

SASCHA BERNHOLT, ILKA PARCHMANN UND MICHAEL L. COMMONS

Kompetenzmodellierung zwischen Forschung und Unterrichtspraxis

Modelling Scientific Competence between Research and Teaching Practice

Zusammenfassung

Die aktuellen bildungspolitischen Reformen stellen einen Paradigmenwechsel in Richtung einer *outcome*-Orientierung des Bildungssystems dar. Neben der Setzung von Bildungsstandards als normative Zielvorgaben durch die KMK wird in der Entwicklung von Kompetenzmodellen ein weiterer Schritt zur Umsetzung dieser Richtungsänderung gesehen. Kompetenzmodelle sollen eine Verbindung zwischen den intendierten Zielen und den tatsächlich erbrachten Schülerleistungen herstellen und so im Sinne eines Monitorings zur Qualitätssicherung und Qualitätsentwicklung im Bildungssystem beitragen. Aus der Forschungsperspektive ist dabei neben der empirischen Absicherung der Modellstrukturen auch die angemessene Passung des Modells zu den angestrebten Bildungsvorgaben nachzuweisen. Von ebenso zentraler Bedeutung sind jedoch auch unterrichtspraktische Notwendigkeiten wie die Anschlussfähigkeit und Nützlichkeit des Kompetenzmodells für Unterrichtsgestaltung und Individualdiagnostik, denen bei der Modellentwicklung Rechnung getragen werden muss.

Ausgehend von dem normativen Modell der Nationalen Bildungsstandards, dem „Model of Hierarchical Complexity“ und weiteren Modellen aus der aktuellen Forschungsliteratur baut dieser Artikel den theoretischen Rahmen eines Kompetenzstrukturmodells für das Fach Chemie auf. Empirische Befunde zur Fundierung der Modellannahmen werden ebenso diskutiert wie weiterführende Forschungsfragen für fachdidaktische Studien und eine mögliche Nutzung des Modells in der Unterrichtspraxis.

Schlüsselwörter: Kompetenzmodelle, Chemie, Bildungsstandards, Unterricht

Abstract

The current reforms in the German education policy pursue a paradigm shift in the educational system towards an outcome orientation. After implementing the National Educational Standards as normative objectives, the development of competence models is considered to be another important step towards this policy shift. Competence models should link the intended objectives to students' actual achievements and thus contribute for the purposes of a system monitoring. From the research perspective the adequate fit of a particular competence model and the corresponding educational goals is to be shown in addition to the empirical validation of the model structure. Also, the model development has to take into account requirements of the teaching practice. In particular, this implies the usefulness of the competence model for teaching arrangements and its connection to individual diagnostics to aid the effective promotion of competence.

This article builds up the theoretical frame for a competence structure model in the field of chemistry based on the “Model of Hierarchical Complexity”, the National Educational Standards, and other similar models from the current research literature. Empirical findings to support the model assumptions are discussed and research questions for didactical studies as well as possible applications of the model in teaching practice are presented.

Keywords: competence models, chemistry, educational standards, education

1 Einleitung

Die öffentliche und politische Reaktion auf die Ergebnisse internationaler Vergleichsstudien (z. B. PISA, TIMSS, IGLU) führte zu einem Wandel des deutschen Schulsystems und einer Neuorientierung der genannten Bildungsziele. Weiterhin ist von einem „Paradigmenwechsel in der Bildungspolitik im Sinne von ‚outcome-Orientierung‘, Rechenschaftslegung und Systemmonitoring“ die Rede (KMK 2005a, 6). Dieser Wechsel soll in einem ersten Schritt durch die von der KMK beschlossene Einführung von Bildungsstandards für den Mittleren Schulabschluss (KMK 2005b) und die damit verbundene Absicht erreicht werden, den Kompetenzerwerb der Schüler/-innen in zentralen Vergleichsarbeiten am Ende der Mittel- und Oberstufe zu erfassen. Die bundesweit geltenden Standards sollen eine Vergleichbarkeit der Qualität schulischer Bildung im föderalen Wettbewerb der Länder herstellen. Die normative Festlegung der Bildungsziele orientiert sich dafür an dem Konzept der naturwissenschaftlichen Grundbildung (vgl. KMK 2005b, 4) bzw. einer *scientific literacy* (vgl. Bybee 1997), das u. a. auch bei der inhaltlichen Ausrichtung der PISA-Studien Pate stand.

Die Vorbereitung und Durchführung von Vergleichsstudien auf Basis dieser bildungstheoretischen Vorgaben erfordern die Entwicklung empirisch abgesicherter Kompetenzmodelle, die hinsichtlich der Testaufgaben sowohl deren curriculare Validität als auch Art und Niveau der Anforderungen prä-

skriptiv sichern. Eine derartige Absicherung stellt hohe Anforderungen an die zu Grunde liegenden Modelle, denen bei der Entwicklung auf unterschiedliche Weise Rechnung getragen werden kann (vgl. Schecker & Parchmann 2006). Dies beinhaltet nicht nur die empirische Absicherung der angenommenen Kompetenzstruktur sondern insbesondere auch die adäquate Reflektion der erwünschten Ziele und *outcomes* des Unterrichts durch das auf dem Modell aufbauende Testsystem (vgl. Osborne & Dillon 2008, 27). Dabei müssen neben den Kriterien der Forschung auch die Ansprüche und Erwartungen der Unterrichtspraxis mitbedacht werden. Dies betrifft zum einen die Charakteristika und Faktoren der Modelle, zum anderen aber auch deren Umsetzung in konkrete Aufgabenstellungen zur Kompetenzförderung und einer begleitenden Kompetenzdiagnostik. Damit ist jedoch keine Gleichsetzung von Lern- und Testaufgaben gemeint. Beide orientieren sich entsprechend ihres Einsatzzwecks an unterschiedlichen Zielstellungen und Anforderungen. Gemäß den Ansprüchen einer neuen Aufgabenkultur im Unterricht sollen Lernaufgaben nach Möglichkeit offen gestaltet sein und somit den Erwerb eines breiten Spektrums an (Teil-)Kompetenzen ermöglichen, während die Konstruktion von Testaufgaben auf eine möglichst präzise und trennscharfe Diagnose einzelner Kompetenzfacetten abzielt. Dazu unterliegen letztere hohen psychometrischen Standards, die eine möglichst objektivierbare und reliable Auswertung und eine Skalenbildung anstreben (vgl. Abb. 1).

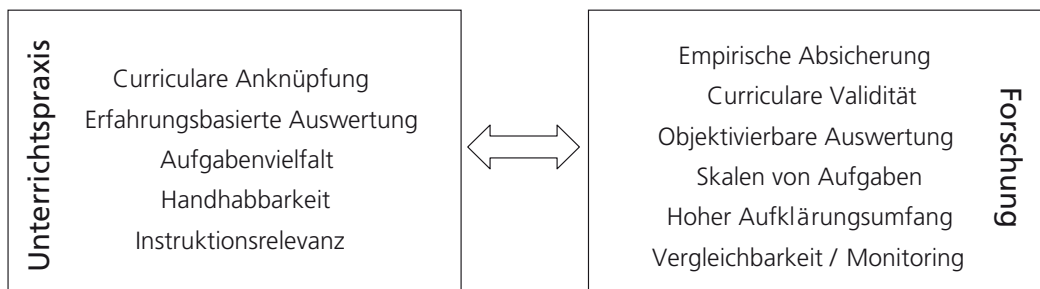


Abb. 1: Spagat zwischen den Interessensfeldern Unterrichtspraxis und Forschung

„Eine provokante These könnte lauten, dass psychometrisch gute Aufgaben fachdidaktisch für die Entwicklung von Unterrichtsqualität uninteressant sind. Tatsächlich muss aber eher gezeigt werden, wie die Rückmeldungen aus psychometrisch validen Tests als Diagnostik für die Gestaltung von Lehr-Lern-Prozessen genutzt werden können“ (Schecker & Parchmann 2006, 62f). Um diese Rückmeldungen zu nutzen, müssen Lehrkräfte jedoch in der Lage sein, die Konstruktionsprinzipien der Testaufgaben und somit das zugrunde liegende Kompetenzmodell zu verstehen und mit ihrem Unterricht in Verbindung zu bringen. Von diesem Verständnis hängt nicht nur die Akzeptanz und somit die erfolgreiche Implementierung des Testsystems ab, sondern auch dessen Wirksamkeit. Schließlich müssen letztlich die Lehrkräfte in der Lage sein, die Testergebnisse (also den gemessenen Output) auf der Unterrichtsebene in zukünftigen, verbesserten Input zu verwandeln. Um dies zu ermöglichen, benötigen die Lehrkräfte nicht nur ein Grundverständnis über statistische Kennwerte, sondern vielmehr eine Einsicht in die Maßstäbe der kriteriumsorientierten Testauswertung. Je weiter diese Kriterien von den Anforderungen der Unterrichtspraxis entfernt sind, desto schwieriger wird es sein, dass sich der Kompetenzaufbau im Unterricht und die Kompetenzmessung durch Tests im gleichen Punkt treffen.

Langfristig muss es also insbesondere darum gehen, Kompetenzentwicklungsmodelle aufzubauen, die es ermöglichen, den diagnostizierten Kompetenzstand einer Schülerin/eines Schülers mit instruktionsrelevanten Informationen zu verknüpfen. Im Hinblick auf den momentanen Stand der Forschung ist jedoch festzuhalten, dass es dazu kaum Ansätze gibt und es momentan aussichtsreicher erscheint, theoretisch und empirisch die Struktur der Kompetenzdimensionen und -facetten zu untersuchen. Dabei sollte jedoch, wie dargestellt, die Anschlussfähigkeit an den Unterricht berücksichtigt werden.

In den nachfolgenden Abschnitten soll somit über die Beschreibung von Ansätzen und Beispielen der Kompetenzmodellierung dargestellt werden, wie ein derartiger Brückenschlag aussehen kann. Anschließend folgt die Diskussion der ersten empirischen Ergebnisse einer Umsetzung für den Chemieunterricht.

2 Ansätze zur Beschreibung naturwissenschaftlicher Kompetenz

„Die Frage danach, wie Wissen überprüft und beurteilt werden kann, ist mit der vorgeordneten Frage verknüpft, welche Art von Wissen überhaupt gewollt ist“ (Krüssel 2001, 138). Der normative Rahmen, den die Bildungsstandards (KMK 2005b) aufspannen, muss weiter gefüllt und konkretisiert werden (wie es bspw. durch Kerncurricula getan wird), um Lehrkräften und Schüler/-innen „eine Orientierung und eine deutliche Vorstellung darüber zu vermitteln, was die Erwartungen an das Lernen und seine Ergebnisse sind“ (Ölkers & Reusser 2009, 308f). Dies betrifft insbesondere die Bereiche, die über Fachwissen hinausgehen. „Würden die bisherigen Standards nicht weiterentwickelt, so ergäbe sich für die Testentwicklung das Problem, dass Standards getestet würden, welche die selbst gesetzten Qualitätsansprüche gar nicht erfüllen und zu deren Aufbau im Unterricht die Schulen bisher keine Gelegenheit hatten“ (ebd., 308). Bei dieser Umsetzung spielen Kompetenzmodelle als Bindeglied zwischen den abstrakten Bildungszielen und konkreten (Test-)Aufgabenstellungen eine zentrale Rolle (vgl. Klieme et al. 2003).

2.1 Nutzen von Kompetenzmodellen

Kompetenzmodelle unterliegen der Grundannahme, dass die enthaltenen Kompetenzen nicht vereinzelt und zusammenhangslos vorliegen, sondern dass sie in einem systematischen Gefüge stehen, das nicht nur durch die Systematik der Fachwis-

senschaft, sondern psychologisch begründet ist. Demnach sollen Kompetenzmodelle im Bildungsbereich dazu dienen, die Struktur der zentralen, spezifischen Fähigkeiten, die im Rahmen von Schule und Unterricht erworben werden sollen, sowie die Stufen ihres Erwerbs zu beschreiben (Klieme et al. 2003). Dabei können Kompetenzstruktur- und Kompetenzentwicklungsmodelle unterschieden werden (Schecker & Parchmann 2006). Sie sollen die vermuteten Zusammenhänge des Kompetenzgefüges aufzeigen und unterschiedliche Ausprägungen der Teilkompetenzen diagnostizierbar nachweisen.

Die meisten Untersuchungen zur Kompetenzmodellierung stützen sich auf das von Weinert (2001) ausgearbeitete, umfassende Verständnis von Kompetenz (vgl. OECD 2003). Einer ähnlichen Auffassung folgend kann kompetentes Handeln als Zusammenspiel von deklarativem und prozeduralem Wissen, metakognitiver Regulation, motivationalen und volitionalen Prozessen sowie Emotionen und Dispositionen aufgefasst werden (vgl. Perkins & Ritchhart 2004). Doch auch wenn Kompetenz als Synthese von kognitiven und nicht-kognitiven Komponenten angesehen wird, konzentriert sich die Erfassung von Kompetenz häufig auf die kognitiven Faktoren und hierbei insbesondere auf klassische Leistungsmerkmale (vgl. Oelkers 2007; Herzog 2007).

