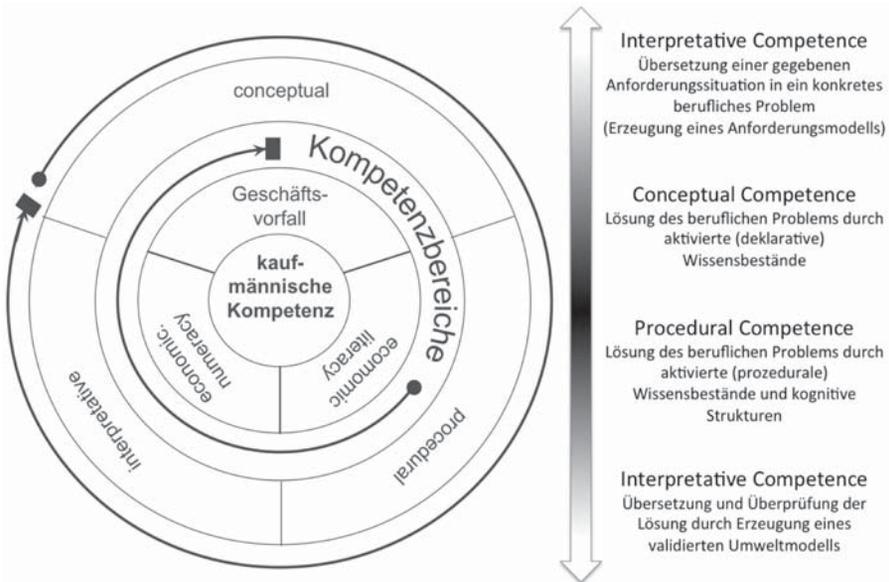


Abb. 1: Kompetenzstrukturmodell für die kaufmännische Bildung  
(WINTHER, in Vorbereitung)

Im Hinblick auf die Intentionen eines Large Scale-Assessment in der kaufmännisch beruflichen Bildung wurden solche Inhalte ausgewählt, die auch für einen internationalen Vergleich geeignet wären: Geschäftsvorfälle aus dem Bereich der betrieblichen Wertschöpfung sowie Geschäftsvorfälle, die betriebliche Steuerungsprozesse abbilden.

## 2 Testinhalte und Testformat

Die zentrale Herausforderung des Projekts war es, ein Testformat zu entwickeln, das den Implikationen des Kompetenzstrukturmodells gerecht werden kann: Zum einen sollten verschiedene berufliche Inhaltsbereiche mit ihren komplexen Zugriffs- und Handlungsebenen abgebildet werden und zum anderen sollte der Aktionsraum der Auszubildenden auf vollständige Arbeits- und Geschäftsprozesse in ihrem ganzheitlich zeitlichem und organisationalem Umfang bezogen sein. Mit der Berücksichtigung dieser Anforderungen sollte der Anspruch abgesichert werden, berufliche Handlungskompetenz in ihren verstehensbasierten (*conceptual*) und handlungsbasierten (*procedural* und *interpretative*) Strukturen abzubilden. Hierfür wurden (1) konzeptuale Anwendungsaufgaben und (2) prozessuale Simulationsaufgaben konstruiert, die in eine computerbasierte Testumgebung – ALUSIM – eingebunden sind.

## 2.1 Auswahl der Testinhalte

Wir sind bei der Auswahl der Testinhalte zur Erhebung des Konstrukts „Berufliche Handlungskompetenz“ davon ausgegangen, dass ein systemisches Verstehen von unternehmensinternen Prozessen und ein entsprechendes Handeln in diesen eine wesentliche Qualifikationsanforderung für den Bereich der Aus- und Weiterbildung von Industriekaufleuten und damit eine zentrale Bedingung für die Ausformung beruflicher Handlungskompetenz darstellt. Für die angemessene Erfassung der entsprechenden Kompetenz ist es erforderlich, ein Testformat zu entwerfen, das im Hinblick auf seine „Authentizität“ theoretisch begründet und fachdidaktisch entsprechend ausgestaltet ist; denn nur so ist es möglich, ein systemisches Verstehen von und ein entsprechendes Handeln in Geschäftsprozessen als eine Basiskompetenz von Industriekaufleuten über entsprechende Teilkompetenzen zu definieren und in Graduierungsschritten auszubilden. Hierzu ist es notwendig, Testinhalte, Testitems und empirisches Testmodell parallel zu entwerfen und wechselseitig aufeinander abzustimmen. Das Ziel ist es, berufliche Anforderungssituationen so zu entwickeln, dass es möglich wird, die zu bearbeitenden Aufgaben auf die zugrunde liegenden Arbeits- und Lernprozesse zu beziehen und zugleich Auskunft über die kognitiven Verarbeitungsschritte während des Lösungsprozesses zu gewinnen. Anders formuliert: Die Konstruktion integriert damit die Modellierungsschritte für Aufgaben, die einerseits betriebliche Arbeitssituationen abbilden und andererseits als Psychometrik Aussagen über die kognitiven Anforderungen bei ihrer Bearbeitung erlauben. Damit ist die traditionelle Trennung aufgehoben, nach der zunächst versucht wurde, angemessene Testaufgaben zu konstruieren und diese dann den Probanden vorzulegen und entsprechend auszuwerten. Über die Aufgabenformulierung wird es möglich – das ist der Vorteil der Item Response-Theorie –, parallel die Schwierigkeitsparameter der Aufgaben und die Fähigkeitsparameter der Testpersonen zu skalieren.

Wie haben wir die relevanten Inhaltsbereiche gewonnen und festgelegt?

- Wir wählten unter mehreren Möglichkeiten als Bezugspunkt einen Industriebetrieb aus, der eine mittlere Größe besitzt und von der Produktstruktur her Produkte fertigt, die eine einfache Stückliste aufweisen und von ihrer Funktionalität her voll verständlich sind. Es handelt sich um ein Unternehmen, das Aluminiumverpackungen vor allem für Kosmetika, Getränke und Chemikalien produziert, z. B. die Dosen für Nivea-Creme.
- In dem Unternehmen führten wir Beobachtungen an den Arbeitsplätzen für Industriekaufleute durch; zugleich interviewten wir die Mitarbeiterinnen und Mitarbeiter bezüglich der Häufigkeit und Schwierigkeit der von ihnen zu erledigenden Aufgaben.
- Die so gewonnenen Ergebnisse wurden mit der Personal- und der Ausbildungsabteilung sowie mit dem Betriebsrat diskutiert und auf ihre Angemessenheit hin beurteilt.
- Wir erhielten von Auszubildenden ihre kontinuierlich auszufüllenden Lernberichte, die wir ebenfalls auswerteten.
- Des Weiteren analysierten wir die Ausbildungsordnung für Industriekaufleute sowie Ausbildungspläne verschiedener Betriebe.
- Für den Bereich der Berufsschule analysierten wir den KMK-Rahmenlehrplan sowie ausgewählte Lehrbücher.

Auf der Grundlage dieser Analysen sowie von Entwicklungen, die am Seminar für Wirtschaftspädagogik der Georg-August-Universität Göttingen seit vielen Jahren vorgenommen und dabei auch in die Niedersächsischen Lehrpläne und Richtlinien aufgenommen worden waren, konstruierten wir Geschäftsvorfälle zu den Gebieten Einkauf, Vertrieb und Arbeitsvorbereitung. Mit diesen Gebieten wird eine große Spannweite der Tätigkeiten von Industriekaufleuten abgedeckt, die es gestatten, sowohl von der Tiefe als auch von der Breite der Anforderungen her das Konstrukt der beruflichen Handlungskompetenz objektiv, reliabel und valide abzubilden. Für den Entwurf der Geschäftsvorfälle war dabei wichtig, welche Art von Arbeits- und Geschäftsprozessen abgebildet werden sollten. Gemäß der Literatur, aber auch aufgrund von Praxiserfahrungen spielen dabei betriebliche Wertschöpfungsprozesse eine entscheidende Rolle. Diese lassen sich für die Bereiche des Vertriebs und des Einkaufs gut modellieren. Zudem lassen sich über Simulationen Handlungen erfassen, die diesen Wertschöpfungsprozessen zugrunde liegen. Neben den Wertschöpfungsprozessen finden wir im Aufgabenbereich von Industriekaufleuten auch Steuerungsprozesse; diese wurden für den Bereich der Arbeitsvorbereitung modelliert. Grundlage für diese Aufteilung war die Annahme, dass Wertschöpfungs- bzw. Steuerungsprozesse eine unterschiedliche kognitive Verarbeitungsstruktur aufweisen. Aufgrund der Unterschiede in den Anforderungen darf vermutet werden, dass die Auszubildenden zu ihrer Bewältigung auf unterschiedliche berufsfachliche Fähigkeiten zurückzugreifen haben. Daneben wurde versucht zu erfassen, welche verstehensbasierten Kompetenzen gegeben sind. Die entsprechenden Items sind auf die Adaptation ausgewählter betriebswirtschaftlicher Konzepte in den drei genannten betrieblichen Bereichen bezogen.

Für die ausgewählten Gebiete haben wir Aufgaben entwickelt, die zu ihrer Lösung eine Testzeit von vier bis fünf Stunden erfordern – und damit weit über dem Limit für den inhaltsbezogenen Teil eines VET-LSA liegen. Anders formuliert: Die Breite und Tiefe der von uns entwickelten Aufgaben übertreffen quantitativ erheblich das Insgesamt an Aufgaben, wie sie im Rahmen eines VET-LSA aus Zeit- und auch aus Motivationsgründen vorgegeben werden können. Mit dieser Breite und Tiefe veranschaulichen die Aufgaben aber auch in qualitativer Hinsicht die Möglichkeiten einer Erfassung beruflicher Handlungskompetenz.

## 2.2 Auswahl des Testformats

In der Machbarkeitsstudie zum Berufsbildungs-PISA (vgl. BAETHGE ET AL., 2006) wurde herausgestellt, dass es notwendig ist, für die Erfassung von berufsfachlicher Kompetenz arbeitsplatzrelevante komplexe Aufgaben einzusetzen. Vor diesem Hintergrund galt es sicherzustellen, dass die Testformate, mit denen die Aufgaben präsentiert werden, als „authentisch“ gelten können (vgl. hierzu ACHTENHAGEN, 2001; ACHTENHAGEN & WEBER, 2003).

### 2.2.1 Inszenierung von Authentizität

Die folgenden Ausführungen verdeutlichen, wie versucht wurde, eine größtmögliche Nähe zur beruflichen Ausbildungsrealität zu wahren. Im Vorgriff kann bereits hier darauf hingewiesen werden, dass berufliche Handlungskompetenz immer im Hinblick auf eine berufstypische Aufgabenbewältigung definiert ist und entsprechend auch in

den Zwischen- und Abschlussprüfungen zu ermitteln versucht wird; denn ohne die Annahme des „Berufstypischen“ machte es keinen Sinn, die Verbände und die Kammern als koordinierende Institutionen einzuschalten. Der ausschließliche Rückgriff und die damit verbundene Einschränkung auf arbeitsplatz- und betriebsspezifische Anforderungen würde gegen die übergeordneten Ziele der Berufsbildung verstoßen: nicht nur im Sinne einer Bereitstellung von ausreichenden Humanressourcen für die Gesellschaft, sondern auch im Sinne einer breit angelegten Persönlichkeitsentwicklung. Von daher waren unsere Aktivitäten bei der Formulierung der Testinhalte und bei der Entwicklung geeigneter Testformate darauf bezogen, die Urteile von Fachleuten einzuholen, die sich sowohl auf den Erfolg am Arbeitsplatz – als Erbringung der Arbeitsleistung – als auch auf den Erfolg der Ausbildungsmaßnahmen zur Ermöglichung des Erfolgs am Arbeitsplatz bezogen. Anders formuliert: Für den gesamten Prozess der Testentwicklung haben wir auf betriebliche und schulische Expertisen zurückgegriffen, mit deren Hilfe wir sicherstellten, dass unsere Vorschläge den berufstypischen Anforderungen an kaufmännischen Arbeitsplätzen entsprechen (hier haben wir den Schulleitern und Fachleitern sowie der AKA zu danken, die unsere Aufgabenvorschläge kritisch durchgesehen haben).

In der neueren Lerntheorie, vor allem im Bereich des „Situational Learning“ (vgl. RESNICK, 1987; GREENO, 1998; BYRNES, 2008), wird darauf verwiesen, dass es darauf ankomme, Lernprozesse zu favorisieren, „wie sie im Leben zu finden sind“. Dabei wird dem Lernen in der Arbeitswelt per se „Authentizität“ zugesprochen (vgl. ACHTENHAGEN & WEBER, 2003, S. 187). Will man diesem Kriterium gerecht werden – und das trifft auch auf die in unserem Projekt vorgenommene Inhaltsmodellierung und die Entwicklung des Testformats zu –, dann gilt es, die Modellierungsprozesse für Lern- und Arbeitsbedingungen sowie das „Authentizitätsparadox“ zu berücksichtigen. Eine entscheidende Einsicht, die aus der Kritik an der „Situational Learning“-Bewegung resultiert, ist es, dass „Authentizität“ nicht per se existiert, sondern jeweils inszeniert werden muss. Wenn „berufliche Handlungskompetenz“ gemessen werden soll, dann sind die Situationen, in denen sie zu zeigen ist, zu modellieren. AKA-Tests, Klassenarbeiten, aber auch betriebliche Arbeitsproben gehen davon aus, dass die zu zeigenden Leistungen sich auf berufstypische Situationen bzw. Handlungen beziehen; denn es wäre nicht möglich, jeweils situationsgebunden betriebsspezifisch und arbeitsplatzspezifisch Aufgaben zu formulieren. Anders herum gilt aber ebenfalls: Auch Arbeitsproben, die im Betrieb zu erledigen sind, treffen nicht übergreifend eine arbeitsplatztypische Situation, wie sie in der Arbeitsprobe vorausgesetzt wird. Die jeweilige Arbeitsprobe ist dann einer Situation anzupassen, die sich in ihrer Spezifität auf bestimmte Kunden oder Lieferanten, auf Preise sowie Lieferkonditionen im gerade aktuellen Fall beziehen muss. Anderenfalls ist die Situation – unter dem Versuch der Wahrung ihrer „Authentizität“ – gesondert zu konstruieren. In Rückmeldungen von Betrieben, die sich mit den ihnen von uns geschickten Arbeitsproben auseinandergesetzt haben, wurde genau dieser Sachverhalt herausgestellt.

