

Überlegungen und Analysen zur Spezifität kaufmännischer Kompetenz

Tade Tramm und Susan Seeber

1 Problemstellung

Kompetenzen sind wissensbasiert und domänenspezifisch – hiervon werden auch alle Ansätze auszugehen haben, die sich um eine Definition kompetenzorientierter Standards in der beruflichen Bildung bemühen. Zugleich jedoch gibt es gute Argumente dafür, etwa im kaufmännischen Bereich berufsübergreifende Grundlagen von Kompetenzen zu vermuten, die sich jenseits aller beruflichen Spezifizierungen auf einen gemeinsamen Kernbereich kaufmännischen Denkens und Handelns beziehen. Ein solcher Kernbereich wäre an der globalen Erklärung individueller Leistungsunterschiede orientiert, böte aber nur wenige Bezugspunkte für curriculare Konstruktionen. Zu fragen wäre in diesem Zusammenhang deshalb, ob kaufmännische Kompetenz als homogener Fähigkeitskomplex zu denken ist oder ob sie nicht angemessener in einem mehrdimensionalen Bezugsrahmen zu konstruieren ist.

Der nachfolgende Beitrag konkretisiert diese Überlegungen und berichtet über erste Versuche, vorliegendes Datenmaterial aus der Hamburger Studie „Untersuchung von Leistungen, Motivation und Einstellungen“ (ULME) von Schülerinnen und Schülern in beruflichen Bildungsgängen zu Effekten dualer Ausbildung unter diesem Aspekt zu analysieren.

2 Ökonomische Bildung und kaufmännische Kompetenz – Probleme der Konzeptualisierung

Während die Entwicklung von Bildungsstandards im allgemein bildenden Bereich dynamisch voranschreitet (vgl. Klieme, Avenarius et al. 2003), schien sich die Berufliche Bildung lange Zeit darin zu genügen, dass mit dem System der Kammerprüfungen bereits Elemente einer outputorientierten Steuerung greifen. In einer Stellungnahme zum Workshop des BIBB zu einem „Berufsbildungs-PISA“ im Jahre 2003 beklagte dagegen Straka (2003, S. 8), dass über die Messgüte dieser Diagnoseinstrumente öffentlich kaum etwas bekannt sei und er kritisierte, dass sie „in ihrer Qualität noch immer mehr Gemeinsamkeiten mit den mittelalterlichen Lossprechungsritualen des Handwerks als mit den heute möglichen Methoden der Diagnostik“ hätten. Der einzige „allgemein zugängliche objektive Test zur Erfassung beruflicher Fachkundigkeit“ sei noch immer der von Klaus Beck und Volker Krumm in Zusammenarbeit mit Rolf Dubs aus dem

Amerikanischen adaptierte (Soper & Walstad 1987) „Wirtschaftskundliche Bildungs-Test“ (1998).

Mit diesem Test verbindet sich die Absicht, den Begriff der ökonomischen Bildung operational im Zuge einer von Beck als subsumtive Strategie bezeichneten Vorgehensweise zu fassen. Ausgangspunkt ist dabei nicht die Operationalisierung eines vorgängig hermeneutisch bestimmten Bildungsbegriffs, sondern die sukzessive Eingrenzung eines solchen Konstrukts durch ihm subsumierbare Items (Beck 1989, S. 580ff.). Beck sympathisiert in seinem Konzept mit einem operationalen Begriffsverständnis, wonach ein fraglicher Begriff durch ihm zugeordnete Testoperatoren definiert wird. Er weist freilich zugleich darauf hin, dass das Credo, „Ökonomische Bildung ist, was der Test für ökonomische Bildung misst“, unschuldiger daherkomme als es wirklich sei: Derjenige der einen solchen Test konstruiert habe, müsse doch eine zumindest implizite und vielleicht noch vage Vorstellung von diesem Konzept haben. „Wie hätte er sonst einen aufgabenhomogenen Test entwickeln können?“ (ebenda, S. 584).

In diesem Sinne verankert Beck seinen Test zu ökonomischem Wissen und Denken in einem dreidimensionalen Konzept ökonomischer Bildung: Von Bildung solle erst dann gesprochen werden, wenn Wissen und geistige Leistungsfähigkeit als Grundlage verantwortlichen Handelns in zweierlei Weise kontrolliert werde. Einerseits nämlich „durch einen Kanon von positiven und negativen Einstellungen gegenüber gegebenen Sachverhalten“ und andererseits durch die Bindung an moralische Prinzipien. Das Niveau ökonomischen Denkens und Wissens korreliere empirisch positiv mit der Entwicklung einer positiven Einstellung zu ökonomischen Fragestellungen (ebenda, S. 588), was für den Zusammenhang der kognitiven Dimension mit der moralischen Reflexionsfähigkeit allerdings nicht festzustellen sei: Aus diesem Grund erweise sich „eine ökonomiespezifische Konzeptualisierung und Erfassung der moralischen Orientierung als sinnvoll und erforderlich“ (ebd. 590).

Mit diesen systematischen Relationierungen gelingt es Beck eine Argumentationsfigur zu entfalten, aus der heraus einem auf die kognitive Dimension konzentrierten, also auf ökonomisches Wissen und Denken gerichteten Test eine hohe Validität auch im Hinblick auf ein mehrdimensional konzipiertes Konzept ökonomischer Bildung zugemessen werden kann. Indem er den Bildungsbegriff dabei auf die Befähigung zu verantwortlichem Handeln bezog, schlug er zudem die Brücke zum Kompetenzbegriff in seiner heute verbreiteten Verwendung (vgl. Weinert 2001, 27f.).

Während es Beck mit dieser Argumentation durchaus überzeugend gelingt, die zentrale Stellung ökonomischen Wissens und Denkens für ein auf verantwortliches Handeln gerichtetes Bildungskonzept zu begründen (und zugleich Desiderate zu benennen, die er in seiner weiteren Forschungsarbeit konsequent angegangen ist),

bleibt er eine vergleichbar plausible Begründung im Hinblick auf die inhaltliche Eingrenzung seines Definitionsbereiches schuldig. Der WBT beschränkt sich inhaltlich auf gesamtwirtschaftliche Aspekte; seine 22 Themenbereiche beziehen sich auf die Komplexe „Grundlagen“ des Wirtschaftens (12 Items, Version A), „Mikroökonomie“ (13), „Makroökonomie“ (13), „Internationale Beziehungen“ (8).

Hier stellt sich die Frage, inwieweit es ein solch enger thematischer Horizont erlaubt, im umfassenden Sinne von ökonomischer Bildung zu sprechen oder gar, wie Straka dies tut, auf berufliche Fachkundigkeit (im kaufmännischen Bereich?) zu schließen. Müsste beides nicht in viel stärkerem Maße auch auf einzelwirtschaftliches Wissen und Denken abheben, worunter hier sowohl Aspekte der Betriebswirtschaftslehre als auch des privaten Haushaltens, des Handelns als Konsument und als Vermarkter der eigenen Arbeitsleistung verstanden sein sollen?