Um den Bereich dieser Leistungsmerkmale abzustecken, die in Unterricht und Schule erwartet werden, weisen Kompetenzmodelle häufig drei Aspekte aus: eine inhaltliche Anbindung, einen Anforderungsbereich und tätigkeitsspezifische Teilkompetenzen (vgl. Abb. 2).

Die inhaltliche Anbindung sollte einen engen Bezug zu curricularen Vorgaben aufweisen, da diese (im Normalfall) die Inhalte des Unterrichts bestimmen. Dies entspräche gängiger Schulpraxis, in der nur überprüft werden darf, was vorher unterrichtet wurde. Soll eine gute Passung zu den Nationalen Bildungsstandards gewährleistet werden, kann bspw. auf die dort vorgeschlagenen Basiskonzepte (bzw. Leitideen in der Ma-

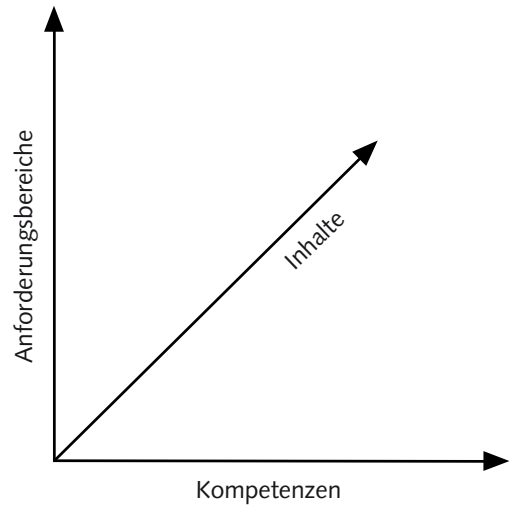


Abb. 2: Kompetenzdimensionen

thematik) zur Strukturierung der Inhalte zurückgegriffen werden (vgl. Neumann et al. 2007; Schecker & Parchmann 2006).

Der Anforderungsbereich dient der Beschreibung eines schwierigkeitskonformen Konstrukts. Zur Charakterisierung der Schwierigkeit von Aufgaben und der zur Lösung notwendigen Fähigkeiten finden sich in der Literatur verschiedene Vorschläge (Klieme 2000; Fischer & Draxler 2001, 2002; Prenzel et al. 2001, Commons et al. 1998 u. a.). Dabei wird davon ausgegangen, dass sich Kompetenzen in einer Hierarchie abbilden lassen: Je anspruchsvoller die Anforderungsmerkmale der Aufgabe, desto höher die notwendige Kompetenzausprägung auf Seiten der Schüler/-innen, um die Aufgabe zu bewältigen.

Der dritte Bereich der (Teil)Kompetenzen beschreibt im Sinne eines Tätigkeitsprofils die Fähigkeiten und Fertigkeiten (auch teilweise als Prozesse oder Aktivitäten benannt), die von den Schüler/-innen erwartet werden. Im Bremen-Oldenburger-Kompetenzmodell werden hier bspw. die Prozesse „Fachwissen nutzen“, „Erkenntnisse gewinnen“, „Kommunizieren“ und „Bewerten“ genannt (Schecker & Parchmann 2006, 58), was die Intention einer direkten Anschlussfähigkeit an die Kompetenzbereiche der Bil-

dungsstandards erkennen lässt. Deren Unterteilung wurde jedoch normativ gesetzt, eine empirische Bestätigung dieser Struktur steht noch aus (vgl. dazu Schmidt 2008; Christiansen 2007).

Doch so intensiv der Bereich der Kompetenzmodellierung und Kompetenzmessung momentan auch beforscht wird, so sehr mangelt es bisweilen an einheitlichen Zielvorstellungen und inhaltlichen Konkretisierungen (vgl. Parchmann 2008; Ditton 2002). Die zahlreichen publizierten Modelle unterscheiden sich sowohl hinsichtlich der Anlage der Kompetenzdimensionen, als auch bezüglich deren Graduierung. Weder hinsichtlich der Anzahl der notwendigen (und sinnvollen) Dimensionen herrscht Übereinkunft, noch bei der inneren Struktur der einzelnen Dimensionen. So werden mitunter hierarchische, sich parallel entwickelnde oder qualitativ unterschiedliche Komponenten angenommen. Nur im Anforderungsbereich wird allgemein eine hierarchische Struktur favorisiert, denn hierin wird der „Schlüssel zur Konstruktion und Auswertung kriteriumsorientierter Tests“ gesehen, da „sie eine Alternative zur willkürlichen Setzung von Leistungsmarken auf einem Kontinuum bieten“ (Klieme et al. 2003, 81). Dazu muss neben der inneren Struktur der einzelnen Dimensionen eines Kompetenzmodells ebenso die Gesamtstruktur des Modells untersucht werden. „Empirische Untersuchungen mithilfe von Tests sind erforderlich, um zu prüfen, ob diese Modelle tatsächlich die Aspekte der Kompetenzen von Lernenden, ihre Niveaustufung und ggf. ihre Entwicklung angemessen widerspiegeln“ (Klieme et al. 2003, 82). Diese „Angemessenheit“ betrifft zum einen die Modellprüfung im Sinne einer Konstruktvalidierung, zum anderen die Überprüfung des Erreichens der Zielvorgaben durch die Lernenden (vgl. Schecker & Parchmann 2006, 51). Testen als selbstreferenzieller Prozess verlangt neben der empirischen Absicherung der angenommenen Kompetenzstruktur ferner eine adäquate Reflektion der erwünschten Ziele und *outcomes* des

Unterrichts durch das auf dem Modell aufbauende Testsystem (vgl. Osborne & Dillon 2008, 27).

Dabei ist weiter zu bedenken, dass die Einführung eines Testsystems in die Schulrealität kein einseitiger Vorgang ist, sondern dass sich Testsystem und Unterricht wechselseitig beeinflussen. „Es ist den meisten Praktikern bekannt, dass Schüler nicht primär Gegenstände studieren, um den Gegenstand näher kennen zu lernen, sondern sie üben sich im prüfungsrelevanten Wissen, um Prüfungen über die Gegenstände zu bestehen“ (Krüssel 2001, 139). Hinsichtlich der Einführung zentraler Überprüfungen ist also anzunehmen, dass sich Lehrer/-innen bei der Gestaltung ihres Unterrichts und eigener Klausuren an dem Modell orientieren, das derartigen Unternehmungen zugrunde liegt: „Teachers increasingly looked not to the curriculum specifications to define what the intentions of the curriculum should be but to the assessment items“ (Osborne & Dillon 2008, 23).

Aus diesem Blickwinkel müssen die „Nebenwirkungen“ des zugrunde liegenden Kompetenzmodells von Anfang an mitgedacht werden. Dies betrifft die Frage danach, in welche Richtung ein „kompetenzmodellorientierter“ Unterricht das Verhältnis von Faktenwissen und problemorientiertem Handlungswissen beeinflusst (vgl. Krüssel 2001) und inwieweit komplexeres Denken angeregt wird (*higher order thinking*; vgl. Osborne & Dillon 2008; Zoller 2002). Bei der Formulierung und Implementierung stellt sich also die Frage: Welche Eigenschaften hätte ein Kompetenzmodell, dessen Einführung eine positive Entwicklung der Unterrichtspraxis nach sich ziehen könnte? Auch wenn der primäre Fokus der Einführung von Vergleichsarbeiten auf einem Systemmonitoring liegt, sollte man diese Verantwortung nicht unterschätzen.

Unter der Maßgabe einer erfolgreichen Implementierung des Modells in die Praxis muss neben der Berücksichtigung der spezifischen Vorgaben der Forschung (z. B. psychometrische Kennwerte) auch eine

Vermittlung zwischen diesen Interessensfeldern angestrebt werden (vgl. das Modell der „learning communities“ in den Projekten SINUS und CHiK: Ostermeier 2004; Gräsel & Parchmann 2004). Dies ist notwendig, da der Unterricht neben der Verfolgung der curricularen Vorgaben und den dort zugrunde liegenden Bildungszielen die Schüler/-innen auf die Prüfungssituation vorbereiten und ihnen ermöglichen muss, diese bei entsprechender persönlicher Anstrengung auch bestehen zu können: „[...] students would have to learn to do the kinds of thinking required by the tests“ (Shavelson 2007, 19). Dies ist insbesondere fundamental, wenn individuelle Laufbahnentscheidungen an das Prüfungsergebnis gekoppelt sind, bspw. bei sog. *high stakes assessments*.

Daher wird natürlich gewünscht, dass derartige Unternehmungen mehr instruktionsrelevante Informationen liefern: „Die Rückmeldefunktion von schulischen Erfolgskontrollen sollte verstärkt Auffordnungscharakter für die nächsten didaktisch-methodischen Entscheidungen des Lehrers haben“ (Krüssel 2001). Hier muss eine Brücke geschlagen werden zwischen den Ansprüchen und Erwartungen auf Seiten der Forschung und der Unterrichtspraxis. Eine erfolgreiche Implementierung eines derartigen Testsystems kann nur erfolgreich sein, wenn sie von allen Beteiligten mitgetragen wird (für eine umfangreiche, mehrperspektivische Darstellung vgl. Labudde 2007). Dies schließt insbesondere die Lehrkräfte mit ein. Somit ist nicht nur eine Objektivitätssteigerung des Testinstrumentariums anzustreben, sondern eine Verbesserung des wechselseitigen Verstehens von Forschung und Unterrichtspraxis: „Leistungsbewertung ist Kommunikation!“ (Krüssel 2001, 141).

2.2 Beispiele für Kompetenzstrukturmodelle in den Naturwissenschaften

In der Forschungsliteratur wurden in den vergangenen Jahren verschiedene Vorschläge zur Beschreibung von Kompetenz

in den Naturwissenschaften vorgelegt (für einen internationalen Überblick vgl. Waddington, Nentwig & Schanze 2007). Kompetenzentwicklungsmodelle versuchen, den zeitlichen Verlauf und die innere Dynamik des Kompetenzerwerbs abzubilden, wobei bislang nur wenige und teilweise nur eingeschränkt verallgemeinerbare Ergebnisse vorliegen (vgl. Eggert & Bögeholz 2006; Liu & Lesniak 2005; Hammann 2004; von Aufschnaiter 2003; QCA 2000; Case 1992).

Zur Beschreibung einer Kompetenzstruktur gibt es, wie oben angesprochen, verschiedene Annahmen, von denen an dieser Stelle vier in aller Kürze vorgestellt werden sollen. Die Beispiele unterscheiden sich dabei nicht nur in der vorgeschlagenen Dimensionierung und Graduierung von Kompetenz sondern auch in dem Vorgehen, durch das die jeweilige Strukturierung entwickelt wurde.

1 TIMSS

Die TIMS-Studie wurde ausgehend von Stufen naturwissenschaftlicher Grundbildung vorgenommen (vgl. Klieme 2000). Diese umfasst am Beispiel von TIMSS/III die Dimensionen ‚Sachgebiete‘ (Geowissenschaftliche Themen, Biologie, Chemie/Physik) und ‚Fähigkeiten‘ (‚Verstehen von Einzelinformationen‘, ‚Verstehen komplexer Informationen‘, ‚Konzeptionalisieren und Anwenden‘ sowie ‚Experimentieren und Beherrschung von Verfahren‘). Aufbauend auf diesem Rahmen nutzen Klieme, Baumert, Köller & Bos (2000) die Daten der TIMS-Studie zur Festlegung von vier Kompetenzstufen, die anschließend durch Analyse entsprechend ihrer Schwierigkeit verorteter Aufgaben inhaltlich konkretisiert wurden: ‚Naturwissenschaftliches Alltagswissen‘, ‚Erklärung einfacher alltagsnaher Phänomene‘, ‚Anwendung elementarer naturwissenschaftlicher Modellvorstellungen‘, ‚Verfügung über grundlegende naturwissenschaftliche Fachkenntnisse (vgl. Klieme et al. 2000).

