

Auf dem Weg zu einem Kompetenzstufenmodell für die berufliche Bildung – Erfahrungen aus dem Projekt ULME

1 Problemstellung

Die Durchführung groß angelegter Lernstands- und Lernentwicklungsuntersuchungen hat in Hamburg bereits eine beachtliche Tradition (vgl. SCHWIPPERT in diesem Band). Mit dem von der Behörde für Bildung und Sport (Kultusministerium) 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 und danach im Zweijahresrhythmus systematisch weiter evaluiert wurde, in der entsprechenden Teilpopulation auch während ihrer schulischen Laufbahn im Berufsbildungssystem weiter verfolgt. In drei Untersuchungsrounden wurden Schülerinnen und Schüler beim Übergang in eine berufliche Schulform (ULME I), kurz vor Abschluss sogenannter teilqualifizierender Berufsfachschulen (ULME II) 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 Intelligenz 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 der Universität Hamburg; die wissenschaftliche Leitung des Gesamtprojekts lag bei Prof. Dr. Rainer Lehmann von der Humboldt Universität zu Berlin.

Die Arbeitsstrategie zur Entwicklung dieser berufsbezogenen Tests und mithin der Versuch einer Annäherung an die Operationalisierung beruflicher Handlungskompetenz in Form von standardisierten Testitems steht im Mittelpunkt dieses Beitrags. Der Argumentationsbogen soll dabei von einigen grundsätzlichen Reflexionen im Hinblick auf das (berufs-)bildungspolitische Relevanz- und Wirkungsspektrum dieser Untersuchung über deren theoretisch konzeptuelle Grundlagen, insbesondere im Hinblick auf den Begriff der Handlungskompetenz und der Kompetenzstufen, bis hin zu unserem Versuch einer Niveaueinstufung von Aufgaben auf der Grundlage kognitionspsychologisch-taxonomischer Überlegungen geschlagen werden. Wir werden uns dabei im wesentlichen auf Ansätze und Erfahrungen aus dem Kontext von ULME III beziehen. Lediglich die abschließenden Überlegungen, die erste Auswertungsergebnisse mit einbeziehen, basieren zumindest bezüglich der empirischen Erfahrungswerte auf den Daten aus ULME II.

2 Das Projekt ULME im Kontext der Hamburger Lernausgangslagen- und Lernstandsuntersuchungen – Versuch einer bildungspolitischen Einordnung

Die „Untersuchung von Leistungen, Motivation und Einstellungen“ soll Lernleistungen und Lernvoraussetzungen der Schüler in ausgewählten Bereichen der beruflichen Ausbildung auf der Grundlage von Tests und Fragebögen systematisch transparent machen. In Zielen und Instrumenten bildet ULME weitgehend eine Fortsetzung der Hamburger Lernausgangslagen-Untersuchung (LAU). Beginnend mit der Studie zu Beginn der gymnasialen Oberstufe (LAU 11) wurde parallel die Längsschnitt-Untersuchung ULME I-III in berufsbildenden Schulen eingeführt und mit ULME III im Frühjahr 2005 abgeschlossen.

Die Erkenntnisse von LAU und ULME dienen zum einen der verbesserten Gestaltung des Unterrichts, zum anderen einer Überprüfung der Positionierung der untersuchten Schulformen und Bildungsgänge und damit bildungspolitischer Entscheidungen. Bezogen auf die duale Berufsausbildung stehen dabei die schulischen und nicht die betrieblichen Lernprozesse im Mittelpunkt, wenngleich hinsichtlich der Lerneffekte sicherlich keine differenzierende Ursachenzuschreibung möglich sein dürfte.

Als nahezu flächendeckende Längsschnittuntersuchungen, die zusammen einen Zeitraum von fast einem Jahrzehnt umgreifen, sind LAU und ULME unter vielen anderen nationalen und internationalen Untersuchungen besonders herausgehoben. In der gegenwärtigen bildungspolitischen Lage führt die Entwicklung eines Untersuchungsinstrumentariums zur Messung von Lernständen unweigerlich zur Frage, ob das Instrumentarium selbst nicht auch eine geeignete Ausgangsbasis für die Entwicklung von Bildungsstandards hergeben kann. Insbesondere im Kontext von ULME ist die Entwicklung von Standards jedoch nicht thematisiert worden. Dennoch lässt der Umgang mit den Ergebnissen der PISA- und anderer Studien befürchten, dass, sobald die Ergebnisse vorliegen, die angewandten Leistungstests unter der Hand in der öffentlichen und politischen Diskussion den Rang von „Standards“ zugeschrieben bekommen. Von der zuständigen Hamburger Senatorin wurden so unmittelbar nach der Veröffentlichung der in einigen Punkten alarmierenden Ergebnisse von LAU 11 weitreichende bildungspolitische Forderungen aus den Befunden abgeleitet (FREIE UND HANSESTADT HAMBURG 2004). Dabei war keineswegs geklärt, wie weit die Studie „Standards“ abbildete, welche Kompetenzen in diesen Ergebnissen beschrieben wurden und welche Bedeutung sie für weitere Lernprozesse und die Bewährung im beruflichen und privaten Leben haben. Wie NEUWEG (in diesem Heft) zeigt, ist auch die Überantwortung der Entwicklung von Bildungsstandards an „Experten“ ein mit weitreichenden Risiken behafteter Schritt. Erfahrungen aus den USA zeigen, wie gesellschaftspolitische bis hin zu privatwirtschaftlichen Interessengruppen Definitionsmacht über Bildungsstandards gewinnen können und wie das dort in einigen Staaten eingeführte Testregime in Verbindung mit Selektionsmechanismen für Collegezugänge den Unterricht kontrolliert. Unter diesen Umständen sind anspruchsvolle pädagogische Projekte zwar z. B. Gegenstand der Weiterbildung von Lehrkräften, aber in der Schule fordern Eltern, Schüler und Schulleiter striktes „teach to the test“, wie eine erfahrene Lehrerin jüngst in einem Interview sagte (WINERIP 2005).

Die Entwicklung der berufsbezogenen Tests von ULME geht von einem alle Ausbildungsberufe übergreifenden Modell von kognitiven Kompetenzniveaus aus, die in den Tests durch Items domänenspezifisch operationalisiert werden. Mit diesem Ansatz wird eine Strategie verfolgt, die kaum in der Gefahr steht, sich in Stoffkatalogen zu verlieren, die nicht selten zu Standards verdichtet werden ohne eine überzeugende Theorie für diesen Prozess zu haben. Für die Aufgabenproduktion in ULME war konstitutiv, dass Experten aus Schulen und Universitäten zusammen arbeiteten und dass ein Klassifikationsraster von Wissensarten und kognitiven Leistungen die Itementwicklung anleitete und kontrollierte (s. u.). Es besteht die Hoffnung, dass mit diesem Ansatz das Risiko verringert wird, vorschnell und rigide für eine Standardisierung von Unterricht in Anspruch genommen zu werden. Dennoch bleibt festzuhalten, dass auch ULME ein diagnostisches Instrumentarium enthält, das – wie jede Diagnose – an Vergleichswerten, auch an Normwerten orientiert sein muss. Wenn ULME dazu führt, dass ein Unterstützungssystem für Berufsschulen, Lehrkräfte und Schüler eingerichtet wird, das Ergebnisse aus ULME und anderen Untersuchungen aufgreift und für gezielte Verbesserungen von Schule und Unterricht nutzt, ist das Potenzial dieser Untersuchung optimal ausgeschöpft.

3 Der konzeptuelle Rahmen des Projekts ULME – Handlungskompetenz und Kompetenzstufen

Für die Entwicklung berufsbezogener Aufgabensätze in ULME II und III kommt dem Begriff der Kompetenz bzw. Handlungskompetenz eine zentrale Bedeutung zu. Über dieses Konzept wird der Zusammenhang der Testentwicklung mit der Diskussion um Bildungsstandards im allgemeinen und mit der Diskussion um kompetenzorientierte Lernfeldcurricula in der beruflichen Bildung im besonderen deutlich. Zugleich erweist sich der Kompetenzbegriff zunehmend als kontextabhängig und demnach in seiner generellen Verwendung als zu unbestimmt.

Im Zusammenhang der großen internationalen Systemevaluationsstudien und der Debatte um nationale Bildungsstandards wird Kompetenz überwiegend verstanden als „eine Disposition, die Personen befähigt, bestimmte Arten von Problemen erfolgreich zu lösen, also konkrete Anforderungssituationen eines bestimmten Typs zu bewältigen“ (KLIEME et al. 2003, 72). WEINERT (2001, 27f.) präzisiert, man verstehe „unter Kompetenzen die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“.