Systematisch ist damit die Frage nach der Bereichs- bzw. Domänenspezifität von Kompetenzen aufgeworfen, die sich hier in doppelter Weise stellt, als einerseits die thematische Kohärenz eines Kompetenzbereichs durch die Homogenität der Items zu sichern ist, als aber andererseits auch der Komplexität und Mehrdimensionalität dieses Kompetenzbereiches entsprochen werden muss. Dieses zweite Postulat stellt sich dann als besonders herausfordernd dar, wenn die Domänenspezifität einer Kompetenz nicht durch den Bezug auf ein abgegrenztes (Unterrichts-)Fach begründet ist, sondern durch den Bezug auf ein konkretes, mehrdimensionales Handlungs- und Problemfeld, wie dies im Kontext berufsbezogener Kompetenzen insbesondere unter dem Signum lernfeldstrukturierter Curricula der Fall ist.

An dieser Stelle wird eine höchst interessante Divergenz zwischen dem Kompetenzbegriff der allgemeinen Diskussion um Bildungsstandards und dem Konzept der Handlungskompetenz im berufsbildenden Bereich deutlich. Pointiert gesagt fokussiert die allgemeine Standarddiskussion auf kognitive Leistungsfähigkeit in Bezug auf eingeführte Schulfächer (vgl. dazu auch Klieme, Avenarius et al. 2003, 25). Der Gegenstandsbereich oder Inhalt dieser Fächer ist curricular definiert, die spezifische Herausforderung für die Testentwicklung besteht darin, diese Lerninhalte in Aufgaben- oder Problemstellungen einzubetten, über die sie der empirischen Erfassung zugänglich sind. Bindeglied zwischen den Inhalten und den Aufgaben müssen dabei Annahmen über kognitive Leistungen sein, die die kompetenzbestimmenden Inhalte quasi prozeduralisieren und die mit den Testaufgaben abgeprüft werden. Kurz gesagt weist der Entwicklungsweg vom mehr oder weniger eindeutig definierten Inhalt über die damit zu verknüpfende kognitive Leistung zur situierten Aufgabe (vgl. Abb. 1).

Im berufsbildenden Bereich könnte nur unter der Voraussetzung analog zur Diskussion um die allgemeinen Bildungsstandards argumentiert werden, dass der Bezugspunkt eine im Wesentlichen materiell definierte ökonomische oder techni-

sche Bildung wäre. Sobald jedoch der Anspruch des Lernfeldansatzes ins Spiel kommt, verkehrt sich der oben skizzierte Argumentationsgang ins Gegenteil. Im Mittelpunkt des Lernfeldansatzes steht nämlich die Idee, Kompetenzen für bestimmte, primär berufliche Handlungs- oder Orientierungsleistungen zu entwickeln, die aus ihrem Begründungszusammenhang heraus bereits situiert sind. Zu klären ist also im Wesentlichen, bis zu welchem Komplexitätsniveau diese Leistungen entwickelt werden sollen, und insbesondere, welche Wissensbasis und welche Denkprozesse diesen Leistungen zugrunde liegen. Der Klärungsweg führt damit also von der (beruflichen) Situation über die geforderte Leistung hin zum Wissen, das in dieser Leistung wirksam wird (vgl. hierzu Tramm 2003; Straka 2005).

<u>Kompetenzen in der allgemeinen Standard-</u> <u>diskussion</u>	<u>Kompetenzen in der Berufs- und Wirt-</u> <u>schaftspädagogik</u>
<ul style="list-style-type: none"> • Fokus auf kognitive Leistungsfähigkeit • Ausgangspunkt: Schulfächer • Problem: Situierung von Lerninhalten <p>➤ <i>Vom Inhalt über die Leistung zur Situierung</i></p>	<ul style="list-style-type: none"> • Fokus auf Handlungs- und Orientierungsfähigkeit • Ausgangspunkt: berufliche Handlungsfelder und Problemsituationen • Problem: Identifikation von kognitiven Leistungen und relevanter Wissensbasis <p>➤ <i>Von der Situation über die Leistung zum Inhalt</i></p>

Abb. 1: Problemkontexte der Kompetenzdefinition

In erster Näherung kann dieser doppelte Bezug berufsbezogener Kompetenzen folgendermaßen illustriert werden:

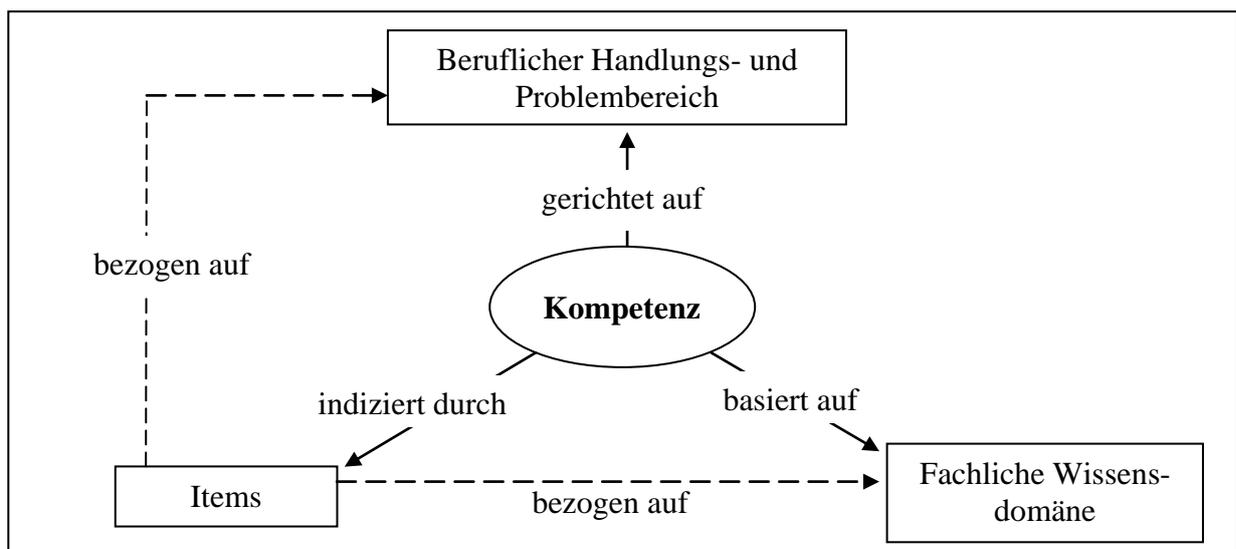


Abb. 2: Referenzsystem der Kompetenzbestimmung im beruflichen Bereich

Berufliche Kompetenzen in diesem Sinne sind auf die Bewältigung beruflicher Problem- und Handlungsanforderungen gerichtet und basieren auf domänenspezifischem „intelligenten Wissen“ (vgl. Klieme, Avenarius et al. 2003, 22).