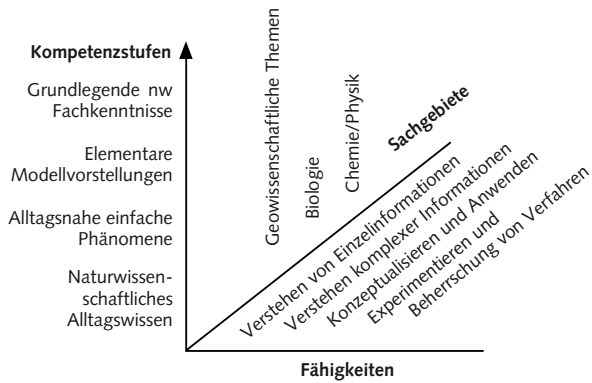


Abb. 3: Kompetenzdimensionen in TIMSS

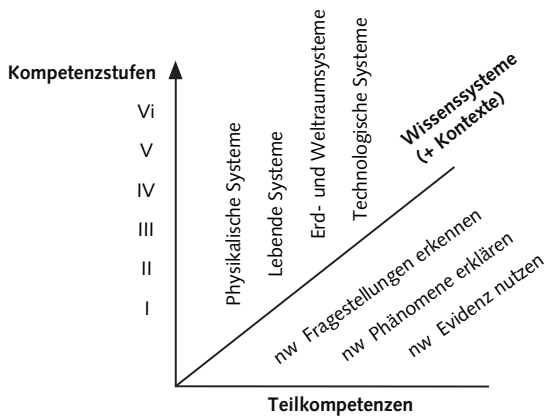


Abb. 4: Kompetenzdimensionen in PISA 2006

2 PISA

Das PISA assessment framework (OECD 2006) basiert auf dem Konzept der *scientific literacy* (vgl. Bybee 1997). Dabei wird Kompetenz sowohl in der internationalen Studie als auch in der deutschen Erweiterung als das Potenzial aufgefasst, „in einem bestimmten Kontext Anforderungen zu bewältigen und Probleme zu lösen“ (OECD 2006, 25; Prenzel et al. 2007, 37). Die naturwissenschaftlichen Testaufgaben werden dementsprechend in Kontexte eingebunden („Gesundheit“, „Natürliche Ressourcen“, „Umwelt“, „Risiken/Gefahren“ und „Grenzen von Naturwissenschaften und Technik“), womit eine Überprüfung der Alltagsauglichkeit und Anwendbarkeit des Wissens der Jugendlichen angestrebt wird.

Über die Tests und Fragebögen werden die Bereiche ‚naturwissenschaftliches Wissen‘, ‚Wissen über die Naturwissenschaften‘, ‚Einfluss von Naturwissenschaften und Technik auf unsere materielle, intellektuelle und kulturelle Umwelt‘, sowie ‚Motivationale Orientierung‘ erfasst. Das naturwissenschaftliche Wissen wird weiter unterteilt in vier Wissenssysteme („Physikalische Systeme“, „Lebende Systeme“, „Erd- und Weltraumsysteme“ und „Technologische System“), die jeweils wiederum in verschiedene Kategorien unterteilt werden (vgl. Prenzel et al. 2007, 69). Die Auswahl der Inhalte orientiert sich dabei ausdrücklich nicht an Schulfächern und Lehrplänen.

Hinsichtlich der Kompetenzdimension, die Handlungs- und Leistungspotenziale der Jugendlichen erfassen soll, wird in PISA 2006 naturwissenschaftliche Kompetenz in drei Teilkompetenzen differenziert, die sich auf unterschiedliche Anforderungen beziehen: ‚naturwissenschaftliche Fragestellungen erkennen‘, ‚naturwissenschaftliche Phänomene erklären‘ sowie ‚naturwissenschaftliche Evidenz nutzen‘ (vgl. Prenzel et al. 2007, 68).

Die Ausprägungen des Anforderungsniveaus wurden wie bei TIMSS post-hoc analysiert. Dieses *proficiency scaling* (vgl. Klieme 2000) ist zwar psychometrisch fundiert, die Festsetzung der Stufenanzahl jedoch weitgehend willkürlich: Während bei TIMSS vier Stufen festgelegt werden, werden bei PISA 2006 sechs Kompetenzstufen beschrieben (Prenzel et al. 2007, 77).

Nach Neumann et al. (2007, 105) haben sich „post hoc konstruierte Modelle für die naturwissenschaftlichen Fächer bisher nicht bewährt“. Die Extraktion von Kompetenzstufen allein aus empirischen Daten und den vorhandenen Aufgabentypen ist

kritisch, da „Kompetenzausprägungen, die nicht in den Aufgaben enthalten waren, auch nicht im Modell erscheinen“ können (Schecker & Parchmann 2006, 52). Inwieweit auf dieser Basis also die Aufgabenschwierigkeit präskriptiv gesichert bzw. unterrichtsrelevante Aufgaben konstruiert und den Stufen verlässlich zugeordnet werden können, ist offen (vgl. Prenzel et al. 2001; Klieme 2000).

Während deskriptive Modelle (wie die Kompetenzstufen aus TIMSS und PISA) der Aufklärung von Varianz in gezeigten Schülerleistungen dienen, versuchen normative Modelle a priori eine Einordnung erwarteter Lernergebnisse vorzunehmen. Darüber hinaus dienen sie der Strukturierung entsprechender Lernangebote.

3 Bremen-Oldenburger Kompetenzmodell

Schecker & Parchmann (2006) schlagen ein fünfdimensionales Rahmenmodell vor, das als Basis für weitere Untersuchungen dienen und insbesondere eine enge Anbindung an die Bildungsstandards der KMK gewährleisten soll. So decken sich die beiden Dimensionen ‚Inhaltsbereich/ Basiskonzept‘ und ‚Handlung/ Prozess‘ mit den Basiskonzepten bzw. Kompetenzbereichen der Bildungsstandards. Des Weiteren werden die Dimensionen ‚Kontext‘ (‚innerfachlich‘, ‚persönlich-gesellschaftlich‘ und ‚professionelle Anwendungen‘) sowie ‚Kognitive Anforderung‘ (‚Divergentes Denken‘, ‚Konvergentes Denken‘, ‚Umgang mit mentalen Modellen‘, ‚Umgang mit Zahlen‘; vgl. Rost et al. 2005) gesetzt. Zur Beschreibung des Anforderungsniveaus dient die Dimension ‚Ausprägung‘ mit den Komponenten ‚lebensweltlich‘, ‚fachlich nominell‘, ‚aktiv anwendend‘ und ‚konzeptuell vertieft‘ (vgl. Schecker & Parchmann 2006, 58).

Die stark ausdifferenzierte Beschreibung soll der Komplexität realer Lehr- und Entwicklungsprozesse annähernd Rechnung tragen und weniger eine direkte psychometrische Messbarkeit suggerieren. Als Rahmenmodell dient es somit als Basis für empirische Untersuchungen, die neben der Überprüfung von Teilstrukturen insbesondere zentrale Einflussfaktoren determinieren sollen. Dabei zeigten sich bislang Schwierigkeiten insbesondere hinsichtlich der postulierten Anforderungsniveaus, deren angenommene Strukturierung sich in den Daten nicht wiederfinden ließ (vgl. Einhaus 2007; Schmidt 2008).

4 Evaluation der Standards in den Naturwissenschaften der Sek I (ESNaS)

Aufbauend auf Vorarbeiten zur vertikalen Vernetzung (Fischer, Glemnitz, Kauertz & Sumfleth 2007) entwickelten Neumann et al. (2007) ein Modell zur Beschreibung physikalischer Kompetenz (vgl. auch Kauertz 2008). Die Umsetzung dieses Modells für die Chemie umfasst die Dimensionen ‚Kompetenzbereiche‘ (und damit eine Anbindung an die Strukturierung der Bildungsstandards), ‚Kognitive Aktivität‘ und ‚Komplexität‘.

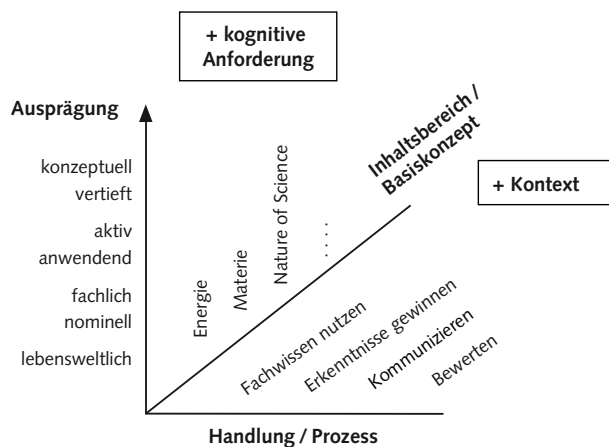


Abb. 5: Kompetenzdimensionen im Bremen-Oldenburger-Kompetenzmodell (3 von 5 Dimensionen ausgeführt)

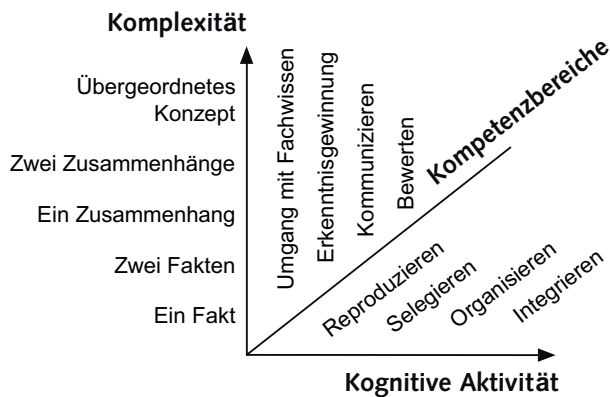


Abb. 6: Kompetenzdimensionen im Bremen-Oldenburger-Kompetenzmodell (3 von 5 Dimensionen ausgeführt)

Die ‚Kognitive Aktivität‘ beschreibt die zum Aufgabenlösen notwendigen kognitiven Anforderungen durch die Komponenten ‚Reproduzieren‘, ‚Selegieren‘, ‚Organisieren‘ und ‚Integrieren‘. Die Dimension ‚Komplexität‘ beschreibt das Vernetzungsniveau der Inhalte über fünf hierarchisch angeordnete Stufen: „Ein Fakt (I), zwei Fakten (II), ein Zusammenhang (III), zwei Zusammenhänge (IV), übergeordnetes Konzept (V)“ (Walpuski et al 2008, 325).

Diese zunächst normative Setzung einer Kompetenzstruktur wurde empirisch überprüft (Kauertz & Fischer 2006) und konnte mit Einschränkung bestätigt werden. So lassen sich die Basiskonzepte hinsichtlich der notwendigen Kompetenz unterscheiden und die Komplexitätsniveaus sind in der Lage, die Schwierigkeit einer Aufgabe vorherzusagen, wenn auch mit „unbefriedigender“ Genauigkeit (Kauertz 2008). Zudem tragen die kognitiven Aktivitäten den bisherigen Untersuchungen zufolge nur marginal zur Aufgabenschwierigkeit bei. Für die anstehende Evaluation der Bildungsstandards in den Naturwissenschaften durch das IQB, für die das beschriebene Modell die Grundlage der Aufgabenkonstruktion liefert, wurde diese Dimension daher neu operationalisiert (vgl. Walpuski et al. 2008; Neumann et al. 2007).

2.3 Desiderata

Auch wenn mit dieser verkürzten Darstellung kein vollständiger Überblick über den aktuellen Stand der Kompetenzmodellierung in den Naturwissenschaften gegeben werden kann, lässt sich dennoch festhalten, dass bislang keine Einigkeit über ein geeignetes Modell besteht. Die beschriebenen Modelle und weitere Vorschläge aus der Literatur (vgl. KMK 2004; Schecker, Fischer & Wiesner 2004; QCA 2000) decken sich zwar weitestgehend in der Berücksichtigung bestimmter Strukturmerkmale (vgl. Abb. 2), doch

gibt es hinsichtlich der Ausdifferenzierung und der genauen Ausrichtung deutliche Unterschiede. Ebenso steht bei vielen Modellen die notwendige empirische Bestätigung der normativen Struktur noch aus. Hinsichtlich der Anzahl der angenommenen Kompetenzdimensionen betrifft dies insbesondere die Bestätigung der Notwendigkeit bzw. des Mehrwerts jeder (zusätzlichen) Dimension durch empirische Ergebnisse. Auch die Nutzbarkeit von post hoc gewonnenen Modellen (oder Modelldimensionen) ist durch entsprechende fachdidaktische und lernpsychologische Forschung zu prüfen.

Des Weiteren wurde die Verknüpfung mit der Unterrichtspraxis noch nicht geleistet. Abgesehen von der Prüfung der Lehrplanvalidität im Rahmen von PISA und ESNaS (wobei hier nur die „Bekanntheit“ der Inhalte und Anforderungen gesichert wurde und nicht eine wiederkehrende, regelhafte und damit verlässliche Behandlung im Unterricht) müssen weitere Untersuchungen sowohl das Vorhandensein als auch den Nutzen jeder Dimension und Komponente eines Kompetenzmodells überprüfen. Ebenso liegen zur Individualdiagnostik und zur gezielten Förderung auf Basis von Testergebnissen kaum Forschungsergebnisse vor. Auch wenn diese nicht im zentralen Fokus von Large-Scale-Assessments (z.B. PISA, TIMSS) bzw. von

zentralen Programmen zum Bildungsmonitoring (z.B. ESNas) stehen, kann dauerhaft der Kompetenzerwerb der Schüler/-innen nicht nur von den Zielen her erfasst werden, sondern muss auch mit den Prozessen der Zielerreichung verknüpft werden. Dies ist notwendig, wenn die Testergebnisse im Sinne einer Schulevaluation über eine reine Beschreibung des status quo hinaus für die Unterrichtspraxis nutzbar sein sollen.