Ohne hier auf Differenzierungen und Kontroversen zum Kompetenzbegriff näher einzugehen (vgl. hierzu WEINERT 1999; ERPENBECK/VON ROSENSTIEL 2003), lassen sich doch weithin konsensuale Merkmale des Konstrukts analytisch herausstellen:

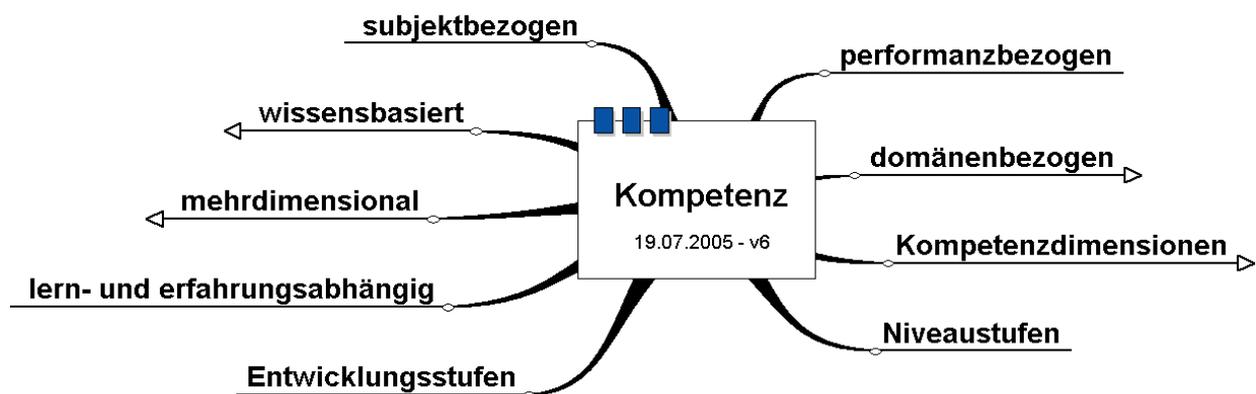


Abb. 1: Aspekte des Kompetenzkonzepts

- Kompetenzen sind **subjektbezogen**, d. h. sie werden in Bezug auf Menschen formuliert, die über diese Kompetenzen verfügen oder sich diese aneignen sollen.
- Kompetenzen werden in Bezug auf abgegrenzte Leistungen oder Leistungsbereiche, also **performanzbezogen** formuliert und erfasst. Der Bezug zur Performanz und damit zu konkretem Verhalten ist konstitutiv für den Kompetenzbegriff im Sinne CHOMSKYs.
- Kompetenzen sind **wissensbasiert**, d. h. sie implizieren eine strukturierte Wissensbasis, aus der heraus die jeweiligen Leistungen situationspezifisch generiert werden. Dieses Wissen kann in vielfältiger Form vorhanden sein und aktiviert werden. Gängig ist die Unterscheidung in deklaratives, prozedurales und konditionales Wissen; neben explizitem (verbalisierbarem, bewusstseinsfähigem) scheint auch implizites Wissen eine große Rolle zu spielen (vgl. NEUWEG 1999). Kompetenzentwicklung ist damit immer auch Wissensentwicklung, und umgekehrt soll sich aus pädagogisch-pragmatischer Perspektive Wissenserwerb in Kompetenzerweiterung niederschlagen.
- In diesem Sinne werden Kompetenzen als **lern- und erfahrungsabhängig** angesehen, ungeachtet der Frage, ob und inwieweit bestimmte elementare Schemata angeboren sein mögen.
- Kompetenzen werden heute weitgehend in dem Sinne **mehrdimensional** konzeptualisiert, als sie nicht nur kognitive und psychomotorische Aspekte, sondern auch motivationale, soziale und volitionale (Willens-) Aspekte umfassen.
- Weiterhin kann aus dieser subjektzentrierten Perspektive Kompetenzentwicklung als ein individueller Lern- und Entwicklungsprozess verstanden werden, und in diesem Sinne kann nach **Entwicklungsstufen oder -phasen** gefragt bzw. kann angenommen werden, dass sich dieser Prozess in solchen Stufen oder Phasen vollzieht.
- Im Hinblick auf den Performanzbereich besteht weitgehend Konsens darüber, dass Kompetenzen im wesentlichen **domänenspezifisch** zu fassen sind, sich also auf abgegrenzte

- Gegenstandsbereiche, Problem- oder Handlungsfelder beziehen und damit auch spezifisches Wissen und Können voraussetzen.
- Quer dazu finden sich verbreitet Ansätze zur Differenzierung von **Kompetenzdimensionen**, etwa im Hinblick auf unterscheidbare Aspekte der Interaktion des Subjekts mit seiner Umwelt. Verbreitet ist insbesondere die durch Heinrich ROTH (1971, 379ff) pädagogisch-anthropologisch fundierte Differenzierung in Sach-, Sozial- und humane Selbstkompetenz.
 - Schließlich ist dem Kompetenzkonzept der Gedanke inhärent, dass Kompetenzen in unterschiedlichen qualitativen Ausprägungen vorhanden sind und dass es damit sinnvoll ist, **Niveaustufen der Kompetenz** zu definieren und auf der Performanzebene zu erfassen. „Jede Kompetenzstufe ist durch kognitive Prozesse und Handlungen von bestimmter Qualität spezifiziert, die Schülerinnen und Schüler auf dieser Stufe bewältigen können, nicht aber auf niedrigeren Stufen“ (KLIEME et al. 2003, 22). Niveaustufen werden mithin hierarchisch (genauer komplexionshierarchisch) gedacht; sie sind nach unserer Einschätzung damit jedoch durchaus nicht gleichzusetzen mit Entwicklungsmodellen der Kompetenz, denn aus der Tatsache der Abstufung ergibt sich weder zwingend, dass der individuelle Entwicklungsprozess der Abfolge dieser Niveaustufen folgt, noch, dass es didaktisch geboten wäre, Sequenzen entlang dieser Stufung zu konzipieren. Dennoch wäre natürlich die Beantwortung der Frage nach der Kohärenz von Niveau- und Entwicklungsstufenmodellen analytisch und didaktisch von hoher Relevanz.

Die Entwicklung von Kompetenzniveau- oder auch Kompetenzstufenmodellen ist ein zentrales Anliegen im Kontext der Forschungen zur Bildungsevaluation und Standardentwicklung (z. B. KLIEME et al. 2003, 74ff.). Nach verbreiteter Einschätzung müssen derartige Modelle in Kooperation von fachdidaktischer und empirischer Bildungsforschung und in Form eines iterativen Zusammenspiels theoretisch-systematischer Ex-ante-Überlegungen und empirischer Ex-Post-Analysen entwickelt werden (HELMKE/HOSENFELD 2003, 4f.). Quasi im Schnittpunkt beider Arbeitsrichtungen liegt die Erarbeitung von Test-Aufgaben, mit denen die Lernenden konfrontiert werden und die performanzseitig Aufschluss über die vorhandenen Kompetenzen und das erreichte Kompetenzniveau geben sollen.

Eher pragmatisch fundierte Kompetenzstufenmodelle liegen mittlerweile etwa für die Bereiche der fremd- und muttersprachlichen Kompetenz, der mathematischen oder naturwissenschaftlichen Kompetenz vor (z. B. BAUMERT et al. 2001); für den berufsbildenden Bereich gibt es solche Modelle bislang noch nicht. Ihre Entwicklung wird jedoch zunehmend als dringlich empfunden und in diesem Kontext stehen auch die Arbeiten von ULME.

Vor diesem Hintergrund ist die Frage zu diskutieren, ob das in Bezug auf die allgemeine Standard- und Kompetenzdiskussion entfaltete Kompetenzverständnis ohne weiteres auf den berufsbildenden Bereich zu übertragen ist. Zur Beantwortung dieser Frage soll am Attribut der „Domänenspezifität“ von Kompetenzen angesetzt werden. „Kompetenzen“ so heißt es in der KLIEME-Expertise (2003, 22), „spiegeln die grundlegenden Handlungsanforderungen, denen Schülerinnen und Schüler in der Domäne ausgesetzt sind“ und der Erwerb von Wissen

müsse „beim systematischen Aufbau von Wissen in einer Domäne beginnen“ (ebenda). Genau an dieser Stelle findet sich eine interessante Abgrenzung, wenn KLIEME et al. (ebenda) in einer Fußnote anmerken: „Der hier verwendete Begriff von ‚Kompetenzen‘ ist daher ausdrücklich abzugrenzen von den aus der Berufspädagogik stammenden und in der Öffentlichkeit viel gebrauchten Konzepten der Sach-, Methoden-, Sozial- und Personalkompetenz. Kompetenzen werden hier verstanden als Leistungsdispositionen in bestimmten Fächern oder ‚Domänen‘“. Der Begriff der „Domäne“ wird auf der gleichen Seite mit Bezug auf die Wissenspsychologie wiederum mit „einem Lernbereich oder einem Fach“ gleichgesetzt.