Was ist unter „Authentizität“ in diesem Zusammenhang zu verstehen? – Ein Blick in die Literatur zeigt, dass dieser Begriff eher schlagwortartig gebraucht und nicht im Hinblick auf empirische Forschungsvorhaben hin formuliert und operationalisiert wird (vgl. ausführlich hierzu ACHTENHAGEN & WEBER, 2003). DOYLE (2000) hat verschiedene Ansätze zur Bestimmung und Realisierung von „Authentizität“ einer tiefgreifenden Kritik unterzogen, wobei er insbesondere die Versuche der Cognition & Technology

Group at Vanderbilt (CTGV, 1991) kritisiert, die durch die „Jasper Woodbury“-Serie weltweit die Diskussion um das Situated Learning dominiert hat. Er schlägt vor, die Problematik in drei Dimensionen zu diskutieren:

- (a) Lernerbezogene Authentizität: Hier geht es um das Anknüpfen an das Vorwissen und die spontanen Interessen der Lernenden.
- (b) Fach- und inhaltsbezogene Authentizität: Hier handelt es sich um reale Daten, Instrumente und Operationen einer Fachdisziplin, wie sie im Curriculum repräsentiert sind; ermöglicht werden soll ein forschendes Lernen mit realen Gegenständen.
- (c) Situationsbezogene Authentizität: Hier ist die Partizipation in realen Situationen einer „Community of Practice“ angesprochen, d. h. es geht um ein Lernen, mit dem eine Enkulturation in eine Expertenkultur angestrebt wird.

Diese Aufteilung nach drei Dimensionen lässt sich auf die Forderung nach Erfassung beruflicher Handlungskompetenz beziehen: In Abhängigkeit von den personalen Voraussetzungen unter Berücksichtigung der Vorgaben von Ausbildungsordnung, Ausbildungsplan und Rahmenlehrplan geht es um das Hineinwachsen in die Anforderungen des jeweiligen Arbeitsplatzes. Hier wird sehr deutlich, worin die Vorteile des deutschen Systems der Dualen Berufsausbildung liegen: in der systematischen Unterfütterung der fallbezogenen Erfahrungen am jeweiligen Arbeitsplatz. An dieser Stelle ist auch die Bedeutung einer kritischen Betrachtung des Authentizitäts-Problems gegeben: Es wäre naiv anzunehmen, dass berufliche Handlungskompetenz allein aus einer naturwüchsigen Erfahrung am Arbeitsplatz entstünde; es kommt vielmehr darauf an, die Authentizität der Lernerfahrung zu inszenieren, um den Lernerfolg systematisch zu sichern. Auch im Dualen System ist die „Realität“ der Berufserfahrung nicht von allein gegeben; sie muss vermittelt werden. Das Paradox dabei ist – und das wird in der politischen Diskussion oft überspielt –, dass „Authentizität“ vermittelt wird, um Unvermitteltheit zu demonstrieren (hier ist an Politikerinterviews anzuknüpfen, die z. B. im Fernsehen mit ihrer „Spontaneität“ zuvor im Detail exakt inszeniert sind). „Realität“ wird erst über Übersetzungsleistungen erschlossen (vgl. hierzu auch die Diskussionen zur Repräsentation von Wissen und Erfahrungen (MARKMAN, 1999; MISLEVY, 2007).

### 2.2.2 Unternehmenssimulation ALUSIM

Die Beziehung zwischen Kontext, Inhalt und Design eines Tests wird in modernen Assessmentmodellen auf der Basis kognitiver Theorien hergestellt (EMBRETSON, 2002; MISLEVY, 2007; WILSON, 2008). Diese Modelle veranschaulichen, dass die Entwicklung eines Tests Modellierungsschritte voraussetzt, die die Vorstellungen von Realität im Hinblick auf die Ideen und Ziele der Testung übersetzen. Erst auf Basis dieser Übersetzungsleistung können die Inhalte identifiziert, ausformuliert und für die Testaufgaben in Abstimmung mit dem Testformat aufbereitet werden. Zusammenfassend sind hierfür vier Entwicklungsschritte notwendig (WINTHER, in Vorbereitung):

- Die Analyse von Inhalten, die als domänentypisch gelten können,
- die Ausformulierung von Arbeits- und Denkstrukturen vor dem Hintergrund der identifizierten Inhalte auf Basis kognitiver Theorien,
- die Abstimmung der Arbeitsmittel und -methoden sowie

- die Wahl eines geeigneten psychometrischen Modells, mit dem die Besonderheiten der Aufgaben und die unterschiedlichen Fähigkeitsausprägungen der Lernenden / Auszubildenden abgebildet werden können.

Ziel des Projekts war es, spezifische berufliche Erfordernisse für das erfolgreiche Ausführen beruflicher Handlungen zum Ausgangspunkt der Kompetenzmessung zu machen. Die Handlungsprodukte stellen damit die Grundlage für die Beschreibung von berufsfachlicher Kompetenz dar. Im kaufmännisch-verwaltenden Bereich sind diese Produkte i. d. R. durch Bearbeitungs- und Entscheidungsprozesse erzeugte Daten bzw. Datenzusammenfassungen, die unterschiedlich komplex sind und verschiedene Formate aufweisen können. Beispiele hierfür sind Daten in Auftragsbestätigungen oder Bestellformularen bzw. auf Daten basierende Kundenkorrespondenz. Die Daten ihrerseits werden auf Basis betrieblicher Arbeits- und Geschäftsprozesse erzeugt, die sich über ein Zusammenspiel verschiedener Arbeitsmittel, unternehmensinterner ERP-Software und kaufmännischer Entscheidungsalgorithmen modellieren lassen. Für die Messung berufsfachlicher Kompetenz wurden diese internen Abläufe und Produkte simuliert, d. h. authentisch für das Testformat übersetzt.

Ausgangspunkt der Entwicklung einer betrieblichen Unternehmenssimulation als Testformat für berufsfachliche Kompetenz war die Novelis Deutschland GmbH – Werk Göttingen, deren reale Arbeits- und Geschäftsprozesse so „authentisch“ wie möglich abzubilden waren. Die Modellierung dient dem Zweck, ausgewählte domänentypische Geschäftsvorfälle als Wertschöpfungs- bzw. Steuerungsprozesse im Werk Göttingen der Novelis GmbH abzubilden. Diese Geschäftsvorfälle sollten eine Struktur aufweisen, die es möglich macht, Items zu konstruieren, mit deren Hilfe berufsfachliche Kompetenz gemessen werden kann. Für das Testformat bedeutet dies, dass eine authentische Beschreibung und Erfassung der Handlungsorientierung sowie des Handlungsvollzuges in den Anforderungssituationen möglich werden. Das jeweilige Testformat muss daher

- reale Arbeitsprozesse (Verhandlungen, Arbeitsverteilung in der Interaktion),
- reale Geschäftsprozesse (z. B. die konkrete Abwicklung einer Bestellung),
- reale kaufmännische Entscheidungen (z. B. Ermittlung eines Liefertermins).

abbilden können. Dies geschieht mit Hilfe der Unternehmenssimulation ALSUSIM. Das Testformat ALUSIM ist web-basiert und enthält

- eine allgemeine Einführung in die Unternehmensstruktur (einschließlich der Unternehmenshistorie);
- einen animierten Arbeitsplatz (Schreibtisch mit Zugriff auf sämtliche Ablagen und die simulierte Software; vgl. Abbildung 2) sowie
- ein animiertes Sideboard, über die die web-basierten Situationen gesteuert werden;
- allgemeine Zusammenstellungen zur geschäftlichen Lage des Unternehmens (textbasierte Analyse der Geschäftszahlen sowie Darstellungen der Bilanz, Erfolgsrechnung und Geldflussrechnung).

In der Unternehmenssimulation sind reale Betriebsprozesse der Bereiche Vertrieb, Einkauf und Arbeitsvorbereitung verarbeitet. Für diese Bereiche wurden typische Arbeitssituationen als Modell zur Abbildung von domänenspezifischen Geschäftsvorfällen herausgegriffen, so z. B. für den Vertrieb der Eingang einer Bestellung per Fax. Eine Erweiterung dieser Bereiche sowie die Hinzunahme zusätzlicher Inhalte



Impressum: Georg-August-Universität Göttingen, Professur für Wirtschaftspädagogik, Platz der Göttinger Sieben 5, 37073 Göttingen

Fotos: 2008, NOVELIS INC.

Abb. 2: Simulierter Arbeitsplatz

sind aufgrund der umfangreichen Modellierung der simulierten Unternehmung jederzeit möglich. Eine Vorbemerkung ist an dieser Stelle allerdings nötig: Die Beschränkung auf die Inhalte Vertrieb, Einkauf und Arbeitsvorbereitung erfolgte aufgrund der Abstimmung der Ausbildungsrichtlinien im internationalen Vergleich. Sowohl Personalfragen als auch Probleme des Rechnungswesens, die im deutschen Ausbildungsplan enthalten sind, bleiben wegen der nur schwierig möglichen internationalen Vergleichbarkeit ausgeklammert, können aber unschwer für weitere Testungen einbezogen werden. Für die Erfassung berufsfachlicher Kompetenz kann die vorgenommene Eingrenzung vernachlässigt werden, da mit der entwickelten Aufgabenstruktur die entsprechenden Dispositionen adäquat gemessen werden können (vgl. Abschnitt 3). Das Testformat gibt für die drei betrieblichen Bereiche den organisationalen und strukturellen Rahmen vor.

- Der einzelne Testbereich wird mit einem Videoclip eingeleitet, der eine Büro-situation zeigt. Dieser Videoclip bezeichnet die zu lösende berufliche Anforderungssituation und gibt z. T. Hinweise, was zu tun sei. Für jeden Bereich gibt es mehrere solcher Videoclips, die in ihrer Abfolge den Sequenzcharakter der Geschäftsprozesse verdeutlichen.
- Die Anforderungssituationen sehen z.B. vor, dass ein eingegangenes Dokument zu bearbeiten ist, indem es in das ERP-System der ALUSIM einzugeben, über Informationen aus Produkt-, Kunden- oder Lieferantendateien anzureichern oder in eine Antwort an einen Kunden oder Lieferanten oder eine Anfrage an eine andere Unternehmensabteilung umzusetzen ist.

- Für die Bearbeitung der Anforderungssituation stehen über den web-basierten Schreibtisch die Informationen, aber auch die Kommunikationsmedien, wie Fax, Brief oder E-Mail, zur Verfügung.
- Die Lösung der Anforderungssituation ist in einem Dokument festzuhalten, das als Handlungs- bzw. Entscheidungsprodukt die Basis der Kompetenzmessung darstellt.
- Zur Ermittlung der berufsfachlichen Kompetenz sind in der Simulation ALUSIM zwei verschiedene Testbereiche implementiert: die beruflichen Simulationsaufgaben zur Abbildung handlungsbasierter Kompetenz und aus den Simulationsaufgaben abgeleitete, eher schulbezogene Anwendungsaufgaben zur Messung verstehensbasierter Kompetenz.

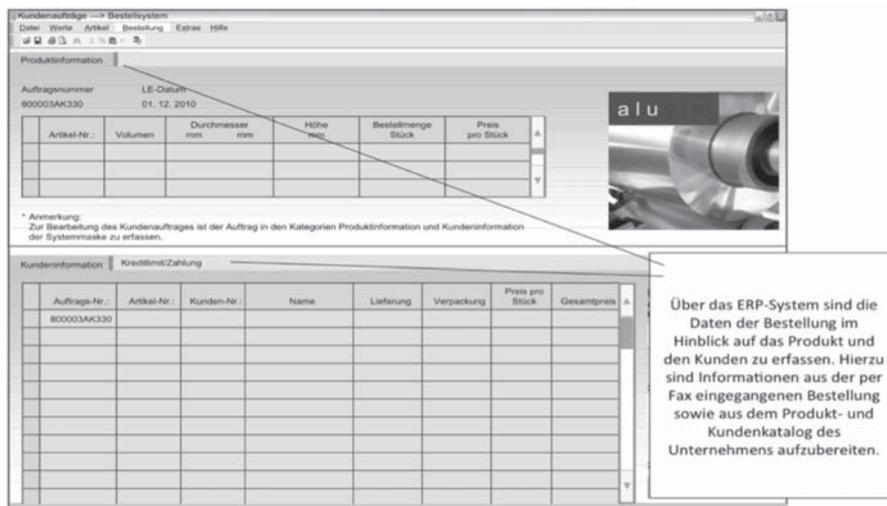


Abb. 3: Betriebliches ERP-System

Anhand von Inhaltsbeschreibungen aus dem Bereich Vertrieb soll demonstriert werden, wie Testformat und Testinhalt vor dem Hintergrund eines passenden Messmodells parallel konstruiert wurden: In einem ersten Videoclip wird der Auszubildenden mitgeteilt, dass als Reaktion auf den halbjährlich verschickten Produktkatalog ein Fax mit einer Bestellung eingegangen sei. Die Informationen aus dem Fax sind in das ERP-System des modellierten Unternehmens ALUSIM einzugeben. In Fortführung des Geschäftsprozesses bezieht sich ein zweiter Videoclip auf eine E-Mail aus der Produktionsabteilung, in der mitgeteilt wird, dass der Produktionstermin für die vom Kunden bestellte Ware nicht eingehalten werden kann. Die Auszubildende erhält die Aufgabe, anhand des ERP-Systems den neuen Termin zu bestimmen und den Kunden entsprechend zu benachrichtigen. Die Aufgabenlösung besteht hier in der Information des Kunden – wobei für die Bewertung der Lösung auch sozial-kompetentes Verhalten mit erfasst wird: die Gestaltung der Nachricht sowie die Art der Begründung.