„Die zentrale Frage ist deshalb, wie der Prozess des kumulativen Kompetenzaufbaus durch schulischen Unterricht nicht nur besser als bisher erfasst, sondern auch individuell gefördert werden kann“ (Ölkers & Reusser 2008, 28). Dazu reichen die für die Überprüfung von Standards konzipierten Diagnoseinstrumente und -verfahren jedoch nicht aus bzw. für diesen Zweck wurden sie auch nicht konzipiert, so dass an dieser Stelle deutlicher Forschungsbedarf zur Verknüpfung von Rückmeldeverfahren, Lernprozess- und Förderdiagnostik besteht (ebd.).

3 Das Modell der hierarchischen Komplexität

Auf Basis der dargestellten Ansätze und der diskutierten Anforderungen und Desiderata zur Kompetenzmodellierung soll nun ein eigener Ansatz zur Beschreibung fachspezifischer Kompetenz vorgestellt werden. Dazu wird zunächst der theoretische Rahmen gespannt, auf dem das Modell aufbaut, bevor dessen Struktur sowie erste empirische Befunde zur Bestätigung dieser Annahmen vorgestellt werden. Die Darstellung der Modellentwicklung und -prüfung wird dabei an geeigneten Stellen um Bezüge zu unterrichtspraktischen Anforderungen ergänzt.

3.1 Lerntheoretische Begründung zur Beschreibung fachspezifischer Kompetenz

Die meisten aktuellen kognitiven Theorien basieren auf oder beinhalten Netzwerkmo-

delle. Diese beschreiben die Struktur deklarativen Wissens in Form von propositionalen Netzen (vgl. Schnotz 1994; Tergan 1986) im Gegensatz zu einer Anhäufung unverbundener Einzelelemente („Knowledge is not a basket of facts“; Anderson 1984). Dem Bild des Netzwerks als Wissensstruktur folgend, steigt die Qualität des Wissens demnach mit dem Grad an Vernetzung der Inhaltselemente (vgl. Peuckert 1999; Steiner 2006).

Die steigende Vernetzung kann über unterschiedliche Teilprozesse interpretiert werden, die sich jedoch in ihrem Prinzip weitestgehend decken. Aebli (1980, 1981) stellt für den Aufbau begrifflichen Wissens die Prozesse Verknüpfen, Verdichten/Objektivieren und Strukturieren in den Mittelpunkt, wobei durch das Verdichten als zentralem Schritt Ketten verknüpfter Elemente zu Elementen höherer Ordnung werden. Diesen Schritt beschreiben Laird, Rosenbloom & Newell (1986) analog als *chunking*. „Dieser Prozess dient vor allem der Entlastung des Arbeitsgedächtnisses, wo die *chunks* weniger Verarbeitungskapazität beanspruchen und weiterverwendet [...] werden können“ (Steiner 2006, 166).

Die zum Verständnisaufbau notwendige Vernetzung von Inhalten über den Schritt der „Objektivierung“ bzw. des *chunkings* und die dafür erforderliche Anbindung an Vorwissensstrukturen (vgl. Süß 1996; Renkl & Stern 1994; Collins & Loftus 1975) im Sinne kumulativen Lernens (vgl. Gagné 1973) sollen eine Auffassung von verständnisorientiertem Lernen aufzeigen. Dabei wird sowohl für den Wissenserwerb als auch für die Wissensüberprüfung der Schlüssel in der Analyse von Aufgaben gesehen. Zur Charakterisierung von Aufgabenattributen gibt es verschiedene Ansätze (Klieme 2000; Fischer & Draxler 2001; Prenzel et al. 2001; Fischer et al. 2007 u. a.), unter denen die Beschreibung durch Komplexitätsstufen zur Erfassung eines schwierigkeitskonformen Konstrukts in der aktuellen Forschungsliteratur unterschiedlicher Inhaltsdomänen einen möglichen Ansatz darstellt (Fischer

1980; Dawson 2003; Neumann et al. 2007; Kauertz 2008).

3.2 Das „Model of Hierarchical Complexity“

„The concept of hierarchical complexity is at the core of most cognitive developmental theories“ (Dawson-Tunik 2006, 114; vgl. Case 1985; Demetriou, Platsidou, Efklides & Metallidou 1991; Fischer & Bidell 2006). Der Grundgedanke ist dabei, dass mit der Handhabung einer steigenden Anzahl von Inhaltselementen auch die Schwierigkeit der Verarbeitung ansteigt. Auch das hier vorgestellte Modell greift auf ein Komplexitätsschema zurück, das *Model of Hierarchical Complexity* (Commons et al. 1998; Commons & Richards 2002, 1984a, 1984b). Dieses basiert auf logisch-mathematischen Prinzipien, stellt also ein inhaltsunabhängiges Modell dar, und soll daher auf alle Inhaltsbereiche angewendet werden können, in denen Informationen in organisierter Form verarbeitet werden.

Die Beschreibung fachspezifischer Kompetenz auf Basis eines domänenunabhängigen Modells ist dabei nicht trivial, sondern impliziert verschiedene Grundannahmen. Dazu gehört die Annahme, dass Elemente der Kognitionsentwicklung und Charakteristika der inhaltlichen Domäne strukturelle Ähnlichkeiten aufweisen. Diese Annahme ist nicht ohne Weiteres zu bestätigen, jedoch stützen die bisherigen Befunde zur Anwendung des *Model of Hierarchical Complexity* in unterschiedlichen Domänen (vgl. zur Übersicht: Commons 2007) und dessen Nutzung als Rating-Verfahren im Vergleich zu anderen domänenspezifischen Verfahren (vgl. Dawson 2002) die Verwendung im Inhaltsbereich Chemie.

Die Nutzung des Modells erfolgt ebenfalls vorwiegend über die Untersuchung von Aufgabenmerkmalen. Diese Untersuchung fokussiert auf inhärente Aufgabenattribute, die objektiv erfasst und damit quantitativ messbar gemacht werden können. Dazu gehören

die Kodierung der Information (*coding*), die syntaktische und lexikalische Schwierigkeit der Aufgabe, die Position relevanter Informationen innerhalb der Aufgabe (*position effects*), die Anzahl gleichwertiger Bearbeitungsschritte (*horizontal complexity; information load*) sowie, als zentrales Konstrukt, die *hierarchical complexity*, bestimmt durch die nicht-willkürliche Verknüpfung der Lösungsschritte (Commons et al. 1998). Auch wenn alle diese genannten Faktoren einen Einfluss auf die Aufgabebearbeitung ausüben, so soll die *hierarchical complexity* doch zum Großteil der Varianzaufklärung beitragen. Aus diesem Grund wird im Folgenden auch ausschließlich auf dieses Konstrukt fokussiert. Daneben beeinflussen selbstredend weitere Faktoren die Aufgabebearbeitung, so z.B. Art und Umfang an Unterstützung oder individuelle Probandeneigenschaften (Vorwissen, Lerngeschichte, Vertrautheit mit dem Aufgabenformat oder Aufgabenkontext etc.). Diese Faktoren sind jedoch nicht objektivierbar, sondern basieren auf der individuellen Konstellation von Aufgabe und Proband. Obwohl die Aufgabenschwierigkeit zweifelsfrei von diesen Faktoren beeinflusst wird, werden sie im Rahmen des *Model of Hierarchical Complexity* nicht quantifiziert (Commons et al. 1998).

Die Stufen des Modells sind nach Konstruktionsprinzipien aufgebaut, die z.T. auf die Stufentheorie Jean Piagets zurückgreifen (vgl. Piaget 1966). So bauen obere Stufen auf unteren Stufen auf, wobei die unteren Stufen durch die oberen Stufen organisiert werden. Jedes Element einer höheren Stufe entsteht durch eine Verknüpfung und Koordination von Elementen der darunter liegende Stufe. Die Verknüpfung darf dabei nicht beliebig sein und keine simple Aneinanderreihung bzw. Hintereinanderausführung von Elementen der unteren Stufe darstellen (Commons et al. 1998).

Die Anlehnung an die Stufentheorie Piagets schließt jedoch nicht dessen restriktive Implikationen für Anordnung und Aufbau der (Entwicklungs-)Stufen auf Lernerseite ein. Dazu gehört insbesondere die Annahme

eines altersabhängigen, konzertierten Fortschreitens durch Entwicklungsstadien. Im Sinne des beschriebenen Neo-Piaget-Modells bilden die Stufen vielmehr ein Muster ab, auf dessen Basis sich Konstruktions- und Koordinationsvorgänge zum Aufbau neuer, domänenspezifischer Fähigkeiten aus bereits erworbenen Fähigkeiten *beschreiben* lassen. Es kommt somit nicht zu einer Entwicklung universeller Fähigkeiten, sondern es handelt sich um die Entwicklung von Fähigkeiten entlang einer universellen Skala (für eine detaillierte Darstellung vgl. Fischer & Bidell 2006).

Die hierarchische Komplexität einer Aufgabe bezieht sich auf Grund des Konstruktionsprinzips der Stufen auf die enthaltene Anzahl von Verknüpfungsoperationen. Mit steigender Komplexitätsstufe ist also eine wachsende Anzahl von Verknüpfungsoperationen zu leisten. Damit unterscheiden sich die Stufen nicht nur auf quantitativer Ebene hinsichtlich der Anzahl an zu verarbeitenden Inhaltselementen, sondern vielmehr auf qualitativer Ebene durch die Reorganisation von einfacheren mentalen Repräsentationen in komplexere (*horizontal vs. hierarchical complexity*; Commons et al. 1998). Diese Sichtweise entspricht Modellen der Informationsverarbeitung (vgl. Anderson & Lebiere 1998) und konstruktivistischen Lerntheorien (Steiner 2006; vgl. Kap. 2.1), nach denen im Zuge fortschreitender Entwicklung von Fähigkeiten nicht nur eine steigende Anzahl an Informationen oder Problemaspekten simultan mental repräsentiert und manipuliert werden können, sondern zudem auch eine Zunahme an Informationstiefenverarbeitung erfolgt (Stichworte: Verdichtung/Objektierung, *chunking*; Laird, Rosenbloom & Newell 1986; vgl. Kap. 2.1). Dabei verläuft der Aufbau von Wissensstrukturen über verschiedene Konstruktions- und Koordinationsvorgänge analog zur Aufbauweise höherer Modellstufen aus Elementen unterer Stufen (vgl. Piaget 1966; Aebli 1981; Commons et al. 1998).

Die Konstruktionsweise des *Model of Hierarchical Complexity* überträgt sich auch auf die Messmethode an sich. So definiert das Modell

die Schwierigkeitsstufen der Aufgabenbearbeitung durch Begriffe, die zur Beschreibung der hierarchischen Komplexität derjenigen Aufgaben genutzt wurden, welche durch die Bearbeitung erfolgreich bewältigt wurden.

Bei einer modellkonformen Auswertung wird zudem grundsätzlich unterschieden zwischen der hierarchischen Komplexität von Aufgaben- oder Problemstellungen und der Ausführung bzw. dem Ergebnis des Aufgabenbearbeiters. Es wird also gerade nicht versucht, die Schwierigkeit einer Aufgabe ausschließlich an den Bearbeitungsergebnissen einer Probandengruppe festzumachen, sondern die Aufgabenschwierigkeit wird theoretisch begründet und den empirischen Ergebnissen vergleichend gegenübergestellt (Commons et al. 1998). Die Aufgabeneigenschaften werden so als teilweise, jedoch nicht als exklusiv verantwortlich für die Aufgabenperformanz angesehen. Vielmehr wird anerkannt, dass auch wichtige Eigenschaften des Probanden das Ergebnis beeinflussen.

3.3 Adaption des *Model of Hierarchical Complexity* für den Inhaltsbereich Chemie

Die Adaption des Modells für den Inhaltsbereich Chemie (MHC-C) fand aufbauend auf dem beschriebenen theoretischen Rahmen in mehreren Schritten statt. So wurden zahlreiche Beispielaufgaben entwickelt oder unterschiedlichen Quellen entnommen (Schulbücher, Materialien mitwirkender Lehrkräfte etc.) und mit Modellaufgaben aus Vorstudien (Dawson-Tunik 2006) abgeglichen, um eine erste Zuordnung von Aufgaben chemischen Inhalts zu den Modellstufen zu erreichen. Dazu wurden die Lösungsschritte sowohl der selbstentwickelten als auch der entnommenen Aufgaben analysiert und bei ähnlichen Vorgehensweisen und Anforderungen entsprechend gruppiert. Durch diesen ersten Schritt entstanden fünf Gruppen an Aufgaben steigender Komplexität.