Im Ergebnis ergibt sich aus diesen Festlegungen eine höchst interessante Divergenz zwischen dem Kompetenzbegriff der allgemeinen Standarddiskussion und dem Konzept der Handlungskompetenz in der Berufs- und Wirtschaftspädagogik. Pointiert gesagt fokussiert die allgemeine Standarddiskussion auf kognitive Leistungsfähigkeit in Bezug auf eingeführte Schulfächer des traditionellen Kanons (vgl. dazu auch KLIEME 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 Inhalt über die damit zu verknüpfende kognitive Leistung zur situierten Aufgabe.

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

Im berufsbildenden Bereich könnte analog argumentiert werden, wenn der Bezugspunkt eine im wesentlichen materiell definierte ökonomische oder technische Bildung wäre. Sobald jedoch der Anspruch des Lernfeldansatzes ins Spiel kommt, verkehrt sich dieser 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

hier im wesentlichen, bis zu welchem Komplexitätsniveau diese Leistungen entwickelt werden sollen und zu klären ist weiterhin, welche Wissensbasis diesen Leistungen zugrunde liegt. Der Klärungsweg führt hier also von der (beruflichen) Situation über die geforderte Leistung hin zum Wissen, das in dieser Leistung wirksam wird (vgl. hierzu TRAMM 2003; STRAKA in diesem Heft) .

Damit wird ein pragmatisch akzentuiertes Konzept von Kompetenz wirksam, wie es, unter Berufung auf CHOMSKY (1970) vor allem in der Arbeitspsychologie entwickelt wurde (HACKER 1998, VOLPERT 1992), das nach unserem Verständnis aber auch mit der Handlungstheorie Hans AEBLIs (1980) vereinbar ist.

Handlungskompetenz wird danach – analog zur Sprachkompetenz im Sinne der Generativen Transformationsgrammatik CHOMSKYs (1970) – als die Fähigkeit verstanden, aus einem begrenzten Elementen- und Regelsystem (Wissensbasis) heraus eine prinzipiell unendliche Vielzahl situationsadäquater Handlungen generieren zu können¹. Analytisch lassen sich dabei zwei Teilleistungen unterscheiden: Einerseits die Fähigkeit zur Orientierung, d. h. zur Wahrnehmung, Deutung und Bewertung von Situationen, zur angemessenen kognitiven Modellierung von Handlungssituationen. Dies ist gefordert im Zuge der Wahrnehmung einer Ausgangssituation, der Modellierung des Zielzustandes, des Abwägens alternativer Handlungswege und der Wahrnehmung und Beurteilung von Zwischenzuständen. HACKER (1998) spricht in diesem Zusammenhang vom Operativen Abbildsystem. In diese Leistung fließen sowohl kognitive als auch affektive und volitionale Aspekte mit ein. Andererseits impliziert der Begriff der Handlungskompetenz die Fähigkeit, Situationen gedanklich und real schrittweise zu verändern, einen Ist-Zustand also in einen Soll-Zustand zu transformieren. Diese operative Kompetenz wird im Problemlösen wie in der Regulation der praktischen Handlung wirksam. Sie kann mit unterschiedlichen Medien vollzogen werden, von hochabstrakten Symbolen über Sprache bis hin zu physischen Objekten (vgl. ausführlicher TRAMM 1996).

Für die Entwicklung berufsbezogener Standards und beruflicher Kompetenzmodelle spielt damit der Bezug auf authentische Lebenssituationen eine konstitutive Rolle. Dies erleichtert die Aufgabenkonstruktion insofern durchaus, als die (ggf. unter systematischen Gesichtspunkten modifizierte) berufliche Realsituation ein geeigneter Bezugspunkt hierfür zu sein verspricht (vgl. hierzu den Beitrag von REETZ in diesem Heft). Zugleich wird man sich hier allerdings (wie im Lernfeldansatz auch) vor einer unreflektiert-utilitaristischen Fixierung auf den Status-quo beruflicher Praxis hüten müssen, will man die Lernenden nicht unter Preisgabe des Bildungsauftrages der Berufsschule von der Chance zur gedanklichen Durchdringung ihrer beruflichen Praxis und damit auch von der Chance zur (Mit-)gestaltung oder auch nur von der Möglichkeit der flexiblen Anpassung der eigenen Kompetenz an veränderte Bedingungen abschneiden (vgl. TRAMM 2003). Unter diesem Aspekt wird nun allerdings auch deutlich, dass auf Verständnis basierende Orientierungs- und Handlungskompetenz die Frage nach den Bildungsinhalten in einem veränderten Sinne neu stellt: Es ist dies die Frage

¹ Analog gilt dies auch für Wahrnehmungsleistungen, Interpretationsleistungen (Deutungen) oder Urteilsleistungen (Wertungen) die im Begriff der Orientierungsleistung zusammengefasst werden können.

nach den kategorialen Zugängen zu einem (beruflichen) Lebens- und Handlungsbereich, die Frage nach den Schlüsselproblemen, den grundlegenden Denkfiguren und den zentralen Begriffen für das Verständnis der ökonomischen oder der technischen Perspektive. Der direkte Weg von der beruflichen Situation zur arbeitsanalogen Testaufgabe wäre also eine unzulässige Abkürzung. Das curriculare Relevanzproblem kann hier keinesfalls dispensiert werden; es gilt vielmehr unter Validitätsaspekten danach zu fragen, für welche Orientierungsleistungen und Handlungen qualifiziert werden soll, welche Probleme, Konzepte und Begriffe, welche operativen Fähigkeiten und welches Faktenwissen diesen Leistungen zugrunde liegt.

4 Pragmatischer Rahmen der Itementwicklung

Der Konstruktionsprozess berufsspezifischer Aufgabensätze für ULME III begann im August 2004 mit einer Informationsveranstaltung. Den für die Itemerstellung berufenen Lehrern und Lehrerinnen aus 17 berücksichtigten Berufen wurde erläutert, welche Aufgabenformate verwendet werden dürfen, welche Strategie zur Auswahl der Inhalte empfohlen wird und wie mit Hilfe einer Klassifikationsmatrix das Anspruchsniveau der Aufgaben überprüft werden kann. Der Pilottest sollte Aufgaben für 170 Minuten Bearbeitungszeit umfassen und innerhalb von 3 Monaten fertig sein. Die Pilotierung fand im Dezember 2004 statt. Entsprechend der Auswertungen der Berliner Humboldt Universität wurden die Tests zwischen Januar und März überarbeitet und ergänzt, sodass im April 2005 die Haupttestung in Hamburg stattfinden konnte. Die Lehrer und Lehrerinnen arbeiteten überwiegend einzeln, nur selten formierten sich Teams. Für die Testerstellung erhielten sie stundenweise Entlastung von der Regelarbeit. Während des gesamten Konstruktionszeitraums stand das IBW-Team beratend zur Seite.

Neben diesen knappen personellen und zeitlichen Ressourcen stand die Einschränkung, dass der Test schriftlich, in standardisierter Form, im Zeitrahmen von 90 Minuten und unter Verwendung geschlossener Antwortalternativen durchgeführt und von fachfremden Personen ausgewertet werden sollte.

All dies machte es erforderlich, die Komplexität der Aufgabenstellung zu reduzieren und eine Bearbeitungsstrategie zu vereinbaren, die ein Optimum an Vergleichbarkeit, Transparenz und Rationalität sichert, die aber doch auch und vor allem unter den vorliegenden Rahmenbedingungen für alle Beteiligten machbar war. Hieraus ergaben sich 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. Motivationale und volitionale Aspekte und auch berufsübergreifende Kompetenzen wurden z. T. mit anderen Instrumenten erfasst.
- Der Test sollte berufsspezifisch angelegt sein und eindeutig spezifische berufliche Kompetenzen und Wissensbestände zum Gegenstand haben.
- Der Test sollte aus einer größeren Zahl voneinander unabhängiger Einzelaufgaben bestehen; komplexe mehrstufige Aufgabenstellungen waren so kaum möglich.

- In den Aufgaben sollte nicht nur Wissen punktuell abgefragt, sondern 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 Niveaustufen sollte über die verschiedenen Berufe hinweg vergleichbar sein.

Hieraus ergab sich, dass die Sicherung der curricularen Validität in Bezug auf Inhaltsbereiche und in Bezug auf das geforderte kognitive Anspruchsniveau zu berücksichtigen war.

Mit dem Kriterium der Validität wird danach gefragt, ob ein Test tatsächlich das erfasst, was mit ihm erfasst werden soll. Curriculare Validität bezieht sich mithin auf die Frage, ob mit einem Test tatsächlich die spezifischen Fähigkeiten, Kenntnisse und Fertigkeiten erfasst werden, die dem Lehrplan gemäß in einem Bildungsgang erworben werden sollen.

Um diese Frage zu beantworten, ist es zunächst wieder sinnvoll nach dem Leistungsniveau zu fragen, das die Schüler zum Schluss der Ausbildung erreicht haben sollen. Die Validität wäre hier verfehlt, wenn ein Test nur die Reproduktion von Wissen in einem Bereich erfassen würde, wo der Lehrplan Verständnis, Anwendung oder kritische Reflexion erfordert.