### 3 Auswertung der Ergebnisse

Im Hinblick auf die Anforderungen eines Large Scale-Assessment in der beruflichen Bildung basieren die Auswertungen des Projekts zur Messung berufsfachlicher Kompetenz auf den Modellen der Item Response-Theorie (IRT). Die Vorteile der IRT kommen insbesondere dann zur Geltung, wenn eine möglichst große Zahl von Testaufgaben, die verschiedene Personeneigenschaften im Sinne von Kompetenzprofilen erfassen sollen, möglichst vielen Testpersonen vorgelegt werden können. Im Rahmen des geplanten VET-LSA beziehen sich diese Testaufgaben nicht nur auf Items zur Messung beruflicher Handlungsfähigkeiten; es werden zusätzliche Testitems enthalten sein müssen, die insbesondere im Hinblick auf Konzepte des Selbst und auf die individuelle Entwicklung ergänzende Informationen bereitstellen können. Das impliziert – vergleichbar mit der Ausgestaltung der PISA-Studien – ein Multi-Matrix-Design. In letzter Konsequenz bedeutet dies, dass für ein solches Testdesign, in dem nicht alle Testpersonen aus Zeitgründen alle Testitems bearbeiten können, Modelle zu formulieren sind, mit denen dennoch eine alle Informationen ausschöpfende Auswertung möglich ist. Die Modelle der IRT – insbesondere die aus der Rasch-Familie – weisen Eigenschaften auf, die für eine Kompetenzmessung unter der Bedingung eines internationalen Vergleichs und im Hinblick auf ein aus Zeitgründen notwendiges Multi-Matrix-Design unabdingbar sind. Vor dem Hintergrund dieser Überlegungen ist auch die Expertise von ROST im Rahmen der Machbarkeitsstudie „Berufsbildungs-PISA“ zu lesen (ROST, 2006). In dieser Expertise kommt ROST aufgrund der zu erwartenden Testparameter – Längsschnittdesign, Verlinkung der Teilstichproben sowie Effekte auf die Itemschwierigkeiten – zu dem Ergebnis, den Einsatz von IRT-Modellen für die Messung berufsbezogener Kompetenzen zu präferieren.

Es sind insbesondere zwei Eigenschaften, die die Verwendung von IRT-Modellen sinnvoll erscheinen lassen (ROST, 2004): (1) IRT-Modelle erlauben es, Personen- und Itemparameter parallel zu skalieren. Das Antwortverhalten auf ein Testitem wird dadurch ausschließlich durch die Itemschwierigkeit und die Personenfähigkeit bestimmt. Damit liegen mit IRT-Modellen der Rasch-Familie komplette Verhaltensmodelle vor, die für die Person und für das Testitem durch differenzierte Parameter zu beschreiben sind. Diese Separabilität von Personen- und Itemmerkmalen hat zur Folge, dass die Aussagen über die Personen, die einen Test bearbeitet haben, und über die Items dieses Tests nur von den Merkmalen der Personen und Items abhängen, nicht jedoch von den Methoden, mit denen diese Merkmale gemessen worden sind – dies setzt in letzter Konsequenz standardisierte Testumgebungen voraus, da Variationen der Situation mit Modellen der IRT bislang nicht modelliert werden können. Mit der Entscheidung für solche Modelle sind damit direkte Messungen an kaufmännischen Arbeitsplätzen, die sich erheblich in Abhängigkeit u. a. der Branche und der Betriebsgröße voneinander unterscheiden, ausgeschlossen, da eine Vergleichbarkeit der Situationen nicht gewährleistet werden kann. (2) Personen- und Itemparameter sind unabhängige Variablen, die separat geschätzt werden können. Rasch beschreibt diese Eigenschaft als spezifische Objektivität und misst ihr eine essentielle messtheoretische Bedeutung bei (RASCH, 1960, p. 104/105). Spezifische Objektivität im Rahmen von IRT-Modellen bedeutet, dass die Personenparameter von der Stichprobe der Items und die Itemparameter von der Stichprobe der Personen unabhängig sind. Nur so sind Vergleiche zwischen Personen sowie Vergleiche zwischen Items zulässig (Stichprobenunabhängigkeit).

Die Personenparameter sind demnach unabhängig (1) von den Items, mit denen sie gemessen worden sind, und (2) von den Merkmalen anderer Personen; die Itemparameter sind hingegen unabhängig (1) von dem Antwortverhalten der Personen und (2) von den Merkmalen anderer Items des Tests.

Für die Praxis von Tests bedeutet dies, dass Unterschiede zwischen den Leistungsfähigkeiten von Personen ohne Störfaktoren geschätzt werden können und dass für eine reliable Festlegung der Personenfähigkeit nur Ausschnitte der latenten Variablen abgebildet werden müssen. Für die vorliegende Studie lässt sich folglich aus den drei verschiedenen Testbereichen „Vertrieb“, „Einkauf“ und „Arbeitsvorbereitung“ ein reliabler und stabiler Schätzer der Kompetenz unter der Bedingung ermitteln, dass die einzelnen Testbereiche durch unterschiedlich schwierige Testitems valide repräsentiert sind. Im Umkehrschluss ließe sich folglich im Hinblick auf ein international vergleichendes VET-LSA die Kompetenz eines Auszubildenden über einen einzigen betrieblichen Bereich erfassen, wenn dieser die Kompetenzskala in ausreichender Breite definiert.

### 3.1 Beschreibung der Stichprobe

Es wurden insgesamt 264 Auszubildende (3. Ausbildungsjahr, sieben Berufsschulen, 61 Betriebe, vier Bundesländer) erfasst; 56,5 Prozent von ihnen sind weiblich. Das Alter der Auszubildenden variiert zwischen 18 und 34 Jahren, wobei 83,4 Prozent der Befragten in der Altersgruppe von 20 bis 23 Jahren liegen – dies ist typisch für eine Abschlussklasse im Rahmen einer drei-jährigen kaufmännischen Berufsausbildung.

Information	Beschreibung	Ergebnis
Geschlecht	weiblich:	56,5%
	männlich:	43,5%
Alter	Streuung:	18 bis 34 Jahre
	Häufung:	20 bis 23 Jahre (Anteil = 83,4%)
Höchster Schulabschluss	allgemeine Hochschulreife:	43,1%
	Fachhochschulreife:	40,5%
	Sekundarabschluss I:	10,7%
	qualifizierter Hauptschulabschluss:	0,8%
Betriebsgröße	bis 50 Beschäftigte:	6,1%
	51 bis 500 Beschäftigte:	38,2%
	501 bis 2500 Beschäftigte:	27,1%
	mehr als 2500 Beschäftigte:	24,8%
Branche (Auswahl)	Metallverarbeitung:	19,9%
	Holzverarbeitung:	10,1%
	Chemie:	10,8%

Abb. 4: Beschreibung der Stichprobe (N = 264)

Mit dem Sample konnte eine angemessene Verteilung der Auszubildenden im Hinblick auf die Branchen und die Betriebsgrößen realisiert werden. Nur sechs Prozent der Auszubildenden absolvieren ihre Ausbildung in Kleinbetrieben (bis 50 Beschäftigte); die Mehrzahl lernt in Unternehmen mit bis zu 500 (38,2 Prozent) und

24,8 Prozent in Unternehmen mit mehr als 2.500 Beschäftigten. 19,9 Prozent der Auszubildenden kommen aus der Metallverarbeitung, 10,8 Prozent aus der Chemie und 10,1 Prozent aus der Holzverarbeitung. Darüber hinaus sind u. a. Auszubildende aus den Bereichen Energieversorgung, Bau sowie Automobilzulieferung im Sample enthalten. Die Stichprobe zeigt, dass die Auszubildenden im Ausbildungsberuf des Industriekaufmanns / der Industriekauffrau höhere allgemeine Voraussetzungen mitbringen – 43,1 Prozent der Auszubildenden verfügen über die allgemeine Hochschulreife und 40,5 Prozent über die Fachhochschulreife; nur zwei Probanden haben einen qualifizierten Hauptschulabschluss.

### 3.2 Fragestellungen der Untersuchung

Die nachfolgenden Aussagen zur Auswertung der Ergebnisse beziehen sich ausschließlich auf den Einsatz der web-basierten Unternehmenssimulation ALUSIM. Das Projekt umfasste zusätzlich im Hinblick auf ALUSIM validierte Arbeitsproben, deren Rücklauf aus den Betrieben bislang jedoch noch keine aussagekräftige Analyse zulässt. Insbesondere im Hinblick auf ein international vergleichendes Large Scale-Assessment können Effekte von Interesse sein, in denen sich die Leistungsfähigkeit der Auszubildenden auf individuelle und institutionale Bedingungen beziehen lässt. Hierfür sind Analysen und Auswertungen im Mehrebenenmodell vorzunehmen. Diese Modellierung setzt jedoch für stabile und aussagekräftige Ergebnisse eine sehr große Stichprobe voraus – die im Rahmen eines VET-LSA realisiert werden könnte; für die vorliegende Studie ist der Stichprobenumfang allerdings zu klein, so dass im Folgenden ausschließlich Ergebnisse zu den Kompetenzskalen und Kompetenzstufen beruflicher Handlungskompetenz auf Basis des Kompetenzstrukturmodells für die kaufmännische Bildung (WINTHER & ACHTENHAGEN, 2008) präsentiert werden.

Insbesondere folgende Fragestellungen standen im Fokus:

- (1) Wie lässt sich berufliche Handlungskompetenz über repräsentative Testitems beschreiben? – In diesem Zusammenhang ist auch zu klären, ob sich berufliche Handlungskompetenz über eine eindimensionale oder mehrdimensionale Kompetenzstruktur abbilden lässt.
- (2) Lässt sich im Hinblick auf die Personenfähigkeit ein Unterschied zwischen handlungsbasierten Simulationsaufgaben und verstehensbasierten Anwendungsaufgaben empirisch sichern? – Die Hypothese ist, dass handlungsbasierte Simulationsaufgaben durch ihren Authentizitätsgehalt und ihre Nähe zur betrieblichen Praxis in einer Ingesamtanalyse von Auszubildenden des Dualen Systems besser gelöst werden können als Aufgaben, die stärker von der betrieblichen Praxis abstrahieren und deren vorrangiger Sinn darin liegt zu testen, ob betriebswirtschaftliche Konzepte verstanden worden sind und ob diese vor dem Hintergrund simulierter Unternehmensprozesse von den Auszubildenden interpretiert und validiert werden können.
- (3) Lässt sich für berufliche Handlungskompetenz ein Kompetenzniveaumodell formulieren? – Dies erfordert Aussagen darüber, welche Itemeigenschaften charakteristisch für mögliche Kompetenzstufen sind und über welche Merkmale ein Item demnach als schwierig oder als leicht zu identifizieren ist. Im Zusammenhang mit dem Kompetenzstrukturmodell für die kaufmännische Bildung (WINTHER & ACHTENHAGEN, 2008) wurden hierfür Anforderungsparameter vorgestellt, mit

denen (a) die Komplexität von Anforderungssituationen systematisch variiert werden kann und anhand derer (b) Testpersonen verschiedenen Kompetenzstufen zuzuordnen sind. Der Anspruchsgehalt einer beruflichen Anforderungssituation ist demnach abhängig von (1) der funktionalen Modellierung, (2) der inhaltlichen Komplexität und (3) von der Art der kognitiven Taxonomierung.

Zur Beantwortung dieser Fragen wurde für die Studie folgendes posttest only-Design gewählt (Abbildung 5):

	Testbereich Einkauf	Testbereich Vertrieb	Testbereich Arbeitsvorbereitung	
Testgruppe 1	182	182	182	182
Testgruppe 2	43	–	–	43
Testgruppe 3	–	39	39	39
INSGESAMT	225	221	221	264

Abb. 5: Design der Erhebung (mit Angabe der Zahl der Testpersonen)

Insgesamt haben 264 Auszubildende an der Erhebung teilgenommen. 43 davon haben ausschließlich den Testbereich „Einkauf“ (Testgruppe 2) und 39 ausschließlich die Testbereiche „Vertrieb“ und „Arbeitsvorbereitung“ (Testgruppe 3) im Rahmen der Simulation ALUSIM gelöst; die von den Auszubildenden in den Testgruppen 2 und 3 jeweils nicht behandelten betrieblichen Bereiche waren in Form von authentisch inszenierten Arbeitsproben in ihrem Ausbildungsbetrieb zu bearbeiten. Beide Testformate sind durch Ankeritems miteinander verbunden, so dass durch eine Verlinkung der jeweiligen Teilstichproben generalisierbare Aussagen möglich sind. Hierfür ist allerdings der Rücklauf der Arbeitsproben aus den Betrieben noch zu gering.

Die einzelnen Testbereiche sind wie folgt über Items der handlungsbasierten Simulationsaufgaben und verstehensbasierten Anwendungsaufgaben repräsentiert (Abbildung 6):

	Testbereich Einkauf	Testbereich Vertrieb	Testbereich Arbeitsvorbereitung	
Itemanzahl der handlungsbasierten Simulationsaufgaben	18	12	4	34
Itemanzahl der verstehensbasierten Anwendungsaufgaben	11	12	3	26
INSGESAMT	29	24	7	60

Abb. 6: Itemanzahl in den zwei verschiedenen Testformen

Zur Erfassung der berufsfachlichen Kompetenz liegen insgesamt 60 Items vor. 34 Items dienen der Erfassung handlungsbasierter Kompetenz über Simulationsaufgaben innerhalb betrieblicher Anforderungssituationen und 26 Items der

Erfassung verstehensbasierter Kompetenz über Anwendungsaufgaben vor dem Hintergrund betrieblicher Anforderungssituationen. Hinsichtlich der verschiedenen Testbereiche fällt auf, dass der Bereich der „Arbeitsvorbereitung“ in Relation zu den zwei anderen Testbereichen unterrepräsentiert ist. Dies liegt daran, dass dieser Testbereich ausschließlich den Umgang mit dem betriebswirtschaftlichen Konzept der Maschinenstundensatzberechnung umfasst – eine Ausweitung dieses auf die Steuerungsprozesse von Unternehmen abzielenden Testbereichs ist jedoch auf Grundlage der umfassend modellierten Datenstruktur jederzeit möglich. Das stärkere Gewicht der Simulation lag auf betrieblichen Wertschöpfungsprozessen. Hierfür sind insbesondere zwei Gründe ausschlaggebend gewesen: (1) Die Analyse der Ausbildungsplatzanforderungen zeigt, dass die Bereiche Vertrieb und Einkauf für einen internationalen Vergleich von besonderem Interesse sind, da mit ihnen die Internationalität der Geschäftsbeziehungen, Aspekte der Kunden- und Lieferantenbindungen und insgesamt ein Verständnis für betriebliche Geschäfts- und Arbeitsprozesse getestet werden können. (2) Die Analyse der deutschen Curricula hat deutlich gemacht, dass – will man weite Teile des Rechnungswesens von einem internationalen Vergleich ausschließen – der inhaltliche Fokus auf dem Bereich der betrieblichen Wertschöpfung liegt.