Die derart gruppierten Aufgaben wurden weiter analysiert und hinsichtlich gemeinsamer Merkmale untersucht. Dabei wurde

versucht, für die fünf Aufgabengruppen möglichst aussagekräftige Label zu finden, die mit den entsprechenden Komplexitätsstufen assoziiert werden sollten. Die originären Bezeichnungen der Stufen aus dem *Model of Hierarchical Complexity* (primary, concrete, abstract, formal, systematic) wirkten (insbesondere in einer deutschen Übersetzung) intuitiv wenig fassbar und bisweilen missverständlich konnotiert, so dass eine direkte Übernahme nicht sinnvoll erschien. Daher entstand ein modellassoziertes, aber im Grunde eigenständiges Komplexitätsschema für den Inhaltsbereich Chemie mit den folgenden fünf hierarchisch angeordneten Stufen (vgl. Abb.7):

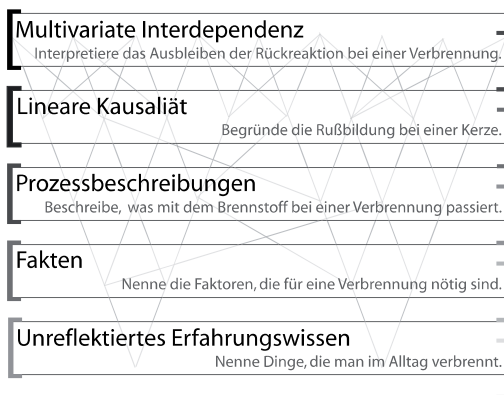


Abb. 7: Adaptiertes Komplexitätsschema mit Beispielaufgaben

Stufe 1: Unreflektiertes Erfahrungswissen



Hierzu zählen Wissensbestände, die Schüler/-innen i. d. R. nicht im Unterricht erworben haben sondern die z. B. aus dem Alltag und der Lebenswelt der Schüler/-innen stammen. Dazu zählt insbesondere die Nennung von Beispielen oder offensichtlicher Beobachtungen.

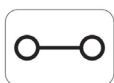
Der Bezug zum Alltag ist dabei nicht als Kontext zu verstehen; bei einer derartigen Zuordnung ließen sich Zusammenhänge auf allen hier dargestellten Komplexitätsstufen denken. Vielmehr sind unter dieser Stufe isolierte Elemente zu fassen, die ohne Bezug zu einem fachsystematischen Gefüge genutzt werden.

Stufe 2: Fakten



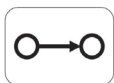
Diese Stufe umfasst die Nennung isolierter Begriffe und Gesetzmäßigkeiten, die im Unterricht erworben wurden. Dabei werden keine Zusammenhänge erläutert, Erklärungen oder Argumentationen formuliert oder Vergleiche herangezogen, sondern lediglich isolierte Fakten chemischen Fachwissens reproduziert.

Stufe 3: Prozessbeschreibungen



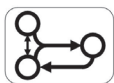
Hierunter fällt die Beschreibung von Vorgängen, die einen zeitlichen Verlauf aufweisen, mit oder ohne Zuhilfenahme von (mental) Modellen. Dies umfasst jedoch keine simplen Vorher/Nachher-Beobachtungen sondern eher die Fokussierung auf den Prozess wie z.B. die Beschreibung einer Mechanismusabfolge.

Stufe 4: Lineare Kausalität



Hierzu zählt die Erfassung und Beschreibung von linearen Ursache-Wirkungs-Zusammenhängen. Dabei ist die Begründung des Zusammenhangs von zentraler Bedeutung, wodurch sich Operationen dieser Stufe von einfachen wenn-dann-Beziehungen abgrenzen lassen.

Stufe 5: Multivariate Interdependenz



Hier sind Begründungen gemeint, die über eindimensionale Zusammenhänge hinausgehen. Dazu gehört die Handhabung mehrerer Variablen, deren Zusammenspiel komplexe Ursachen-/Wirkungszusammenhänge hervorrufen. Bei Zusammenhängen, in denen verschiedene lineare Beziehungen eine Rolle spielen, steht die Analyse der Überlagerung und der wechselseitigen Beeinflussung dieser Beziehungen im Fokus.

Analysen der Überlagerung und der wechselseitigen Beeinflussung dieser Beziehungen im Fokus.

In der Adaption des MHC-C sind die Kernelemente des Modells erhalten geblieben. Die fünf Stufen bauen hierarchisch aufeinander auf, Zusammenhänge auf höheren Stufen liefern immer auch Beziehungen darunter liegender Stufen und organisieren deren Elemente in gesetzmäßiger Weise. Die Stufen unterscheiden sich damit weiterhin durch einen Anstieg der Komplexität auf quantitativer und qualitativer Ebene.

Das MHC-C bietet die Möglichkeit, Unterrichtsinhalte systematisch zu strukturieren. Auch wenn die Struktur durch die kognitionspsychologische Vorlage bestimmt wird, wurde somit eine enge fachliche Anbindung erreicht. Diese sollte eine gute Basis für die Entwicklung und Zuordnung von Aufgaben darstellen. Die Strukturierung von Unterrichtsinhalten offenbart zudem eine deutliche Nähe zum üblichen Vorgehen der Unterrichtsgestaltung im Sinne zunehmender Komplexität, so dass an dieser Stelle auch eine Anschlussfähigkeit zu einem Kompetenzentwicklungsmodell gegeben scheint. Derartige Möglichkeiten bedürfen jedoch weiterer Untersuchungen. Die inhaltsstrukturierende Ausrichtung der Stufen rückt zudem von einer Einschätzung der Aufgabenschwierigkeit in Abhängigkeit von konkreten Beispielen im vorangegangenen Unterricht ab (vgl. KMK 2004, 2005b) und entspricht somit eher einer kriterium-orientierten Bezugsnormierung. Dies trifft mit Einschränkung auch auf die o.g. Modelle zu (vgl. Kap. 2.2), wobei es insbesondere zum ESNaS-Modell eine gewisse Nähe aufweist. Dort findet sich neben einer qualitativen Abstufung (Fakt – Zusammenhang – Konzept) auch eine quantitative Abstufung (1 Fakt – 2 Fakten etc.). Die quantitative Begrenzung der Elemente im ESNaS-Modell (max. 2 Fakten bzw. Zusammenhänge) bedeutet eine bessere Kontrolle der Aufgabenschwierigkeit, geht aber auch mit einem Verlust an Freiheit und Vielfalt bei der Aufgabenkonstruktion einher.

Diese alternative Art der Komplexitätsstruktur und die Differenzierung von Komplexität und kognitiver Aktivität als schwierig-

keitsbestimmende Dimensionen im ESNaS-Modell lassen sich dahingehend deuten, dass bei der Konstruktion (und der derzeitigen Weiterentwicklung) eine detaillierte Strukturaufklärung im Fokus stand (und steht). Zusätzliche Dimensionen erleichtern und objektivieren die Ausrichtung von Messinstrumenten und führen so u. U. zu einer Präzisierung der Aufgabeneinstufung, bedingen aber auch eine aufwändigere Handhabung. Das adaptierte Modell hierarchischer Komplexität verzichtet auf die Ausweisung eines Tätigkeitsprofils und orientiert sich bei der Komplexitätsbeschreibung vorwiegend an einer Hierarchisierung durch die Verknüpfungsqualität der Inhaltselemente. Diese Fokussierung auf einen „high-impact“-Faktor führt zu einem weniger umfangreichen Modell, lässt sich aber nur rechtfertigen, wenn dies nicht zu Lasten einer verlässlichen Einschätzung der Aufgabenschwierigkeit führt. Dazu sollen im Folgenden empirische Befunde zur Nutzbarkeit des Modells angeführt werden.

3.4 Empirische Befunde zum MHC-C

Um die angenommene hierarchische Struktur des Modells empirisch zu überprüfen, wurde ein Testinstrument entwickelt. Durch die Verwendung geeigneter Auswertungsverfahren sollten die empirischen Bearbeitungsergebnisse der theoretischen Struktur gegenübergestellt werden. Um die Anwendbarkeit des Modells in allen Jahrgangsstufen nachzuweisen, sollte ein möglichst umfassender Querschnitt der Schülerschaft Berücksichtigung finden. Dazu musste ein Thema ausgewählt werden, das beginnend vom Anfangsunterricht ein Thema der Schulchemie darstellt.

Bei dieser Herangehensweise wird ein erweitertes Verständnis der Inhaltsdimension angelegt. Entwicklung und Anwendung der Basiskonzepte bleiben eine eigene Kompetenz (wie in den Nationalen Bildungsstandards für Chemie vorgesehen; vgl. KMK 2005b), Kernthemen des Unterrichts stellen

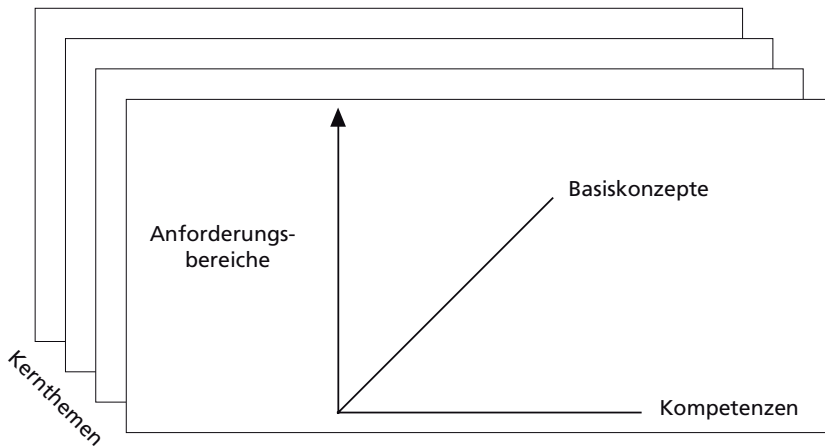


Abb. 8: Aufspaltung der Inhaltsdimension

die Basis für die Entwicklung von Testheften dar, die in nachfolgenden Studien auch vergleichend eingesetzt werden sollen (vgl. Abb. 8). Diese Kernthemen, die spiralcurricular im Unterricht fortlaufend wieder aufgegriffen werden, bieten eher die Möglichkeit, einen Kompetenzzuwachs vergleichbar aufzeigen zu können. Die Basiskonzepte werden dementsprechend zur Strukturierung der Inhalte unter jeweils konzeptspezifischer Sichtweise herangezogen. Bei dieser Aufspaltung der Inhaltsdimension in zentrale Themenfelder auf der einen Seite (z. B. Verbrennung, Säuren und Basen etc.) und den Basiskonzepten auf der anderen Seite, können unterschiedlichen Aspekten Rechnung getragen werden, z. B. dem Kontexteinfluss auf Lernen und Interesse oder unterschiedlichen Schwerpunkten und Schwierigkeitsmerkmalen seitens der Basiskonzepte bei unterschiedlichen Themen.

Vor diesem Hintergrund fiel die Wahl auf den Themenbereich „Verbrennung“. Dieser inhaltliche Rahmen wurde durch die Entwicklung modellbasierter Aufgaben bzw. durch die Einstufung vorhandener Aufgaben operationalisiert. Bei der Aufgabenentwicklung wurde insbesondere auf fachdidaktische Literatur zur Schülervorstellungen im Bereich Verbrennungen zurückgegriffen (deVos, Wobbe & Verdonk 1987; BouJaoude

1991; Kesidou & Duit 1993; Mulford 1996; Schmidt 1997; Kind 2004).

Interviewstudie

Die Aufgaben wurden im Rahmen einer Pilotstudie in problemzentrierten Interviews mit Schüler/-innen der Klassenstufen 9-11 eingesetzt. Die Bearbeitungsergebnisse wurden qualitativ analysiert (vgl. Dawson-Tunik 2006), um die modellbasierte Einstufung der Aufgaben den tatsächlich gegebenen Antworten differenziert gegenüberstellen zu können. Die qualitative Auswertung der Interviewdaten ermöglichte genauere Einblicke in unterschiedliche Bearbeitungsstrategien, individuelle „Stolpersteine“ und in die Bedeutung anderer Faktoren, die eine Aufgabenbearbeitung und eine darauf aufbauende Kompetenzausprägung beeinflussen können (vgl. Abb. 9; Bernholt & Parchmann 2008).