Ein zweiter Aspekt curricularer Validität bezieht sich darauf, ob mit dem Test die verschiedenen relevanten Inhaltsbereiche vollständig und mit dem Gewicht erfasst worden sind, das ihnen jeweils im Lehrplan beigemessen wurde. Die Überprüfung dieser inhaltlichen Dimension der Validität setzt voraus, dass zunächst berufsspezifisch ein Überblick über die verschiedenen Inhaltsbereiche gewonnen wird, wobei die Kategorien weder zu differenziert noch zu grob sein sollten (ca. 10-15 Inhaltsbereiche). Im zweiten Schritt sollte dann eine Gewichtung dieser Bereiche vorgenommen werden und schließlich sollte versucht werden, dass die Verteilung der Items auf diese Inhaltsbereiche der Gewichtung möglichst nahe kommt.

Bei der Identifikation und Gewichtung der Inhaltsbereiche sind grundsätzlich verschiedene pragmatische Vorgehensweisen und Bezugnahmen denkbar:

- Bezug auf den Lehrplan, Bildungsplan bzw. Rahmenlehrplan, insbesondere dann, wenn ein solcher in kompetenzorientierter Form vorliegt und Hinweise auf die zeitliche Gewichtung einzelner Inhalts- bzw. Leistungsbereiche gibt.
- Bezug auf Gliederung und Gewichtung in Stoffplänen bzw. in vorliegenden Kammerprüfungssätzen. Bei dieser Vorgehensweise besteht die starke Gefahr, dass der Bezug auf anzustrebende Kompetenzen weitgehend durch die Orientierung an einer fachlichen Stoffsystematik ersetzt wird.
- Denkbar wäre auch eine Strategie, eine Inhaltssystematik aus einer Systematik relevanter beruflicher Situationen abzuleiten oder zumindest in der Gewichtung hierauf zu beziehen. Ausgangspunkt hierfür müsste dann die Identifikation einer begrenzten Anzahl beruflicher Orientierungs- und Handlungssituationen sein, die im Zentrum des jeweiligen

Berufsbildes stehen. Hierfür wäre auch die Bezugnahme auf die Ausbildungsordnung sinnvoll, ggf. auch die Rücksprache mit betrieblichen Experten.

Obgleich diese Vorgehensweisen bei der einführenden Informationsveranstaltung ausführlich erläutert wurden und eine Handreichung an die Aufgabenersteller verteilt wurde, musste in ULME III aus zeitlichen Gründen im wesentlichen darauf gesetzt werden, dass die als Experten eingebundenen Lehrer auf der Grundlage der je spezifischen Situation in den 17 einbezogenen Berufen eine vertretbare inhaltliche Gewichtung vornahmen. Im Entwicklungsprozess und insbesondere auch in der Überarbeitungsphase auf Grundlage der Pilotierungsergebnisse wurde der Aspekt der inhaltlichen Gewichtung von Seiten der wissenschaftlichen Begleitung in besonderem Maße beachtet. Dennoch bleibt der Aspekt der curricularen Validität in dieser Hinsicht sicher problematisch, ihm muss bei der notwendigen Weiterentwicklung der Instrumente besonderes Gewicht beigemessen werden.

Das besondere Augenmerk des IBW-Teams galt im Entwicklungsprozess dem kognitiven Anspruchsniveau der Aufgabensätze. Die strategische Schlüsselrolle zur Verbindung pragmatischer Möglichkeiten und konzeptueller Ansprüche spielte die im Folgenden vorzustellende Klassifikationsmatrix zur Differenzierung des inhaltlichen und formalen Anspruchsniveaus von Testaufgaben.

Leistung		Reproduzieren	Anwenden/ Verstehen	Kritisieren/ Reflektieren
Wissensarten				
Fakten		A B		
Konzepte		A B		
Prozeduren		A B		

Abb. 2: Klassifikationsmatrix für ULME

Die Matrix ist dadurch charakterisiert, dass die **kognitive Leistung** und die in beruflichen Leistungen repräsentierten **Wissensarten** aufeinander bezogen werden. Somit werden zwei wesentliche Aspekte von Kompetenz gleichzeitig berücksichtigt. Die Matrix diene also bereits bei der Aufgabenerstellung zur Orientierung mit dem Ziel, in den verschiedenen Inhaltsbereichen eine angemessene Streuung in Hinblick auf das Anspruchsniveau zu erreichen. Da der Schwerpunkt der Aufgaben im Leistungsbereich „Verstehen/Anwenden“ sowie in den Inhaltsbereichen „Konzeptwissen“ und „prozedurales Wissen“ liegen sollte, dienten folgende Werte als Orientierung, d. h. der Test sollte Aufgaben enthalten, die annäherungsweise diese Niveaueverteilung erreichen:

kognitive Leistung	Soll-Wert	Wissensarten	Soll-Wert
Reproduktion:	30 %	Faktenwissen:	20 %
Verstehen/Anwenden:	60 %	Konzeptwissen:	50 %
Kritisieren/Reflektieren	10 %	Prozeduren:	30 %

Bevor im folgenden die Klassifikationsmatrix erläutert wird, ist der hier skizzierte Arbeitskontext zusammenfassend dargestellt:

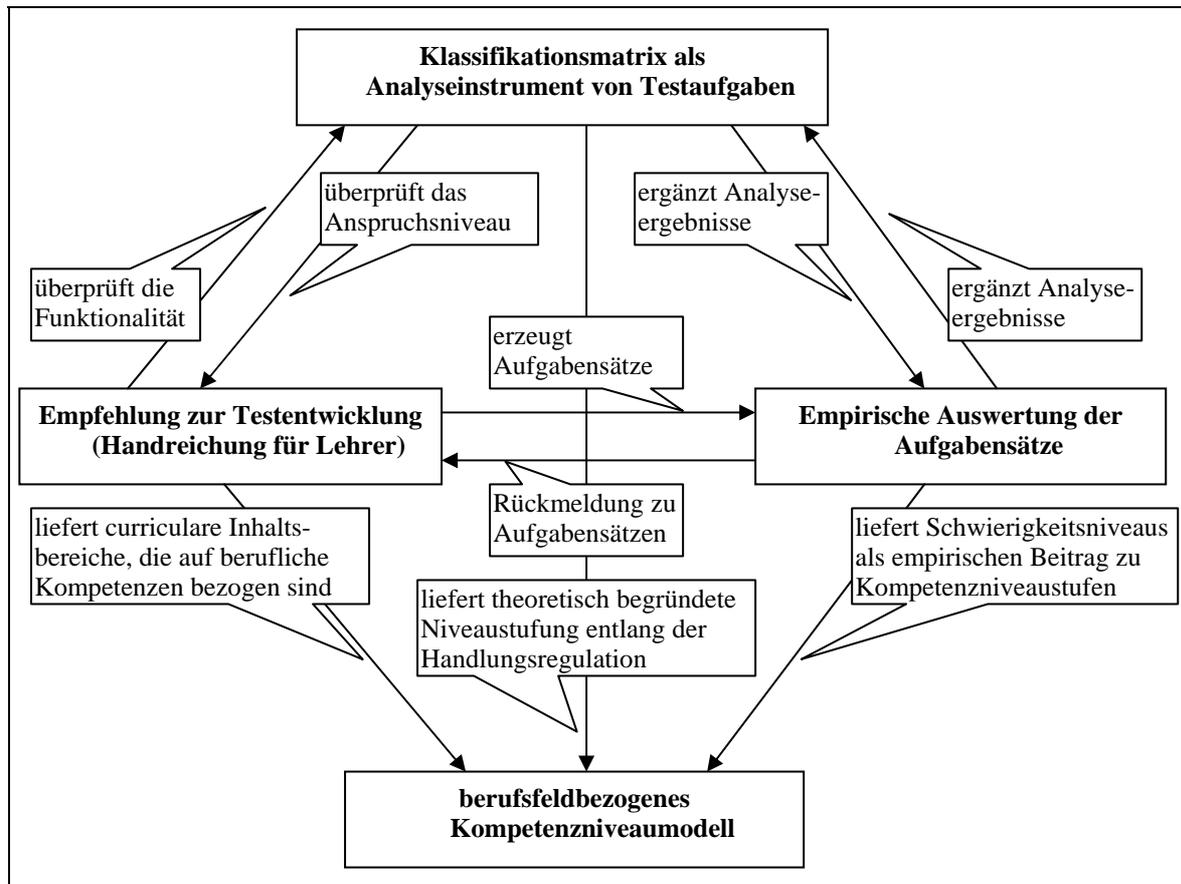


Abb. 3: Zusammenfassende Darstellung des Arbeitskontextes

5 Systematischer Kontext der Itementwicklung – die Aufgabenklassifikationsmatrix

5.1 Funktion und Struktur der Matrix

Die im vorigen Abschnitt dargestellte Matrix ist das Ergebnis eines Entwicklungsprozesses. Bereits zur Erstellung des Fachtestes Wirtschaft und Verwaltung im Rahmen von ULME II wurde eine Klassifikationsmatrix verwendet, wobei dieser erste Versuch mehrfach überarbeitet wurde. Die im Kontext von ULME III verwendete und hier dargestellte Matrix musste auch dem Anspruch der Eindeutigkeit, Plausibilität und Handhabbarkeit genügen, da die beschriebenen knappen Ressourcen die Bereitschaft, ein solches Analyseinstrument einzusetzen, einschränkten. Es galt nämlich, jede einzelne Aufgabe in Hinblick auf das kognitive Leistungsniveau und das in der Leistung repräsentierte Wissen zu analysieren und entsprechend einer Zelle der Matrix zuzuordnen. Auf diese Weise konnten Überarbeitungsempfehlungen entweder zu einzelnen Aufgaben (häufig ist es möglich, durch Änderungen im Aufgabentext oder bei den Antwortalternativen ein höheres Anspruchsniveau zu erzielen) oder

