

Das ist doch nicht fair! – Mehrdimensionalität und Testfairness in kaufmännischen Assessments¹

KURZFASSUNG: Im Beitrag wird ein Modell beruflicher Kompetenzentwicklung vorgestellt. Unter Rückgriff auf dieses Modell werden vor dem Hintergrund der Kompetenzmessung in der kaufmännischen Domäne insbesondere die theoretischen und methodischen Aspekte von psychometrischer Struktur und Testfairness diskutiert. Hierzu werden zwei kaufmännische Assessments skizziert, die sich auf unterschiedliche Zeitpunkte und Phasen der beruflichen Kompetenzentwicklung beziehen lassen. Auf Basis der gewonnenen Erkenntnisse gelingt es, Übergangsprozesse in der beruflichen Bildung differenzierter zu beschreiben. So wird unter Nutzung differenzieller Itemanalysen (DIF) geprüft, ob in kaufmännischen Assessments zwischen Auszubildenden mit spezifischen individuellen Merkmalen oder im Hinblick auf eine ausgewählte Charakteristik der Ausbildungsbetriebe systematische Unterschiede existieren und wie sich diese interpretieren lassen. Für den Übergang in berufliche Ausbildungsgänge ist z.B. relevant, inwieweit berufliche Vorerfahrungen und fachspezifisches Wissen die weitere Kompetenzentwicklung beeinflussen und wie Unterschiede zwischen den Lernenden eines Ausbildungsganges lokalisiert und ausgeglichen werden können.

ABSTRACT: The paper presents a model of professional competence and is dealing with two aspects of interest regarding the measurement of competence in the domain of Business and Administration: The psychometric structure of competence as well as aspects of test fairness. Two assessments, which are settled in different phases of the competence development, will be described in order to picture transition processes in VET. Therefore, based on DIF analyses, aspects of fairness in competence-based testing as well as for the use in instructional processes to foster learning and teaching in the specific domain of economic sciences will be discussed. The DIF analyses are done in different ways: comparing individual characteristics as well as comparing companies' properties. To describe the transition from school to apprenticeship it may be relevant to identify substantial bias comparing the prior knowledge of apprentices as well as their experiences in VET programs to face differential processes of competence development.

1 Modell beruflicher Kompetenzentwicklung

1.1 Ausgangslage in der beruflichen Bildung

Ein wesentliches Desiderat der berufs- und wirtschaftspädagogischen Forschung ist die empirische Fundierung ihrer Kompetenzmodellierungen. Es mangelt nicht an inhaltlich ausgearbeiteten Konzepten, die ihren Ursprung in weiten Teilen in der pädagogischen Anthropologie HEINRICH ROTH'S (1971) haben, jedoch aber an einer Umsetzung dieser Konzepte in operationalisierbare Modelle. ROTH'S Ansatz ist der

1 Der Beitrag ist entstanden im Rahmen des Teilprojekts „Kompetenzorientierte Assessments in der beruflichen Bildung“ (Wi 3597/1-1) des DFG-Schwerpunktprogramms „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen (SPP 1293). Die empirischen Daten wurden im Rahmen des DFG-Projekts „Systemisches Verstehen von Geschäftsprozessen als kaufmännische Kompetenz“ (Ac 35/26-1) erhoben.

einer ganzheitlichen Handlungsfähigkeit, die sich über Sach-, Sozial- und personale Kompetenz ausdrücken lässt. Sacheinsichtige, sozialeinsichtige und moraleinsichtige Handlungsfähigkeit entwickelt sich in Abhängigkeit von der Lernumwelt, den Lern-, Sozialisations- und Erziehungsprozessen. Diese Annahmen charakterisieren das normative Leitbild der beruflichen Handlungskompetenz. Handlungskompetenz wird als die berufliche Fähigkeit bezeichnet, die es Individuen erlaubt, in spezifischen Anforderungssituationen angemessen zu handeln. Der Erwerb von Kompetenzen ist – im Sinne PIAGETS – das Resultat von Entwicklungs- und Lernprozessen (vgl. hierzu REETZ, 1999; BADER & MÜLLER, 2002). Das Modell beruflicher Handlungskompetenz ist jedoch ein Gestaltungsparadigma für berufliche Curricula und kein Modell, das als psychometrische Grundlage für die berufliche Kompetenzmessung herangezogen werden könnte. Die Schwierigkeit ist, dass die für eine erfolgreiche Ausübung beruflicher Tätigkeiten notwendigen Kompetenzen unzureichend erforscht sind und dass für eine Erfassung beruflicher Handlungskompetenz im Sinne einer komplexen, mehrdimensionalen, personalen Genese die forschungsmethodischen Voraussetzungen erst geschaffen werden müssen (WINTHER, 2010a; 2010b).

Als Alternative werden Kompetenzmodelle in der beruflichen Bildung – hier unter dem Anspruch einer empirischen Validierung – als Modelle potentieller Handlungsfähigkeiten und -möglichkeiten interpretiert. Der Fokus dieser Kompetenzmodelle liegt auf funktionalen und kontextspezifischen Operationalisierungs- und Messansätzen ohne einen allgemeinen Generalisierungsanspruch, aber valide im Hinblick auf ähnliche Aufgaben bzw. Situationen. Die Grundannahme besteht darin, dass die Struktur der konstruierten Anforderungssituationen zur Messung der Kompetenz der Struktur der zu erfassenden Kompetenz des Individuums entspricht und vice versa (vgl. hierzu SEEBER, 2008; SEEBER, NICKOLAUS, WINTHER ET AL., 2010; WINTHER & ACHTENHAGEN, 2009; 2010). Das Beschreiben der relationalen Beziehung zwischen einer beruflicher Anforderungssituation und der personalen Fähigkeit, diese und hinreichend ähnliche Anforderungssituationen bewältigen zu können, stellt für die Kompetenzmessung die zentrale Voraussetzung dar. Formal lässt sich diese Beziehung als mathematische Relation beschreiben, die den Zusammenhang zwischen Personenfähigkeit θ_v und Itemschwierigkeit δ_i im Hinblick auf das Antwortverhalten X_{vi} abbildet (Modelle der Item Response-Theorie, vgl. u. a. WRIGHT & STONE, 1979; WILSON, 2004):

$$P(X_{vi} = 1 | \theta_v, \delta_i) = \frac{\exp(\theta_v - \delta_i)}{1 + \exp(\theta_v - \delta_i)} = \frac{e^{(\theta_v - \delta_i)}}{1 + e^{(\theta_v - \delta_i)}}$$

Am Beispiel der Exponentialfunktion des 1-parametrischen logistischen Modells oder Rasch-Modells wird deutlich, dass das Antwortverhalten auf ein Testitem ausschließlich durch die Itemschwierigkeit und die Personenfähigkeit bestimmt wird. Diese Eigenschaft der Modelle der Rasch-Familie wird als Separabilität von Personen- und Itemmerkmalen beschrieben und 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. Die Wahl der Methoden der Kompetenzmessung spielt jedoch vor dem Hintergrund der Validierung beruflicher Kompetenzmessung – und folglich im Rahmen der theoretischen Modellierung von Kompetenzen in einer Domäne – eine wesentliche Rolle. Es sind insbesondere drei Perspektiven, die in der beruflichen Bildung eingenommen wer-

den und die sich in einem Koordinatensystem mit den Achsen (1) zunehmender Generalisierungsgrad des Kompetenzkonzepts sowie (2) zunehmender Abstraktionsgrad der (beruflichen) Anforderungssituation darstellen lassen: behavioristische und kognitivistische Auffassung von Kompetenz sowie Kompetenz als Sammlung allgemeiner Fähigkeiten (vgl. *Abbildung 1*).

Während berufliche Kompetenzen im Rahmen behavioristischer Auffassungen als die effektive Anwendung verfügbaren Wissens und Könnens in spezifischen Situationen definiert werden und damit in der Anwendung direkt beobachtbar sind, werden mit zunehmendem Grad an Generalisierung und Abstraktion Kompetenzen verstärkt als umfassende Konzepte und zusammenhängende Systeme von Kenntnissen, Fertigkeiten, Fähigkeiten und Einstellungen interpretiert, über die eine Person bei der Ausübung beruflicher Tätigkeiten verfügen sollte. Je genereller die Kompetenzvorstellungen sind, umso vielfältiger, aber zugleich auch abstrakter werden die Konstruktionen der (beruflichen) Anforderungen: Zur Abbildung generischer Kompetenzen werden berufliche Anforderungssituationen so zerlegt, dass sie sich als Bündel allgemeiner Kenntnisse, Fertigkeiten und Einstellungen beschreiben lassen, die nicht auf eine spezifische Situation oder eine Domäne beschränkt bleiben, sondern über verschiedene Berufsbereiche hinweg einsetzbar sind (z. B. Kommunikationsfähigkeiten). Diese Universalitätsannahme hat fast logisch zwingend zur Folge, dass valide Messungen als äußerst schwierig oder kaum realisierbar eingeschätzt werden (u. a. NORIS, 1991; SCHELLEN, 2004).