Schriftlicher Test

Die Aufgaben wurden aufbauend auf dieser Vorstudie überarbeitet, ergänzt und dienen als Grundlage zur Entwicklung eines Paper & Pencil-Tests, der sowohl Daten zur reliablen Einstufung der Aufgaben als auch der Modellprüfung liefern sollte. Um möglichst

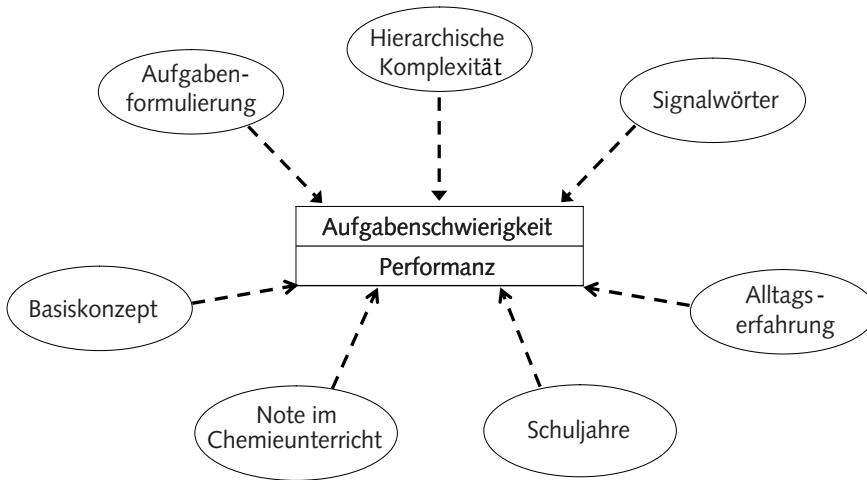


Abb. 9: Vermutete Einflussfaktoren auf Basis der Interviewstudie (Bernholt & Parchmann 2008)

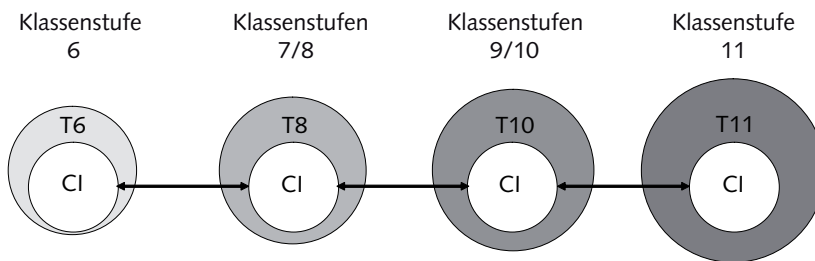


Abb. 10: Common-Item Nonequivalent Groups Design (vgl. Kolen & Brennan 2004, 19)

allgemeine Aussagen über die Anwendbarkeit des adaptierten Modells machen zu können, erschien eine Beschränkung auf einen bestimmten Aufgabentypus wenig sinnvoll. Daher fanden sowohl *selected-response* als auch *constructed-response* Aufgaben Verwendung.

Neben den Chemie-Aufgaben wurde ein bereits erprobtes Instrument zur Erfassung hierarchischer Komplexitätsstufen an den Test angegliedert werden („Laundry Problem Series“; Commons, Miller & Kuhn 1982; Inhelder & Piaget 1958). Die zentrale Anforderung dieser Aufgabenserie ist die Isolierung von Variablen, die durch logisches Denken und Schlussfolgern identifiziert werden können. Dieser getrennt

dargebotene Aufgabenabschnitt diente zum Abgleich der fachbezogenen Ergebnisse, bleibt aber in der folgenden Darstellung unberücksichtigt.

Der Einsatz der zusammengestellten Aufgaben ist in einer Querschnittsstudie über die Jahrgänge 6 bis 11 erfolgt. Dabei musste jedoch berücksichtigt werden, dass es sich um eine äußerst heterogene Probandengruppe handelt. Um generelle Unterschiede (Lesefähigkeit und -geschwindigkeit, Vorwissen, Erfahrungen mit Tests etc.) zu berücksichtigen, wurden vier unterschiedliche Testhefte erstellt (T6, T8, T10 und T11), wobei sich ein Teil der Aufgaben (*common items*: CI) in jedem Testheft deckt (vgl. Abb. 10). Dieses als *Common-Item Nonequivalent Groups-*

Design bezeichnete Vorgehen (Kolen & Brennan 2004, 19ff) wird auch in anderen Querschnittsstudien eingesetzt.

Mit steigender Jahrgangsstufe nimmt dabei sowohl der Umfang an Testaufgaben (angedeutet durch die größeren Kreise) als auch der Anteil schwieriger Aufgaben (angedeutet durch die dunklere Färbung) zu. Ein Großteil der Aufgaben wird jedoch von den Schüler/-innen aller Jahrgangsstufen bearbeitet und kann so für einen Vergleich der Jahrgänge herangezogen werden.

Alle Testhefte wurden dabei von den Probanden im Zeitrahmen einer Schulstunde bearbeitet. Diese Vorgabe wurde aus pragmatischen Gründen getroffen, um die Bereitschaft von Lehrer/-innen und Schüler/-innen zur Teilnahme an dieser Untersuchung nicht übermäßig zu strapazieren. Zudem sind Schüler/-innen aller Jahrgangsstufen diese Testlänge gewöhnt, so dass es nicht zu massiven Ermüdungserscheinungen während der Testdurchführung kommen sollte. Für die Teilnahme an der Studie konnten schließlich 19 Klassen (N=460) gewonnen werden.

Zur Auswertung der offenen Aufgaben ($n_o = 16$, – 10–14 pro Testheft) wurde ein ausführliches Kodiermanual entwickelt, bei den geschlossenen Aufgaben ($n_g = 52$; 30–50 pro Testheft) wurden die ausgewählten Antworten entsprechend ihrer Position erfasst. Die Auswertung der offenen Aufgaben ist nicht vollständig objektivierbar, sondern stets von subjektiven Deutungsmustern geprägt, dennoch konnte mit einem Wert von $\kappa_o = 0,77$ eine hohe Interrater-Reliabilität nachgewiesen werden (berechnet anhand von 14% (65 von 460) des Datensatzes; Cohen 1960). Alle Antwortcodes wurden abschließend in eine richtig/falsch-Kategorie umkodiert.

Die kodierten Antworten wurden unter Verwendung des dichotomen Rasch-Modells ausgewertet (Rasch 1960). Dieses Modell der probabilistischen Testtheorie unterliegt der Grundannahme, dass der Bearbeitung aller Items ein gemeinsames latentes Konstrukt zu Grunde liegt. Die Lösungswahrschein-

lichkeit einer Aufgabe hängt dann nur von deren Schwierigkeit und der Fähigkeit der Person ab. Dieses Modell erscheint für die vorliegende Studie am ehesten geeignet, da es den Grundannahmen des *Model of Hierarchical Complexity* weitgehend entspricht. Da das Konstrukt der hierarchischen Komplexität als maßgeblicher Faktor der Aufgabenschwierigkeit angenommen wird und auch das Fähigkeitsniveau der Probanden mit Hilfe dieses Konstrukts beschrieben werden kann (vgl. Commons & White 2003), sollte das eindimensionale Rasch-Modell diese Theorieannahmen am ehesten repräsentieren. Diese Vermutung wird durch eine ausreichend hohe Reliabilität (EAP/PV-Reliabilität: 0,64–0,84, je nach Testheft) und einer Überprüfung der Item-Kennwerte gestützt (64 von 66 Items erfüllen die Bedingungen $0,8 < \text{weighted MNSQ} < 1,2$ bzw. $T < 2,0$; bestimmt durch ACER ConQuest, vgl. Wu, Adams & Wilson 1998).

Ziel war es aufzudecken, inwieweit sich in den empirischen Daten die theoretische Einstufung der Aufgaben wiederfindet bzw. inwieweit die Einstufung in der Lage ist, die Aufgabenschwierigkeit vorherzusagen. Diese Auswertung wurde zunächst für jedes Testheft vorgenommen, um die Anwendbarkeit des Modells in den entsprechenden Jahrgangsstufen, also innerhalb eines Doppeljahrgangs, zu überprüfen. Mittels einfaktorierter Varianzanalyse (UniANOVA) konnte mit dem Item-Parameter der Rasch-Skalierung als Abhängige Variable und der Komplexitätsstufe der Aufgabe als Unabhängige Variable eine Varianzaufklärung durch die theoretische Einstufung der Aufgaben von über 65% berechnet werden. Die Auswertung des Zusammenhangs zwischen theoretischer Einstufung und empirischer Aufgabenschwierigkeit (also dem Item-Parameter der Rasch-Skalierung) wurde für jedes Testheft und dort wiederum für jeden Testteil (A – geschlossen/Chemie; B – offen/Chemie) einzeln bestimmt. Dies ergab für die Testteile A und B jeweils folgende Minimalwerte der UniAnova-Berechnung hinsichtlich der vier Jahrgangstesthefte: A (geschlossene

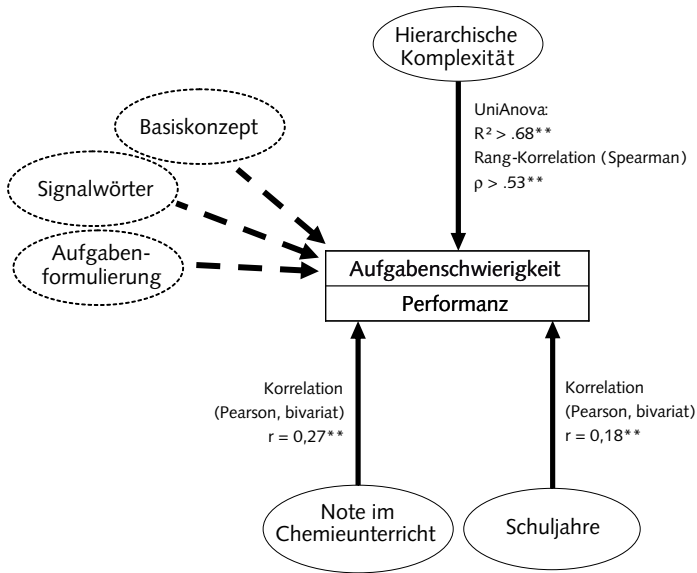


Abb. 11: Übersicht über nachgewiesene (schwarze Pfeile) und vermutete (gestrichelte Pfeile) Einflussfaktoren

Aufgaben/Chemie): $R^2 = 0,68^{**}$; B (offene Aufgaben/Chemie): $R^2 = 0,70^{**}$ (Da die Werte der Varianzaufklärung für die vier Testheft jeweils äußerst nah beieinander lagen, wurde an dieser Stelle auf eine getrennte Angabe verzichtet). Der Wert für die Rang-Korrelation der Modellstufen (Spearman's Rho) lag je nach Testteil zwischen $\rho = 0,53^{**}$ und $\rho = 0,64^{**}$ (vgl. Abb. 11).

Des Weiteren wurde zur Untersuchung des Einflusses von Alter und Unterrichtserfahrung der Probanden eine Auswertung über den Jahrgangsquerschnitt der Probandengruppe vorgenommen. Für diesen Jahrgangvergleich wurden die Ergebnisse der Rasch-Skalierung der einzelnen vier Testhefte mittels eines Test-Equating-Verfahrens auf eine gemeinsame Skala transformiert, um die Ergebnisse miteinander in Beziehung setzen zu können (*concurrent calibration*; vgl. Kolen & Brennan 2004; Morrison & Fitzpatrick 1992). Bei der weiteren Auswertung ließen sich Korrelationen zwischen Personen-Parameter und Chemie-Note der Schüler/-innen (Pearson, bivariat: $r = 0,27^{**}$, $p < 0,01$) bzw. zwischen Personen-Parameter und Jahrgangsstufe der Schüler/-

innen (Pearson, bivariat: $r = 0,18^{**}$, $p < 0,01$) feststellen (vgl. Abb. 11). Gendereinflüsse ließen sich in den Daten ebenso wenig finden wie nennenswerte Korrelationen zur Deutsch- oder Physiknote.

Eine qualitative Auswertung der Aufgaben (insbesondere bei Abweichungen von theoretischer Einstufung und dem Item-Parameter der Rasch-Skalierung) ergab verschiedene Aufgabenattribute, die sich zu wenigen Einflussfaktoren systematisieren ließen. So ließ sich der schwierikeitssteigernde Einfluss des Basiskonzepts Energie, der bereits in der Interviewstudie beobachtet wurde (vgl. Bernholt & Parchmann 2008), erneut finden. Zudem konnte ein Effekt auf die Aufgabebearbeitung in den Multiple-Choice-Fragen auch anhand bestimmter „Signalwörter“ festgemacht werden, wie z. B. „Sauerstoff“ oder „Wasserstoff“, bei denen die Schüler/-innen tendenziell entschieden, dass es zu einer Verbrennung kommen müsste, auch wenn die Konstellation der weiteren Brandbedingungen (z. B. kein adäquater Verbrennungspartner oder mangelnde Aktivierungsenergie) dies verhinderten. Hier schienen sich die Schüler/-innen leicht verunsichern zu

lassen. Auf Grund der Aufgabenkonstellation ließen sich diese Effekte jedoch nicht quantifizieren.