Die Beantwortung der Frage nach der Struktur solcher Kompetenzen und der ihnen korrespondierenden Wissenssysteme ist Gegenstand domänenspezifischer (oder hier besser: berufsspezifischer) Kompetenzmodelle, mit denen einerseits **funktionale Teildimensionen** innerhalb einer Domäne (eines Berufes) unterschieden werden (z. B. Rezeption und Produktion von Texten) und die es andererseits erlauben, **unterschiedliche Niveaustufen** auf solchen Dimensionen zu unterscheiden (Klieme, Avenarius et al. 2003, S. 22). Die Erarbeitung derartiger Modelle muss nach verbreiteter Einschätzung in Kooperation von fachdidaktischer und empirischer Bildungsforschung in Form eines iterativen Zusammenspiels theoretisch-systematischer Ex-Ante-Überlegungen und empirischer Ex-Post-Analysen erfolgen (Helmke & Hosenfeld 2003, 4f.). Quasi im Schnittpunkt beider Arbeitsrichtungen liegt die Entwicklung von Aufgabensätzen, die subjektbezogenen Aufschluss über das individuell erreichte Kompetenzniveau geben sollen und deren empirisch ermittelte unterschiedliche Schwierigkeitsgrade bzw. Lösungswahrscheinlichkeiten zugleich Hinweise auf die funktional und/oder inhaltlich bedingte hierarchische Stufung dieser Kompetenzen geben können.

Im Folgenden soll mit Blick auf die eingangs erwähnte Hamburger Schulleistungsuntersuchung der Frage nachgegangen werden, ob und inwieweit in einem auf die Anforderungen eines kaufmännischen Curriculums bezogenen kompetenzorientierten Test inhaltlich unterschiedlich begründeten Kompetenzdimensionen erkennbar sind, oder ob die Annahme eines Generalfaktors „ökonomische bzw. kaufmännische Kompetenz“ besser mit der Datenstruktur vereinbar ist.

3 Das Hamburger Projekt ULME

3.1 Der bildungspolitisch-pragmatische Kontext des Projekts

Die Durchführung groß angelegter Lernstands- und Lernentwicklungsuntersuchungen hat in Hamburg Tradition (vgl. Schwippert 2005). Mit dem von der Hamburger Schulbehörde durchgeführten Projekt ULME ("Untersuchung von Leistungen, Motivation und Einstellungen in der beruflichen Bildung") wurde dieser Ansatz erstmals auch auf den berufsbildenden Bereich und damit auf den Erwerb beruflicher Kompetenzen ausgedehnt. Hierbei wurde die Jahrgangskohorte, die erstmals im Schuljahr 1996/97 im Rahmen des Projekts LAU in der fünften Jahrgangsstufe systematisch und danach im Zweijahresrhythmus weiter evaluiert wurde, in der entsprechenden Teilpopulation mindestens zum Teil auch während ihrer schulischen Laufbahn im Berufsbildungssystem weiter verfolgt. In drei Untersuchungsrounden wurden Schülerinnen und Schüler beim Übergang in eine berufliche Schulform (ULME I, vgl. Lehmann, Ivanov et al. 2005), kurz

vor Abschluss so genannter teilqualifizierender Berufsfachschulen (ULME II, vgl. Lehmann, Seeber & Hunger 2006) und schließlich kurz vor Abschluss einer regulär dreijährigen Berufsausbildung im Dualen System (ULME III) untersucht. Neben Fachleistungstests in Deutsch, Mathematik und Englisch und weiteren Tests zu Interessen, Lernstrategien und allgemeinen Fähigkeiten zum schlussfolgernden Denken wurden mit ULME II erstmals und dann spezifischer noch mit ULME III auch berufsbereichs- bzw. berufsbezogene Tests eingesetzt. Deren Entwicklung erfolgte unter Beratung und Mitwirkung eines Teams des Instituts für Berufs- und Wirtschaftspädagogik (IBW) der Universität Hamburg; die wissenschaftliche Leitung des Gesamtprojekts lag bei Prof. Dr. Rainer Lehmann von der Humboldt Universität zu Berlin.

3.2 Pragmatischer Kontext der Aufgabenentwicklung

Die Konstruktion berufsspezifischer Aufgabensätze musste für insgesamt 17 duale Ausbildungsberufe unter erheblichem Zeitdruck in einem Zeitraum von knapp zwei Monaten bis zur ersten Pilotierung geleistet werden (vgl. Brand, Hofmeister & Tramm 2005). Die jeweiligen Tests sollten in standardisierter Form, im Zeitrahmen von 90 Minuten und unter Verwendung weitestgehend geschlossener Antwortformate durchgeführt und von fachfremden Personen ausgewertet werden können.

Für die Konstruktion der Tests ergaben sich mit Blick auf die systematischen Ansprüche und unter Berücksichtigung der Rahmenbedingungen die folgenden Konsequenzen:

- Der Test musste sich auf die kognitive Dimension der Kompetenz konzentrieren, d. h. auf die kognitiven Fähigkeiten, Fertigkeiten und Kenntnisse zur Lösung beruflich relevanter Aufgaben und Probleme.
- Der Test sollte berufsspezifisch angelegt sein und eindeutig spezifische berufliche Kompetenzen und Wissensbestände zum Gegenstand haben.
- Der Test musste aus einer größeren Zahl analytisch unabhängiger Einzelaufgaben bestehen; komplexe mehrstufige Aufgabenstellungen konnten nicht integriert werden.
- In den Aufgaben sollte Wissen nicht nur abgefragt, sondern vor allem in berufsbezogenen Leistungszusammenhängen abgefordert werden.
- Die Aufgaben sollten das Spektrum der beruflich relevanten Leistungssituationen und der einschlägigen Wissensbereiche abdecken.
- Die Aufgaben sollten im Anspruchsniveau unterschiedlich sein; die Verteilung auf verschiedene Niveaus sollte, wo immer möglich und sinnvoll, über die verschiedenen Berufe hinweg vergleichbar sein.

Unter dem Aspekt curricularer Validität galt es, mit den Tests die jeweiligen relevanten Inhaltsbereiche mit Blick auf den Lehrplan möglichst angemessen zu berücksichtigen. Die Überprüfung dieser Inhaltsvalidität setzte voraus, dass zu-

nächst berufsspezifisch ein Überblick über die verschiedenen Inhaltsbereiche gewonnen wurde (ca. 10 - 15 Inhaltsbereiche). Im zweiten Schritt wurde eine Gewichtung dieser Bereiche vorgenommen und anschließend die Verteilung der Items auf diese Inhaltsbereiche beurteilt.