Es sei darauf hingewiesen, dass sich die Studie vorrangig mit der Erfassung domänenspezifischer Kompetenz in betrieblichen Arbeits- und Geschäftsprozessen befasst. Die theoretische Konzeption des Kompetenzstrukturmodells für die kaufmännische Bildung geht darüber hinaus jedoch auch von domänenverbundenen Fähigkeitsstrukturen aus (WINTHER, 2007; WINTHER & ACHTENHAGEN, 2008). Domänenverbundene Kompetenzen in Abgrenzung zu den domänenspezifischen Kompetenzen lassen sich als verfügbare Repräsentationen und/oder Verhaltensmuster charakterisieren, die stark ausführungsbezogen sind, um die Bewältigung einer beruflichen Anforderungssituation zu unterstützen (GREENO, 1998). Die Anforderungen der vorliegenden Studie liegen jedoch vorrangig darin zu prüfen, wie domänenspezifische Situationen in Testitems zu transformieren sind.

Darüber hinaus sollte eine Testform entwickelt werden, die in der Lage ist, handlungsrelevante Kenntnisse und Fähigkeiten in beruflichen Anforderungssituationen zu testen. Eine der zentralen Fragen vor diesem Hintergrund ist, wie es die Probanden verstanden haben, mit der ungewohnten Testumgebung umzugehen. In diesem Zusammenhang ist anzumerken, dass die Verfasser dieser Studie keinen Zweifel daran hegten, dass Auszubildende im Beruf des Industriekaufmanns / der Industriekauffrau über ausreichende Computer Literacy verfügen. Es wurde im Gegenteil davon ausgegangen, dass der Einsatz einer web-basierten Unternehmenssimulation mit den enthaltenen arbeitsplatzrelevanten Komponenten wie ERP-Software, Programme der Datenverarbeitung und E-Mail client-Software die Bearbeitung von Testsituationen erleichtert und zudem motivationsförderlich wirkt. Um dennoch in diesem Bereich Fragen auszuschließen, enthalten die handlungsbasierten Simulationaufgaben Items, mit denen sichergestellt werden soll, dass die Probanden die Funktionsweise der Simulation verstanden haben und zugleich wissen, wie die zur Anforderungsbewältigung notwendigen Informationen aus den einzelnen Elementen der Simulation beschafft werden können. Die Analyse in diesem Punkt zeigt, dass die Items, die einen sicheren Umgang mit dem Testformat anzeigen, Lösungsquoten von mehr als 95 Prozent aufweisen, so dass davon ausgegangen werden kann, dass die Probanden den Umgang mit dem Testformat beherrschen.

### 3.2 Die Dimensionen beruflicher Handlungskompetenz

Der Test zur Erfassung berufsfachlicher Kompetenz wurde auf Basis des Kompetenzstrukturmodells für die kaufmännische Bildung konstruiert (WINTHER & ACHTENHAGEN, 2008; 2009; ACHTENHAGEN & WINTHER, 2008). Das Kompetenzstrukturmodell weist zwei differenzierte Ebenen auf: (1) Auf der Zugriffsebene wird beschrieben, welche kognitiven Prozesse zur Bearbeitung eines Inhaltsbereichs vorrangig zu aktivieren sind. Das Modell unterscheidet hier zwischen drei verschiedenen Kompetenzstrukturen: „conceptual“, „procedural“ und „interpretative“. Vor diesem Hintergrund setzt sich der Test zur Messung berufsfachlicher Kompetenz aus zwei zentralen Testbereichen zusammen, die zugleich die verschiedenen Dimensionen berufsfachlicher Kompetenz repräsentieren sollen: Einem Testteil, der handlungsbasierte Kompetenz über Simulationsaufgaben innerhalb betrieblicher Situationen misst und damit auf die Kompetenzdimensionen „procedural“ und „interpretative“ abstellt, und einem ergänzenden Testteil, der verstehensbasierte konzeptuale Kompetenz vor dem Hintergrund betrieblicher Situationen erfasst.

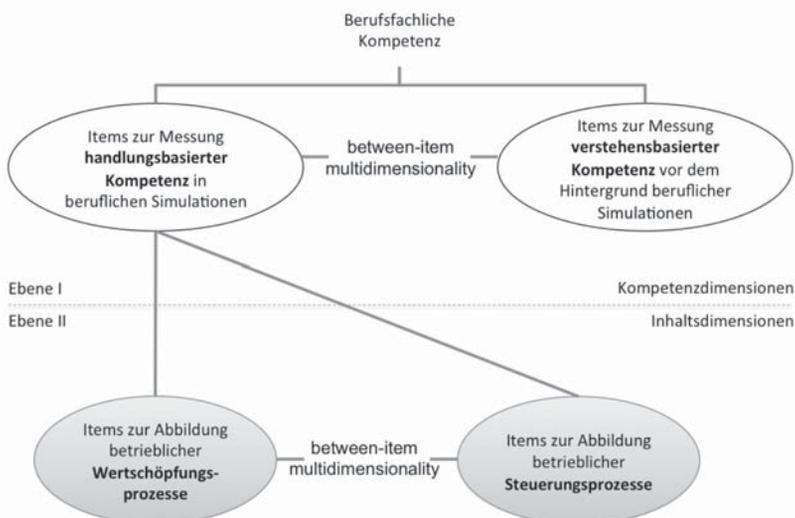


Abb. 7: Dimensionalität berufsfachlicher Kompetenz

(2) Auf der Inhaltsebene wird zentral zwischen domänenverbundenen und domänenspezifischen Kompetenzdimensionen unterschieden, wobei sich domänenverbundene Kompetenz über die Bereiche „economic literacy“ und „economic numeracy“ beschreiben lässt. Diese Bereiche sind keine bzw. nur marginale Bestandteile des vorliegenden Projekts und werden an dieser Stelle nicht weiter ausgeführt. Für den domänenspezifischen Bereich sind variierende Geschäftsvorfälle vorgesehen. Hierfür wurden in dem vorliegenden Projekt Geschäftsvorfälle im Hinblick auf zwei verschiedene betriebliche Geschäftsprozesse modelliert: Die Bereiche Einkauf und Vertrieb zur Abbildung betrieblicher Wertschöpfungsprozesse und der Bereich der Arbeitsvorbereitung als Beispiel für betriebliche Steuerungsprozesse. Vor diesem

Hintergrund wird angenommen, dass die Simulationen in den betrieblichen Bereichen „Einkauf“ und „Vertrieb“ eine andere kognitive Verarbeitungsstruktur voraussetzen als die Anforderungen des Bereiches „Arbeitsvorbereitung“. Die hinter den Wertschöpfungs- und Steuerungsprozessen stehenden betriebswirtschaftlichen Konzepte einerseits und der Umgang mit diesen Geschäftsprozessen in der Praxis andererseits unterscheiden sich. Diese Unterschiede sind als Anforderungsmerkmale der Situationen in die Ausdifferenzierung der Bereiche eingegangen, so dass vermutet werden kann, dass die Auszubildenden zur Bewältigung der Anforderungssituationen auf verschiedene berufsfachliche Fähigkeiten zurückgreifen müssen. Für den Testteil zur Messung verstehensbasierter Kompetenz werden diese Unterschiede in den Fähigkeitsstrukturen der Auszubildenden nicht angenommen – es wird eher von einem generellen Verstehenswert ausgegangen, da sich die Testitems ausschließlich auf die Anwendungen ausgewählter betriebswirtschaftlicher Konzepte einschließlich der notwendigen Validierungen betrieblicher Entscheidungen vor dem Hintergrund wechselnder Inhalte beziehen (verstehensbasierte Anwendungsaufgaben).

Der Test enthält folglich auf zwei verschiedenen Ebenen Items im Hinblick auf eine between-item Mehrdimensionalität (vgl. EMBRETSON & REISE, 2001).<sup>2</sup> Die Ebene der generellen Teststruktur gibt die möglichen Kompetenzdimensionen vor: handlungsbasierte und verstehensbasierte berufsfachliche Kompetenzen. Die Ebene der inhaltlichen Ausdifferenzierung zeigt mögliche Unterschiede in betrieblichen Teilbereichen auf. Diese Mehrdimensionalität impliziert auch für die Leistungsfähigkeiten der Auszubildenden eine mehrdimensionale Struktur. Bevor jedoch die Personenfähigkeiten valide geschätzt werden können, sind die einzelnen Dimensionen zu modellieren, und es ist zu prüfen, inwieweit berufsfachliche Kompetenz tatsächlich eine mehrdimensionale Disposition ist. Die Prüfung der Dimensionalität des Tests erfolgt auf Basis des Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM; ADAMS, WILSON & WANG, 1997; WILSON, 2005). Die Fit-Statistik für die Ebene 1: Kompetenzdimensionen, ist in *Abbildung 8* dargestellt. In das Modell sind alle 60 Items zur Messung berufsfachlicher Kompetenz eingegangen; zur Modellierung der Dimensionen wurden die 34 Items der handlungsbasierten Kompetenz der Dimension I und die 26 Items der verstehensbasierten Kompetenz der Dimension II zugeordnet. Das Modell wurde aufgrund der theoretischen Vorarbeiten zur Struktur von kaufmännischer Kompetenz (vgl. WINTHER & ACHTENHAGEN, 2008; 2009) konfirmatorisch berechnet.

Die handlungsbasierten Simulationsaufgaben und die verstehensbasierten Anwendungsaufgaben messen unterschiedliche latente Dimensionen. Oder anders formuliert: Zur Bewältigung der betrieblichen Simulationen werden von den Probanden andere Fähigkeitsstrukturen eingesetzt als zur Bewältigung der Anwendungsaufgaben. Während die Simulationsaufgaben eher eine prozedurale Struktur aufweisen, werden in den Anwendungsaufgaben vorrangig deklarative Wissensbestände benötigt. Sowohl der Signifikanztest als auch der Test der Effektstärke sprechen für eine zweidimensionale Struktur. Die EAP/PV-Reliabilität für beide Kompetenzdimensionen deutet auf eine sehr gute Qualität der Skalen hin; die Reliabilität für die handlungsbasierte Kompetenz (.762) ist leicht höher als die

2 Zur Differenzierung zwischen „procedural competence“ und „interpretative competence“ im Rahmen der handlungsbasierten Kompetenzdimension sind Analysen auf Basis einer within-item-Mehrdimensionalität notwendig. Diese werden im Rahmen des vorliegenden Beitrags nicht berichtet (vgl. hierzu WINTHER, in Vorbereitung).

	Eindimensionales Modell	Zweidimensionales Modell	Differenz
Deviance (-2Log-Likelihood)	14.505,50	14.403,59	101,91
Anzahl der geschätzten Parameter	80	83	3
EAP/PV-Reliability Dimension I	.794	.762	–
EAP/PV-Reliability Dimension II	–	.705	–

Anmerkung: Die Anzahl der geschätzten Parameter umfasst die Modellparameter und die einzelnen Itemsteps.

Abb. 8: Fit-Statistik für die Kompetenzdimensionen

der verstehensbasierten Kompetenz (.705). Die Annahme eines zweidimensionalen Modells lässt sich mit Hilfe der Chi-Square Statistik absichern: Der Unterschied zwischen den Devianzen der beiden Modelloptionen beträgt 101,91 zugunsten der zweidimensionalen Lösung und ist damit hoch signifikant ( $df = 3; p < .001$ ). Auch die Korrelation zwischen den beiden Kompetenzdimensionen spricht mit .587 für einen signifikant von 1 verschiedenen Zusammenhang zwischen den Dimensionen.

In die Dimensionsprüfung auf der inhaltlichen Ebene und in die entsprechenden Modellberechnungen sind nur die Items zur Erfassung der handlungsbasierten Kompetenz mittels Simulationsaufgaben eingegangen. Auch auf der Ebene der Inhaltsdimensionen sind die Tendenzen für ein zweidimensionales Modell – Wert schöpungsprozesse in Abgrenzung zu betrieblichen Steuerungsprozessen – statistisch signifikant (Chi-Square = 59,51;  $df = 3; p < .001$ ). Der Vergleich der Devianzen spräche in diesem Fall sogar für eine dreidimensionale Lösung, in der jeder inhaltliche Testbereich eine separate Fähigkeitsstruktur abbilden würde (Chi-Square = 124,28;  $df = 7; p < .001$ ). Jedoch ist aufgrund der geringen Itemanzahl und der großen Anzahl fehlender Werte im Bereich „Arbeitsvorbereitung“ diese Dimension sowohl in der zweidimensionalen als auch in der dreidimensionalen Lösung nicht stabil genug, so dass im Folgenden von nur einer inhaltlichen Dimension ausgegangen wird.

### 5.3 Unterschiede zwischen den Dimensionen beruflicher Handlungskompetenz

Zur Abbildung der Unterschiede zwischen den Dimensionen beruflicher Handlungskompetenz werden folglich zwei latente Variablen miteinander verglichen: die Personenfähigkeit, handlungsbasierte Simulationen innerhalb betrieblicher Situationen lösen zu können (handlungsbasierte Kompetenz), und die Personenfähigkeit, verstehensbasierte Anwendungsaufgaben vor dem Hintergrund betrieblicher Situationen zu bewältigen (verstehensbasierte Kompetenz). Die zwei Skalen der beruflichen Handlungskompetenz sind dann umfassend modelliert, wenn die Testitems, die diese Skalen repräsentieren, auf dem Kontinuum des Personenmerkmals ausreichend streuen. Hierbei ist zu beachten, dass möglichst Items unterschiedlichster Schwierigkeitsgrade gefunden werden sollten. Abbildung 9 stellt die Items sowie die Verteilung der Personenschätzer der zwei Skalen dar. Die Abbildung ist wie folgt