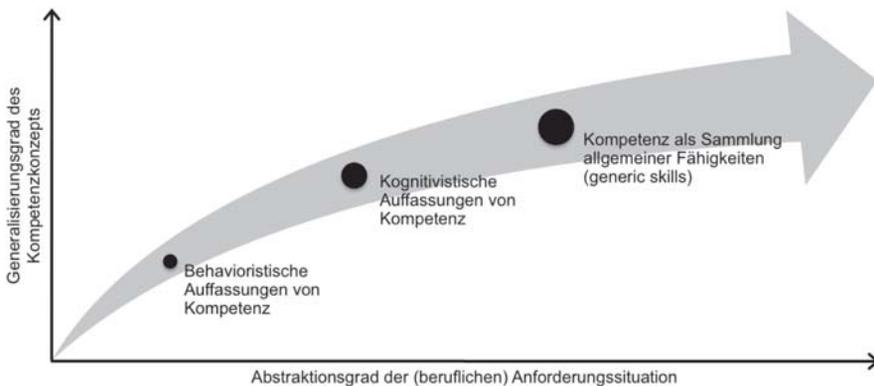


Abb. 1: Theoretische Auffassungen von Kompetenz

In der beruflichen Bildung ist mit der Annahme eines auf allgemeine, generische Fähigkeiten basierenden Kompetenzmodells zudem ein zentrales bildungspolitisches Problem berührt: Eine „Über“betonung der Bedeutung basaler Fähigkeiten (z. B. Rechnen, Lesen, Schreiben) in aktuellen Kompetenzmodellen steht der strukturellen Bewertung gegenüber, dass mit der zunehmenden Einbindung der beruflichen Bildung in das Berechtigungssystem der Stellenwert allgemeinbildender Fächer stärker stieg als der der beruflichen Inhalte (Debatte über die Gleichwertigkeit von allgemeiner und beruflicher Bildung sowie die Debatte über den Wert der beruflichen Bildung; vgl. in diesem Zusammenhang auch die aktuelle Debatte

über die Einordnung beruflicher Bildungsabschlüsse in den DQR). Damit wird man aber den jeweils domänenspezifischen Anforderungen im beruflichen Bereich nicht gerecht. Es werden folglich für die berufliche Bildung alternative Modelle benötigt, die Kompetenzen an das berufliche Handeln in spezifischen Situationen binden und mit denen sich die Binnenstruktur beruflicher Kompetenz, deren Graduierungen und Entwicklungen beschreiben lassen. Hiermit wird eine kognitivistische Auffassung von Kompetenz eingenommen: Kompetenz ist danach eine kognitive Funktion im Sinne des Potenzials, in spezifischen (auch neuen) beruflichen Anforderungssituationen (mit unterschiedlichem Anforderungsgehalt) angemessen handeln zu können. Kompetenz ist dabei nicht direkt beobachtbar, sondern erschließt sich aus der Performanz in einer Vielzahl hinreichend ähnlicher Situationen (vgl. WINTHER & ACHTENHAGEN, 2009; WINTHER, 2010a).

1.2 Von kaufmännischer zu professioneller Kompetenz

In der Expertise „Zur Entwicklung nationaler Bildungsstandards“ werden Kompetenzmodelle als Mittler zwischen abstrakten (curricularen) Bildungszielen und konkreten Aufgabenstellungen definiert (KLIEME, AVENARIUS, BLUM, DÖBRICH, GRUBER ET AL., 2003, S. 71), deren Ziel es ist, zu einer theoretischen Beschreibung der Struktur spezifischer Fähigkeiten und der Stufen ihres Erwerbs beizutragen (ebd., S. 65). So werden zur Beschreibung von Kompetenzentwicklung verschiedene theoretisch-konzeptionelle Ansätze genutzt, die von der Experten-Novizen-Forschung über Fragen der individuellen Förderung bis zu Entwicklungen lebenslangen Lernens reichen. Insgesamt lässt sich jedoch festhalten, dass momentan kaum elaborierte Kompetenzentwicklungsmodelle und geprüfte Testverfahren zur Diagnostik der Kompetenzentwicklung vorliegen. Zudem zeigt sich im Forschungsfeld der Kompetenzmessung der Trend, Kompetenzentwicklungsmodelle in inhaltlich sehr stark reduzierten Bereichen, z. B. „Experimentieren“ im Fach Physik (u. a. IPN, 2003) oder „Installation einer netzgekoppelten Photovoltaikanlage“ für den Beruf des Elektrikers / der Elektronikerin (RAUNER ET AL., 2009) zu erproben. Diese Ansätze lassen sich kaum plausibel auf Entwicklungsstufen der Kompetenz in einer spezifischen Domäne beziehen. Sie zeigen allerdings im Umkehrschluss, welche Herausforderungen und Anforderungen an die Annahmen bezüglich der Entwicklung von Kompetenz zu stellen sind: Kompetenzentwicklungsmodelle müssen davon ausgehen, dass Kompetenzerwerb und -verlauf von vielseitigen Interaktionen des Lernenden mit der Umwelt geprägt sind; sie können damit Fragen der Erlern- und Vermittelbarkeit von Kompetenzen beantworten und lassen sich folglich insbesondere als Planungsmodelle für Lehr-Lernprozesse nutzen. Sie sind als empirisch fundierte Modelle zu definieren, die Strukturen und Entwicklungsverläufe von Kompetenzen unter Berücksichtigung spezifischer Inhalte aufführen. Das Ziel von Kompetenzentwicklungsmodellen besteht darin, Kompetenzstufen bzw. -niveaus nicht allein aus der Analyse der Anforderungssituation abzuleiten (vgl. Kompetenzniveaumodelle), sondern die Stufen als entwicklungspsychologisch begründete qualitative Unterschiede zu interpretieren (vgl. IPN, 2003, S. 18). Vorbild hierfür sind in erster Linie die Arbeiten PIAGETS zum Modell der kognitiven Entwicklung (u. a. 1959). Zentral ist die Annahme, dass in jeder Entwicklungsstufe die vorhergehenden einfacheren Stufen differenziert erfasst und neu integriert werden. Dies impliziert, dass die Vorgänge beim Lernenden sehr detailliert zu erheben sind. So sind Art und Umfang der kog-

nitiven Repräsentation, Lernschwierigkeiten bzw. -hindernisse, Lerngelegenheiten oder auch vorunterrichtliche Vorstellungen diejenigen Faktoren, die die qualitativen Unterschiede der Denk- und Problemlösemodi hinsichtlich gleicher Probleme in den verschiedenen Alters- bzw. Entwicklungsstufen begründen können. Ein Beispiel eines Kompetenzmodells, das sich an den Arbeiten PIAGETS orientiert, befasst sich mit Entwicklungsstudien zur mathematischen Grundbildung (VAN DE RIJT, VAN LUIT & PENNING, 1999). Für acht theoretisch abgeleitete und fachbezogene Aspekte mathematischen Wissens – z.B. „concept of comparison“, „seriation“, „structured counting“ – wurden Items konstruiert und in eine unidimensionale Skala transformiert, die mathematische Kompetenz differenziert abzubilden vermag (vgl. auch VAN DE RIJT, GODFREY, AUBREY, VAN LUIT, GHESQUIERE ET AL., 2003).

Kompetenzentwicklung kann sich als Veränderung der Kompetenzstruktur sowohl im Sinne einer Differenzierung nach Teilkompetenzen als auch im Sinne einer Integration verschiedener Teilkompetenzen zeigen – hierzu existieren qualitative wie quantitative Befunde, die sich wiederum in Kompetenzmodelle integrieren lassen (vgl. SCHECKER & PARCHMANN, 2006). Ein Beispiel für diese Sichtweise ist WINTHER (2010a) entnommen (vgl. *Abbildung 2*): Für die unterschiedlichen Stationen der professionellen Entwicklung von kaufmännischen Auszubildenden und Angestellten lassen sich drei verschiedene theoriebasierte Kompetenz(structur)modelle unterscheiden, die im Sinne einer Entwicklungsperspektive im Rahmen beruflicher Karrieren aufeinander bezogen werden können: Kaufmännische Kompetenz zu Ausbildungsbeginn und im Ausbildungsverlauf sowie professionelle Kompetenz im Rahmen eines Beschäftigungsverhältnisses in kaufmännischen Berufsfeldern.

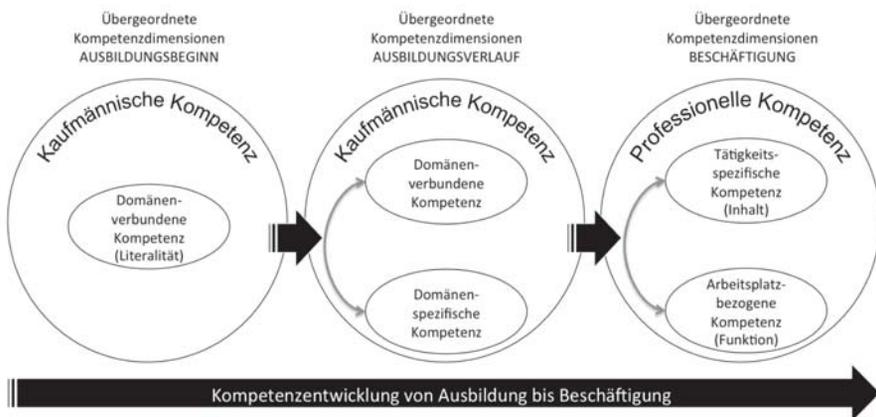


Abb. 2: Veränderung der Kompetenzstruktur im Verlauf professioneller Entwicklung (WINTHER, 2010a, S. 259)

Die Veränderung der Kompetenzstrukturen in den verschiedenen Entwicklungsstadien berücksichtigt, dass sich im Zuge der beruflichen Kompetenzentwicklung Strategien der Anforderungsbewältigung herausbilden, die von dem einzelnen Auszubildenden/Arbeitnehmer als effektiv bewertet werden und die in der Folge zu anderen Kompetenzkonstrukten führen. Im Verlauf der Ausbildung werden

domänenverbundene Fähigkeitsstrukturen zunehmend um domänenspezifische erweitert. Domänenverbundene und domänenspezifische Kompetenzen sind damit spezifische Dimensionen kaufmännischer Kompetenz im Rahmen beruflicher Ausbildungsgänge. Es wird darüber hinaus angenommen, dass sich die domänenspezifische Kompetenz im Verlauf der beruflichen Kompetenzentwicklung im Rahmen eines Beschäftigungsverhältnisses durch eine Kategorie ablösen lässt, die als professionelle Kompetenz überwiegend über tätigkeitsspezifische und arbeitsplatzbezogene Faktoren charakterisiert werden kann. Die Kompetenzentwicklung wird in diesem Modell folglich über das Konzept der Literalität sowie über das Konzept der (zunehmenden) Professionalität in kaufmännisch relevanten Anforderungssituationen beschrieben und dabei an eine zentrale Fragestellung gebunden: Wie beeinflusst domänenverbundene Literalität den Auf- und Ausbau domänenspezifischer, professioneller Kompetenz? Literalität wird unter Bezug auf internationale Leistungsvergleiche (IALS, ALL, PISA) als Fähigkeit formuliert, text-, bild- und/oder zahlssprachige Informationen in alltagsnahen Kontexten zu verarbeiten und zu nutzen. Für den Bereich der kaufmännischen Kompetenz zeigen sich diese Kontexte insbesondere im Hinblick auf mathematische und sprachliche Facetten eines mit allgemeinen Fähigkeiten zu bewältigenden ökonomischen Konzepts. Professionelle Kompetenz in einer spezifischen Domäne wird hingegen im Sinne persönlich fachlicher Voraussetzung definiert, um in beruflichen Anforderungen angemessen agieren zu können. Professionelle Kompetenz wird dabei in einem berufsbiographischen Prozess erworben und lässt sich insbesondere in Kontexten sichern, die nicht durch Routinen geprägt sind (zur professionellen Kompetenz beispielsweise von Lehrkräften vgl. u.a. HOYLE, 2001; EVANS, 2002; zur professionellen Kompetenz von Managern vgl. EGGER-SUBOTITSCH & KREIML, 2009; zur Beziehung zwischen domänenverbundenen und domänenspezifischen Kompetenzen vgl. insbesondere GELMAN & GREENO, 1989, GREENO, 1998).