4 Klassifikation und Dimensionierung des Tests am Beispiel des Ausbildungsberufes Bürokaufmann/-frau

Der Ausbildungsberuf Bürokaufmann/-kauffrau war einer von sieben kaufmännischen Ausbildungsberufen in ULME III. Die Entwicklung dieses Tests erfolgte in enger Zusammenarbeit zwischen einem in Unterricht, Lehrplanentwicklung und Prüfungsarbeit sehr erfahrenen Lehrer einer Hamburger Berufsschule und einem Mitglied des IBW-Teams.

Curricularer Bezugspunkt war dabei der Hamburger Bildungsplan für diesen Ausbildungsberuf von 2002. Dieser strukturiert das Berufsschulcurriculum im berufsbezogenen Lernbereich in elf Lernfelder und bündelt diese wiederum zu drei im Abschlusszeugnis notenmäßig auszuweisenden Fächern:

- Organisation und Personalmanagement 260 Stunden
- Auftragsbearbeitung: 260 Stunden
- Leistungsprozesse und Controlling: 360 Stunden

Unter dem Aspekt der curricularen Validität wurde der Itemkonstruktion und der abschließenden Itemauswahl eine Klassifikation der Aufgaben unterlegt, mit der der Test in dreifacher Weise ausbalanciert wurde, nämlich im Hinblick

- auf die spezifischen inhaltlich-thematischen Schwerpunkte bezogen auf die zu Fächern gebündelten Lernfelder,
- auf die mit den Aufgaben angesprochenen formalen Inhaltsklassen in Anlehnung an die Klassifikation von Anderson & Krathwohl (2001) (Fakten, Konzepte, Prozeduren) und
- auf die Art der geforderten Leistung in Anlehnung an eine von Metzger et al. (1993) vorgeschlagene Reduktion der Bloomschen Taxonomie (Reproduzieren; Anwenden/Verstehen; Kritisieren/Reflektieren).

In inhaltlich-formaler Hinsicht dominieren Aufgaben, die konzeptuelles Wissen und algorithmisches prozedurales Wissen erfordern (ca. 83%). In formaler Hinsicht enthält der Test dominant Aufgaben die auf Wissensanwendung bzw. Verstehen zielen (ca. 77%). Bei der Suche nach Kompetenzdimensionen, die sich untereinander im Hinblick auf den Charakter der jeweiligen kognitiven Leistungsanforderungen unterscheiden, schied die curricular vorgegebene Differenzierung der Lernfeldbündel zu den drei genannten Fächern deshalb aus, weil diese, der Lernfeldkonzeption folgend, fächerübergreifend angelegt waren und

damit hinsichtlich der jeweils einbezogenen Inhaltsbereiche ausgesprochen heterogen waren und zudem erhebliche Überschneidungen aufwiesen. Plausibler schien uns eine Differenzierung der Items im Hinblick auf die jeweils angesprochenen inhaltlichen Bezüge und Referenzsysteme, was zu einer Unterscheidung von vier Aufgabenklassen führte:

1. „Aufgaben mit Bezug auf die Durchführung und Optimierung betrieblicher Prozesse und die Gestaltung der dahinter stehenden Organisationsstruktur (*betriebswirtschaftliche Organisation und Leistungsprozesse*). Diese Aufgaben zielen auf betriebliches Erfahrungswissen sowie damit verknüpft betriebswirtschaftliches Hintergrundwissen.
2. Aufgaben mit Bezug auf die rechtliche Normierung kaufmännischen Handelns auf der einzelbetrieblichen Ebene (*rechtliche Dimension*). Gefordert sind hier weitgehend unabhängig von betriebswirtschaftlichem Prozesswissen und ökonomischen Kalkülen Kenntnisse rechtlicher Regelungen sowie die Fähigkeit zur Subsumtion konkreter Sachverhalte unter bestimmte Rechtsnormen.
3. Aufgaben mit Bezug auf die Modellierung ökonomischer Strukturen und Prozesse auf der Wertebene (*Wertschöpfungsdimension*). Diese Aufgaben zielen auf Konzepte und Prozeduren aus dem Bereich des Rechnungswesens, die aufgrund des stark axiomatischen Charakters des traditionellen Rechnungswesenunterrichts häufig nur schwach mit betriebswirtschaftlichem Inhaltswissen verknüpft sind (vgl. Preiss & Tramm 1996).
4. Aufgaben mit Bezug auf gesamtwirtschaftliche Strukturen und Prozesse und den volkswirtschaftlichen Ordnungsrahmen (*volkswirtschaftliche Dimension*). Diese Items sind wenig mit betrieblichem Prozesswissen verknüpft und entsprechen weitgehend den auch im WBT angesprochenen Inhaltsbereichen.

5 Empirische Befunde zur Dimensionierung kaufmännischer Kompetenz

5.1 Zur Skalierung der beruflichen Leistungstests in ULME-III

Die getesteten Jugendlichen bearbeiteten im Rahmen der berufsbezogenen Tests insgesamt 51 Aufgaben (107 Einzel-Items), die ein relativ breites Spektrum beruflicher Anforderungen repräsentieren. Sollen nun die Befunde der beruflichen Leistungstests im Hinblick auf curriculare, didaktische und psychometrische Fragestellungen interpretiert werden, so ist in jedem Fall eine inhaltliche Interpretation der erworbenen berufsbezogenen Kompetenzen erforderlich. Dabei wird die jeweilige Kompetenz an einem „Kriterium wohldefinierter Operationen gemessen, die ihrerseits bestimmte Lernziele repräsentieren können“ (Klieme, Baumert et al. 2000, 114). Zudem setzt dieses Vorgehen Skalierungsmethoden voraus, die Fähigkeitsschätzungen nach Maßgabe fachlich bestimmter Anforder-

rungsstufen ermöglichen (vgl. Baumert, Bos & Lehmann 2000, 27). Dementsprechend wurde zunächst für die in ULME III eingesetzten beruflichen und allgemeinen Fachleistungstests geprüft, inwiefern diese unter Nutzung von Methoden der Item-Response-Theory (vgl. Fischer & Molenaar 1995) skalierbar sind. Bei diesem Verfahren werden die Fähigkeiten der Probanden ermittelt, während gleichzeitig der Schwierigkeitsgrad der Testaufgaben auf demselben Maßstab bestimmt wird. Aufgabenschwierigkeiten und Schülerfähigkeiten sind über die Kategorie der Lösungswahrscheinlichkeit miteinander verknüpft. Die Schwierigkeit eines Items ist durch die Lage des Wendepunktes der so genannten Item-Charakteristik-Kurven definiert, an dem die von der Personenfähigkeit abhängige Lösungswahrscheinlichkeit 0,5 beträgt (vgl. Rost 2004, 115ff.). Durch eine einfache Transformation der Itemschwierigkeiten wurde erreicht, dass Aufgabenschwierigkeit und Personenfähigkeit einer Lösungswahrscheinlichkeit von 0,65 entsprechen.