Übertragung auf andere Kernthemen

Um die Aussagekraft der bisherigen Ergebnisse zu stützen, wurde das MHC-C auf weitere zentrale Themengebiete des Chemieunterrichts übertragen. Zum Thema „Säuren und Basen“ wurde ein schriftlicher Test mit offenen und geschlossenen Aufgaben in der Jahrgangsstufe 10 eingesetzt (N=126; Scholz 2008), zum Themengebiet „Redox-Reaktionen“ wurde ebenfalls ein schriftlicher Test (offene und geschlossene Aufgaben) in den Jahrgängen 11 und 12 eingesetzt (N=113; Klostermann 2008). Die Auswertung beider Testdurchführungen ergab vergleichbare Werte in der Varianzaufklärung wie der beschriebene Test zu Verbrennungsprozessen. Die zentralen Testspezifikationen und Auswertungsergebnisse aller drei Untersuchungen sind in Tabelle 1 nochmals zusammengefasst. Mit dem geringen Modellumfang konnte somit eine sehr hohe Varianzaufklärung erreicht werden. Eine detaillierte qualitative Analyse der dennoch natürlich vorhandenen

Abweichungen ließ zudem vermuten, dass diese Abweichungen durchaus systematisierbar sind. Hier könnten sich Hinweise für die Ausdifferenzierung eines umfangreicheren Kompetenzmodells finden lassen. Die Notwendigkeit und der Mehrwert eines umfangreicheren Modells sind dabei aber noch zu klären, denn das Ziel einer Weiterentwicklung des Modells ist insbesondere eine gesicherte Nutzbarkeit in der Unterrichtspraxis.

Trotz der vielversprechenden Untersuchungsergebnisse sei nochmals darauf hingewiesen, dass Aufgabenkomplexität und Aufgabenschwierigkeit nicht gleichzusetzen sind. Die hierarchische Komplexität einer Aufgabenstellung trägt (wie oben bereits theoretisch beschrieben und empirisch bestätigt) zu einem Großteil zur Varianzaufklärung bei, ist mit dem Konstrukt der Aufgabenschwierigkeit jedoch nicht identisch. Diese ist sicherlich nicht unabhängig von Inhaltsbereich und anderen nicht-stuffbaren Aufgabefaktoren, wie auch andere Modelle vermuten (vgl. Kauertz 2008, Fischer & Draxler 2001, Klieme 2000). Insbesondere der Einfluss des Inhaltsbereiches ist mit den bisher zu diesem Modell durchgeführten Studien nicht zu präzisieren. Hier kann

Tab. 1: Übersicht über die bisherigen empirischen Ergebnisse im Bereich Fachwissen

	Verbrennungsprozesse	Säuren und Basen (Scholz 2008)	Redox-Reaktionen (Klostermann 2008)
Jahrgangstufen	6-11	10	11-12
Probanden (N)	460	126	113
Aufgabenformat (Anzahl n)	geschlossen (52) offen (16)	geschlossen (10) offen (14)	geschlossen (6) offen (16)
Reliabilität (EAP/PV)	0,65	0,74	0,85
UniAnova (R ²)	0,68** – 0,70**	0,57**	0,54**
Rang-Korrelation (ρ)	0,53** – 0,64**	0,75**	0,65**

(Bei den Ergebnissen der beiden Erweiterungsstudien sei darauf hingewiesen, dass die Stichprobengröße jeweils am unteren Rand der für eine reliable Parameterschätzung notwendigen Größe liegt.)

weiteres Varianzpotenzial vermutet werden (vgl. Einhaus 2007), dessen Wirkung auf die Vorhersagekraft der Modellstufen noch abgeklärt werden muss.

Untersuchungen zu anderen Kompetenzbereichen

In einer ersten explorativen Studie wurde die Übertragbarkeit des adaptierten Modells auf den Kompetenzbereich „Bewertung“ analysiert (Wienert 2008). Dazu wurde die Beschreibung der Modellstufen zum Bereich Fachwissen (vgl. Abb. 7) mit dem Göttinger Modell zur Bewertungskompetenz (vgl. Eggert & Bögeholz 2006) abgeglichen. Der Vergleich zeigte viele Gemeinsamkeiten zwischen den Modellen (z. B. die Relationstruktur der Inhaltelemente auf den jeweiligen Stufen), wobei sie sich in einigen Punkten jedoch auch unterscheiden, was durch die unterschiedlichen Anforderungen der Kompetenzbereiche bedingt scheint. Unterschiede zeigten sich beispielsweise darin, dass im Göttinger Modell zur Bewertungskompetenz die Anforderungen unterschiedliche Facetten aufweisen (z. B. Normbezug und Perspektivenübernahme), was sich nicht direkt im oben beschriebenen Modell zum Fachwissen wiederfinden ließ.

Das für die empirische Untersuchung angewandte Modell beinhaltet Grundideen der beiden Modelle und zeigt eine Graduierung der Kompetenz in fünf Stufen: Von Stufe 1

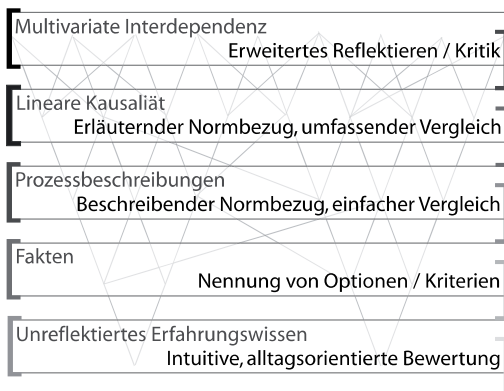


Abb. 12: Adaptierte Modellstufen zu den Kompetenzbereichen "Fachwissen" und "Bewertung" (verändert nach Wienert 2008)

bis Stufe 5 ist eine steigende Komplexität und Strukturierung der Antworten, eine steigende Perspektivenübernahme der Schüler/-innen und eine vermehrte Reflexion der Aufgabeninhalte zu erkennen (vgl. Abb. 12).

Zur Untersuchung der Modellstufen wurden Aufgaben zu Bewertungsaspekten am Beispiel alternativer Kraftstoffe entwickelt. Diese drei offenen und mehrschrittigen Aufgaben zu den Themen Treibstoffe, Biodiesel und BtL-Diesel wurden in vier Klassen (N=119) eingesetzt und die Bearbeitungsergebnisse u. a. mit Hilfe der Modellstufen ausgewertet (Interrater-Reliabilität Cohens $\kappa=0,82$). Dabei zeigte sich, dass sich eine Zuordnung der Schülerantworten vorwiegend auf die unteren drei Modellstufen beschränkte, mit wenigen Ausnahmen auf Stufe 4. Dies kann u. U. auf die Aufgabenstellungen zurückgeführt werden, legt aber durchaus auch den nicht unerwarteten Schluss nahe, dass Anforderungen von Bewertungsaufgaben für die Schüler/-innen ungewohnt und dementsprechend schwierig sind.

Des Weiteren wurde das Modell zur Aufgabenkonstruktion in einer Studie zum Kompetenzbereich „Kommunikation“ herangezogen. In dem Projekt „F₂L“ wird untersucht, welche Vorstellungen Schüler/-innen mit der chemischen Formelsprache verknüpfen und mit welchen Hilfestellungen (z. B. Analogiebildungen) sie sich das System der chemischen Formelsprache erarbeiten können (vgl. Heuer & Parchmann 2008). Die Untersuchung der Formelsprache als chemietypische Kommunikationsform soll dabei erste Hinweise für eine Strukturierung dieses Bereiches liefern.

3.5 Mögliche Ansätze zur Förderung systematischer Komplexitätsentwicklung im Unterricht

Die präsentierten Ergebnisse unterstützen die Annahme, dass sich das zu Grunde liegende Modell eignet, die Fähigkeitsstruktur von Schüler/-innen abzubilden. Der zu Beginn eingeleiteten Diskussion folgend

(vgl. Kap. 2) ist weiterhin zu klären, welche Ansätze das Modell für den Unterricht zur Umsetzung curricularer Bildungsziele liefert. Dazu sollen zwei Möglichkeiten aufgezeigt werden.

Neben der Diagnose stellen Mapping-Methoden auch eine Möglichkeit dar, Vernetzung von Inhalten und kumulatives Lernen zu fördern, wie verschiedene Studien bestätigen (vgl. Übersicht nach Renkl & Nückles 2006, 141ff). Dabei lassen sich prototypische Strukturen in einen direkten Zusammenhang mit den o.g. adaptierten Modellstufen bringen: Abstraktions- (Überordnungs-Unterordnungs-Relationen) und Komplexionshierarchien (Teil-Ganzes-Relationen; vgl. Stufe 2: Fakten); zeitliche Abläufe (vgl. Stufe 3: Prozessbeschreibungen); deterministische und probabilistische Kausalmodelle (vgl. Stufe 4: Lineare Kausalität); institutionalisierte Handlungsstrukturen (vgl. Stufe 5: Multivariate Interdependenz; vgl. Renkl & Nückles 2006). Die explizite Berücksichtigung der o.g. Strukturtypen von Concept Maps bzw. deren Verknüpfung mit den Modellstufen liefert an dieser Stelle Ansätze sowohl für Lerner als auch Lehrkräfte. Ausgehend von selbst-erstellten Concept Maps mit eher assoziativen Strukturen zu Beginn einer Unterrichtseinheit, können Lerner versuchen, gezielt weitere Verknüpfungen herzustellen und insbesondere vorhandene Verknüpfungen zu vertiefen (im Sinne des Aufbaus von Kausal- und Interdependenzbeziehungen), wobei die Gesamtstruktur der Concept Maps natürlich jeweils angepasst und die Konstellation der Knoten und Relationen neu abgestimmt werden müssen. Dennoch bietet sich Lernern durch dieses Vorgehen ein Anknüpfungspunkt, der ihnen nicht nur aufzeigt, wo sie ansetzen können, sondern auch wie. Die empirischen Ergebnisse dazu sind noch nicht eindeutig, so dass an dieser Stelle weiterer Forschungsbedarf besteht.

Des Weiteren scheint auch die Kombination der Verwendung spezifischer naturwissenschaftlicher Denkweisen mit bewusster Reflexion des Vorgehens (Metakognition) auf Basis des vorgestellten Modells möglich. Da-

bei müsste den Schüler/-innen ermöglicht werden, die Qualität ihres Wissens einzuschätzen, um darauf aufbauend eigenständig Lernwege erarbeiten zu können, die zu einer tieferen Verarbeitung des Lernstoffs führen (im Sinne des Verstehens von Kausalität und Interdependenz). Erste Erprobungen mit entsprechend gestalteten Aufgaben werden zurzeit durchgeführt.

Damit bietet das vorgestellte Modell einen Anker zu einem lernzielorientierten Testverfahren, was zu einer größeren Überlappung von unterrichtlichen Lernzielen und modellbasierten Testanforderungen führen könnte. "Teaching to the test" would, then, turn out to be positive; it would involve teaching the kinds of knowledge, understanding, and reasoning skills that constitute the outcomes of [...] scholarship" (Shavelson 2007, 20). Für eine Darstellung erster Ansätze und Erfahrungen zur Anwendung des Modells im Unterricht vgl. Bernholt et al. (2009).

4 Zusammenfassung und Ausblick

Das vorgestellte Modell hierarchischer Komplexität hat sich im Rahmen der bisherigen Untersuchungen als reliables und gewinnbringendes Diagnoseinstrument im Kompetenzbereich Fachwissen erwiesen. Die potenziellen Möglichkeiten zur Ableitung instruktionsrelevanter Informationen und für eine Nutzbarkeit im Unterricht wurden angedeutet, ebenso wurden erste Schritte zur Ausweitung der Modellnutzung im Kompetenzbereich Bewertung dargestellt. Das genaue Potenzial muss selbstredend in weiteren Untersuchungen erprobt werden. Dazu gehört auch eine Analyse der Möglichkeiten, fachübergreifende Lernziele (z. B. Problemlösen, *higher-order thinking*; vgl. Zoller 2002) stärker zu forcieren.

Ob das MHC-C neben der Beschreibung eines Kompetenzstands auch zur Analyse von Kompetenzentwicklung geeignet ist, kann noch nicht beantwortet werden. In diesem Bereich besteht insgesamt ein starker Forschungsbedarf, denn zur Erfas-

sung der zeitlichen und inneren Dynamik des Kompetenzaufbaus und zur Entwicklung von Prozessmodellen und explikativen Theorien mangelt es an umfangreich angelegten Längsschnittstudien.

In beiden Modellierungsbereichen sind jedoch nicht nur die Forschungsperspektive, sondern auch verstärkt die Anforderungen der Unterrichtspraxis zu berücksichtigen. Diese Felder sind vernetzt und von daher nicht isoliert voneinander zu betrachten. Dementsprechend bedarf es engerer Kooperationsbemühungen zwischen Forschung und Praxis und einer kritischen Auseinandersetzung mit der Notwendigkeit bzw. dem Mehrwert umfangreicher, mehrdimensionaler Modelle. Jede Präzisierung durch zusätzliche Dimensionen und Komponenten erleichtert und objektiviert die Ausrichtung von Messinstrumenten, geht aber (ganz abgesehen von der Praktikabilität) immer auch mit Einengung und dem Verlust von Vielfalt auf Seiten des Zielbereichs und somit insgesamt der Aussagekraft der Messergebnisse einher. „Das Dilemma besteht überall, wo eine präzise Messung angestrebt wird, die unausweichlich mit der (praktischen) Relevanz des Gemessenen in Konflikt gerät“ (Herzog 2007, 62; vgl. das Bandbreite-Zuverlässigkeitsdilemma in Cronbach 1971, 179ff).