In diesem Zusammenhang lässt sich das in *Abbildung 2* dargestellte Modell der Kompetenzentwicklung in der kaufmännischen Domäne auch als ein Modell der Beschreibung von Übergangsprozessen interpretieren: den Übergängen von Schule in Ausbildung einerseits, die vorrangig durch domänenverbundene Kompetenzvorstellungen im Sinne der Literacy-Konzeption geprägt sind (vgl. ausführlich Kapitel 3.1), und den Übergängen von Ausbildung in Beschäftigung andererseits bei denen zwischen domänenverbundenen und domänenspezifischen Kompetenzbereichen differenziert wird (hierzu Kapitel 3.2). Im Folgenden werden für diese zwei Übergänge Assessments in der kaufmännischen Domäne vorgestellt, die im Verständnis der *Abbildung 1* einer kognitiven Kompetenzauffassung folgen. Unter Rückgriff auf das Modell beruflicher Kompetenzentwicklung werden für die zwei Übergangsprozesse entsprechende psychometrische Kompetenzmodellierungen skizziert. Die Auswertung der Assessments wird insbesondere vor dem Hintergrund der Verwertbarkeit dieser Assessments als Instrumente der Kompetenzdiagnostik diskutiert, wobei hier vorrangig Fragen der Fairness betrachtet werden. Die theoretischen und methodischen Grundlagen zur Beschreibung von Test- und Itemfairness werden im Kapitel 2 vorangestellt.

2 Verwertbarkeit von Assessments in der beruflichen Bildung

2.1 Fairness von Assessments

Zentrale Merkmale von Assessments in Übergangsprozessen sollten deren anwendungsorientierte Nutzung sowie im Hinblick auf zertifizierende Prüfungsverfahren deren Fairness sein (vgl. hierzu Beiträge in RYAN & SHEPARD, 2008). Hierfür ist notwendig, dass die Testaufgaben der Prüfungsverfahren die Anforderungen erfüllen, die an ein kompetenzbasiertes, d. h. am Output der Lernenden orientiertes Prüfen, wie es in den Curricula sowie Richtlinien der beruflichen Bildung eingefordert, jedoch bislang nicht umgesetzt wird, gestellt werden: Sie sollen valide im Hinblick auf spezifische berufliche Anforderungssituationen sein und eine Erfassung berufsfachlicher Kompetenzen auf Basis eines empirisch validierten Kompetenzmodells erlauben, mit dessen Hilfe sowohl Struktur als auch Graduierung beruflicher Handlungen beschrieben werden können. Testfairness bedeutet, dass Probandengruppen mit spezifischen Merkmalen aber gleicher Fähigkeitsausprägung im Hinblick auf einzelne Bereiche eines Tests nicht systematisch bevorzugt oder benachteiligt werden. Die Merkmale, anhand derer die Fairness eines Prüfungsverfahrens beschrieben wird, sind in der Regel funktionale Gradienten des Bildungsausgangs, wie beispielsweise das Geschlecht, die ethnische Herkunft oder der soziale Status. Für die berufliche Bildung, deren Bildungsgänge sich durch eine hohe Heterogenität der Auszubildenden bzw. der Schülerschaft sowie durch eine weitreichende Durchlässigkeit zwischen den Bildungsgängen kennzeichnen lassen, werden darüber hinaus weitere Faktoren der Testfairness relevant, die sich insbesondere an den Bedingungen der Übergangsmärkte und der nachfolgenden Lern- und Bildungskarrieren festmachen lassen. Für den Übergang von Schule in Ausbildung ist vor dem Hintergrund dieser Überlegungen z. B. von Interesse, wie sich unterschiedliche Bildungskarrieren vor dem Eintritt in einen spezifischen beruflichen Bildungsgang auf den Umgang mit domänenverbundenen Anforderungssituationen und damit auf das konzeptionelle Verständnis alltagsbezogener ökonomischer Konzepte auswirken und inwieweit berufliche Vorerfahrungen und fachspezifisches Wissen die weitere Kompetenzentwicklung in den Bildungsgängen beeinflussen.² Für den Übergang von der Ausbildung in eine Beschäftigung werden Effekte relevant, die sich zusätzlich auf Merkmale des Ausbildungsbetriebes beziehen lassen. So ist vor allem mit Blick auf die kaufmännischen Abschlussprüfungen von Interesse, ob die unterschiedlichen Branchen oder auch die Betriebsgrößen der Ausbildungsbetriebe zu Bevorzugungen bzw. Benachteiligungen der Auszubildenden im Prüfungsverfahren führen und ob dies letztlich einen Einfluss auf die Übernahme in ein Beschäftigungsverhältnis hat.³

2 Von besonderem Interesse sind in diesem Zusammenhang auch Fragen der Lokalisierung von Misskonzepten sowie der Umgang mit Misskonzepten innerhalb der Lehr-Lernprozesse eines Ausbildungsganges. Empirische Befunde zeigen, dass ein Überwinden von misskonzeptbehafteten Vorstellungen kaum gelingt (u. a. WEBER, 1994) und insbesondere dadurch ein Ausgleichen der Leistungen von Lernenden besondere Schwierigkeiten aufweist (für eine Leistungssteigerung bei Leistungsausgleich vgl. u. a. HELMKE, 1988) – dies mit den entsprechenden Konsequenzen für die berufliche Kompetenzentwicklung.

3 Für die berufliche Bildung gibt es keine empirischen Evidenzen dafür, dass die Übernahme in ein Beschäftigungsverhältnis von den Leistungen in den Abschlussprüfungen abhängig ist. Es ist jedoch gängige Praxis, dass die Verhandlungen über Beschäftigungsverträge vor Abschluss des

Zur Klärung dieser Fragen sind Forschungsergebnisse notwendig, auf deren Basis sich Eigenschaften „fairer“ Items und Testverfahren beschreiben lassen.

2.2 Differenzielle Itemanalysen zur Prüfung der Testfairness

Anhand differenzieller Itemanalysen (Differential Item Functioning Analysis (DIF); vgl. u. a. HOLLAND & THAYER, 1986; MOORE, 1996) wird geprüft, ob sich bei einzelnen Items eines Tests Unterschiede zwischen Subgruppen ergeben, die verschiedene Merkmale aufweisen. Zentrales Merkmal der DIF-Analyse ist, dass keine generellen differenziellen Effekte zwischen Gruppen untersucht werden, sondern Unterschiede im Antwortverhalten von Personen mit gleicher Fähigkeit im Vergleich über die differenziell zu unterscheidenden Gruppen: „If the responses to an item have different frequencies for different sub-groups, that is evidence of differential impact of the item on those sub-groups [...] Instead DIF studies focus on whether respondents at the same locations give similar responses across the different subgroups“ (WILSON, 2005, p. 165). Bei der DIF-Analyse wird folglich geprüft, ob die Testitems „fair“ für alle Probanden sind oder ob Items im Test enthalten sind, die eine Subgruppe bei Annahme gleicher Fähigkeiten systematisch bevorzugen oder benachteiligen. Empirisches Kriterium für die Existenz differenzieller Unterschiede ist die Differenz zwischen gruppenspezifischen Itemschwierigkeiten. Übertragen auf das Rasch-Modell ergibt sich folgende formale Struktur:

Rasch-Modell (Logarithmierter Wettquotient):

$$\ln \frac{P(X_{vi} = 1 | \theta_v, \delta_i)}{P(X_{vi} = 0 | \theta_v, \delta_i)} = \theta_v - \delta_i$$

Rasch-Modell für DIF:

$$\ln \frac{P(X_{vi} = 1 | \theta_v, \delta_i)}{P(X_{vi} = 0 | \theta_v, \delta_i)} = \theta_v - \delta_i - \gamma * X_i Z_v$$

Die Personenfähigkeit ist mit θ_v und die Itemschwierigkeit mit δ_i gekennzeichnet. Der DIF-Effekt beschreibt die Item*Gruppe-Interaktion (γ). X_i ist dabei der Indikator für das Item und Z_v der Indikator für die Persongruppen. Werden systematische Effekte der Bevorzugungen bzw. Benachteiligung auf Itemebene untersucht, ist Z_v ein dichotomer Gruppenindikator, bei dem die Referenzgruppe (z.B. Gruppe von Lernenden ohne kaufmännische Vorbildung oder Gruppe von Auszubildenden in Ausbildungsbetrieben mit weniger als 500 Beschäftigten) mit 0 und die Vergleichsgruppe (z.B. Gruppe von Lernenden mit kaufmännischer Vorbildung oder Gruppe von Auszubildenden in Ausbildungsbetrieben ab 500 Beschäftigten) mit 1 kodiert ist.