In einem ersten Schritt erfolgte auch für die Gruppe der Bürokaufleute die Skalierung des beruflichen Fachleistungstests auf Basis des einparametrischen dichotomen Raschmodells, das ein Latent-Trait-Modell darstellt, bei dem die manifeste Reaktion auf jedes Item in der geschilderten Weise auf eine spezifische Personenfähigkeit und eine ebenso bestimmte Itemschwierigkeit zurückgeführt wird (vgl. dazu Baumert, Köller et al. 2000, 60ff.). Die Skalierbarkeit nach dem Rasch-Modell impliziert paarweise Unabhängigkeit der einzelnen Aufgaben, ferner die Abwesenheit (bedeutsamer) Rate-Tendenzen und das Fehlen von Effekten, nach denen einzelne Probandengruppen und/oder einzelne Items durch spezielle Antwortmuster charakterisiert sind. Demnach muss eine eindeutige und in allen Teilpopulationen gültige Rangfolge der Aufgaben in Bezug auf die Schwierigkeit existieren, was in der Regel an unterschiedliche Anforderungsmerkmale der Aufgaben gebunden ist (vgl. dazu Klieme, Baumert et al. 2000, 111ff.). Sofern das Rasch-Modell gilt, erlaubt die dann gültige Annahme der spezifischen Objektivität eine Bestimmung der Personenparameter unabhängig von der Itemstichprobe.

5.2 Zur Datenbasis im Ausbildungsberuf Bürokaufmann/Bürokauffrau

An der Hamburger Leistungsstudie ULME III, die für jeden Ausbildungsberuf insgesamt sechs Testkomponenten umfasste, nahmen 160 Schülerinnen und Schüler aus neun Abschlussklassen der beiden Berufsschulen teil, die in Hamburg Bürokaufleute ausbilden. Den beruflichen Kompetenztest absolvierten 156 Jugendliche. Für 135 Schülerinnen und Schüler dieses Ausbildungsberufs liegen Daten aus dem Längsschnitt vor, d. h. hier ist es prinzipiell möglich, die Lernentwicklung abzubilden und den Zusammenhängen zwischen den Eingangsleistungen in den allgemein bildenden Domänen und der beruflichen Fach- und Methodenkompetenz am Ende der Ausbildung nachzugehen.

5.3 *Latente Zusammenhänge und Hintergrundmodelle*

Die Inhalte der beruflichen Aufgabenanforderungen im Test für die Bürokaufleute beziehen sich, wie bereits im vorangegangenen Abschnitt ausgeführt, auf unterschiedliche Inhaltsbereiche, denen wiederum jeweils bestimmte Prinzipien, Begriffe etc. zugrunde liegen. Ihre erfolgreiche Bewältigung erfordert spezifische Fähigkeiten und Fertigkeiten. Insofern erscheint es geboten, die Interpretierbarkeit eines Gesamtestwertes für den beruflichen Leistungstest kritisch zu prüfen. Die Güte eines Modells lässt sich daran ablesen, „wie ‚genau‘ es die vorgegebenen Daten (Kovarianz- und Korrelationsmatrizen) reproduzieren kann“ (Schneider 1986).

Nachfolgend werden drei Modelle vorgestellt, deren Angemessenheit empirisch vergleichend geprüft werden soll. Im ersten Modell wird angenommen, dass sich die Leistungsscores im beruflichen Test für Bürokaufleute auf ein relativ komplexes, gleichwohl homogenes Fähigkeitsbündel ökonomischen Verständnisses sensu Beck zurückführen lassen (zur Struktur des Modells vgl. die weiter hinten diskutierte Abb. 3). Es wird hierbei angenommen, dass eine relativ komplexe latente Dimension, hier ein ökonomisches Grundverständnis ausreicht, die Leistung im Test zu bestimmen. Die über die einzelnen Testitems berechneten Erfolgsquoten bilden dann die Grundlage für die Schätzung der so definierten latenten Grundfähigkeit.

Folgt man hingegen der Annahme, dass die Performanz im hier betrachteten Test durch unterschiedliche Fähigkeitsdimensionen determiniert wird, wobei ggf. ein statistischer Zusammenhang zwischen den verschiedenen latenten Dimensionen bestehen kann, so ist a priori zu klären und zu begründen, welche Dimensionen geprüft werden, welches Item welche Dimension mit welchem Gewicht misst. Dabei ist es auch möglich, dass einzelnen Items sich nicht nur auf *eine* inhaltliche Dimension beziehen, sondern mehreren Dimensionen zugehören. In einer ersten Annäherung an dieses Thema haben wir eine Strukturprüfung nach fachinhaltsbezogenen Bereichen vorgenommen (s. Kap. 4).

Als Beispiele für diese zweite Variante wurde Ansätze gewählt, bei denen unterstellt wird, dass sich die Testleistungen am besten durch zwei bzw. vier spezifische, aber untereinander korrelierte inhaltsbezogene Fähigkeitsfaktoren erklären lassen (zur zweidimensionalen Struktur vgl. Abb. 4). Bei den beiden mehrdimensionalen Ansätzen wurden die insgesamt 107 Items den vier bzw. zwei Subdimensionen zugeordnet. Der vierdimensionale Ansatz folgte der in Kap. 4 erläuterten Differenzierung; beim zweidimensionalen Ansatz wurde zwischen der Dimension des Rechnungswesens einerseits und volks-, betriebswirtschaftlichen sowie rechtlichen Aspekten beruflichen Handelns andererseits differenziert.

5.4 Modellprüfung unter Nutzung des einparametrischen Rasch-Modells

Unter Verwendung des Computerprogramms *ConQuest* wurden über das einparametrische logistische Modell die drei zur Diskussion stehenden Modellvarianten geprüft und miteinander verglichen. Die berechneten latenten Korrelationen zwischen den Dimensionen wurden bei den beiden mehrdimensionalen Ansätzen über die Maximum-Likelihood-Methode (z. B. Rost 2004, 309ff.) direkt als Modellparameter geschätzt. Auf diesem Wege erhält man für jeden Probanden eine Schätzung des Personenparameters in jeder berücksichtigten Dimension. Ein Nachteil dieses Analyseverfahrens liegt allerdings darin, dass die allgemeine Anpassung der Modellgüte nur unzureichend erkennbar wird.

Der *eindimensionale Ansatz* wies bei 107 Items zufrieden stellende Eigenschaften auf, insbesondere eine hohe interne Konsistenz ($\alpha_{wle} = 0,87$).