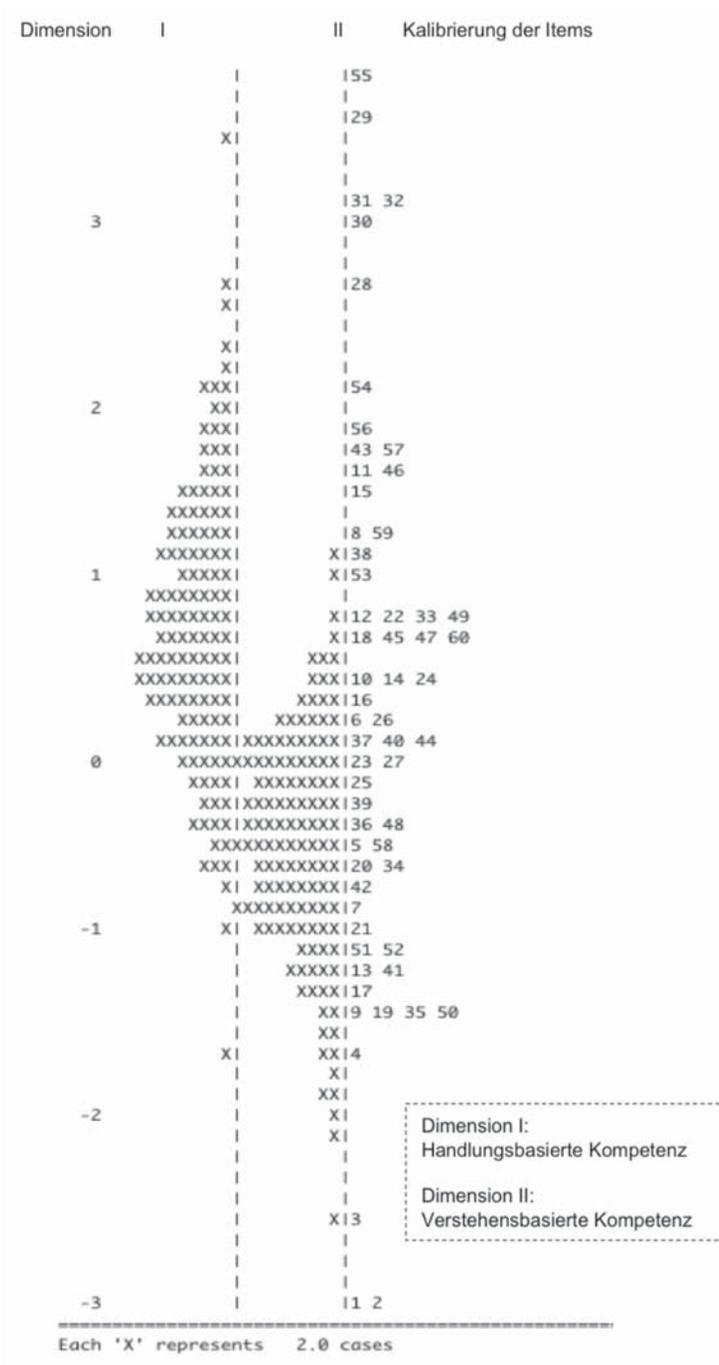


Abb. 9: Skalen beruflicher Handlungskompetenz

zu lesen: Die 60 Items des Tests zur Messung berufsfachlicher Kompetenz sind auf zwei Dimensionen aufgeteilt. Die Dimension I: Handlungsbasierte Kompetenz, wird über 34 Items und die Dimension II: Verstehensbasierte Kompetenz, über 26 Items verschiedener Schwierigkeitsgrade repräsentiert. Die Kalibrierung der Items erfolgt dabei auf einer Logit-Skala. Diese Skala stellt die Relation zwischen Personenfähigkeit und Itemschwierigkeit dar.

Für den vorliegenden Test ist das Item 2 mit einem Itemparameter von -4,297 das leichteste Item und das Item 55 mit einem Schwierigkeitsschätzer von 4,084 das anspruchsvollste Item im Test (aus Darstellungsgründen ist die Logit-Skala der Abbildung 9 auf den Bereich  $-3 \leq x \leq 3,5$  begrenzt; die Items mit geringeren bzw. höheren Schwierigkeitsausprägungen sind jedoch an den Skalenrändern abgebildet). Für die unterschiedlichen Anwendungsbereiche von Tests gibt es verschiedene Aussagen darüber, wie breit die Items eine latente Variable repräsentieren sollten. Für die Messung von Leistungsfähigkeiten hat sich eine Skala von  $\pm 2$  Logits bewährt. Vergleichbare Studien (LEHMANN & SEEBER, 2007; BAUMERT, BOS & LEHMANN, 2000) zeigen, dass diese Skalenbreite genügt, um Unterschiede zwischen den Probanden valide und reliabel schätzen zu können.

Abbildung 9 zeigt, dass die beiden Dimensionen berufsfachlicher Kompetenz vor diesem Hintergrund umfassend repräsentiert sind. Für beide Skalen liegen sowohl verhältnismäßig leichte als auch schwierigere Items vor. Tendenziell ist der Testteil der handlungsbasierten Kompetenz (mittlere Itemschwierigkeit = 0,342 Logits) leicht schwieriger als der Testteil der verstehensbasierten Kompetenz (mittlere Itemschwierigkeit = 0,156 Logits;  $t = 0,423$ ,  $df = 55$ ;  $p = 0,674$ ). Die Skala der verstehensbasierten Kompetenz wird von Items im Bereich  $-1,409 \leq x \leq 1,730$  beschrieben; die Skala der handlungsbasierten Kompetenz über den Schwierigkeitsbereich  $-4,297 \leq x \leq 4,084$ . Die Werte der zentralen Tendenz belegen, dass für beide Skalen der berufsfachlichen Kompetenz sowohl sehr leichte als auch schwierige Items vorliegen und die Skalen damit umfangreich repräsentiert sind, wobei der Test zur Messung der verstehensbasierten Kompetenz deutlich enger konstruiert ist.

Die Anzahl der Items im Bereich der mittleren Schwierigkeitsgrade sowie die dargestellten Verteilungen der Personenparameter deuten auf eine angemessene Diskrimination zwischen den Testpersonen hin. Gerade in diesem Zusammenhang werden die Vorteile von Modellen der IRT besonders deutlich: Dadurch, dass die Itemparameter und die Personenparameter auf einer gemeinsamen Skala, der Logit Skala, transformiert werden, ist das Testverhalten der Personen direkt ersichtlich. Der Mittelwert der Verteilung der handlungsbasierten Kompetenz beträgt 0,750 (Varianz = .612). Dieser Wert besagt, dass Items, die einen ebensolchen Schwierigkeitsparameter aufweisen, mit einer Wahrscheinlichkeit von 50 Prozent gelöst werden können. Verfügt eine Person über einen höheren Fähigkeitsparameter, nimmt die Lösungswahrscheinlichkeit zu. Umgekehrt gilt, dass bei Items, die leichter als 0,750 sind, im Mittel eine höhere Wahrscheinlichkeit als  $p = 0,5$  besteht, diese Items richtig lösen zu können. Die verstehensbasierte Kompetenz der Auszubildenden ist im Mittel niedriger ausgeprägt. Der Mittelwert der Verteilung beträgt -0,481 (Varianz = 0,427). Dies besagt, dass Items mit einem höheren Schwierigkeitsgrad im Mittel mit einer geringeren Wahrscheinlichkeit als  $p = 0,5$  gelöst werden können. In den PISA- und TIMS-Studien sind diese Wahrscheinlichkeiten aufgrund des Mastery-Gedankens von Lernen auf einem höheren Niveau fixiert. Dieses Vorgehen halten wir für die vorliegende Studie nicht für zweckmäßig, da (1) diese Studie keine cur-

ricular gesetzten Lernstände erfasst und sich (2) das Testformat der Studie in der Erprobung befindet.

Zur genaueren Definition der Skalen berufsfachlicher Kompetenz sind sowohl die Itemparameter als auch die Personenparameter zu schätzen. Die Parameterschätzung erfolgt auf Basis des Maximum-Likelihood-Prinzips. Personen- und Itemparameter werden gemeinsam so geschätzt, dass die Likelihood für die beobachteten Daten möglichst groß wird (iteratives Vorgehen; für einen thematischen Überblick vgl. EMBRETSON & REISE, 2001). Die Personenparameter wurden mit dem WLE-Schätzer (weighted likelihood estimator) ermittelt; ihre Verteilung in Abbildung 9 zeigt, dass sich Lage und Ausprägung der zwei Kompetenzstrukturen unterscheiden. Dieser Unterschied ist signifikant ( $t = 21,143$ ;  $df = 261$ ;  $p = 0,000$ ). Die Effektstärke liegt mit  $d = 1,39$  im sehr hohen Bereich. Das Leistungsvermögen der Auszubildenden ist im handlungsbasierten Kompetenzbereich über eine Standardabweichung höher ausgeprägt als im verstehensbasierten Kompetenzbereich: Die Auszubildenden zeigen bessere Leistungen in den betriebsbezogenen Aufgaben als in den eher schulbezogenen.

Die Fragen, welche Items das Modell fiten und welchen Wert der Itemparameter annimmt, lassen sich mit Hilfe der Software ConQuest (WU, ADAMS, WILSON & HALDANE, 2007) beantworten. Zur Veranschaulichung der Itemparameterschätzung sind exemplarisch die Werte der Items 5 und 41 aus dem Bereich der handlungsbasierten Simulationsaufgaben sowie die Items 15 und 20 aus dem Bereich der verstehensbasierten Anwendungsaufgaben in Abbildung 10 wiedergegeben.

Item	Schwierigkeits-schätzer	weighted MNSQ	Konfidenz-intervall	T-Wert	Item-diskrimination
5	-0,536	1,05	(0,84; 1,16)	0,6	0,30
41	-1,136	0,92	(0,76; 1,24)	-0,6	0,38
15	1,481	0,97	(0,83; 1,17)	-0,3	0,40
20	-0,416	0,91	(0,92; 1,08)	-2,4	0,15

Abb. 10: Schätzung der Itemparameter

Zur Beurteilung des Item-Fits sind drei Kriterien zu beachten: (1) Ein zuverlässiger Schätzer für den Item-Fit sind die gewichteten Abweichungsquadrate (weighted mean square) zwischen erwarteten und beobachtbaren Häufigkeiten. Die weighted MNSQ haben einen Erwartungswert von 1; ein Wert exakt 1 zeigt einen idealen Fit an, Werte größer und kleiner 1 sprechen für einen Unter- respektive Überfit. Um ein Item in den Test aufzunehmen, sollten die Werte des weighted MNSQ innerhalb des Intervalls  $0,75 \leq wMNSQ \leq 1,33$  liegen (ADAMS & KHOO, 1996). Ergänzend gilt, (2) dass der T-Wert idealerweise nicht signifikant und (3) die Trennschärfe der Items größer .20 sein sollten. Für Item 20 in Abbildung 10 gilt, dass zwar der weighted MNSQ im Toleranzbereich liegt, der T-Wert und der Wert der Itemdiskrimination jedoch nicht, so dass dieses Item aus dem Test ausgeschlossen wird (vgl. auch ADAMS & WU, 2002).

Insgesamt erfüllen zwei Items aus dem Bereich der verstehensbasierten Anwendungsaufgaben (Items 20 und 48) diese Bedingungen nicht; aus dem Bereich der handlungsbasierten Simulationsaufgaben fittet ein Item das Modell nicht. Die

Skala: Handlungsbasierte Kompetenz, wird folglich über 33 Items<sup>3</sup> und die Skala: Verstehensbasierte Kompetenz, über 24 Items, die alle innerhalb eines konservativ festgelegten weighted MNSQ-Wertebereichs von 0,90 bis 1,10 liegen, valide und reliabel repräsentiert.

#### 5.4 Kompetenzstufen beruflicher Handlungskompetenz

Zur Ermittlung von Kompetenzstufen werden Impulse aus dem beruflichen Arbeitsumfeld verarbeitet, denen differenzierende Kraft im Hinblick auf die Charakterisierung von Anforderungssituationen zugeschrieben werden kann. Die zentrale Idee besteht darin, dass über geeignete Kriterien spezifische Annahmen über die Wirkung einzelner Schwierigkeitsparameter getroffen werden können, um die Frage zu beantworten, wie sich Lösungsprozesse bei einer systematischen Variation des Anspruchsniveaus der Anforderungssituationen verändern. Das Ziel ist es, solche Einflussfaktoren zu identifizieren, die das Anspruchsniveau direkt erhöhen bzw. reduzieren können und mit deren Hilfe sich Veränderungen im Lösungsverhalten in Folge der Variation des Anspruchsniveaus möglichst detailliert beschreiben lassen. Der Zusammenhang zwischen Anforderungssituation und Antwortverhalten wird mit Hilfe einer Construct Map (WILSON, 2005) modelliert – ein Zusammenhang der nach der qualitativen Analyse später in das psychometrische Modell eingeht (WINTHER, in Vorbereitung).

Bei der Entwicklung von Testaufgaben und der entsprechenden Scoring-Verfahren für die Testauswertung ist folglich darauf zu achten, dass die erarbeiteten Anforderungssituationen unterschiedliche Komplexitätsgrade aufweisen müssen: Sind sie so einfach, dass alle Probanden sie sofort oder nach kurzer Einarbeitung ausführen können, oder so anspruchsvoll, dass längere Lern- und Übungsphasen erforderlich sind? Zur Beantwortung dieser Fragen sind kohärente Beziehungen zwischen curricularen Anforderungen, fachspezifischen Ausdifferenzierungen und kognitionspsychologischen Modellen herzustellen (vgl. WILSON, 2005; WINTHER & ACHTENHAGEN, 2008). Mögliche Kriterien, anhand derer der Anspruchsgehalt einer beruflichen Anforderungssituation variiert werden kann, sind: (1) funktionale Modellierung (modeling), (2) inhaltliche Komplexität (content) und (3) die Art der kognitiven Taxonomierung (cognition).

Die Dimension des funktionalen Modellierens versucht vor allem, die Aspekte der Vertrautheit mit der Situation sowie der praktischen Erfahrung im Umgang mit ihr zu berücksichtigen: Wenn ich z.B. dem Kunden mitteilen muss, dass unser Unternehmen eine Lieferzusage nicht einhalten kann, bin ich dann in der Lage, mir die gesamte Situation modellhaft vorzustellen – bis hin zu den Produktionsproblemen des Kunden und der langfristigen weiteren Entwicklung der Geschäftsbeziehungen? Wie habe ich mich auch unter der Perspektive der Sozialkompetenz in der (schriftlichen) Kommunikation mit dem Kunden zu verhalten? – Es leuchtet ein, dass der Prozess der funktionalen Modellierung vor dem Hintergrund spezifischer Anforderungssituationen entscheidend von den Hinweisen beeinflusst wird, die in der jeweiligen Situation gegeben sind. Das Aufgabenniveau hängt damit von den gemachten Erfahrungen und dem Umfang und der Qualität der Informationen ab, über die ich

3 Hier sind auch jene Testitems enthalten, die den sicheren Umgang mit der Testform anzeigen.

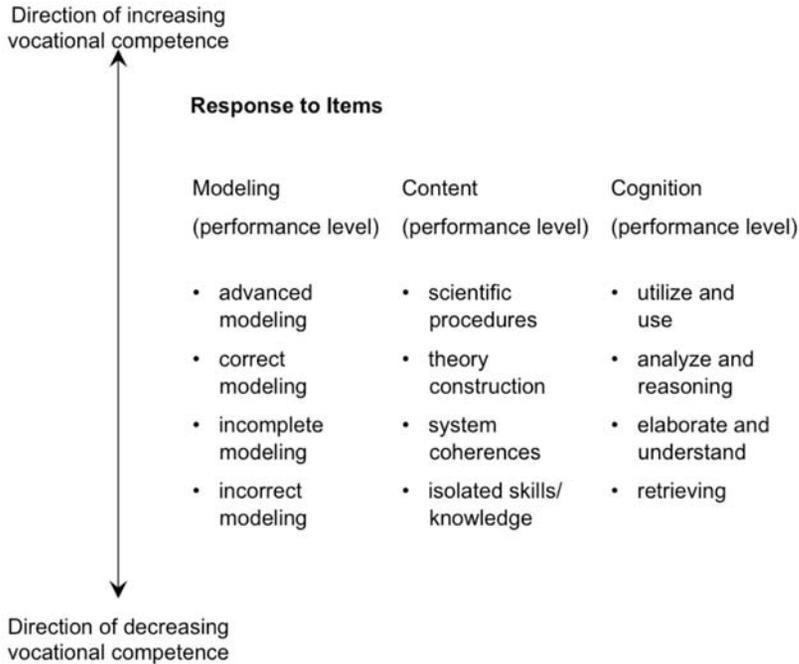


Abb. 11: Construct Map (WINTHER, in Vorbereitung)

bei der Aufgabenbewältigung verfüge. Die Antworten der Probanden können dabei von inkorrekt bis umfassender funktionaler Modellierung variieren.