zum gesamten Aufgabensatz gegeben werden, wenn das repräsentierte Niveau stark von den Soll-Werten abwich. Im späteren Verlauf der Testentwicklung ging es darum, aus einem Aufgabenpool, der sich aufgrund der Pilotierungsergebnisse als funktional erwies, diejenigen Aufgaben herauszusuchen, die dem angestrebten Niveau entsprachen, oder auch umgekehrt, einen Mangel an Aufgaben eines bestimmten Niveaus aufzudecken und entsprechend nachzuliefern.

Die Entwicklung der Struktur der Matrix erfolgte in Anlehnung an ANDERSON & KRATHWOHL (2001). Diese Autoren haben die BLOOM'sche Lernzieltaxonomie dahingehend revidiert, dass sie die Wissensdimension systematisch ausdifferenzierten und von der kognitiven Prozessdimension trennten:

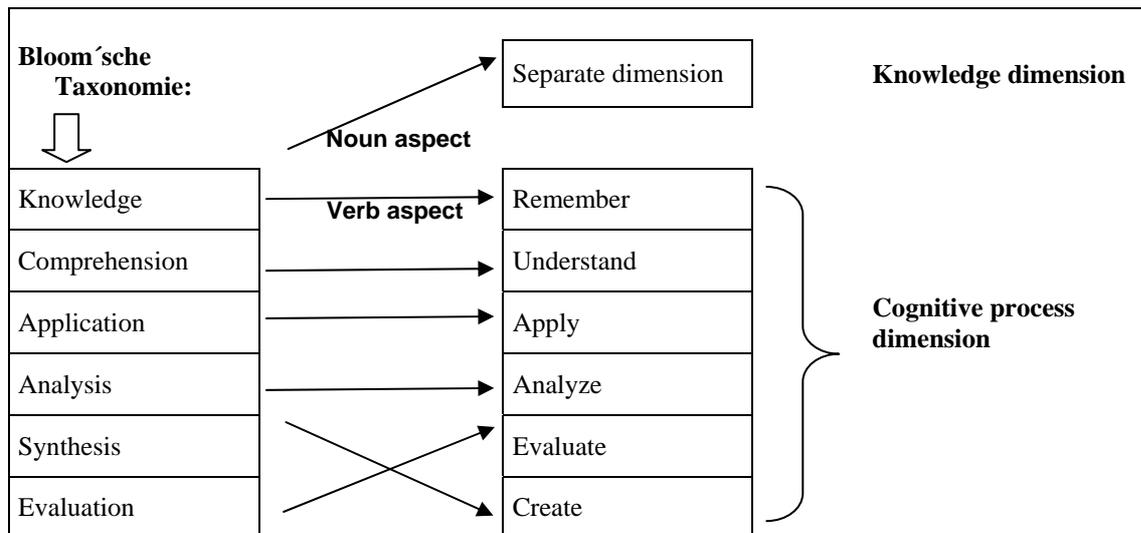


Abb. 4: Summary of the structural changes from the original framework to the revision (ANDERSON & KRATHWOHL 2001, 310)

Das vollständige von ANDERSON & KRATHWOHL entwickelte Tableau enthält neben den sechs Verhaltensdimensionen vier Wissenskategorien mit jeweils zwei bis drei Unterkategorien. Um den Umfang des Analyseinstruments einzugrenzen ohne dabei aber wesentliche Aspekte, die in unserem Kontext relevant sind, zu verlieren, fassten wir Kategorien zusammen oder schlossen andere aus.

Mit Blick auf den Anwendungsbereich der Klassifikationsmatrix entschieden wir uns, auf die Kategorie „meta-cognitive-knowledge“ zu verzichten. Bestandteil des ULME Testinstrumentariums war in allen drei Erhebungen ein Schülerfragebogen, mit dessen Hilfe u. a. Lerngewohnheiten (Lernstrategien und Reflexionsvermögen über den Lernfortschritt) erfasst werden. Mit den berufsbezogenen Fachtests sollte dieser Bereich explizit nicht erfasst werden.

Die Bezeichnung der Verhaltenskategorien „analyze, evaluate, create“ änderten wir in „Kritisieren/ Reflektieren“. Dies hängt mit den in ULME favorisierten Aufgabentypen zusammen. Es ist nur schwer (vielleicht überhaupt nicht) möglich, in einem gebundenen Aufgabenformat einen Sachverhalt so zu repräsentieren, dass die Antwort eine neue, vom Schüler erarbeitete

Lösung darstellt; deshalb verzichten wir an dieser Stelle darauf, wesentliche Aspekte kognitiver Leistung zu erfassen, die vor allem in die Kategorie `create` fallen würden. Es sollte aber möglich sein, geschlossene Aufgaben so zu erstellen, dass die Schüler über Strukturen, Zusammenhänge, Kriterien o.ä. kritisch reflektieren und eigene vorhandene oder in der Situation selbst erst zu entwickelnde Kriterien verwenden müssen, um die richtige Antwort zu wählen.²

Im Folgenden werden zunächst die Wissens- dann die Leistungskategorien näher beschrieben. In dem mit diesem Beitrag verlinkten Artikel von [Wiebke HOFMEISTER](#) sind die Kategorien ausführlicher beschrieben. Dort finden sich auch Anwendungsbeispiele aus ULME II.

5.2 Die Inhalts- oder Wissensdimension der Aufgaben

Ziel ist, das Wissen, das mit den Aufgaben erfasst werden soll, systematisch zu beschreiben, um Aussagen über das Niveau des in der Leistung repräsentierten Wissens machen zu können. Leitfrage bei der Aufgabenanalyse ist also stets: Was ist die Intention der Frage? Wie ist das Wissen strukturiert, das zur Beantwortung der Frage herangezogen wird?

In der folgenden Tabelle sind alle Kategorien charakterisiert:

Wissenskategorie	Unterkategorie	Charakterisierung
Faktenwissen	terminologisch	<ul style="list-style-type: none"> – Aussagen über spezifische Bezeichnungen, Namen, Symbole – Definitionen, i. S. v. Worterklärungen durch Umschreibung – Aussagen zu sprachlichen und symbolischen Konventionen
	empirisch	<ul style="list-style-type: none"> – Aussagen über Normen sowie über die Existenz, Quantität, Qualität und Zusammensetzung von Phänomenen – Die Aussagen über Phänomene sind falsifizierbar
Konzeptwissen	Klassifikationen und Kategorien	<ul style="list-style-type: none"> – Sacherklärungen (Bildung von Oberbegriffen, multiple Diskrimination) – Begriffshierarchien ermöglichen Kategorien übergreifende Schlussfolgerungen
	Verallgemeinerungen und Prinzipien	<ul style="list-style-type: none"> – Aussagen zu Prozessen und Zusammenhängen zwischen empirischen Phänomenen – Aussagen zu Ursache-Wirkung-Beziehungen (juristisch auch Ursache-Folge-Beziehungen), stochastischen Beziehungen...
Prozedurales Wissen	Fertigkeiten und Verfahren	<ul style="list-style-type: none"> – Zur Bewältigung berufstypischer Handlungssituationen – Lösungsweg und Handlungsergebnis sind bekannte Schemata

² Eine ausführliche Beschreibung der Entwicklung der Matrix und ihres Begründungszusammenhangs enthält der in diesem Heft veröffentlichte Beitrag von [Wiebke HOFMEISTER: Erläuterung der Klassifikationsmatrix zum ULME-Kompetenzstufenmodell](#). Dort sind die Unterkategorien beschrieben und die Zusammenfassungen erläutert. Im Anhang dieses Beitrags befindet sich die Matrix mit kurzen Erläuterungen und prototypischen Beispielen.