Die *vierdimensionale Variante* lieferte Anhaltspunkte für zusätzliche Dimensionen. Bei Annahme dieser Lösung zeigen sich eine relativ enge latente Korrelation zwischen den Dimensionen „Rechtliche Aspekte“ und „Volkswirtschaftliche Aspekte“ sowie ein mäßiger Zusammenhang zwischen „Rechtlichen Aspekten“ und dem Bereich des „Rechnungswesens“. Es wurde jedoch bei den Subdimensionen „Rechtliche Aspekte“ und „Volkswirtschaftliche Aspekte“ eine interne Konsistenz erreicht, die $\alpha_{wle} = 0,58$ bzw. $\alpha_{wle} = 0,55$ nicht überschritt und insoweit unbefriedigend blieb. Die Anpassungsgüte des Modells an die empirischen Daten war gegenüber dem eindimensionalen Modell deutlich ungünstiger ($\text{Chi}^2=8054.266$, $\text{df}=103$ gegenüber $\text{Chi}^2=5264.611$, $\text{df}=106$), so dass es einstweilen für den hier vorliegenden beruflichen Leistungstests nicht angezeigt scheint, mit der vierdimensionalen Variante zu arbeiten.

Eine bessere Passung zwischen empirischen Daten und Testmodell wird beim *zweidimensionalen Ansatz* erreicht. Es wurden insgesamt 56 Items den betriebs- und volkswirtschaftlichen sowie rechtlichen Aspekten beruflichen Handelns zugeordnet, demgemäß entfielen 51 Items auf das ökonomische Modellieren, Aufbereiten und Umgehen mit Daten unter Anwendung weitgehend normierter Instrumente und Regeln des Rechnungswesens. In diesem Ansatz wurde ersichtlich, dass die zweite Dimension deutlich anspruchsvollere Aufgaben enthält als die erste. Beide Dimensionen erlangen eine noch zufrieden stellende interne Konsistenz von $\alpha_{wle} = 0,80$ bzw. $\alpha_{wle} = 0,73$; die mäßige latente Korrelation von 0,51 zwischen den beiden Konstrukten bzw. Subtests deutet darauf hin, dass die Ausdifferenzierung einer eigenständigen Rechnungswesen-Komponente zusätzliche substanzielle Einsichten ermöglichen könnte. Allerdings ist der gesamte Modellfit immer noch etwas ungünstiger als bei der eindimensionalen Variante ($\text{Chi}^2=5366.603$, $\text{df}=105$), so dass eine weitere Prüfung der Dimensionalität auf der Grundlage von Strukturgleichungsmodellen zwar notwendig, zugleich aber aussichtsreich erscheint.

5.5 Modellprüfung unter Nutzung von Strukturgleichungsmodellen

Auf der Grundlage so genannter Strukturgleichungsmodelle unter Nutzung des Programms *Amos 4.0* wurden deshalb der ein- und der zweidimensionale Ansatz geprüft. Die Fallzahlen im Datensatz und die relativ hohe Zahl an beobachteten Variablen machten eine Vorauswahl der Items erforderlich. Kriterium für die Auswahl der Items waren eine möglichst eindeutige Zuordnung zu einer der beiden Dimensionen und die Messgüte der Items selbst. Insgesamt wurden 21 der 107 Items in die Modellprüfung aufgenommen. Insofern waren hier günstigere Prüfbedingungen gegeben als im Falle des Vergleichs über das Raschmodell im Programm *ConQuest*. Das Programm *Amos* stellt verschiedene Gütestatistiken zur Verfügung, um die Qualität der Schätzergebnisse zu beurteilen. Die Tabelle 2 enthält die herangezogenen Kennwerte zur Beurteilung der Modellanpassung.

<i>Modell</i>	χ^2	<i>df</i>	<i>RMSEA</i>	<i>TLI</i>	<i>IFI</i>
General-Faktor-Modell	235,60	188	0,04	0,98	0,98
zwei korrelierende spezifische Faktoren	225,841	187	0,04	0,98	0,99

Tab. 2: Statistiken für die Anpassungsgüte konkurrierender Modelle

Wie aus den Statistiken hervorgeht, weisen beide Modelle sehr ähnliche Anpassungskennwerte auf. Zwar wären nach dem χ^2 -Kriterium die Voraussetzungen für beide Modellvarianten streng genommen nicht gegeben (bei leichtem Vorteil für den zweidimensionalen Ansatz), doch die übrigen Kriterien (*RMSEA*=Root mean square error of approximation; *TLI*=Tucker-Lewis index, *IFI*=Index of fit), die relativ unabhängig von der Stichprobengröße sind, legen den Schluss nahe, dass in beiden Fällen die Anpassungsgüte ausreichend ist, wiederum mit geringfügig besserem Wert für die zweidimensionale Struktur (zur Diskussion der Goodness-of-Fit Statistics vgl. Byrne 2000, 79ff.).

In den Abbildungen 3 sind die geschätzten Modellparameter für die zweidimensionale Variante dargestellt. Die Faktorladungen der endogenen Variablen auf die beiden latenten Konstrukte variieren zwischen 0,76 (BU482) und 0,16 (BU11). In einer Reihe von Items zeigte sich eine markante Verbesserung der Korrelation bei einer Homogenisierung der Konstrukte im zweidimensionalen Ansatz gegenüber dem eindimensionalen Modell.

Neben dem etwas günstigeren Modell-Fit der zweidimensionalen Variante (Abb. 3) hat dieser den zusätzlichen Vorteil, differenziertere Einblicke in die Leistungsstände der Hamburger Schülerinnen und Schüler in den Abschlussklassen der Bürokaufleute zuzulassen. Die Doppelpfeile zwischen den Sub-Dimensionen symbolisieren die latenten, d. h. messfehlerbereinigten Interkorrelationen.