Die inhaltliche Dimension macht deutlich, in welcher Weise relevante Informationen zusammenzufügen sind. Ein Beispiel ist hier das Heraussuchen von Informationen zu einem möglichen Lieferanten – verglichen mit der selbstständigen Erarbeitung eines Lieferantenvergleichs unter Heranziehung quantitativer und qualitativer Daten, deren Gewichtung sowie der abschließenden Begründung der Auswahl eines bestimmten Lieferanten. Die Antworten der Probanden reichen in dieser Kategorie von isolierten Wissensbeständen bis hin zu begründeten (wissenschaftlich adäquaten) Theorieformulierungen.

Die kognitionspsychologische Dimension schließlich lässt sich danach gliedern, wie umfangreich die kognitive Beanspruchung einer betrieblichen Anforderungssituation ist. Zur Systematisierung kognitiver Beanspruchung können kognitionspsychologische Taxonomien herangezogen werden (BLOOM, ENGELHART, FURST, HILL & KRATHWOHL, 1956; ANDERSON & KRATHWOHL, 2001; MARZANO & KENDALL, 2007). Ergebnisse aus früheren Projekten belegen, dass die kognitive Beanspruchung ausgehend von reproduktiven Tätigkeiten zu solchen des Problemlösens ansteigt (WINTHER, 2006) – entsprechend sind die Aufgaben zu formulieren. Bezüglich der Bewertung der Antwortleistungen sind Ausprägungen zwischen Abrufen von Informationen und der sinnvollen Verwendung kognitiver Ressourcen zur Bewältigung von Problemstellungen vorstellbar.

Die schwierigkeitsklassifizierenden Parameter helfen dabei, die Skalen der beruflichen Handlungskompetenz in inhalts- und kriterienbezogene Kompetenzstufen zu

zerlegen. Hierfür wurde in unserem Projekt jedes der schwierigkeitsklassifizierenden Merkmale durch fünf Rater den einzelnen Aufgaben des Tests der berufsfachlichen Kompetenz zugeordnet. Die Reliabilität ist dabei an den ICC (intra-class correlation coefficient; vgl. u. a. MISLEVY, 2007) gebunden. In den Auswertungen liegt dieser für alle 57 Items, die letztlich in den Test aufgenommen wurden, im Bereich  $0.818 \leq r_{ICC} \leq 0.946$ . Die Annahmen über die unterschiedlichen Schwierigkeiten der Items lassen sich dazu nutzen, den empirisch geschätzten Itemparameter zu prognostizieren. Im Rahmen einer Regressionsanalyse werden aus den drei schwierigkeitsbestimmenden Merkmalen, die jeweils vier Ausprägungen aufweisen, neun operationale Itemparameter (Dummy-Variablen) gebildet. Das Ergebnis der Regressionsanalyse – abgebildet über die unstandardisierten Regressionsgewichte – ist in Abbildung 12 dargestellt.

Vier Prädiktoren sind hoch bedeutsam für die Vorhersage der Itemschwierigkeit. Im Hinblick auf die funktionale Modellierung sind Anforderungssituationen, die von den Lernenden eine eigenständige Übersetzung der Situation in die notwendigen Schritte des Lösungsprozesses verlangen, besonders anspruchsvoll. Zwei Stufen der funktionalen Modellierung zeigen sich vor diesem Hintergrund im Regressionsmodell signifikant. Besonders stark ist die Itemschwierigkeit von der Notwendigkeit umfangreicher Modellierungsleistungen ( $\beta = .654$ ;  $p = .000$ ) abhängig. Doch bereits teilweise Modellierungsanforderungen ( $\beta = .256$ ;  $p = .001$ ) beeinflussen den Anspruchsgehalt eines Items. Die Bedeutung der inhaltlichen Komplexität für die Itemschwierigkeit wird im Prädiktor „Systemkohärenzen“ ( $\beta = .228$ ;  $p = .013$ ) deutlich. Items, mit denen Zusammenhänge zwischen verschiedenen betriebswirtschaftlichen Konzepten abgebildet werden sollen, sind für die Probanden folglich schwieriger zu lösen. Der kognitive Prozess wird durch den Prädiktor „Analysieren und Validieren“ ( $\beta = .362$ ;  $p = .000$ ) bedeutsam für die Komplexität einer betrieblichen Anforderungssituation. Die vier Merkmale der Itemschwierigkeit erklären zusammen 74,9 Prozent der Varianz ( $F = 45,021$ ;  $p = .000$ ).

adj. R <sup>2</sup> = .749	Regressionsgewicht	Standardfehler	Beta-gewicht	Irrtumswahrscheinlichkeit
Konstante	-1,479	.202	–	.000
Umfangreiche Modellierungsleistung (Funktionale Modellierung 3)	2,128	.334	.654	.000
Korrekte Modellierungsleistung (Funktionale Modellierung 2)	0,840	.246	.256	.001
Systemkohärenz (Inhaltliche Komplexität 2)	0,756	.296	.228	.013
Analysieren und Validieren (Kognitive Taxonomierung 3)	1,293	.327	.362	.000

n = 57 (Anzahl der Testitems für beide Kompetenzdimensionen)

Abb. 12: Unstandardisierte Regressionsgewichte der Prädiktoren der Itemschwierigkeit

In Anlehnung an HARTIG (2007) lassen sich die unstandardisierten Regressionsgewichte als Rangwerte der Itemschwierigkeit interpretieren. Ein Item, für das ein

Schwierigkeitsmerkmal geratet wurde, ist um den Betrag des Regressionsgewichts (gemessen in Logits) schwieriger als ein Item, das dieses Schwierigkeitsmerkmal nicht trägt. Die Summe der Logits stellt folglich die erwartete Schwierigkeit eines Items im Hinblick auf die schwierigkeitsbeschreibenden Merkmale dar. Der Wert, bei dem ein oder mehrere als signifikant identifizierte Schwierigkeitsmerkmale erstmalig systematisch auftreten, kann als untere Schwelle einer Kompetenzstufe interpretiert werden. Für die Skalen zur Erfassung berufsfachlicher Kompetenz können wir vier Kompetenzstufen ermitteln, die sich als Kaufmännisches Grund- und Regelwissen, Kaufmännisches Handlungs- und Aktionswissen, Kaufmännisches Analysewissen und Kaufmännisches Entscheidungswissen umschreiben lassen.

In Abbildung 13 sind die Kompetenzstufen über die Logit-Grenzen definiert; zusätzlich zu den Schwellenwerten ist die prozentuale Zuordnung der Auszubildenden auf die einzelnen Kompetenzstufen angegeben. Der Anteil der Auszubildenden in den jeweiligen Kompetenzstufen im Hinblick auf die zwei verschiedenen Kompetenzdimensionen unterscheidet sich deutlich: Wenn die Kompetenzstufe 2: Kaufmännisches Handlungs- und Aktionswissen, als Anspruch des Arbeitsmarktes an Absolventen des Dualen Systems in Anlehnung an die Vorgaben anzusehen ist, können im Bereich der handlungsbasierten Kompetenz 22,6 Prozent und im Bereich der verstehensbasierten Kompetenz 78,2 Prozent der getesteten Auszubildenden diesen Anspruch nicht adäquat erfüllen. Zu einer Relativierung der Ergebnisse sei an dieser Stelle erwähnt, dass (1) die Stichprobe ausschließlich Auszubildende im Rahmen des 3-jährigen Ausbildungsganges enthält, so dass die erwarteten leistungsstarken Auszubildenden insbesondere im Bereich der verstehensbasierten Kompetenz nicht in der Masse erfasst werden konnten, und dass (2) die Ergebnisse im handlungsbasierten Kompetenzbereich weit positiver ausgefallen sind, als dies vor dem Hintergrund vergleichbarer Studien (ULME III: LEHMANN & SEEBER; TOSCA: KÖLLER, WATERMANN, TRAUTWEIN & LÜDTKE, 2004) hätte angenommen werden können.

Im Hinblick auf die **verstehensbasierte Kompetenz** ist das Ergebnis absolut erwartungskonform; es repliziert in hohem Maße die Befunde des ULME III-Projektes (Untersuchung von Leistungen, Motivation und Einstellungen in der beruflichen Bildung; LEHMANN & SEEBER, 2007). Die Befunde zeigen, dass die Leistungen z. B. im Ausbildungsberuf Industriekaufmann / Industriekauffrau tendenziell eine bimodale Verteilung mit einem Gipfel im unteren Leistungsbereich und einem flachen Kurvenverlauf zum oberen Leistungsspektrum hin aufweisen. Die Leistungsschwächen beziehen sich dabei insbesondere auf Testaufgaben, die das Verstehen und Interpretieren ökonomischer Beziehungen zum Inhalt haben (LEHMANN & SEEBER, 2007, S. 140). Auch für die vorliegende Studie zeigt sich, dass ein Großteil der Auszubildenden (42,7 Prozent) zwar über kaufmännisches Grund- und Regelwissen verfügt, das jedoch zu unflexibel ausgebildet ist, um es variabel im Rahmen beruflicher Anforderungssituationen einsetzen zu können. Hinzu tritt sicher auch ein Effekt eines „trägen“ Wissens, das durch gezieltes Üben kurz vor der Abschlussprüfung aktiviert werden soll.

Für den Testbereich der **handlungsbasierten Kompetenz** fällt insbesondere positiv auf, dass nur drei Prozent der Auszubildenden unterhalb der Kompetenzstufe I liegen und dass 52,6 Prozent die Kompetenzstufe II, die von den Autoren als Mindestanspruchsstufe des Arbeitsmarktes definiert wurde, und weitere 24,8 Prozent die darüber liegenden Kompetenzstufen erreichen konnten.

Kompetenzstufe	Stufenschwelle der Kompetenzbereiche	Anteil der Auszubildenden auf den Stufen der handlungsbasierten Kompetenz (in%)	Anteil der Auszubildenden auf den Stufen der verstehensbasierten Kompetenz (in%)
unter Kompetenzstufe I	-1,479	3,10	35,50
Kompetenzstufe I: Kaufmännisches Grund- und Regelwissen	- 0,723	19,50	42,70
Kompetenzstufe II: Kaufmännisches Handlungswissen	0,117	52,60	21,40
Kompetenzstufe III: Kaufmännisches Analysewissen	1,410	23,30	0,40
Kompetenzstufe IV: Kaufmännisches Entscheidungswissen	3,538	1,50	–

Abb. 13: Kompetenzstufen und prozentuale Verteilung der Auszubildenden

#### 5.4.1 Beispiel für Kompetenzstufe I: Kaufmännisches Grund- und Regelwissen

Die Kompetenzstufe I wird vorrangig über den Faktor der inhaltlichen Kompetenz definiert. Der schwierigkeitsbeschreibende Faktor „Systemkohärenz“ weist das geringste Regressionsgewicht auf, das ausgehend von der Konstanten den unteren Schwellenwert der Kompetenzstufe I: Kaufmännisches Grund- und Regelwissen, definiert. Auszubildende auf dieser Kompetenzstufe verfügen über grundlegende ökonomische Begriffs- und Konzeptstrukturen und können Systemkohärenzen (Prädiktor der Itemschwierigkeit) herstellen. Hierzu benötigen die Auszubildenden jedoch Hinweise in Form von Modellierungshilfen der Anforderungssituation, da sie nur so in der Lage sind, auf ihre Wissensbestände und prozeduralen Fertigkeiten zurückzugreifen.

Folgendes Beispielitem repräsentiert die Kompetenzstufe I für die Skala der verstehensbasierten Kompetenz (Item 21 in Abbildung 9):<sup>4</sup> Das Item bezieht sich auf rechtlich-organisatorische Fragen des Kaufvertrags auf Basis eines verbindlichen Angebots. Dieses Item ist typisch für die Skala der verstehensbasierten Kompetenz. Zur Beantwortung werden deklarative Wissensbestände benötigt, die miteinander in Beziehung zu setzen sind: Liegt ein verbindliches Angebot vor? Welche Fristen gelten für die Annahme eines verbindlichen Angebots? Welche Konsequenzen ergeben sich im Hinblick auf den Kaufvertrag?

Der Itemfit des Beispiels ist mit einem weighted MNSQ von 1,04 und einer Itemdiskrimination von .35 sehr gut. Das gewählte Beispielitem weist einen Itemschwie-

4 Alle Items der Skala der verstehensbasierten Kompetenz beziehen sich auf die Sequenzen der Simulation insgesamt: Für das exemplarische Item wurde als Begründung akzeptiert, dass – vorbehaltlich der Verfügbarkeit – der Kaufvertrag zustande gekommen sei, da auf das verbindliche Angebot binnen einer Woche reagiert wurde.

rigkeitsparameter von  $-.960$  Logits auf. Dieser liegt leicht unterhalb des mittleren Personenschätzers der verstehensbasierten Kompetenzskala ( $-.481$ ). Das Item wurde von 219 Personen bearbeitet, von denen 129 die richtige Lösung gefunden haben (Lösungsquote: 60 Prozent).

---

Beispielitem: Kompetenzstufe I – verstehensbasierte Kompetenz

Die ALUSIM GmbH hat am 29. November 2010 an ihre Stammkunden eine Liste mit Restbeständen der Sparte Industrieprodukte versendet. Diese Restbestände wurden den Kunden in einem beiliegenden Schreiben „mit einer Lieferung innerhalb von zwei Wochen“ und „mit einer Vergünstigung von 50%“ angeboten. Geben Sie in den nachfolgenden Fällen bitte jeweils an, ob ein Kaufvertrag zustande gekommen ist und begründen Sie Ihre Meinung kurz.