	Techniken und Methoden	<ul style="list-style-type: none"> - Zur Bewältigung beruflicher Handlungssituationen, für die keine Standardlösung passt - Lösungswege und Anordnung der Handlungsschritte müssen erst gefunden werden.
--	------------------------	--

Im Unterschied zum Konzeptwissen können beim **Faktenwissen** Wissenselemente isoliert betrachtet werden. Einzelne Informationseinheiten lassen sich ausmachen, die ihre eigene Bedeutung haben. Sie können beschrieben werden, ohne dass Vergleiche, Zusammenhänge oder Abgrenzungen zu anderen Informationseinheiten hergestellt werden müssen. Die Unterkategorie „**terminologisches Wissen**“ ist wichtig, weil die Aneignung und Verwendung spezifischer Terminologie unerlässlicher Teil des Kompetenzerwerbs innerhalb eines Sachgebietes ist. Terminologien erleichtern eine präzise und unvermissverständliche Ausdrucksweise. Der Unterschied zum **empirischen Wissen** besteht darin, dass hier nicht nur der Name, sondern auch die Ausprägung oder spezifische Eigenschaften eines Phänomens thematisiert werden. Dies sind Erscheinungen, wie sie in der Realität beobachtbar sind, sodass empirische Aussagen falsifizierbar sind. Terminologien sind es nicht; bei ihnen kann es lediglich Verstöße gegen sprachliche Konventionen geben.

Das wesentliche Unterscheidungsmerkmal zwischen Aussagen zum Faktenwissen und Aussagen zum **Konzeptwissen** ist, dass bei letzteren immer eine Erklärungsintention verfolgt wird. Es handelt sich um Aussagen, die eher auf die Ebene der Begriffsbildung abheben und die die Intention haben, Begriffe voneinander abzugrenzen, entweder horizontal (multiple Diskrimination) oder vertikal (Bildung von Oberbegriffen). Werden Attribute eines Begriffs lediglich beschrieben, um den Begriff selbst zu verdeutlichen, dann handelt es sich um Faktenwissen. Werden dagegen Attribute beschrieben, um den Begriff von benachbarten Begriffen oder Kategorien abzugrenzen oder zu unterscheiden, dann handelt es sich um **Konzeptwissen auf der Ebene von Klassifikationen und Kategorien**. Die zweite Unterkategorie erklärt Prozesse und Zusammenhänge zwischen Begriffen bzw. differenzierteren Wissensstrukturen. Symbolisch gesprochen besteht der Unterschied zwischen den Unterkategorien des Konzeptwissens darin, dass bei den Kategorien und Klassifikationen die Ebenen einer Hierarchie oder die Knoten einer Netzstruktur thematisiert sind, während bei den **Verallgemeinerungen** und **Prinzipien** die Verbindungslinien erklärt werden. In diese zweite Kategorie fallen auch Aussagen zu wissenschaftlichen Theorien und zu komplexen Modellen (bei ANDERSON & KRATHWOHL ist dies eine eigene Unterkategorie). Dies kommt in unserem Anwendungszusammenhang jedoch nur selten vor. Meist geht es um Ausschnitte aus Modellen oder um aus Theorien generiertes Wissen und das ist dann mit Verallgemeinerungen und Prinzipien treffend umschrieben.

Das **prozedurale Wissen** hebt sich von den anderen Kategorien ab, da es hier um Handlungen geht. Diese Handlungen, Handlungssequenzen oder -abfolgen können jedoch so abstrakt und zusammengefasst in Aufgaben thematisiert werden, dass es nicht mehr um den Vollzug der Handlung selbst geht, sondern die Gesamthandlung in andere Kontexte eingebettet ist. Dann wäre die Handlung als Konzept thematisiert. Um hier ein pragmatisches Unterscheidungsmerkmal an der Hand zu haben, legten wir fest, dass eine Handlung aus mindestens

zwei Handlungsschritten besteht, die in einer Aufgabe vollzogen oder gedanklich nachvollzogen werden müssen. Ebenso ist zu überprüfen, ob ein Handlungsergebnis Teil der Aufgabe ist.

Der Unterschied zwischen den **Unterkategorien des prozeduralen Wissens** lässt sich auch mit dem Unterschied zwischen Algorithmen und Heuristiken umschreiben. Während bei Algorithmen der Lösungsweg und das Ziel bekannt sind, müssen bei Heuristiken erst die richtigen Verfahren zur Lösung, gegebenenfalls auch das Ziel selbst gesucht werden. Aufgaben, die Heuristiken thematisieren, kommen, ganz gleich auf welche kognitive Leistung sie sich beziehen, nur selten vor.

Die Stärke dieser Wissensdimensionen liegt darin, dass sie sich auf einem Kontinuum von konkret zu abstrakt verorten lassen, wobei es durchaus Überlappungen gibt. Wichtig ist, dass die jeweils höhere Kategorie auch höher organisierte Wissensstrukturen darstellt. Dies gilt auch für die jeweiligen Unterkategorien. Die Stufen geben keine Auskunft über Lern- oder Entwicklungsprozesse; jede Stufe kann erreicht werden, ohne dass die davor liegenden Stufen hätten durchlaufen werden müssen. Die Matrix kann jedoch das in den Testaufgaben thematisierte Wissen systematisch in Hinblick auf die Wissensstrukturen hin überprüfen und diese als Niveaustufen beschreiben. Schwierigkeiten bereiten mitunter noch die Abgrenzungen der einzelnen Kategorien, sodass unterschiedliche Personen zu unterschiedlichen Klassifizierungen kommen. Es scheint auch, dass andere Schwierigkeitsparameter (Komplexität, Intransparenz u. ä.) den Blick auf die Wissensstruktur beeinflussen.

5.3 Die kognitive Leistungsdimension der Aufgaben

Die Leistungsdimensionen sind so angeordnet, dass sie von links nach rechts eine ansteigende Komplexität repräsentieren. Sie sind eine aggregierte Form der BLOOM'schen Kategorien, wobei wir bei der Zusammenfassung einem Vorschlag von METZGER (1993) gefolgt sind (näheres dazu im Beitrag von [Wiebke HOFMEISTER](#) in diesem Heft). Die einzelnen Kategorien lassen sich wie folgt charakterisieren:

Leistungsdimension	Charakterisierung
Reproduzieren	<ul style="list-style-type: none"> – Wiedergeben/-erkennen erlernten Wissens – Strukturgleichheit von Lernsituation und Aufgabe anzunehmen – keine situative Einbindung erforderlich – Typisch: nennen, wiedergeben, beschreiben
Verstehen/Anwenden	<ul style="list-style-type: none"> – erlerntes Schema auf neue Situation übertragen (Abstraktionshierarchie) – erlerntes Schema auf strukturell gleiche aber sprachlich anders gefasste Aufgabe beziehen – Assimilation im Sinne PIAGETs – Typisch: erklären, begründen, ermitteln, beurteilen, vergleichen
Kritisieren/Reflektieren	<ul style="list-style-type: none"> – Grenzen von Schemata bewusst reflektieren – Anforderungen an eine Veränderung der Struktur sind zu identifizieren – Anbahnung der Akkomodation im Sinne PIAGETs – Typisch: Lösung beurteilen, Probleme aufzeigen, Leistungsfähigkeit abwägen

Das **Reproduzieren** ist die Fähigkeit, eine gelernte Information, eine Struktur oder ein Schema in einem unveränderten Umfeld wiederzuerkennen oder die Identität der Information mit vorher gelernten Informationen, Strukturen oder Schemata festzustellen. Wichtigstes Abgrenzungsmerkmal zur nächst höheren Kategorie ist, dass beim Reproduzieren die Intention der Aufgabe nicht darin besteht, zu überprüfen, ob ein Sachverhalt auch verstanden wurde. Deswegen wird auch meist auf eine situative Einbindung verzichtet. Es wird nur Wert darauf gelegt, einen bestimmten Inhalt wiederzuerkennen.

Bei der Kategorie **Verstehen/Anwenden** geht es um die Fähigkeit, eine gelernte Information, eine Struktur oder ein Schema mit veränderter Formulierung oder Darstellung (d. h. „sinngemäß“), aber unverändertem inhaltlichen Gehalt abzubilden, d. h. zu zeigen, dass man die Lerninhalte „verstanden“ hat. Es ist ebenso die Fähigkeit gemeint, das Gelernte auf eine strukturell übereinstimmende oder in Bezug auf den sprachlichen Inhalt neuartige Aufgabe zu übertragen. Der Handelnde prüft die vorliegenden Informationen und Umweltbedingungen und realisiert, dass sie in ein bereits gelerntes, also vorhandenes Handlungsschema passen, welches er dann unverändert anwenden kann. Es handelt sich also um Assimilation im PIAGETSchen Sinne. Bei der Aufgabenerstellung ist auf eine situative Einbindung des Sachverhalts zu achten, da sonst keine Transferleistung stattfinden kann. Es muss sich um echte Situationsaufgaben handeln, da sonst die Gefahr bestünde, doch nur eine Reproduktionsleistung auszulösen (vgl. den Beitrag von REETZ in diesem Heft).