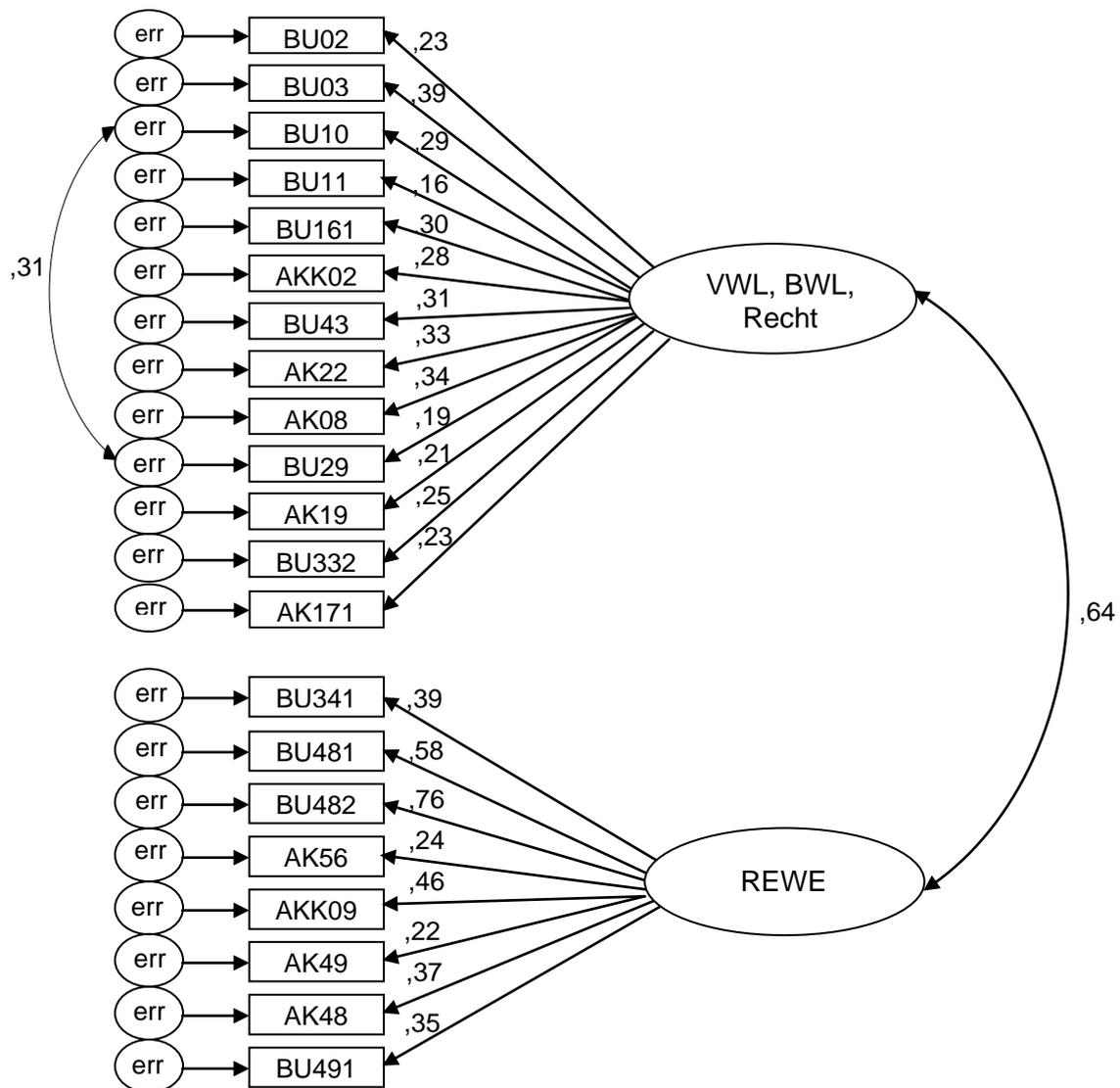


Abb. 3: Modell mit zwei korrelierenden Faktoren zur Erklärung der Testleistungen

Im Rahmen der zweidimensionalen Rasch-Skalierung war es möglich, die Testwerte für jede Sub-Dimension auf einer Metrik darzustellen und damit im unmittelbaren Vergleich zu interpretieren. So zeigt sich beispielsweise, dass die Lernstände im Bereich „Betriebliche Leistungsprozesse und Controlling“ (MW=100, SD=25) markant unterhalb der Leistungen in der anderen Dimension liegen (MW=129,01; SD=23,23). Die Differenz zwischen den beiden Bereichen beträgt damit mehr als eine ganze Standardabweichung. Auf der Klassenebene wird deutlich, dass derartig ausgeprägte Unterschiede zwischen den beiden Sub-Dimensionen nicht für alle Lerngruppen gleichermaßen festzustellen sind, sondern innerhalb der beiden Schulen jeweils zwischen den Klassen bedeutsam variieren. Zwar bleibt für alle Klassen eine Differenz zwischen den Leistungswerten der beiden Sub-Dimensionen zugunsten des beruflichen Handelns und Entscheidens im Kontext allgemeiner betriebs-, volkswirtschaftlicher und wirtschaftsrechtlicher Zusammenhänge zu berichten, jedoch fallen die Differenzmaße

sehr unterschiedlich aus: sie schwanken auf Klassenebene zwischen 0,3 und 2,1 Standardabweichungen. D. h. in einzelnen Klassen sind relativ homogene Leistungsprofile festzustellen, während in anderen Lerngruppen sehr auffällige Leistungsdifferenzen zwischen den beiden Teildimensionen erkennbar werden.

6 Diskussion der Befunde

Dieser erste Ansatz zur Ausdifferenzierung von Dimensionen der in ULME III eingesetzten beruflichen Leistungstests verdeutlicht die Notwendigkeit, differenzierter die hinter den Aufgabenlösungen von Probanden stehenden kognitiven Fähigkeiten und Verständnisleistungen zu analysieren. Dabei wird deutlich, dass sowohl in der theoretischen Ausdifferenzierung der latenten Konstrukte über Item-Analysen als auch in der empirischen Prüfung der Modelle noch intensive Entwicklungsarbeit zu leisten ist. So wurde beispielsweise bei den Dimensionsanalysen mit dem Rasch-Modell vorerst ein „between-item“-Modell gewählt, bei dem unterstellt wird, dass dem Test eine multidimensionale Struktur zugrunde liegt, bei der eindimensionale Sub-Skalen identifizierbar sind. Nicht immer ließen sich jedoch die Items im hier geprüften Beispiel zweifelsfrei einer oder anderen Dimension zuordnen, so dass auch ein „within-Item“-Ansatz zu prüfen wäre, bei dem einzelne Items mehr als nur einer latenten Dimension zugeordnet werden (zu den Modellen vgl. Adams, Wilson & Wang 1997).

Zusammenfassend kann vorerst festgehalten werden, dass es zu rechtfertigen ist, beim Leistungsvergleich für den Ausbildungsberuf Bürokaufmann/-frau, wie er hier im Rahmen der ULME-III-Studie durchgeführt wurde, sowohl einen Gesamtestwert als auch differenzielle Leistungswerte heranzuziehen. Wir konnten für diesen Ausbildungsberuf zeigen, dass differenziertere Analysen durchaus zusätzliche Informationen über die Umsetzung der Curricula und über unterrichtliche Schwerpunktsetzungen an den Hamburger Schulen liefern, die einer ausschließlichen Analyse mit einem Gesamtestwert verborgen geblieben wären.

Die hier berichteten Ergebnisse zu den Strukturanalysen kaufmännischer Kompetenz stehen allerdings unter der Einschränkung einer relativ kleinen, regional beschränkten Stichprobe. Insofern ist eine zufallskritisch breiter abgesicherte Überprüfung der hier diskutierten Modelle sinnvoll.