- a) Am 1.12. 2010 ruft Frau Schridde, Delter Farben GmbH, an und gibt ihre Bestellung ab.
- Kaufvertrag ist zustande gekommen.
  - Kaufvertrag ist nicht zustande gekommen.

Begründung:

---

#### 5.4.2 Beispiel für Kompetenzstufe II: Kaufmännisches Handlungswissen

Die Kompetenzstufe II beinhaltet einen sicheren Umgang mit betriebswirtschaftlichen Konzepten mindestens auf dem Niveau der Systemkohärenz; das Erreichen der Kompetenzstufe II setzt darüber hinaus eine hohe Adaptationsfähigkeit der kognitiven Strukturen der Auszubildenden voraus, um wechselnde betriebliche Anforderungssituationen adäquat ohne weitgehende Modellierungshilfen (→ Prädiktor der Itemschwierigkeit) so zu bewerten, dass anwendbare Lösungen entstehen. Dies setzt eine angemessene Arbeitsorganisation und eine selbständige Arbeitsplanung voraus. Das nachfolgende Beispielitem dient der Veranschaulichung der Kompetenzstufe II: Kaufmännisches Handlungswissen im Rahmen der handlungsbasierten Kompetenzskala: In der videobasierten Situationsbeschreibung des Testbereichs „Vertrieb“ erfahren die Auszubildenden, dass die ALUSIM GmbH aufgrund des schwierigen Marktfeldes versucht, durch eine halbjährliche Erstellung individueller Preislisten für die wichtigsten Kunden Preisvorteile direkt weiterzugeben, um so die Absatzzahlen zu sichern. Die Auszubildenden erfahren weiter, dass die individuellen Preislisten jeweils zum Oktober und April jeden Jahres an die ausgewählten Kunden zusammen mit einer Übersicht der angebotenen Industrieprodukte versendet werden und dass auf Basis dieser versendeten Preislisten am Mittwoch, 1. 12. 2010, eine Bestellung des Kunden FACT Dosiertechnik GmbH per Fax eingeht. Die Auszubildenden sind im Folgenden aufgefordert, diese Bestellung unter Nutzung des ERP-Systems zu bearbeiten. Die besondere Schwierigkeit liegt darin, den richtigen Preis (der Preis aus der Preisliste; der Preis aus dem Produktkatalog) für die bestellte Ware zu ermitteln: Sie müssen (1) erkennen, dass der Preis aus der versendeten Preisliste relevant sein könnte, und sie müssen (2) begründet darüber entscheiden, unter welchen Bedingungen welcher Preis tatsächlich in das Bestellsystem einzugeben ist (Item 6 in Abbildung 9).

Das Item erfüllt mit einem weighted MNSQ von 1,10 und einer Itemdiskrimination von .20 die Anforderungen zur Aufnahme in den Test der handlungsbasierten Kompetenz. Der Schwierigkeitsparameter ist mit .206 Logits geringer als die mittlere Personenfähigkeit der Probanden (.750). Insgesamt wurde das Beispielitem von 214 Personen mit einer Lösungsquote von 59,81 Prozent bearbeitet.

The screenshot shows an ERP system window titled 'Kundenaufträge -> Bestellsystem'. It is divided into two main sections: 'Produktinformation' and 'Kundeninformation'.

**Produktinformation:**

Artikel-Nr.	Volumen	Durchmesser mm	Höhe mm	Bestellmenge Stück	Preis pro Stück

**Kundeninformation:**

Auftrags-Nr.	Artikel-Nr.	Kunden-Nr.	Name	Lieferung	Verpackung	Preis pro Stück	Gesamtpreis
800003AK330							

**Callout Box Text:**

Der gewählte Preis ist von zeitlichen Fristen im Rahmen der rechtlichen Kaufvertragsbestimmungen abhängig. Der Auszubildende muss alle notwendigen Informationen eigenständig sammeln und im Hinblick auf ihre Lösungsrelevanz bewerten.

#### 5.4.3 Beispiel für Kompetenzstufe III: Kaufmännisches Analysewissen

Knapp 25 Prozent der Auszubildenden weisen auf der Skala der handlungsbasierten Kompetenz Fähigkeitsstrukturen auf, die als kaufmännisches Analyse- und Entscheidungswissen beschrieben werden können (Kompetenzstufen III und IV). Die Kompetenzstufe III: Kaufmännisches Analysewissen, definiert sich in Abgrenzung zur Kompetenzstufe II hauptsächlich über den Grad der kognitiven Beanspruchung. Der Prädiktor „Analysieren und Validieren“ weist ein Regressionsgewicht von 1,293 auf, das addiert zur Konstanten und den davorliegenden Kompetenzstufen den unteren Schwellenwert von 1,41 Logits für die Kompetenzstufe III ergibt. Von Auszubildenden auf dieser Kompetenzstufe wird erwartet, dass sie über die Fähigkeit verfügen, die Plausibilität einer berufsspezifischen Lösung analysieren und validieren zu können. Die kognitive Taxonomierung stellt folglich für diese Kompetenzstufe den zentralen Schwierigkeitsparameter dar. Dies setzt eine schlussfolgernde Auslegung von Wissen und damit Tranferfähigkeit voraus. Das nachfolgende Beispielitem illustriert diese Fähigkeitsstruktur: Die ALUSIM GmbH benötigt eine Druckfarbe, die bislang nicht im Sortiment enthalten ist. Die Auszubildenden haben in diesem Zusammenhang Einstandspreise im Rahmen einer vergleichenden Preiskalkulation errechnet. Nachdem die Entscheidung für einen Lieferanten gefallen ist, ist die Einkaufsorder im ERP-System zu erfassen. Die Schwierigkeit besteht darin, dass die Auszubilden-

den in einem kognitiven Validierungsprozess ermitteln müssen, ob der von ihnen im Vorfeld berechnete Einstandspreis oder der Listen-/Angebotspreis zu verarbeiten ist (Item 43 in Abbildung 9). Von 215 Personen, die dieses Item bearbeitet haben konnten nur 23 die korrekte Lösung finden (Lösungsquote: 10,70 Prozent).

Die Itemschwierigkeit liegt mit 1,730 Logits deutlich oberhalb der mittleren handlungsbasierten Kompetenz der Probanden. Das Item fittet die Skala der

Die Auszubildenden haben im Vorfeld verschiedene Preiskalkulationen durchgeführt. Sie müssen nun in einem kognitiven Validierungsprozess ermitteln, ob der von ihnen berechnete Einstandspreis tatsächlich relevant für die Eingabe im ERP-System ist. Es wird erwartet, dass die Auszubildenden erkennen, dass hier der Listen-/Angebotspreis der Farbe zu verarbeiten ist und nicht der Einstandspreis der Preiskalkulation.

handlungsbasierten Kompetenz mit einem weighted MNSQ von 1,03 und einer Itemdiskrimination von .24 sehr gut.

#### 5.4.4 Beispiel für Kompetenzstufe IV: Kaufmännisches Entscheidungswissen

Verfügen Auszubildende über kaufmännisches Entscheidungswissen (Kompetenzstufe IV), sind sie in der Lage, im Sinne eines umfassenden Managementprozesses, die berufliche Anforderungssituation im Hinblick auf die Problemformulierung, die Zielsetzungen, die Lösungsmöglichkeiten und die resultierenden Entscheidungen umfassend zu modellieren (LAUX, 2007). Die Kompetenzstufe IV enthält nur Items des handlungsbasierten Kompetenzbereichs. Items, die dieser Kompetenzstufe zugeordnet sind, lassen offen, welche Bearbeitungsschritte zur Lösungsfindung gegangen werden müssen und welchen Umfang die Lösung letztlich haben soll. Diese Charakteristik eines unpräzise formulierten Anforderungsgehalts wird in der Kreativitätsforschung beispielsweise als blind pool bezeichnet (SIMONTON, 2003). Die Schwierigkeit solcher Situationen liegt darin, zunächst auf Basis bestehender Wissensbestände und prozeduraler Fähigkeiten eine so umfassende Modellierung der Situation vorzunehmen (→ Prädiktor der Itemschwierigkeit), dass Lösungsmöglichkeiten ersichtlich werden.

Die notwendigen Modellierungsschritte können dabei entsprechend der Situation sehr unterschiedlich beschaffen sein. Vor dem Hintergrund betrieblicher Arbeitsprozesse sind beispielsweise die Sammlung, Komprimierung und Aufbereitung von Informationen, Effizienzkalkulationen einzelner Arbeitsprozessschritte oder die Entscheidung für die optimale Kombination von Arbeitsmitteln möglich. Das nachfolgende Itembeispiel illustriert für die Skala der handlungsbasierten Kompetenz, wie eine Anforderungssituation zunehmend komplexer werden kann und wie dadurch die Lösungschancen der Auszubildenden mit einem geringeren Fähigkeitslevel erheblich reduziert werden können. Die Auszubildenden sollen auf der Basis von Angeboten und weiteren Zusatzinformationen eine vergleichende Lieferantenauswahl durchführen. Die Hintergründe des Lieferantenvergleichs sind den Auszubildenden bekannt; zwei Angebote für den Vergleich wurden von den Auszubildenden selbst eingeholt. Die Schwierigkeit liegt darin, die extrem große Menge an Einzelinformationen so zu komprimieren, dass alle notwendigen Daten für einen Lieferantenvergleich vorliegen: Die Informationsquellen unterscheiden sich, die Lieferbedingungen sind verschieden kodiert, die Abpackungen des Produkts variieren, unterschiedliche Währungen sind zu berücksichtigen etc. Auch das Arbeitsprodukt wird nicht präzise formuliert; als Arbeitsmittel wird ausschließlich eine Excel-Datei ohne Inhalt zur Verfügung gestellt (Item 29 in Abbildung 9).

Die Informationen in der Lieferantendatei des ERP-Systems des Unternehmens stellen nur einen Ausschnitt aller notwendigen Informationen dar. Insgesamt müssen für den Lieferantenvergleich acht verschiedene Informationsquellen ausgewertet werden.

Kurz	Lieferbedingungen	Zahlungskonditionen	Qualität	Terminreue	Ökoadult (EMAS)
Büro	ab Werk	008 20 045	1	2	2
Colour	frei Haus	014 20 030	1	1	2
Dralle	frei Haus	008 20 020	2	1	2
Farb	ab Werk	007 20 030	1	1	2
KG	Kunst	000 00 014	1	3	2
Life	ab Werk	007 20 030	1	1	1
MPK	ab Werk	010 20 060	1	1	2
Müll	frei Lager	008 20 045	1	1	2
Papier	frei Haus	010 20 020	1	2	3
Plast	frei Haus	000 00 045	1	2	1

Eingabe der Zahlungsbedingungen entsprechend KHK-FIBU:  
008 20 045  
= 8 Tage 2,0 % Skonto, 45 Tage netto Kasse

Qualität und Terminreue der Lieferanten sind wie Schulnoten vergeben:  
1 = ausgezeichnet

Umfang unternehmensinterner Verfahren und Maßnahmen, die die negativen Umweltauswirkungen betrieblicher Tätigkeit reduzieren.  
Die Ökoadult-Verordnung entspricht dem Environmental Management and Audit Scheme (EMAS).  
Der Ökoadult der Lieferanten ist wie Schulnoten vergeben.  
1 = ausgezeichnet

al u

Das Item (weighted MNSQ = .95; d = .29) liegt mit einer Itemschwierigkeit von 3,602 Logits deutlich oberhalb der mittleren Personenfähigkeit (.750). Die Lösungsquote des Items beträgt 6,32 Prozent; 190 Probanden haben die Bearbeitung dieses Items aufgenommen.

Die Beispielimis illustrieren, wie die Prädiktoren der Itemschwierigkeit als inhaltlicher Anker zur Beschreibung einzelner Kompetenzstufen interpretiert werden können. Im Umkehrschluss zeigen diese Beispiele jedoch auch, dass für eine Testkonstruktion im Vorfeld sehr elaborierte Analysen (1) der Anforderungsbedingungen, (2) der kognitiven Theorien und (3) der inhaltlichen sowie didaktischen Bedingungen

durchzuführen sind. Hierbei ist es hilfreich, dass die neuen Assessmentmodelle der angewandten Kognitionswissenschaft Umgebungsanalysen direkt auf psychometrische Messmodelle beziehen (vgl. u. a. MISLEVY, 2008; WILSON, 2008; EMBRETSON & GORIN, 2001). Oder anders formuliert: Der Einsatz psychometrischer Modelle setzt eine inhaltliche didaktische Vorarbeit im Hinblick auf die Itemkonstruktion zwingend voraus, um im Rahmen der Analyse die Ergebnisse sinnvoll interpretieren zu können – auch für die Interpretation ist fachdidaktische Expertise einzufordern.

#### 4 Schlussfolgerungen

Das Projekt hatte die Aufgabe, die „Konstruktvalidität von Simulationsaufgaben“ zu ermitteln, um bezogen auf die Erfassung von berufsfachlicher Kompetenz eine Alternative zu Papier- und Bleistift-Tests zu entwickeln. Der Anspruch bestand einerseits darin, auf der Ebene der inhaltlichen Ausgestaltung den realen betrieblichen Arbeits- und Geschäftsprozessen durch Authentizitätsmodellierungen gerecht zu werden. Hierzu zählt auch, dass Arbeitsmittel in Entsprechung zur betrieblichen Praxis zur Testbearbeitung eingesetzt werden und dass die konstruierten Anspruchssituationen auf einem umfassend modellierten Datensatz, mit dem ein Simulationsunternehmen exakt beschrieben werden kann, basieren. Andererseits wurde das Ziel verfolgt, auf der Ebene der Person auch solche Fähigkeitsstrukturen erheben zu können, die sich auf Handlungsprozesse beziehen lassen. Dies umfasst für eine als authentisch geltende Situationsbewältigung insbesondere auch die notwendigen Interaktionen mit Kollegen, Kunden und Lieferanten. Aspekte wie Informationsgehalt der Kommunikation, Höflichkeit oder auch die Kompatibilität mit den Unternehmenszielen lassen sich als Unterkategorien der Sozialkompetenz auf diesem Wege beschreiben.