Mit **Kritisieren/Reflektieren** ist die Fähigkeit gemeint, einen Gegenstand oder Sachverhalt umfassend und systematisch (d. h. anhand von Kriterien) hinsichtlich der ihn bestimmenden relevanten Komponenten zu untersuchen. Die Lösung des Problems bedarf einer spezifisch neuen Kriteriumsstruktur. Das Wissen über einen Gegenstand oder einen Sachverhalt stößt aufgrund eines Perspektiv- oder Situationswechsels an Grenzen und muss in eine andere, passende Struktur transferiert werden. Gemeint ist auch die Fähigkeit, einzelne Teile zu einem umfassenden höheren Ganzen zusammenzufügen oder bestimmtes Material (z.B. Sachverhalte, Situationen, Aussagen, Vorschläge, Konzepte, Theorien, Verfahren, usw.) systematisch anhand von stimmigen Kriterien zu bewerten. Zur Beurteilung gehört, dass die Kriterien selbständig gewichtet werden und eine eigenständige Einschätzung getroffen wird. Werden dagegen sowohl Kriterien als auch deren Gewichtung vorgegeben, handelt es sich um „Anwenden“.

Auch bei dieser Dimensionierung der Leistungen handelt es sich um eine Niveaustufung von links nach rechts. Wesentliche kognitive Leistungen können erfasst und unterschieden werden. Problematisch ist, dass nur wenige Aufgaben das Niveau Kritisieren/Reflektieren erreichen. Möglicherweise interpretieren wir die hier zu erbringende Leistung zu anspruchsvoll. Vielleicht suggeriert auch das gebundene Aufgabenformat, dass keine eigenständige kognitive Leistung möglich sei. Diese Kategorie muss weiter präzisiert werden, sodass sie sich zur Analyse der in ULME typischen Aufgabenformate eignet.

Bei der Unterscheidung der kognitiven Leistungsdimensionen werden implizit Annahmen über die erforderlichen vorgelagerten Lernprozesse getroffen. Dies sind immer subjektive

Einschätzungen desjenigen, der die Aufgaben analysiert. Es wäre zu überlegen, ob mit der Methode des lauten Denkens die Denkprozesse der Schüler beobachtet und so die Leistungsdimensionen empirisch überprüft werden könnten.

6 Erfahrungen, erste Auswertungsergebnisse und Perspektiven der weiteren Arbeit

Die Phasen der Itemerstellung im Rahmen von ULME III sind abgeschlossen. Die Untersuchung wurde im Frühjahr 2005 durchgeführt. Mit den empirischen Ergebnissen seitens der Humboldt Universität zu Berlin kann frühestens zum Winter gerechnet werden. Sobald die empirischen Ergebnisse aus ULME II vorliegen (im Spätsommer des Jahres), können diese verglichen werden mit den Ex-ante-Analysen, die mit Hilfe der oben dargestellten Matrix vorgenommen wurden. Im Rahmen von ULME II gab es sechs berufsfeldbezogene Aufgabensätze, wobei nur der Fachtest Wirtschaft und Verwaltung durch das IBW-Team erstellt wurde. Die anderen Tests wurden ohne Klassifikationsmatrix entworfen. Dennoch haben wir die Aufgaben mit Hilfe dieses Rasters klassifiziert, um diese Verteilung mit den Ergebnissen der statistischen Analyse auf der Grundlage des Rasch-Modells vergleichen zu können. Die folgende Tabelle zeigt, wie sich die Aufgaben auf die Niveaustufen verteilen:

Wissensart	Wirtschaft/ Verwaltung		Sozial- wesen		Gesund- heit		Ernährung		Metall- technik		Elektro- technik		Soll- %
	N	%	N	%	N	%	N	%	N	%	N	%	
Faktenwissen	17	20,2	12	50,0	17	65,4	17	58,6	17	68,0	2	16,7	20
Konzeptwissen	60	71,4	12	50,0	7	26,9	3	10,3	7	28,0	7	58,3	50
prozedurales Wissen	7	8,3	0	0,0	2	7,7	9	31,0	1	4,0	3	25,0	30
Summe	84	100,0	24	100,0	26	100,0	29	100,0	25	100,0	12	100,0	100
kognitive Leistung	N	%	N	%	N	%	N	%	N	%	N	%	Soll- %
Reproduktion	18	21,4	10	41,7	17	65,4	15	51,7	22	88,0	2	16,7	30
Anwendung/ Verstehen	66	78,6	14	58,3	9	34,6	14	48,3	3	12,0	10	83,3	60
Kritik/ Reflektieren	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	0	0,0	10
Summe	84	100,0	24	100,0	26	100,0	29	100,0	25	100,0	12	100,0	100

Der Fachtest Wirtschaft und Verwaltung sowie der Elektrotechnik-Test erreichen ein gutes Ergebnis in Hinblick auf die Sollwerte. Obgleich auch diese Tests die Niveaustufe Kritisieren/Reflektieren nicht erreichen (mögliche Gründe vgl. voriger Abschnitt), liegt hier der Schwerpunkt im Bereich Anwenden/Verstehen. Die prozeduralen Aufgaben aus den Tests Gesundheit, Ernährung und Elektrotechnik sind zum größten Teil Rechenaufgaben. Jeder anzuwendende Rechenalgorithmus führt dazu, dass eine Aufgabe als Anwenden im Bereich

Verfahren/Fertigkeiten eingestuft wird. Hier zeigt sich eine bereits erwähnte Schwäche der Matrix: Das erfasste Anspruchsniveau der Aufgaben impliziert keine Aussage in Hinblick auf das Schwierigkeitsniveau. Gerade die hier klassifizierten Rechenaufgaben waren eher leicht und erforderten kein spezielles Fachwissen (zu dieser Problematik vgl. auch den Beitrag von SEEBER in diesem Heft).

Vier der Tests prüften zum großen Teil Faktenwissen auf dem Niveau Reproduzieren. Dies ist u. E. ein Indiz dafür, dass die Klassifikationsmatrix bereits bei der Itemerstellung eingesetzt werden sollte, um solche einseitigen Verteilungen zu vermeiden.

Die unterschiedlich hohe Anzahl der Aufgaben ist darauf zurückzuführen, dass beim Test Wirtschaft und Verwaltung häufig Wahr-Falsch-Aufgaben vorkamen, bei denen jede Aussage (meist 4 verschiedene) einzeln bewertet wurde, da durchaus unterschiedliche Niveaustufen angesprochen wurden. Eine Multiple Choice-Aufgabe, bei der eine von vier Antwortalternativen zu wählen ist, taucht in diesem Test nur einmal auf. Auch im Elektrotechniktest wurde vom Multiple Choice-Format abgewichen (viele Zuordnungsaufgaben), während in den anderen Tests ausschließlich mit Multiple Choice-Aufgaben gearbeitet wurde. Möglicherweise geht die hohe Auswertungsökonomie dieser Aufgaben zu Lasten der Möglichkeit, mit anderen gebundenen Aufgabenformaten ein anspruchsvolles Leistungsniveau zu erreichen.

Aussagen zum **Schwierigkeitsniveau** der Aufgaben werden die Auswertungen der Humboldt Universität zu Berlin erlauben. Mit den dort angewandten Methoden aus der probabilistischen Testtheorie werden entsprechend der Wahrscheinlichkeit richtiger Antworten Aussagen zur Reliabilität des Tests und zum Schwierigkeitsgrad der Items gemacht. Die Schülerfähigkeiten definieren sich über die Aufgabenschwierigkeiten, die wiederum von der Anzahl richtiger Lösungen in der Stichprobe abhängen. Die Leistung also, eine Aufgabe eines bestimmten Schwierigkeitsgrades zu lösen, entspricht der Leistung des Schülers oder der Schülerin bei dieser Aufgabe eine 50-prozentige Trefferquote zu erbringen. Übersteigt eine so geschätzte Fähigkeit die Schwierigkeit der Aufgabe, so wird der Schüler oder die Schülerin die Aufgabe um so eher lösen, je größer die Differenz zwischen Fähigkeits- und Schwierigkeitsparameter ist, bzw. umgekehrt wird die Aufgabe umso eher nicht gelöst, je weiter das Fähigkeitsniveau hinter dem Schwierigkeitsgrad zurückbleibt (LEHMANN et. al. 2005, 20).