Zusammenfassend ergeben sich die in *Abbildung 3* dargestellten Funktionsgleichungen für Items ohne DIF-Effekt und für Items mit systematisch angenommenem DIF-Effekt. Liegen DIF-Effekte vor, weisen die Items für verschiedene Gruppen differenzielle Verläufe auf. Oder anders formuliert: Die item-charakteristischen Funktionen

Prüfungsverfahrens geführt werden. Dies ist ein deutlicher Indikator dafür, dass den Abschlussprüfungen in betrieblichen Kontexten nicht die ihnen zugeordnete Bedeutung zugemessen wird.

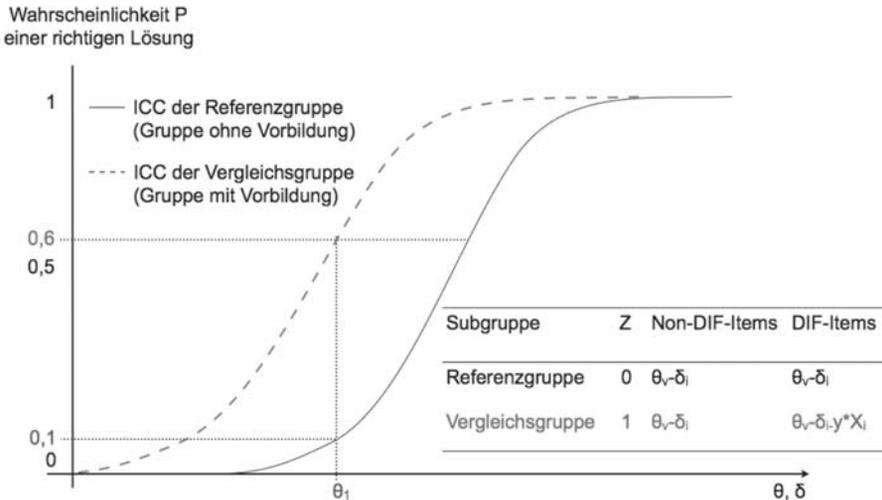


Abb. 3: DIF-Analyse auf Basis des Rasch-Modells sowie Verlauf der ICCs

(Item Characteristic Curves, ICC) unterscheiden sich bei Vorlage eines DIF-Effekts substantiell voneinander. In den meisten Anwendungskontexten lässt sich ein DIF-Effekt als Unterschied in den Schwierigkeiten der Items und damit als Lageunterschied zwischen den ICCs der Vergleichs- und Referenzgruppen interpretieren (vgl. *Abbildung 3*): Liegt die ICC der Vergleichsgruppe beispielsweise links von der ICC der Referenzgruppe, haben bei gleicher Fähigkeit Mitglieder der Vergleichsgruppe ein höhere Wahrscheinlichkeit, das Item zu lösen (0,6 im Vergleich zu 0,1).

3 Befunde zur psychometrischen Struktur und zur Testfairness von Kompetenzmessungen in Übergangsprozessen

3.1 Kaufmännische Kompetenz im Übergang von allgemein bildenden in berufliche Bildungsgänge

Eine besondere Herausforderung bei der Konstruktion von Assessments für berufliche Kontexte stellt die Heterogenität der Akteure dar. Insbesondere die formal in allgemein bildenden oder beruflichen Vollzeitbildungsgängen erworbenen fachspezifischen Vorwissensleistungen sind hierbei von besonderer Bedeutung. Das Modell kaufmännischer Kompetenzentwicklung (vgl. *Abbildung 2*) geht davon, dass in einer ersten berufsbezogenen Übergangsphase die Ausprägungen allgemeiner domänenverbundener Kompetenzen den weiteren beruflichen Kompetenzerwerb maßgeblich beeinflussen. Unter domänenverbundener Kompetenz werden aus kognitionspsychologischer Perspektive „Pick-up-Schemata“ verstanden, mit denen verfügbare Repräsentationen und/oder Verhaltensmuster charakterisiert werden, die stark ausführungsbezogen sind, um die Bewältigung einer gegebenen (beruflichen bzw. berufsbezogenen) Anforderungssituation zu unterstützen (GELMAN & GREENO, 1989; GREENO, 1998). Für die Konstruktion von Anforderungssituationen, in denen

domänenverbundene Kompetenz erfasst werden soll, hat diese Differenzierung insbesondere mit Blick auf die Reichweite des Kompetenzbegriffs Konsequenzen. Es ist (1) nach spezifischen kognitiven Strukturen zu fragen, die in einer gegebenen Anforderungssituation als aktives Zugriffswissen verfügbar sein müssen (vgl. ANDERSON, 1996), und (2) sind unterstützende Verhaltensmuster und Repräsentationen zu ermitteln, die die Bewältigung der Anforderungssituation begünstigen. Vor diesem Hintergrund ist es sinnvoll, zur Erfassung domänenverbundener Kompetenz auf die Literacy/Numeracy-Konzeptionen der internationalen Vergleichsstudien zurückzugreifen. Die Literacy/Numeracy-Konzeption abstrahiert von der Fachwissenschaft und bindet Anforderungssituationen an Kontexte und Anwendungen des Alltags. Hierdurch wird deutlich gemacht, dass nicht Kenntnisse und Fähigkeiten, wie sie traditionell in den Curricula definiert werden, im Vordergrund stehen, sondern dass es um die funktionale Anwendung von Kenntnissen in der kaufmännischen Domäne geht.

VW korrigiert Gewinnziel nach unten

Mehr als 500 Millionen Euro unter Plan / Neues Oberklassemodell auf Eis gelegt

rit. HAMBURG, 20. Juli. Der Volkswagen-Konzern wird sein Ergebnisziel auch in diesem Jahr verfehlen. Nach Informationen dieser Zeitung rechnet Europas größter Automobilhersteller für 2004 nur noch mit einem operativen Ergebnis vor Sonderinflüssen von 1,5 bis 2 Milliarden Euro. Ursprünglich wollte Volkswagen das Vorjahresergebnis von 2,5 Milliarden Euro übertreffen. Die revidierte Prognose soll am Freitag im Halbjahresbericht veröffentlicht werden, heißt es. Ein VW-Sprecher wollte diese Informationen nicht kommentieren. Den Informationen zufolge ist im ersten Halbjahr – trotz der stabilen Erträge aus der Finanzdienstleistungssparte – mit einem kräftigen Gewinneinbruch zu rechnen. Als Konsequenz aus der abermals verschlechterten Ertragslage hat der Vorstand die Pläne für den Bau des „C1“ vorerst auf Eis gelegt. Mit dem Modell, das 2007 in Serie gehen sollte, wollte VW die Lücke zwischen Passat und Phaeton schließen.

Trotz der Gewinnrevision wird VW am Freitag voraussichtlich keine Verschärfung des im Frühjahr eingeleiteten Restrukturierungskurses ankündigen. Damals wurde das Sparziel für 2004 und 2005 auf insgesamt 4 Milliarden Euro verdoppelt. Die jetzige Zurückhaltung hängt womöglich damit zusammen, daß man vor den im September startenden Tarifverhandlungen keine Drohkulisse errichten will. Gleichwohl steigt der Druck auf die Gewerkschaften, bei den Personalkosten kompromissbereiter zu sein. VW will die Personalkosten bis 2011 um 30 Prozent senken.

Schon in den ersten drei Monaten des Jahres war das operative Konzernergebnis

um 46 Prozent auf 329 Millionen Euro eingebrochen. Seither hat sich das von Konjunktur- und Konsumschwäche geprägte Marktumfeld kaum verbessert. Der Verband der Automobilindustrie erwartet allenfalls noch einen stagnierenden Inlandsabsatz. Auf die europaweit zögerliche Nachfrage reagieren alle Hersteller mit zum Teil kräftigen Preisnachlässen; das hinterläßt deutliche Spuren in den Bilanzen von Ford und General Motors (mitsamt der Tochtergesellschaft Opel). Auch die französischen Hersteller Renault und PSA Peugeot Citroën sowie die italienische Fiat bekommen den Preisdruck zu spüren. Unter den Massenherstellern schwimmt Toyota indes gegen den Strom. Dank der hohen Produktivität eilt der größte japanische Autobauer

bei Absatz und Ergebnis von Rekord zu Rekord.

VW hat die Kaufanreize im Verlauf dieses Jahres stetig erhöht. Das gilt vor allem für den neuen Golf. Nach der kostenlosen Zugabe einer Klimaanlage gibt es seit Juni zusätzlich eine Händlerprämie von 928 Euro für die Inzahlungnahme eines Gebrauchtwagens. Dies stabilisiert den Absatz, verengt aber die Gewinnmargen – und zwar nicht nur beim Golf. Denn um die Differenzierung der Konzernmodelle untereinander aufrechtzuerhalten, müssen auch die Preise der übrigen Modelle wie VW Polo, Skoda Octavia oder Seat Ibiza (direkt oder indirekt) gesenkt werden.

Zudem gerät VW auf wichtigen Auslandsmärkten unter Druck. In Amerika macht den Wolfsburgern die Dollar-Schwäche zu schaffen: Bei Durchschnittskursen zwischen 1,20 und 1,25 Dollar je Euro dürfte das operative Ergebnis im Gesamtjahr um rund 1,2 Milliarden Euro belastet werden. Auch toben in Amerika, wo der VW-Absatz unter dem Modellwechsel bei Jetta und Passat leidet, noch härtere Rabattschlachten als in Europa. All dies beschränkt allein im ersten Quartal in Nordamerika einen Verlust von 235 Millionen Euro. Der chinesische Markt, auf dem VW in der Vergangenheit blendend verdiente, hat stark an Dynamik verloren. Dort versuchen die Konkurrenten – allen voran General Motors –, VW mit Kampfpreisen die Marktführerschaft zu entreißen. Die Wolfsburgern müssen wohl mit einem geringeren Ergebnisbeitrag aus China rechnen.