Bezogen auf die Entwicklung der Testbereiche und der Testitems ist Folgendes zusammenfassend festzuhalten:

- Im Zentrum der Arbeit stand die Konstruktion von Testsituationen, deren Angemessenheit sowohl im Hinblick auf betriebliche als auch auf schulische Anforderungen mehrfach geprüft wurde.
- Diese Anforderungen sind in der Simulation angelehnt an Bedingungen, wie sie im Hinblick auf die Erfüllung betriebs- und arbeitsspezifischer Vorgaben als berufstypisch gelten. Zudem wurden die generellen Zielsetzungen für die Berufsausbildung berücksichtigt.

Im Hinblick auf die Dualität der deutschen Berufsausbildung wurden die Testitems zum einen so konstruiert, dass sie handlungsbasierte Kompetenz in beruflichen Situationen messen. Das ist genau der Punkt, der von den Sozialpartnern als grundlegende Bedingung für die Teilnahme an einem VET-LSA eingefordert wird. Zum anderen wurden Aufgaben konstruiert, um die verstehensbasierte Kompetenz vor dem Hintergrund beruflicher Situationen zu messen. Hier wird vor allem die Leistung erfasst, die sich auf ein konzeptionelles Verständnis der betriebswirtschaftlichen Theorie in berufstypischen Anforderungssituationen beziehen lässt – ein Anspruch, der durchaus dem schulischen Teil der Dualen Berufsausbildung zugeschrieben werden kann. Um es nochmals zu betonen: Alle Aufgaben wurden in Entsprechung zur betrieblichen Praxis und der Vorgaben für die Industriekaufleuteausbildung ent-

wickelt – und von Experten als angemessen im Hinblick auf den Aufgabenbereich und das Aufgabenniveau beurteilt.

Zur Auswertung der Ergebnisse wurden Modelle der Item Response-Theorie gewählt. Die Auftragnehmer der Studie sind davon überzeugt, dass nur über probabilistische Verfahren ein international vergleichendes Large Scale-Assessment zu realisieren ist. Diese Einschätzung bezieht sich zum einen auf das zu erwartende Design eines VET-LSA und zum anderen auf die Fragen der Testkonstruktion und der Analyse der Ergebnisse. Ein zentrales Anliegen der Studie war es daher zu prüfen, ob psychometrische Testmodelle – wie sie auch in den PISA- und TIMS-Studien eingesetzt wurden – für Aufgaben praktikabel sind, mit denen speziell Handlungsprozesse abgebildet werden sollen. Die Analysen haben gezeigt, dass die konstruierten Testitems eine sehr umfassende Skala zur Erfassung handlungsbasierter Kompetenzen repräsentieren und dass sich handlungsbasierte Kompetenz empirisch von verstehensbasierter Kompetenz trennen lässt. Es ist folglich möglich, empirisch gesichert unterschiedliche Fähigkeitsstrukturen der Auszubildenden in Entsprechung zu den Zielen der Dualen Ausbildung separiert zu erfassen und aufeinander zu beziehen. Dies konnte beispielsweise anhand der Definition von Kompetenzstufen gezeigt werden, die sich auf den handlungsbasierten und auf den verstehensbasierten Kompetenzbereich anwenden lassen:

- Die Hälfte der Befragten (52,6 Prozent) befindet sich bezüglich der handlungsbasierten Kompetenz auf einer Kompetenzstufe, auf die hin die Ausbildung der Industriekaufleute von den Sozialpartnern mindestens konzipiert ist; 24,8 Prozent konnten eine höhere Kompetenzstufe erreichen.
- Bezüglich der verstehensbasierten Kompetenz liegen die ermittelten Werte hingegen deutlich niedriger.

Die Studie verfolgte darüber hinaus das Ziel, den Einsatz eines neuen Testformats zu validieren. Hierzu war vorgesehen, die Anforderungssituationen der web-basierten Unternehmenssimulation ALUSIM mit betrieblichen Arbeitsproben in den Ausbildungsbetrieben zu vergleichen. Der Rücklauf aus den Betrieben ist zum Zeitpunkt der Berichterstellung noch so gering, dass hierzu keine aussagekräftigen Ergebnisse vorgelegt werden können. Allein diese Schwierigkeiten zeigen jedoch bereits deutlich, dass eine Direkterfassung von beruflicher Handlungskompetenz in betrieblichen Arbeitsumgebungen für ein international vergleichendes VET-LSA nicht das Mittel der Wahl sein kann. Folgende Argumente kommen hinzu:

- Betriebliche Arbeitsproben stellen ebenso wie betriebliche Simulationen Inszenierungen von Authentizität dar.
- Für betriebliche Arbeitsproben im internationalen Vergleich – die Rücksprachen mit den deutschen Unternehmen machen dies deutlich – besteht ein zentrales Problem darin, wie mit sensiblen Daten und unterschiedlicher Betriebsausstattung umzugehen wäre.
- Die Gewinnung der Stichprobe wäre für den kaufmännischen Bereich überaus aufwändig und kostenintensiv.
- Ein sinnvoller Vergleich der Leistungsfähigkeiten der Auszubildenden setzt voraus, dass die Tests zur Erfassung berufsfachlicher Kompetenz exakt beschrieben werden können, um Objektivität, Reliabilität und Validität des Verfahrens zu gewährleisten. Im Hinblick auf betriebliche Arbeitsproben besteht diese Kontrolle nur bedingt.

Der Einsatz des web-basierten Testformats ALUSIM hat gezeigt, dass die Testbedingungen sehr stark normiert werden können und dass die Auszubildenden eine simulierte Betriebsrealität als interessant und herausfordernd empfinden. Die vorgelegten und evaluierten Testbereiche und Testitems erlauben es, die berufliche Handlungskompetenz angehender Industriekaufleute sowohl mit ihrer (eher betriebsbezogenen) handlungsbasierten Kompetenz als auch mit ihrer (eher schulbezogenen) verstehensbasierten Kompetenz objektiv, reliabel und valide abzubilden und zugleich begründet nach Kompetenzstufen zu differenzieren. Bei der Auswertung kam es darauf an zu prüfen, ob sich die Anforderungssituationen mit dem gewählten Format dafür eignen, in einem international vergleichenden VET-LSA Verwendung zu finden. Die Aussage, die wir aufgrund des gesamten Vorgehens sowie der erhaltenen Daten treffen können, lautet uneingeschränkt: Ja!

## Literatur

- Achtenhagen, F. (2001). Criteria for the Development of Complex Teaching-Learning Environments. *Instructional Science*, 29, 361–380.
- Achtenhagen, F. (2004). Prüfung von Leistungsindikatoren für die Berufsbildung sowie zur Ausdifferenzierung beruflicher Leistungsprofile nach Wissensarten. In M. Baethge, K.-P. Buss & C. Lanfer (Hrsg.), *Expertisen zu den konzeptionellen Grundlagen für einen Nationalen Bildungsbericht – Berufliche Bildung und Weiterbildung/Lebenslanges Lernen* (S. 11–32). Berlin: BMBF.
- Achtenhagen, F. & Weber, S. (2003). „Authentizität“ in der Gestaltung beruflicher Lernumgebungen. In A. Bredow, R. Dobischat & J. Rottmann (Hrsg.), *Berufs- und Wirtschaftspädagogik von A bis Z* (S. 185–199). Baltmannsweiler: Schneider.
- Achtenhagen, F. & Winther, E. (2008). Wirtschaftspädagogische Forschung zur beruflichen Kompetenzentwicklung. In Bundesministerium für Bildung und Forschung (Ed.), *Kompetenzerfassung in pädagogischen Handlungsfelder: Theorien, Konzepte und Methoden*. Bildungsreform-Reihe des BMBF Band 26 (S.117–140). Bonn, Berlin: BMBF.
- Adams, R. J. & Khoo, S. T. (1996). *Quest*. Melbourne, Australia: Australian Council for Educational Research.
- Adams, R. J., Wilson, M. & Wang, W.-C. (1997). The Multidimensional Random Coefficients Multinomial Logit Model. *Applied Psychological Measurement*, 21 (1), 1–23.
- Adams, R. J. & Wu, M. (Eds.) (2002). *PISA 2000 Technical Report*. Paris: OECD.
- Anderson, L. W. & Krathwohl, D. R. (with Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., et al.) (Eds.) (2001). *A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Baethge, M., Achtenhagen, F., Arends, L., Babic, E., Baethge-Kinsky, V. & Weber, S. (2006). *Berufsbildungs-PISA – Machbarkeitsstudie*. Stuttgart: Steiner.
- Baumert, J., Bos, W. & Lehmann, R. (2000). *Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit*. Opladen: Leske und Budrich.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H. & Krathwohl, D. R. (Eds.) (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. New York: David McKay.
- Brand, W., Hofmeister, W. & Tramm, T. (2005). Auf dem Weg zu einem Kompetenzstufenmodell für die berufliche Bildung – Erfahrungen aus dem Projekt ULME. *bwp@ – Berufs- und Wirtschaftspädagogik online*, 8.

- Byrnes, J. P. (2008). *Cognitive Development and Learning in Instructional Contexts*. Boston et al.: Pearson.
- Cognition and Technology Group at Vanderbilt (CTGV) (1991). Technology and the Design of Generative Learning Environments. *Educational Technology*, 31 (5), 34–44.
- Embretson, S. E. (2002). Generating Abstract Reasoning Items with Cognitive Theory. In S. Irvine & P. Kyllonen (Eds.), *Generating Items for Cognitive Tests: Theory and Practice* (pp. 35–60). Mahwah, NJ: Lawrence Erlbaum.
- Embretson, S. E. & Gorin, J. S. (2001). Improving Construct Validity with Cognitive Psychology Principles. *Journal of Educational Measurement*, 38 (4), 343–368.
- Embretson, S. E. & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Gelman, R. & Greeno, J. G. (1989). On the Nature of Competence: Principles for Understanding in a Domain. In L. B. Resnick (Ed.), *Knowing, Learning, and Instruction. Essays in Honor of Robert Glaser* (pp. 125–186). Hillsdale, NJ: Erlbaum.
- Greeno, J. G. (with the Middle School Mathematics Through Applications Project Group) (1998). The Situativity of Knowing, Learning, and Research. *American Psychologist*, 53, 5–26.
- Doyle, W. (2000). *Authenticity*. Paper presented at the AERA Annual Meeting, New Orleans.
- Hartig, J. (2007). Skalierung und Definition von Kompetenzniveaus. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung* (S. 83–99). Weinheim: Beltz.
- Laux, H. (2007). *Entscheidungstheorie*. 7. Aufl. Berlin, Heidelberg, New York: Springer.
- Kilpatrick, I., Swafford, J. & Findell, B. (Eds.) (2001). *Adding it up: Helping Children Learn Mathematics*. National Research Council. Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- Köller, O., Watermann, R., Trautwein, U. & Lüdtke, O. (Hrsg.) (2004). *Wege zur Hochschulreife in Baden-Württemberg. TOSCA – Eine Untersuchung an allgemein bildenden und beruflichen Gymnasien*. Opladen: Leske und Budrich.
- Lehmann, R. & Seeber, S. (Hrsg.) (2007). *ULME III. Untersuchungen von Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in den Abschlussklassen der Berufsschulen*. Hamburg: HIBB.
- Markman, A. B. (1999). *Knowledge Representation*. Mahwah, NJ: Lawrence Erlbaum.
- Marzano, R. J. & Kendall, J. S. (2007). *The New Taxonomy of Educational Objectives*. Thousand Oaks, CA: Corwin Press.
- Mislevy, R. J. (2007). Cognitive Psychology and Educational Assessment. *Educational Measurement*, 128, 257–305.
- Mislevy, R. J. (2008). Issues of Structure and Issues of Scale in Assessment from a Situative/Sociocultural Perspective. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel & L. J. Young (Eds.), *Assessment, Equity, and Opportunity to Learn* (pp. 259–294). Cambridge: Cambridge University Press.
- Rasch, G. (1960). *Probabilistic Models for some Intelligence and Attainment Tests*. Copenhagen: The Danish Institute of Educational Research.
- Resnick, L. B. (1987). Learning in School and Out. *Educational Researcher*, 16 (9), 13–20.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. 2. Aufl. Bern: Huber.
- Rost, J. (2006). Zum Einsatz der Item-Response-Theorie für die Messung berufsbezogener Kompetenzen im Rahmen der Studie „Berufsbildungs-PISA“ – Expertise. In M. Baethge, F. Achtenhagen, L. Arends, E. Babic, V. Baethge-Kinsky & S. Weber, *Berufsbildungs-PISA – Machbarkeitsstudie* (S. XXXIV–XXXVII). Stuttgart: Steiner.
- Simonton, D. K. (2003). Scientific Creativity as Constrained Stochastic Behavior: The Integration of Product, Person, and Process Perspectives. *Psychological Bulletin*, 129, 475–494.

- Wilson, M. (2005). *Constructing Measures: An Item-response Modeling Approach*. Mahwah: Erlbaum.
- Wilson, M. (2008). Cognitive Diagnosis Using Item Response Models, *Journal of Psychology*, 216 (2), 74–88.
- Winther, E. (2006). *Motivation in Lernprozessen. Konzepte in der Unterrichtspraxis von Wirtschaftsgymnasien*. Wiesbaden: DUV.
- Winther, E. (2007). Performanz messen – Kompetenz diagnostizieren. In D. Lemmermöhle, M. Rothnagel, S. Bögeholz, M. Hasselhorn & R. Watermann (Eds.), *professionell lehren – erfolgreich lernen* (S. 303–316). Münster: Waxmann.
- Winther, E. (2008). *Vocational Competence – Constructing Measures in the Field of Business and Administration*. Paper presented at the Berkeley Evaluation and Assessment Research Center, University of California, Berkeley. November, 17th, 2008.
- Winther, E. (in Vorbereitung). *Kompetenzmodellierung und Kompetenzmessung in der beruflichen Bildung*. Habilitationsschrift.
- Winther, E. & Achtenhagen, F. (2008). Kompetenzstrukturmodell für die kaufmännische Bildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 104, 511–538.
- Winther, E. & Achtenhagen, F. (2009). Measurement of Vocational Competencies – A Contribution to an International Large-Scale-Assessment on Vocational Education and Training. *Empirical Research in Vocational Education and Training*, 1, 88–106.
- Wu, M. L., Adams, R. J., Wilson, M. & Haldane, S. A. (2007) *ACER ConQuest. Version 2.0. Generalised Item Response Software*. Camberwell: ACER Press.

Anschrift der Autoren: Dr. Esther Winther, Prof. Dr. Dr. h. c. mult. Frank Achtenhagen, Professur für Wirtschaftspädagogik der Georg-August-Universität Göttingen, Platz der Göttinger Sieben 5, 37073 Göttingen (ewinthe@uni-goettingen.de; fachten@uni-goettingen.de)