Problematisch erscheint es uns jedoch, Befunde aus dem Kontext volkswirtschaftlicher Aufgabenstellungen ohne weiteres auf den Gesamtbereich kaufmännischer Kompetenz zu übertragen. Das Verhältnis von ökonomischer Bildung und kaufmännischer Kompetenz bedarf weiterer konzeptueller und empirischer Aufklärung; Klaus Beck hat mit seinen Forschungen und Diskussionsbeiträgen hierfür wichtige Impulse geliefert.

*Literatur**1. Monographien*

- ANDERSON, L. W./ KRATHWOHL, D. R. (2001): A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. New York: Addison & Wesley.
- BECK, K./ KRUMM, V./ DUBS, R. (1998): Wirtschaftskundlicher Bildungs-Test (WBT). Göttingen et al.: Hogrefe.
- BYRNE, B. M. (2001): Structural Equation Modeling with AMOS. Basis Concepts, Applications and Programming. Mahwah: Lawrence Erlbaum Associates.
- FISCHER, G. H./ MOLENAAR, I. W. (1995): Rasch models – Foundations, recent developments, and applications. New York: Springer.
- HELMKE, A./ HOSENFELD, I. (2003): Vergleichsarbeiten – Standards – Kompetenzstufen: Begriffliche Klärung und Perspektiven für VERA. Manuskript vom 27. August 2003. http://www.mbjs.brandenburg.de/media/lbm1.a.4365.de/vera_standards_kompetenzstufen.pdf Download 1.5.2006.
- 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.
- LEHMANN, R. H./ IVANOV, S./ HUNGER, S./ GÄNSFUSS, R. (2005): ULME I. Untersuchung der Leistungen, Motivationen und Einstellungen zu Beginn der beruflichen Ausbildung. Hamburg: Behörde für Bildung und Sport, Amt für Berufliche Bildung und Weiterbildung.
- LEHMANN, R. H./ SEEBER, S./ HUNGER, S. (2006): Untersuchung der Leistungen, Motivation und Einstellungen von Schülerinnen und Schülern in den Abschlussklassen der teilqualifizierenden Berufsfachschulen (ULME II). Hamburg: Behörde für Bildung und Sport.
- METZGER, C./ WAIBEL, R./ HENNING, C./ HODEL, M./ LUZI, R. (1993): Anspruchsniveau von Lernzielen und Prüfungen im kognitiven Bereich. St. Gallen: IWP.
- ROST, J. (2004): Lehrbuch Testtheorie – Testkonstruktion. Zweite, vollständig überarbeitete und erweiterte Auflage. Bern, Göttingen: Hans Huber.
- SOPER, J. C./ WALSTAD, W. B. (1987): Test of Economic Literacy. Second Edition. New York: Joint Council on Economic Education.
- STRAKA, G. A. (2003): Verfahren und Instrumente der Kompetenzdiagnostik – der Engpass für ein Berufsbildung-PISA? Überarbeiteter Beitrag zum workshop „Berufsbildung-PISA“ des Bundesinstituts für Berufsbildung vom 30. Juni 2003 in Bonn.
- WU, M. L./ ADAMS, R. J./ WILSON, M. R. (1998): ACER ConQuest. Generalised Item Response Modelling Software. Australia: Australian Council for Educational Research.

2. Beiträge in Sammelwerken

- BAUMERT, J./ BOS, W./ LEHMANN, R. H. (Hg.) (2000): TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 1. Opladen: Leske + Budrich.
- BAUMERT, J./ KÖLLER, O./ LEHRKE, M./ BROCKMANN, J. (2000): Anlage und Durchführung der Dritten Internationalen Mathematik- und Naturwissenschaftsstudie zur Sekundarstufe II (TIMSS/III – Technische Grundlagen). In: BAUMERT, J./ BOS, W./ LEHMANN, R. (Hg.): TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Grundbildung am Ende der Schullaufbahn. Band 1. Opladen: Leske + Budrich, 31-84.
- KLIEME, E./ BAUMERT, J./ KÖLLER, O./ BOS, W. (2000): Mathematische und naturwissenschaftliche Grundbildung: Konzeptuelle Grundlagen und die Erfassung und Skalierung von Kompetenzen. In: BAUMERT, J./ BOS, W./ LEHMANN, R. (Hg.): TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Grundbildung am Ende der Schullaufbahn. Band 1. Opladen: Leske + Budrich, 85-133.
- PREIß, P./ TRAMM, T. (1996): Die Göttinger Unterrichtskonzeption des wirtschaftsinstrumentellen Rechnungswesens. In: PREIß, P./ TRAMM, T. (Hg.): Rechnungswesenunterricht und ökonomisches Denken. Wiesbaden: Gabler, 222-323.
- SCHNEIDER, W. (1986): Strukturgleichungsmodelle der zweiten Generation: Eine Einführung. In: MÖBUS, C./ SCHNEIDER, W. (Hg.): Strukturmodelle für Längsschnittdaten und Zeitreihen. LISREL, Pfad- und Varianzanalyse. Bern: Hans Huber, 13 – 26.
- WEINERT, F. E. (2001): Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In: WEINERT, F. E. (Hg.): Leistungsmessung in Schulen. Weinheim und Basel: Beltz Verlag, 17-31.

3. Beiträge in Zeitschriften

- ADAMS, R. J./ WILSON, M. R./ Wang, M. L. (1997): The multidimensional random coefficients multinomial logit. *Applied Psychological Measurement*, 21, 1-24.
- BECK, K. (1989): „Ökonomische Bildung“ – Zur Anatomie eines wirtschaftspädagogischen Begriffs. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 85, 579-596.
- 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*, Heft 8/2005. http://www.bwpat.de/ausgabe8/brand_et al_bwpat8.shtml
- SCHWIPPERT, K. (2005): Vergleichende Lernstandsuntersuchungen, Bildungsstandards und die Steuerung von schulischen Bildungsprozessen. *bwp@ - Berufs- und Wirtschaftspädagogik online*, Heft 8/2005. http://www.bwpat.de/ausgabe8/schwippert_bwpat8.shtml
- STRAKA, G. A. (2005): Die KMK-Handreichungen zur Erarbeitung von Rahmenlehrplänen - eine kritische Reflexion zum zehnten Jahrestag. *bwp@ - Berufs- und Wirtschaftspädagogik online*, Heft 8/2005. http://www.bwpat.de/ausgabe8/straka_bwpat8.shtml
- TRAMM, T. (2003): Prozess, System und Systematik als Schlüsselkategorien lernfeldorientierter Curriculumentwicklung. *bwp@ - Berufs- und Wirtschaftspädagogik online*, Heft 4/2003. http://www.ibw.uni-hamburg.de/bwpat/ausgabe4/tramm_bwpat4.html