Toyota erhöht Absatzziel, Seite 16.
Ford profitiert von seiner Finanzsparte, Seite 16.

Volkswagen im Vergleich

Kurs 31. Dezember 2003=Index 100



Abb. 4: Realer Kontext des Assessments zur Erfassung domänenverbundener Kompetenz (FAZ, 2004; vgl. WINTHER 2010a, S. 170)

Das im hier vorzustellenden Fall gewählte Assessment zur Erfassung kaufmännischer Kompetenz in der Eingangsphase beruflicher Ausbildungsgänge wurde auf

der Grundlage eines Artikels der Frankfurter Allgemeinen Zeitung (FAZ; vgl. *Abbildung 4*) erhoben und fokussiert auf text- und bildsprachliche Kenntnisse sowie auf das Verständnis von quantitativen Werten und Verhältnissen, wobei der zugrunde gelegte Kontext verschiedene kaufmännische Inhaltsbereiche auf Alltagsniveau erfasst. So werden beispielsweise Vorstellungen zum betriebswirtschaftlichen Gewinnkonzept, zum Anleger- und Gläubigerschutz, zu marktwirtschaftlichen Unternehmensstrategien und zu den Anspruchs- und Interessengruppen einer Unternehmung berücksichtigt. Der ökonomische Alltagsbezug wird dadurch realisiert, dass die Items des Assessments an einen Artikel des Wirtschaftsressorts gebunden sind. Der Zeitungsartikel stellt einen realen Kontext zur Verfügung, über den (1) ökonomische Fragestellungen bearbeitet werden können und der (2) eine Quelle realer Zahlen und Fakten darstellt, die von den Lernenden/Auszubildenden erfasst und interpretiert werden müssen.

Die nachfolgenden Ausführungen sind Präzisierungen des in *Abbildung 2* dargestellten Modells beruflicher Kompetenzentwicklung in der kaufmännischen Domäne. Für die Präzisierungen ist es notwendig, die heuristischen Modellvorstellungen (a) in kognitive Kompetenzstrukturen zu übersetzen, die (b) in Form psychometrischer Modellierungen einer Kompetenzerfassung zugänglich werden. Vor diesem Hintergrund geht das kognitive Kompetenzmodell für den ersten Entwicklungsabschnitt: der Übergang von Schule in berufliche Ausbildungsgänge, davon aus, dass sich domänenverbundene Kompetenz als mehrdimensionales Konstrukt interpretieren lässt (vgl. *Abbildung 6*). Für kaufmännische Anforderungssituationen lassen sich zwei Zugänge zur Domäne identifizieren: ein sprachlich-argumentativer sowie ein mathematisch-analytischer Zugang. Diese Zugänge spiegeln sich sowohl in den Ausdifferenzierungen der Fachwissenschaft wider als auch in den angenommenen kognitiven Fähigkeitsstrukturen, die einer Kompetenzmessung zugrunde gelegt werden müssen. Im psychometrischen Kompetenzmodell werden die kognitiven Fähigkeiten über zwei Kompetenzdimensionen abgebildet, die über zwei Skalen erfasst werden: Die Items der Skala „economic literacy“ beziehen sich vorrangig auf den Fließtext des Artikels, mit dessen Hilfe Anforderungssituationen konstruiert werden, deren Bewältigung **sprachlich-argumentative Kompetenzen** voraussetzen. Über die Skala „economic numeracy“ werden **mathematisch-analytische Kompetenzen** erfasst. Die Items der Skala beziehen sich auf grundlegende mathematische Kenntnisse und Fertigkeiten vor dem Hintergrund der konkreten, im Artikel über Graphiken und Zahlen präsentierten Unternehmenswerte.

Das Assessment zur Erfassung domänenverbundener Kompetenz wurde bei 340 Lernenden des Fachgymnasiums Wirtschaft eingesetzt, um kaufmännisches Vorwissen zu erheben. 29 Prozent der Lernenden verfügten bereits über eine Vorbildung in der kaufmännischen Domäne. Insbesondere die Höhere Handelsschule ist ein bedeutender Zubringer für diesen kaufmännischen Ausbildungsgang (77 Prozent); abgeschlossene Berufsausbildungen im kaufmännisch-verwaltenden Bereich lagen nicht vor. Der Test umfasst insgesamt 27 Items, von denen 21 Items die Skala „economic literacy“ und sechs Items die Skala „economic numeracy“ repräsentieren. Die Skalen- und Itemkennwerte der klassischen Testtheorie weisen akzeptable Reliabilitätschätzer ($\alpha_{\text{Literacy}} = 0,771$; $\alpha_{\text{Numeracy}} = 0,785$) auf. Zudem ist die Korrelation zwischen den Skalen ($r = 0,326$) nicht bedeutend hoch. Die im Folgenden dargestellten empirischen Analysen beziehen sich auf zwei zentrale Fragestellungen: (1) Wird mit dem domänenverbundenen Assessment eine Fähigkeitsstruktur

Skala „economic literacy“

Item Itemstamm

- | Item | Itemstamm |
|------|--|
| 01 | Welche Gründe für den Gewinneinbruch bei der Volkswagen AG erkennen Sie in dem Zeitungsartikel? |
| 16 | Das Volkswagenwerk in Wolfsburg erhält vom Land Thüringen den Auftrag, möglichst kurzfristig 2.800 VW Passat zu liefern. Die Pkw sollen den Fuhrpark der Polizei an einigen zentralen Standorten erweitern. Warum ist es für die Volkswagen AG wichtig, kurzfristig liefern zu können? |

Skala „economic numeracy“

Item Itemstamm

- | Item | Itemstamm |
|------|--|
| 26 | Sie sind Aktionär der Volkswagen AG. Der Wert Ihrer VW-Aktien betrug am 31.12.2003 17.500,- €. Verwenden Sie für die folgenden Berechnungen bitte die Angaben aus der Grafik des Zeitungsartikels. Berechnen Sie bitte Ihren Verlust im Vergleich zu einer Investition in den DAX 30. Wie hoch ist Ihr Verlust in Prozent? |
| 27 | Bevor die 2800 VW Passat an die Polizei des Landes Thüringen ausgeliefert werden können, müssen die Pkw mit den für Polizeiwagen typischen grünen Folienstreifen beklebt werden. Es gibt zwei Zulieferer dieser Folien, die zu den unten genannten fixen Bruttopreisen zu liefern bereit sind. Der Bedarf an Folie pro Wagen beträgt 7 m Folie am Stück. Der Zulieferer A liefert für 14,11 Euro pro laufenden Meter (14,11 €/lfm) in 1.000 Meter Rollen. Der Zulieferer B liefert für 14,29 Euro pro laufendem Meter (14,29 €/lfm) in 140 Meter Rollen. Ermitteln Sie für diesen Auftrag das preisgünstigere Angebot! |

Abb. 5: Itembeispiele der Skalen „economic literacy“ und „economic numeracy“

erfasst, oder werden über die Skalen „economic literacy“ und „economic numeracy“ zwei verschiedene Fähigkeitsstrukturen angesprochen, die sich als sprachlich-argumentative und mathematisch-analytische Kompetenzbereiche identifizieren lassen? – Dies ist die Frage nach der Dimensionalität bzw. Struktur von Kompetenz. (2) Gelingt es, mittels der Literacy/Numeracy-Konzeption die Heterogenität der Lernenden in beruflichen Ausbildungsgängen insbesondere im Hinblick auf ihre berufsbezogene kaufmännische Vorbildung so zu berücksichtigen, dass keine systematischen Vor- oder Nachteile für eine der Subgruppen resultieren? – Dies ist eine Frage der Testfairness.⁴

Zur Dimensionalität domänenverbundener Kompetenz

Das kognitive und das psychometrische Kompetenzmodell zur Beschreibung domänenverbundener Kompetenz differenzieren zwischen zwei Kompetenzdimensionen bzw. Fähigkeitsstrukturen (vgl. *Abbildung 6*): sprachlich-argumentative Kompetenz sowie mathematisch-analytische Kompetenz.

Um zu prüfen, ob das theoretisch formulierte Modell für die erhobenen Daten passend ist, werden ein eindimensionales sowie ein zweidimensionales Modell miteinander verglichen. Für das eindimensionale Modell gilt, dass domänenverbundene Kompetenz im Übergang von der Allgemeinbildung in einen beruflichen Ausbildungsgang als latente Variable höherer Ordnung betrachtet wird und der Testscore die suffiziente Statistik für den Schätzer einer übergeordneten domänen-

4 Für einen Überblick über alle Kennwerte und Auswertungen siehe WINTHER (2010a, S. 175ff.).

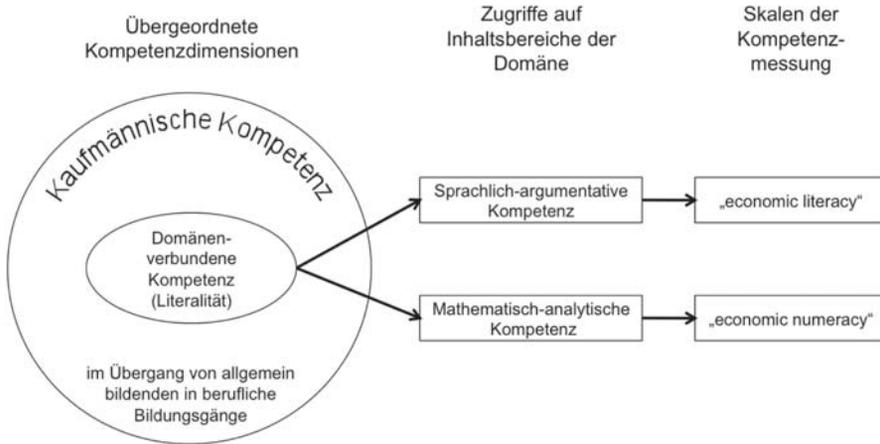


Abb. 6: Theoretisches Kompetenzstrukturmodell domänenverbundener Kompetenz

verbundenen Kompetenz darstellt. Für das zweidimensionale Modell gilt hingegen, dass domänenverbundene Kompetenz über zwei voneinander unterscheidbare Fähigkeitsstrukturen gemessen wird. Die Schätzung der latenten Fähigkeiten erfolgt hierbei simultan über die Skalen „economic literacy“ und „economic numeracy“. Die simultane Schätzung der Item- und Personenparameter ermöglicht es, Messwerte für beide Skalen unabhängig voneinander zu gewinnen. Zudem kann die Korrelation zwischen beiden Skalen als Effektstärke der Dimensionalität interpretiert werden (EMBRETSON & REISE, 2001, p. 75).