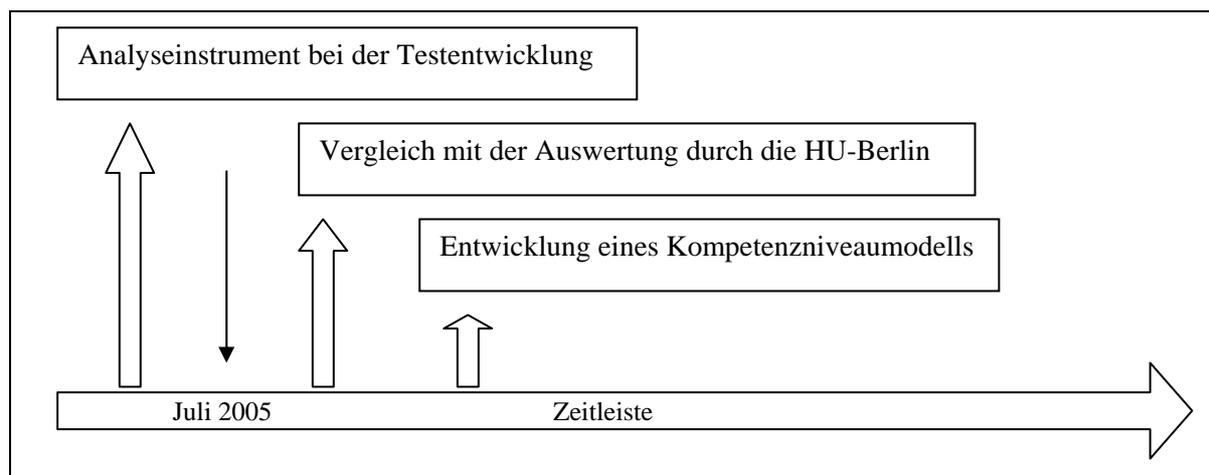
Diese ermittelten Fähigkeitsniveaus diskriminieren in gute und schlechte Schüler in dem Sinne, dass gute Schüler eher in der Lage waren, die Aufgaben zu lösen und schlechte, die dies eher nicht geschafft haben. Die Ergebnisse können keine Aussagen über die kognitiven Leistungen bezogen auf bestimmte Inhaltsbereiche treffen. Die empirischen Untersuchungsergebnisse wären somit zu ergänzen, um die mit Hilfe der Klassifikationsmatrix ermittelte Einschätzungen, die aus theoretischer Perspektive die Aufgabensätze analysieren.

Dabei wäre dann der Frage nachzugehen, ob die empirisch ermittelten Niveaustufen, welche Schwierigkeitsniveaustufen repräsentieren, kompatibel sind mit den in der Klassifikationsmatrix vorgenommenen Abstufungen kognitiver Anspruchsniveaus – oder anders formuliert: *Welche weiteren oder anderen Merkmale als die in der Klassifikationsmatrix zu Grunde gelegten Dimensionen determinieren die Schwierigkeit einer Aufgabe?* Hierzu wäre es sinnvoll,

die Aufgaben einer Schwierigkeitsstufe itemweise nach den begangenen Fehlern zu analysieren, um Kompetenzdefizite und Problembereiche identifizieren zu können (HELMKE & HOSENFELD 2003, 2). Es ist dann zu überprüfen, welche weiteren Dimensionen von Kompetenz hier berührt werden oder ob die Probleme mit der Aufgabenstruktur oder möglicherweise mit dem Aufgabentypus zusammenhängen. Antworten auf diese Fragen könnten erneut zu einer Überarbeitung der Klassifikationsmatrix führen bzw. könnten Anlass geben, die Methoden für die Analyse der ULME-III-Aufgabensätze zu variieren, mit dem Ziel Kompetenzniveaus beschreiben zu können, die empirisch validiert und theoretisch begründet sind.

Wenn dies gelänge, dann könnte – und das wäre die zweite Zielperspektive im Rahmen des ULME-Projekts – auf der Grundlage der Rationale der Testentwicklung (pragmatische Handlungsempfehlung für die Entwicklung von Tests) und des Klassifizierungsrasters (theoretisch begründetes Analyseinstrument des Anspruchsniveaus von Aufgaben) und evaluierter Leistungsniveaus (empirisch ermittelt aufgrund bewährter Aufgabensätze) ein **Kompetenzniveaumodell** für den berufsbildenden Bereich entwickelt werden.

Aus der folgenden Darstellung ist ersichtlich, auf welchem Arbeitsstand (dünner Pfeil nach unten) wir uns im Rahmen des ULME-Projektes befinden, und welche Arbeitsphasen noch vor uns liegen:



Literatur

AEBLI, H. (1980): Denken: Das Ordnen des Tuns, Band I: Kognitive Aspekte der Handlungstheorie. Stuttgart: (Klett-Cotta).

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.

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

- CHOMSKY, N. (1970): Aspekte der Syntaxtheorie. Berlin.
- DÖRNER, D. (1979): Problemlösen als Informationsverarbeitung. 2. Aufl., Stuttgart: Kohlhammer.
- ERPENBECK, J./ VON ROSENSTIEL, L. (2003): Einführung. In: dieselben (Hrsg.): Handbuch Kompetenzmessung. Stuttgart: Schäffer-Poeschel, IX- XL.
- FREIE UND HANSESTADT HAMBURG (2004): Schulleistungstest LAU 11: Hamburg legt Leistungsbilanz für die Sekundarstufe ! vor. (Mitteilung der Staatlichen Pressestelle vom 17. November 2004).
- HACKER, W.(1998): Allgemeine Arbeitspsychologie. Bern: Huber.
- HELMKE, A./ HOSENFELD, I. (2003): Vergleichsarbeiten – Standards – Kompetenzstufen: Begriffliche Klärung und Perspektiven für VERA. Manuskript vom 27. August 2003. http://www.landesschulamt-berlin.de/gvgl/vera%5Chelmke-ua_vera_standards_kompetenzmodelle.pdf (28.07.2005).
- HOFMEISTER, Wiebke (2005): Erläuterung der Klassifikationsmatrix zum ULME-Kompetenzstufenmodell. In: bwp@ – Berufs- und Wirtschaftspädagogik online, Heft 8/2005. http://www.bwpat.de/ausgabe8/hofmeister_bwpat8.shtml
- 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, Motivation und Einstellungen zu Beginn der beruflichen Ausbildung. Hrsg. v. Freie und Hansestadt Hamburg, Behörde für Bildung und Sport, Amt für Schule. Hamburg.
- METZGER, C./WAIBEL, R./HENNING, C./ HODEL, M./ LUZI, R. (1993): Anspruchsniveau von Lernzielen und Prüfungen im kognitiven Bereich. IWP St. Gallen.
- NEUWEG, G. (1999): Könnerschaft und implizites Wissen. Münster u.a.: Waxmann.
- ROTH, H. (1972): Pädagogische Anthropologie. Band II: Entwicklung und Erziehung. Grundlagen einer Entwicklungspädagogik. Hannover: Schroedel.
- REETZ, L. (2005): Situierete Prüfungsaufgaben. Die Funktion von Situationsaufgaben in Abschlussprüfungen des Dualen Systems der Berufsausbildung. In: bwp@ – Berufs- und Wirtschaftspädagogik online, Heft 8/2005. http://www.bwpat.de/ausgabe8/reetz_bwpat8.shtml
- SEEBER, S. (2005): Zur Erfassung und Vermittlung berufsbezogener Kompetenzen im teilqualifizierenden Bildungsgang „Wirtschaft und Verwaltung“ an Hamburger Berufsfachschulen. In: bwp@ – Berufs- und Wirtschaftspädagogik online, Heft 8/2005. http://www.bwpat.de/ausgabe8/seeber_bwpat8.shtml

STRAKA, G. A. (2005): Die KMK-Handreichungen zur Erarbeitung von Rahmenlehrplänen – eine kritische Reflexion zum zehnten Jahrestag. In: bwp@ – Berufs- und Wirtschaftspädagogik online, Heft 8/2005.

http://www.bwpat.de/ausgabe8/straka_bwpat8.shtml

TRAMM, T. (1996): Lernprozesse in der Übungsfirma. Rekonstruktion und Weiterentwicklung schulischer Übungsfirmenarbeit als Anwendungsfall einer evaluativ-konstruktiven und handlungsorientierten Curriculumstrategie. Habilitationsschrift Göttingen 1996.

http://www.ibw.uni-hamburg.de/personen/mitarbeiter/tramm/texte_tt/Habil.pdf

TRAMM, T.: Prozess, System und Systematik als Schlüsselkategorien lernfeldorientierter Curriculumentwicklung. In: bwp@ – Berufs- und Wirtschaftspädagogik online, Heft 4/2003.

http://www.ibw.uni-hamburg.de/bwpat/ausgabe4/tramm_bwpat4.html

VOLPERT, W. (1992): Wie wir handeln – was wir können. Ein Disput als Einführung in die Handlungspsychologie. Heidelberg: Asanger.

WEINERT, F. E. (1999): Concepts of competence. Neuchatel: DeSeCo.

WEINERT, F. E. (2001): Vergleichende Leistungsmessungen in Schulen – eine umstrittene Selbstverständlichkeit. In: WEINERT, F. E. (Hrsg.): Leistungsmessungen in Schulen. Weinheim und Basel: Beltz, 17-31.

WINERIP, M. (2005): Study Great Ideas, but Teach to the Test. In: The New York Times vom 13. Juli 2005. <http://www.nytimes.com/2005/07/13/education/13education.html> (16.07.2005).