Um Modelle im Rahmen der Item Response-Theorie direkt miteinander vergleichen zu können, sind Ansätze der IRT zu wählen, die sich hierarchisch aufeinander beziehen lassen. Für das eindimensionale Modell wird auf das Rasch-Modell (RASCH, 1960) und für das zweidimensionale Modell auf das Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM; ADAMS, WILSON & WANG, 1997) zurückgegriffen. Beide Modelle lassen sich mit Hilfe der Software ConQuest berechnen (WU, ADAMS, WILSON & HALDANE, 2007). Das MRCML-Modell ist ein generalisiertes IRT-Modell der Rasch-Familie und kann folglich auch das 1-parametrische logistische Modell repräsentieren. Zur Bestimmung der Signifikanz wird ein Likelihood-Ratio-Test auf Basis der Chi-Square-Statistik durchgeführt. Die Fit-Statistik für die Kompetenzdimensionen ist in *Abbildung 7* dargestellt.

	Eindimensionales Modell	Zweidimensionales Modell	Differenz
Anzahl der Iterationen	21	21	0
Deviance (-2Log-Likelihood)	6.143,84	5.853,57	290,27
Anzahl der geschätzten Parameter	29	32	3

Abb. 7: Fit-Statistik der Modellprüfung (vgl. WINTHER, 2010a, S. 184)

Sowohl das eindimensionale als auch das zweidimensionale Modell wurden in 21 Iterationen geschätzt. Die Güte des Modellfits kann relativ zu der Veränderung im

Deviance-Wert verglichen werden. Die Differenz zwischen den Deviance-Werten der beiden Modelle ist als Chi-Square mit drei Freiheitsgraden verteilt. Die Freiheitsgrade ergeben sich aus der Differenz der geschätzten Parameter. Während im eindimensionalen Modell 27 Itemparameter sowie ein Populationsmittelwert und ein Varianzschätzer die Parameteranzahl ergeben, werden im zweidimensionalen Modell neben den Itemparametern zwei Populationsmittelwerte für jede latente Dimension, zwei Varianzschätzer und ein Kovarianzschätzer ermittelt. Die geschätzte Differenz zwischen den Deviance-Werten von 290,27 ist signifikant ($df = 3$; $p < ,001$), so dass gefolgert werden kann, dass das eindimensionale Modell die Daten weniger gut fittet als das zweidimensionale Modell. Zur Beurteilung der Effektstärke wird die Korrelation zwischen den latenten Dimensionen herangezogen: Je niedriger die latenten Dimensionen korrelieren, desto höher ist die praktische Signifikanz eines mehrdimensionalen Modells. Sprachlich-argumentative und mathematisch-analytische Kompetenz korrelieren mit $r = 0,209$. Beide Befunde erlauben die Annahme, dass über die Skalen der „economic literacy“ und der „economic numeracy“ unterschiedliche latente Dimensionen gemessen werden. Für den Bereich der domänenverbundenen Kompetenz lässt sich schlussfolgern, dass sowohl sprachlich-argumentative als auch mathematisch-analytische Kompetenzen separat erfasst werden können und beide Kompetenzbereiche verschiedene kognitive Fähigkeitsstrukturen repräsentieren. Wie sich die domänenverbundenen Kompetenzen jedoch in berufsspezifischen Handlungen zeigen – die für die Beschreibung der Kompetenzentwicklung insbesondere im Ausbildungsverlauf sowie im Rahmen einer beruflichen Beschäftigung relevant werden (vgl. *Abbildung 2*) – ist ein offenes Forschungsdesiderat.

Zur Testfairness des Assessments zur Erfassung domänenverbundener Kompetenz
Mit der Ausrichtung des kaufmännischen Assessments zur Erfassung domänenverbundener Kompetenz an grundlegenden ökonomischen Konzepten und wirtschaftsbezogenen Alltagskontexten kann die Annahme vertreten werden, dass Lernende mit entsprechenden beruflichen Vorerfahrungen Vorteile bei der Bearbeitung des Tests haben, da sie (1) mit den fachlichen Inhalten des Tests in einem höheren Maße vertraut sind und (2) die Sprache sowie die Arbeitsmethoden in der beruflichen Domäne erlernt haben. Eine zentrale Frage in diesem Zusammenhang ist, ob sich systematische Unterschiede zwischen Lernenden mit und ohne kaufmännische Vorbildung, die jedoch über gleiche Fähigkeitsausprägungen in den zwei Kompetenzdimensionen verfügen, bestätigen lassen. Die empirische Prüfung wurde mittels differenzieller Itemanalysen (Differential Item Functioning Analysis (DIF); vgl. HOLLAND & THAYER, 1986; MOORE, 1996) durchgeführt. Für beide Kompetenzbereiche gilt, dass Lernende mit kaufmännischer Vorbildung (Vergleichsgruppe) eine signifikant höhere Leistung erbracht haben als Lernende ohne kaufmännische Vorbildung (Referenzgruppe). Im Bereich der sprachlich-argumentativen Kompetenz beträgt der mittlere Unterschied 0,288 Logits und im Bereich der mathematisch-analytischen Kompetenz 0,284 Logits zugunsten der Vergleichsgruppe ($p < 0,05$). Insbesondere im Hinblick auf die Skala der „economic literacy“ sind die differenziellen Effekte praktisch hoch bedeutsam; der Fähigkeitschätzer ist achtmal so groß wie sein Standardfehler und der p-Wert des Chi-Squares ist kleiner 0,005 (vgl. *Abbildung 8*).

Dieses Ergebnis lässt darauf schließen, dass die Skala der „economic literacy“ Items mit substantiellen DIF enthält. In Anlehnung an PÆK (2002; zit. n. WILSON, 2005, p. 167) sind für die Beschreibung von Logit-Differenzen einzelner Items folgende Effektgrößen anzunehmen: Ein Differenzwert kleiner als 0,426 ist vernachlässigbar,

	Fähigkeitsschätzer	Standardfehler	Chi-Square (df)
Sprachlich-argumentative Kompetenz („economic literacy“)			
Kaufmännische Vorbildung	0,144	0,018	9,63 (1)
Keine kaufmännische Vorbildung	-0,144		
Mathematisch-analytische Kompetenz („economic numeracy“)			
Kaufmännische Vorbildung	0,142	0,070	4,12 (1)
Keine kaufmännische Vorbildung	-0,142		

Abb. 8: Mittlere sprachlich-argumentative und mathematisch-analytische Kompetenzen im Gruppenvergleich (vgl. WINTHER, 2010a, S. 192)

Differenzen zwischen 0,426 und 0,638 sprechen für mittlere und Werte größer als 0,638 für hohe Effekte. Von den 21 Items der „economic literacy“-Skala weisen neun Items (42,86%) mittlere oder hohe DIF-Effekte auf. In *Abbildung 9* sind aus Sicht der Vergleichsgruppe die Interaktionen zwischen den Items und der Gruppenzugehörigkeit dargestellt.

Item	Schätzer	Standardfehler	Chi-Square (df)	p-Wert
Item 01	0,288	0,116	103,28 (20)	0,000
Item 05	0,311	0,110		
Item 07	0,388	0,146		
Item 08	0,452	0,174		
Item 13	0,466	0,163		
Item 15	0,783	0,192		
Item 16	-0,774	0,247		
Item 17	-1,472	0,247		
Item 21	-0,254	0,775		

Abb. 9: DIF-Analyse der Skala „economic literacy“ (Item*Vorbildung; vgl. WINTHER, 2010a, S. 193)

Ein positiver Interaktionseffekt bedeutet, dass sich die Itemschwierigkeit für die Vergleichsgruppe um den Schätzwert der Interaktion erhöht; ein negativer Schätzwert reduziert hingegen die Itemschwierigkeit aus Perspektive der Lernenden mit kaufmännischer Vorbildung. Die Abbildung zeigt zusammenfassend, dass sechs der neun DIF-Items Lernende mit kaufmännischer Vorbildung systematisch benachteiligen (positiver Schätzwert der Interaktion), während bei drei Items die Itemschwierigkeit für die Referenzgruppe erhöht ist (negativer Schätzwert der Interaktion). Von entscheidender Bedeutung ist, dass sich die differenziellen Unterschiede im Bereich der sprachlich-argumentativen Kompetenz auf das unterschiedliche Antwortverhalten in den Subgruppen zurückführen lassen: Während Items, die einen stärkeren curricularen Bezug aufweisen, von Lernenden mit entsprechender Vorbildung einfacher gelöst werden können, erweisen sich Testaufgaben, die auf allgemeine betriebswirtschaftliche Konzepte in Alltagskontexten zurückgreifen, für diese Subgruppe im Vergleich zu Lernenden ohne kaufmännische Vorbildung als

anspruchsvoller. Im Hinblick auf die Skala „economic literacy“ lassen sich die DIF-Effekte inhaltlich durch die jeweilige Distanz zu Alltags- respektive curricularen Kontexten interpretieren. So steht das Item 01 (vgl. *Abbildung 5*; mittlerer DIF-Effekt, Logit-Differenz = 0,576) exemplarisch dafür, dass in Anlehnung an das Konzept der „economic literacy“ insbesondere text- und bildsprachliche kaufmännische Aufgaben mit konkreter Alltagsverknüpfung bearbeitet werden. Zur Beantwortung der Frage, welche Gründe für den Gewinneinbruch bei der Volkswagen AG im Zeitungsartikel genannt werden, genügt ein sorgfältiges, verständnisgeleitetes Lesen, ohne auf spezifische kaufmännische Kenntnisse zurückgreifen zu müssen. Lernende mit kaufmännischer Vorbildung – so die Vermutung – greifen für das Herausfiltern von Gründen und Zusammenhängen des Gewinneinbruchs jedoch verstärkt auf das von ihnen erlernte betriebswirtschaftliche Konzept zurück und nutzen dieses zur Beantwortung der Aufgabenstellung. Hierbei bleiben sie unvollständig und erreichen im Vergleich zwischen den Subgruppen eine geringere Antwortqualität. Item 16 (vgl. *Abbildung 5*, hoher DIF-Effekt, Logit-Differenz = 1,548) ist hingegen ein Beispiel für ein stärker am Curriculum orientiertes Item. Die Lernenden sollen diskutieren, warum eine fristgerechte Lieferung für das produzierende Unternehmen von entscheidender Bedeutung ist. Das Item erfasst das konzeptuale Verständnis der Lernenden im Hinblick auf die Innen- und Außenbeziehungen eines Unternehmens im Rahmen der Auftragsabwicklung. Der Lerninhalt wird im kaufmännischen Eingangunterricht intensiv vermittelt, so dass bei dieser Aufgabenstellung Lernende mit kaufmännischer Vorbildung systematische Vorteile haben. Identische Befunde lassen sich für die Skala „economic numeracy“ feststellen. Items, die allgemeine ökonomische Prozeduren (im Sinne basaler mathematischer Verfahren) bezogen auf Alltagskontexte zum Inhalt haben, sind für Lernende mit kaufmännischer Vorbildung schwieriger als für Lernende ohne kaufmännische Vorbildung – das gilt vice versa für Items, die zunehmend curriculare Bezüge aufweisen.

Diese Befunde sind aus zumindest zwei Perspektiven bedeutsam: Zum einen verdeutlichen sie, dass domänenverbundene Kenntnisse als wesentliche Voraussetzung für den Erwerb domänenspezifischer Kompetenzen und deren Anwendung in beruflichen Kontexten zu sehen sind. Zum anderen wird deutlich, dass es mit zunehmendem Fachbezug in der Anforderungssituation und mit zunehmenden domänenspezifischen Fähigkeiten zu einer Überlagerung allgemeiner Konzepte kommen kann, selbst wenn mit diesen eine effizientere Lösung für die Anforderungssituation gefunden werden könnte. Oder anders formuliert: Allgemeine domänenverbundene Fähigkeiten werden blockiert, da die domänenspezifischen Fähigkeiten nicht in ausreichendem Maße internalisiert sind. Für die Konstruktion von Testumgebungen in beruflichen Kontexten sind folglich Anforderungssituationen zu modellieren, die den unterschiedlichen Subgruppen gerecht werden. Vor allem im Hinblick auf die unterschiedlichen Bildungsbiographien und beruflichen Vorerfahrungen stellt sich für Assessments im beruflichen Bereich die Herausforderung, Tests zu konstruieren, die für jeden Probanden zu validen Ergebnissen führen.

3.2 Kaufmännische Kompetenz im Übergang von der Ausbildung in eine Beschäftigung

Das Modell kaufmännischer Kompetenzentwicklung (vgl. *Abbildung 2*) geht davon aus, dass in einer zweiten berufsbezogenen Übergangsphase insbesondere domänenspezifische Kompetenzen für die weitere Kompetenzentwicklung von Bedeutung sind. Domänenspezifische Kompetenz wird als sinnhaftes Handeln in beruflichen Situationen definiert, das sich im Rahmen beruflicher Beschäftigung sowohl in tätigkeitsspezifischen als auch konkret arbeitsplatzbezogenen Anforderungssituationen zeigt. Insbesondere die arbeitsplatzbezogene Dimension wird unterschiedlichen beruflichen Positionen und Karriereverläufen im Sinne professioneller Kompetenzentwicklung gerecht, während die tätigkeitsspezifische Dimension die Inhaltsbereiche der Domäne bzw. eines spezifischen Berufsfeldes abbildet. Im Rahmen der VET-LSA-Studie wurde die tätigkeitsspezifische Dimension als domänenspezifische Kompetenz über zwei kognitive Teilkompetenzen beschrieben (vgl. *Abbildung 10*): verstehensbasierte Kompetenz als das sinnhafte Erschließen beruflicher Anforderungssituationen (erfasst über die Skala „Prozessverstehen“) und handlungsbasierte Kompetenz im Sinne eines sinnhaften Handelns in solchen Anforderungssituationen (erfasst über die Skala „Prozesssimulation“). Diese zweidimensionale Kompetenzstruktur wurde empirisch validiert (Deviance-Differenz = 101,91; $df = 3$; $p < ,001$; vgl. hierzu WINTHER & ACHTENHAGEN, 2010; WINTHER, 2010a).

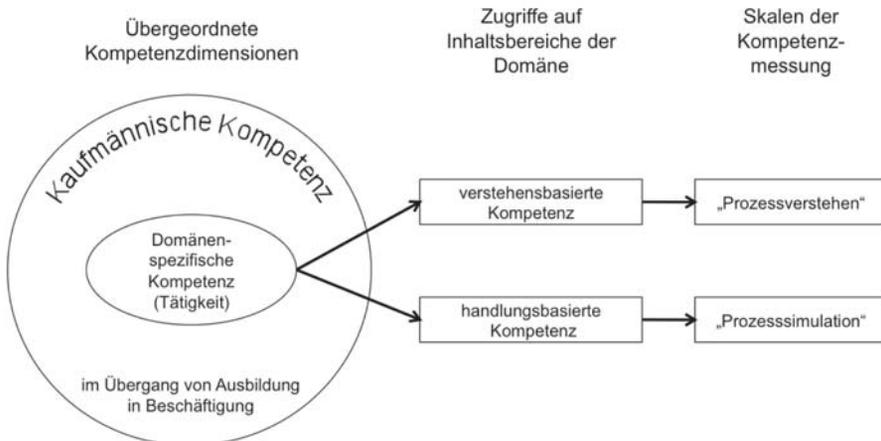


Abb. 10: Theoretisches Kompetenzstrukturmodell domänenspezifischer Kompetenz (EIGENE DARSTELLUNG)

Die zwei Skalen zur Erfassung domänenspezifischer Kompetenz wurden auch im Hinblick auf DIF-Effekte analysiert. Neben persönlichen Eigenschaften wurden insbesondere Merkmale des Ausbildungsbetriebes berücksichtigt, um Aussagen darüber treffen zu können, ob mit der Wahl des Ausbildungsplatzes bereits eine Vorselektion im Hinblick auf die erfolgreiche Bewältigung tätigkeitsspezifischer Anforderungssituationen erfolgt. Als individuelle Bedingungen wurden das Ge-

schlecht, der Bildungsabschluss sowie die schulischen Leistungen in den Fächern Deutsch und Englisch erhoben; als Merkmale des Ausbildungsbetriebes wurden Dummy-Variablen der Betriebsgröße sowie der Branche erzeugt. Beide Skalen domänenspezifischer Kompetenz weisen keine Items mit substantiellem DIF-Effekt auf – der Test ist im Hinblick auf die erhobenen individuellen und institutionellen Bedingungen fair. Eine zentrale Begründung dafür ist, dass der Test konsequent auf der Basis authentischer Geschäftsprozesse konstruiert wurde und dadurch für alle Lernenden eine plausible berufstypische Simulation betrieblicher Realität zur Verfügung stand. In der Folge – so die Annahme – bestehen für Lernende mit gleicher Fähigkeitsausprägung im Vergleich der differenziell zu unterscheidenden Gruppen keine Unterschiede im Antwortverhalten. Ein an betrieblichen Geschäftsprozessen ausgerichtetes Assessment unterdrückt jedoch nicht die generellen differenziellen Effekte zwischen einzelnen Subgruppen. Die Analyse genereller differenzieller Effekte klärt, inwieweit individuelle und institutionelle Faktoren die Leistungsfähigkeit prognostizieren können. Im Rahmen der VET-LSA-Studie erweist sich aus dem Pool der erhobenen Individual- und Institutionenmerkmale bei Annahme einer fünfprozentigen Irrtumswahrscheinlichkeit nur die allgemeine Hochschulreife als Prädiktor der Personenfähigkeit (vgl. *Abbildung 11*).

Allgemeine Hochschulreife (unabhängige Variable; n = 264)	β	p	F	R ²
abhängige Variable: Personenparameter (WLE) der handlungsbasierten Kompetenz	0,301	0,025	5,048 (0,002)	0,055
abhängige Variable: Personenparameter (WLE) der verstehensbasierten Kompetenz	0,309	0,023	2,319 (0,046)	0,026

Abb.11: Prädiktoren der Personenfähigkeit (WINTHER, 2010a, S. 244)

Es zeigen sich darüber hinaus Tendenzen, dass Auszubildende in Großbetrieben (über 500 Mitarbeitende) über bessere Leistungen sowohl im handlungsbasierten ($\beta = 0,132$; $p = 0,055$) als auch im verstehensbasierten Kompetenzbereich ($\beta = 0,131$; $p = 0,090$) verfügen. Die praktischen Signifikanzen aller ermittelten Effekte sind für die vorliegende Stichprobe ($n = 264$) jedoch sehr gering.

4 Zusammenfassung

Die vorgestellten Ergebnisse lassen sich für die Analyse und Interpretation (aber auch die Konstruktion) weiterer Assessments in der kaufmännischen Domäne nutzen. Die Möglichkeiten einer anwendungsorientierten Nutzung können darauf zurückgeführt werden, dass in den theoretisch begründeten und empirisch validierten Kompetenzmodellen Annahmen über den Zusammenhang von zu erfassender Kompetenz und situationalem Verhalten in spezifischen Entwicklungsphasen des beruflichen Kompetenzaufbaus und -erwerbs formuliert werden. Die Kompetenzmodelle zur Beschreibung der Übergangsprozesse erfüllen dabei nicht den Anspruch nach Generalisierbarkeit, wohl aber werden sie der Forderung nach hoher Adaptivität gerecht. Grundlage der Modellierungen ist eine kognitive Kompetenzauffassung,

die im Kern davon ausgeht, dass Kompetenzen als Potentiale zu interpretieren sind und sich damit weder über einzelne personenbezogene Leistungsmerkmale noch auf der Basis isolierter Leistungssituationen erfassen lassen. Die Abbildung beruflicher Kompetenz in den verschiedenen Phasen der Kompetenzentwicklung erfordert hingegen Anforderungssituationen, die sowohl Arbeitsabläufe im Sinne beobachtbarer beruflicher Handlungen als auch wissensbasierte Entscheidungsprozesse als Grundlage der Handlungen beinhalten. Um Aussagen darüber treffen zu können, wie Auszubildende/Lernende in komplexen beruflichen Situationen agieren, sind zudem Vorstellungen davon zu entwickeln, welche personalen und institutionalen Bedingungen ein erfolgreiches Bearbeiten beruflicher Anforderungssituationen stützen oder aber beeinträchtigen. Hierbei ist zwischen generellen differenziellen Effekten und Aspekten der Item- bzw. Testfairness zu unterscheiden: Während ein genereller Vergleich verschiedener Personengruppen Hinweise auf die prognostischen Möglichkeiten eines Assessments geben kann, werden über die differenziellen Unterschiede in den Lösungswahrscheinlichkeiten bei gleicher Fähigkeitsausprägung systematische Verzerrungen im Hinblick auf einzelne Items und in der Folge auf den Gesamttest aufgedeckt.

Für die berufliche Bildung stellen insbesondere die über DIF-Effekte charakterisierten systematischen Bevorzugungen bzw. Benachteiligungen einzelner Subgruppen eine der zentralen Herausforderungen für die Konstruktion kompetenzorientierter Assessments dar. Bedingt durch die ausgeprägten Heterogenitäten sowohl hinsichtlich persönlicher Eigenschaften, Sozialisationen und Bildungskarrieren als auch bezogen auf die deutlichen Unterschiede zwischen den auszubildenden Organisationen, werden Testentwickler vor die Schwierigkeit gestellt, einen für alle Personen gültigen Test zu entwickeln. Hierbei kann kaum auf evidenzbasierte Forschungsergebnisse zurückgegriffen werden, da bislang Untersuchungen fehlen, in denen die personalen Ressourcen und internalen Bedingungen sowie die institutionalen Einflussfaktoren für ein erfolgreiches berufliches Handeln und Verstehen in konkreten berufstypischen Anforderungssituationen analysiert wurden.

Die hier vorgestellten Assessments im Bereich der kaufmännischen Bildung weisen eine zentrale Einschränkung auf: Die Ergebnisse der Assessments zur domänenverbundenen und domänenspezifischen Kompetenzmessung lassen sich nicht aufeinander beziehen. So können zwar charakteristische Übergangsprozesse der kaufmännischen Kompetenzentwicklung auf inter-individueller Ebene beschrieben werden, intra-individuelle Entwicklungen bleiben jedoch unberücksichtigt. Das Modell der kaufmännischen Kompetenzentwicklung geht davon aus, dass über domänenverbundene Literalität der Auf- und Ausbau domänenspezifischer kognitiver Strukturen und Prozesse gestützt wird. Damit wird eine direkte Beziehung zwischen domänenverbundenen und domänenspezifischen Kompetenzen angenommen. Oder anders formuliert: In berufsspezifischen Handlungen wird immer auch auf zuvor erworbene allgemeine Kenntnisse und Fertigkeiten zurückgegriffen. Die Frage allerdings, wie und in welchem Umfang allgemeine und domänenverbundene Literalität den Auf- und Ausbau domänenspezifischer Kompetenzen beeinflusst, ist bislang unbeantwortet.

Literatur

- Anderson, J. R. (1996). ACT. A Simple Theory of Complex Cognition. *American Psychologist*, 51 (4), 355–365.
- Bader, R. & Müller, M. (2002): Leitziel der Berufsbildung: Handlungskompetenz: Anregungen zur Ausdifferenzierung des Begriffs. *Die berufsbildende Schule*, 54 (6), 176–182.
- Adams, R. J., Wilson, M. & Wang, W.-C. (1997). The Multidimensional Random Coefficients Multinomial Logit Model. *Applied Psychological Measurement*, 21 (1), 1–23.
- Egger-Subotitsch, A. & Kreiml, T. (2009). Kompetent für's Management – Das Leonardo-Projekt »Strategical Individual Competencies«. In A. Egger-Subotitsch & R. Sturm (Hrsg.), *Kompetenzen im Brennpunkt von Arbeitsmarkt und Bildung* (S. 38–49). Wien: Arbeitsmarktservice Österreich.
- Embretson, S. E. & Reise, S. P. (2001). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Evans, L. (2002). What is Teacher Development? *Oxford Review of Education*, 28 (1), 123–137.
- 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: Lawrence 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.
- Helmke, A. (1988). Leistungssteigerung und Leistungsausgleich von Leistungsunterschieden in Schulklassen: Unvereinbare Ziele? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 20, 45–76.
- Holland, P. W. & Thayer, D. T. (1986). *Differential Item Performance and the Mantel-Haenszel Procedure* (Research Report 86-31). Princeton, NJ: Educational Testing Service.
- Hoyle, E. (2001). Teaching: Prestige, Status and Esteem. *Educational Management & Administration*, 29 (2), 139–152.
- IPN (Leibniz-Institut für die Pädagogik der Naturwissenschaften) (Hrsg.) (2003). *Ziele und Perspektiven naturwissenschaftlicher Bildung*. Kiel: IPN.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. J. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Frankfurt am Main: DIPF.
- Moore, S. (1996). Estimating DIF with the RCML Model. In G. Engelhard & M. Wilson (Eds.), *Objective Measurement III: Theory into Practice* (pp. 219–238). Norwood, NJ: Ablex.
- Norris, N. (1991). The Trouble with Competence. *Cambridge Journal of Education*, 21 (3), 330–341.
- Rauner, F., Haasler, B., Heinemann, L., Maurer, A. & Piening, D. (Hrsg.) (2009). *Messen beruflicher Kompetenzen. Bd. 2: Die Ergebnisse KOMET 2008*. Münster: Lit-Verlag.
- Reetz, L. (1999). Zum Zusammenhang von Schlüsselqualifikationen – Kompetenzen – Bildung. In T. Tramm, D. Sembill, F. Klauser & E. G. John (Hrsg.), *Professionalisierung kaufmännischer Berufsbildung* (S. 32–51). Frankfurt: Peter Lang.
- Roth, H. (1971). *Pädagogische Anthropologie. Bd. 2: Entwicklung und Erziehung. Grundlagen einer Entwicklungspädagogik*. Hannover: Schroedel.
- Ryan, K. E. & Shepard, L. A. (Eds.) (2008). *The Future of Test-Based Educational Accountability*. New York, NY: Routledge.
- Schecker, H. & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 45–66.
- Schelten, A. (2004). Schlüsselqualifikationen. *Wirtschaft und Berufserziehung. Zeitschrift für Berufsbildung*, 56 (4), 11–13.
- Seeber, S. (2008). Ansätze zur Modellierung beruflicher Fachkompetenz in kaufmännischen Ausbildungsberufen. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 104, 74–97.

- Seeber, S., Nickolaus, R., Winther, E., Achtenhagen, F., Breuer, K. et al. (2010). Kompetenzdiagnostik in der Berufsbildung Begründung und Ausgestaltung eines Forschungsprogramms. *Berufsbildung in Wissenschaft und Praxis, Beilage zu 1/2010*, 1–15.
- Van de Rijt, B., Godfrey, R., Aubrey, C., van Luit, J. E. H., Ghesquiere, P., Hasemann, K., Tancing, S., Kavkler, M. & Tzouriado, M. (2003). The Development of Early Numeracy in Europe. *Journal of Early Childhood Research*, 1 (2), 155–180.
- Van de Rijt, B., van Luit, J. E. H. & Pennings, A. H. (1999). The Construction of the Utrecht Early Mathematical Competence Scales. *Educational and Psychological Measurement*, 59, 289–309.
- Weber, S. (1994). *Vorwissen in der betriebswirtschaftlichen Ausbildung. Eine struktur- und inhaltsanalytische Studie*. Wiesbaden: Gabler.
- Wilson, M. (2004). On Choosing a Model for Measuring. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch-Measurement* (pp. 11–42). Maple Grove, MN: JAM Press.
- Wilson, M. (2005). *Constructing Measures: An Item-response Modeling Approach*. Mahwah, NJ: Lawrence Erlbaum.
- Winther, E. (2010a). *Kompetenzmessung in der beruflichen Bildung*. Bielefeld: wbv.
- Winther, E. (2010b). Systemorientierte Evaluation des Lernfeldkonzepts: Strategien für das berufliche Übergangssystem. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 106 (3), 427–442.
- Winther, E. & Achtenhagen, F. (2009). Skalen und Stufen kaufmännischer Kompetenz. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 105 (4), 521–556.
- Winther, E. & Achtenhagen, F. (2010). Berufsfachliche Kompetenz: Messinstrumente und empirische Befunde zur Mehrdimensionalität beruflicher Handlungskompetenz. *Berufsbildung in Wissenschaft und Praxis, 1/2010*, 18–21.
- Wright, B. D. & Stone, M. H. (1979). *Best Test Design*. Chicago, IL: MESA Press.
- 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 Autorin: Prof. Dr. Esther Winther, Professur für Wirtschaftspädagogik, Universität Paderborn, Warburgerstr. 100, 33098 Paderborn (esther.winther@wiwi.upb.de)