Danksagung

Wir danken den zahlreichen Schülerinnen, Schülern, Lehrerinnen und Lehrern für die Teilnahme an den Erprobungen und den konstruktiven Rückmeldungen. Darüber hinaus gilt unserer besonderer Dank Julia Michaelis, Mareike Klostermann, Andrew M. Richardson, Sara N. Ross, Anja Scholz und Beeke Wienert für ihre engagierte Mitarbeit im Rahmen der Modell- und Aufgabenentwicklung. Ebenso danken wir den Gutachtern für den konstruktiv kritischen Diskurs im Rahmen des Review-Verfahrens.

Literatur

- Adey, P. (1991). Cognitive Acceleration through Science Education. In S. Maclure & P. Davies (Eds.), *Learning to think* (pp. 79-93). Oxford: Pergamon Press.
- Adey, P. & Shayer, M. (1994). *Really raising standards: Cognitive intervention and academic achievement*. London: Routledge.
- Adey, P., Shayer, M. & Yates, C. (1995). *Thinking Science: The curriculum materials of the Cognitive Acceleration through Science Education (CASE) project*. 2nd edition. London: Thomas Nelson and Sons Ltd.
- Aebli, H. (1980). *Denken: Das Ordnen des Tuns. Band I*. Stuttgart: Klett-Cotta.
- Aebli, H. (1981). *Denken: Das Ordnen des Tuns. Band II*. Stuttgart: Klett-Cotta.
- Anderson, R. C. (1984). Some reflections on the acquisition of knowledge. *Educational Researcher*, 5-10.
- Anderson, J. R. & Lebiere, C. (1998). The atomic components of thought. Hillsdale, NJ: Erlbaum.
- Aufschnaiter, C. v. (2003). *Videobasierte Analysen von Lern- und Lehrprozessen in physikalischen Kontexten*. Kumulative Habilitationsschrift. Universität Hannover 2003.
- Bernholt, S. & Parchmann, I. (2008). Lösungsstrategien bei der Bearbeitung von Aufgaben. In D. Höttercke (Eds.), *Kompetenzen, Kompetenzmodelle, Kompetenzentwicklung*, Tagungsband der Gesellschaft für Didaktik der Chemie und Physik (pp. 215-217). Münster: Lit.
- Bernholt, S., Walpuski, M., Sumfleth, E. & Parchmann, I. (2009). Kompetenzentwicklung im Chemieunterricht – mit welchen Modellen lassen sich Kompetenzen und Aufgaben differenzieren? *Naturwissenschaften im Unterricht – Chemie*, Heft 111/112, 78-85.
- Bloom, B. S. (Eds.) (1956). *Taxonomy of Educational Objectives, the Classification of Educational Goals – Handbook I: Cognitive Domain*. New York: McKay.
- BouJaoude, S. B. (1991). A study of the nature of students' understandings about the concept of burning. *Journal of Research in Science Teaching*, 28(8), 689-704.
- Burrmann, U. (1999). Das Interventionsprogramm CASE von Adey und Shayer im Lichte bisheriger Evaluationen. *Zeitschrift für Pädagogische Psychologie*, 13(1/2), 17-23.
- Bybee, R. (1997). Toward an Understanding of Scientific Literacy. In W. Gräber & C. Bolte (Eds.), *Scientific Literacy* (pp. 37-68). Kiel: IPN.

- Case, R. (1985). *Intellectual development: A systematic reinterpretation*. New York: Academic Press.
- Case, R. (1992). Neo-Pagetian theories of intellectual development. In H. Beilin & P. B. Pufall (Eds.), *Piaget's theory: Prospects and possibilities* (pp. 61-104). Hillsdale, NJ: Lawrence Erlbaum.
- Christiansen, D. (2007). *Entwicklung und Erprobung von Aufgaben zur Erfassung zentraler Kompetenzen im Chemieunterricht am Beispiel Säuren und Basen*. Dissertation zur Erlangung des Doktorgrades der Mathematisch-Naturwissenschaftlichen Fakultät der Christian-Albrechts-Universität zu Kiel.
- Cohen, J. (1960). A coefficient for agreement for nominal scales. *Education and Psychological Measurement*, 20, 37-46.
- Collins, A. M. & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407-428.
- Collins, A. M. & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240-247.
- Commons, M. L., Miller, P. M. & Kuhn, D. (1982). The Relation Between Formal Operational Reasoning and Academic Course Selection and Performance Among College Freshmen and Sophomores. *Journal of Applied Developmental Psychology*, 3, 1-10.
- Commons, M. L. & Richards, F. A. (1984a). A general model of stage theory. In M. L. Commons, F. A. Richards, & C. Armon (Eds.), *Beyond formal operations* (Vol. 1: Late adolescent and adult cognitive development) (pp. 120-140). New York: Praeger.
- Commons, M. L. & Richards, F. A. (1984b). Applying the general stage model. In M. L. Commons, F. A. Richards, & C. Armon (Eds.), *Beyond formal operations* (Vol. 1: Late adolescent and adult cognitive development) (pp. 141-157). New York: Praeger.
- Commons, M. L., Trudeau, E. J., Stein, S. A., Richards, F. A. & Krause, S. R. (1998). Hierarchical Complexity of Tasks Shows the Existence of Developmental Stages. *Developmental Review*, 18(3), 237-278.
- Commons, M. L. & White, M. S. (2003). A Complete Theory of Tests for a Theory of Mind Must Consider Hierarchical Complexity And Stage. *Behavioral And Brain Science*, 26(5), 606-607.
- Cronbach, L. J. (1971). *Essentials of Psychological Testing*. 3rd Edition. New York, NY: Harper & Row.
- Dawson, T. L. (2003). A Stage is a Stage is a Stage. *The Journal of Genetic Psychology*, 164(3), 335-364.
- Dawson-Tunik, T. L. (2006). Stage-like patterns in the development of conceptions of energy. In X. Liu & W. Boone (Eds.), *Applications of Rasch measurement in science education* (pp. 111-136). Maple Grove, MN: JAM Press.
- Demetriou, A., Platsidou, M., Efklides, A., and Metallidou, Y. (1991). The development of quantitative-relational abilities from childhood to adolescence: Structure, scaling, and individual differences. *Learning and Instruction*, 1, 19-43.
- deVos, W. & Verdonk, A. H. (1987). A New Road to Reactions, Part 4. The Substance and its Molecules. *Journal of Chemical Education*, 64(8), 692-694.
- Ditton, H. (2002). *Evaluation und Qualitätssicherung*. In R. Tippelt (Eds.), *Handbuch Bildungsforschung* (pp. 775-790). Opladen: Leske + Budrich.
- Eggert, S. & Bögeholz, S. (2006). Göttinger Modell der Bewertungskompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 177-197.
- Einhaus, E. (2007). Schülerkompetenzen im Bereich Wärmelehre – Entwicklung eines Testinstruments zur Überprüfung und Weiterentwicklung eines normativen Modells fachbezogener Kompetenzen. In: *Studien zum Physik- und Chemielernen*, Band 63, Berlin: Logos.
- Fischer, H. E. & Draxler, D. (2002). Konstruktion und Bewertung von Physikaufgaben. In E. Kircher & W. B. Schneider (Eds.), *Physikdidaktik in der Praxis* (pp. 300-322). Berlin: Springer.
- Fischer, H. E. & Draxler, D. (2001). Aufgaben und naturwissenschaftlicher Unterricht. *Der mathematische und naturwissenschaftliche Unterricht* (MNU), 54(7), 388-393.
- Fischer, H. E., Glemnitz, I., Kauertz, A., Sumfleth, E. (2007). Auf Wissen aufbauen – kumulatives Lernen in Chemie und Physik. In E. Kircher, R. Girwidz & P. Häußler (Eds.), *Physikdidaktik – Theorie und Praxis* (pp. 657-678). Berlin: Springer.
- Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, 87, 477-531.
- Fischer, K. W. & Bidell, T. R. (2006). Dynamic development of action, thought, and emotion. In W. Damon & R. M. Lerner (Eds.), *Theoretical models of human development. Handbook of child psychology* (6th Edition., Vol. 1, pp. 313-399). New York: Wiley.

- Gagné, R. M. (1973). *Die Bedingungen des menschlichen Lernens*. 3. Auflage. Hannover: Schroedel.
- Gräsel, C. & Parchmann, I. (2004). Die Entwicklung und Implementation von Konzepten situierter, selbstgesteuerten Lernens. *Zeitschrift für Erziehungswissenschaft*, 7(3), 171-184.
- Hammann, M. (2004). Kompetenzentwicklungsmodelle: Merkmale und ihre Bedeutung - dargestellt anhand von Kompetenzen beim Experimentieren. *Der mathematische und naturwissenschaftliche Unterricht* (MNU), 57(4), 196-203.
- Herzog, W. (2007). Pro und Kontra Bildungsstandards. Die Perspektive eines Skeptikers. In P. Labbude (Eds.), *Bildungsstandards am Gymnasium – Korsett oder Katalysator?* (pp. 57-64). Bern: hep.
- Heuer, S. & Parchmann, I. (2008). SO_2 oder Fußbal_2 – wie Sechstklässler die chemische Formelsprache interpretieren. *Naturwissenschaften im Unterricht – Chemie*, 19(3), Heft 106/107, 20-24.
- Holley, C. D., Dansereau, D. F., McDonald, B. A., Garland, J. C. & Collins, K. W. (1979). Evaluation of hierarchical mapping technique as an aid to prose processing. *Contemporary Educational Psychology*, 4, 227-237.
- Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York, Basic Books.
- Jonassen, D. H., Ionas, I. G. (2008). Designing effective support for causal reasoning. *Educational Technology Research and Development*, 56, 287-308.
- Kauertz, A. (2008). Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben. In: *Studien zum Physik- und Chemielernen*. Band 79. Berlin: Logos.
- Kauertz, A. & Fischer, H. E. (2006). Assessing Students' Level of Knowledge and Analysing the Reasons for Learning Difficulties in Physics by Rasch Analysis. In X. Liu & W. Boone (Eds.), *Applications of Rasch Measurement in Science Education* (pp. 212-246). Maple Grove, MA: Jam Press.
- Kesidou, S. & Duit, R. (1993). Students' Conceptions of the Second Law of Thermodynamics, an Interpretive Study. *Journal of Research in Science Teaching*, 30(1), 85-106.
- Kind, V. (2004). *Beyond Appearances: Students' misconceptions about basic chemical ideas*, 2nd Edition, School of Education, Durham University, UK (self-published).
- Klieme, E. (2000). Fachleistungen im voruniversitären Mathematik- und Physikunterricht: Theoretische Grundlagen, Kompetenzstufen und Unterrichtsschwerpunkte. In J. Baumert et al. (Eds.), *TIMSS – Mathematisch-naturwissenschaftliche Bildung am Ende der Sekundarstufe II* (pp. 57-128). Opladen: Leske & Budrich.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M. et al. (2003). *Zur Entwicklung nationaler Bildungsstandards*. Bonn: Bundesministerium für Bildung und Forschung.
- Klieme, E., Baumert, J., Köller, O. & Bos, W. (2000). Mathematische und naturwissenschaftliche Grundbildung: Konzeptuelle Grundlagen und die Erfassung und Skalierung von Kompetenzen. In J. Baumert, W. Bos & R. Lehmann (Eds.), *Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* (Band 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit) (pp. 85-133). Opladen: Leske + Budrich.
- Klostermann, M. (2008). *Entwicklung und Einstufung von Aufgaben zum Themengebiet Redoxreaktionen*. Unveröffentlichte Bachelor-Arbeit. Universität Oldenburg.
- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Eds.) (2005a). *Bildungsstandards der Kultusministerkonferenz; Erläuterungen zur Konzeption und Entwicklung*. München: Luchterhand.
- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Eds.) (2005b). *Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss*. München: Luchterhand.
- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Eds.) (2004). *Einheitliche Prüfungsanforderungen in der Abiturprüfung (Chemie)*. München: Luchterhand.
- Kolen, M. J. & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking. Methods and Practices*. 2nd Edition. New York: Springer.
- Krüssel, H. (2001). Lernen und Leistungsbewertung als Elemente einer konstruktivistischen Lernkultur. In C. Solzbacher & C. Freitag (Eds.), *Anpassen, verändern, abschaffen? Schulische Leistungsbewertung in der Diskussion* (pp. 123-146). Bad Heilbrunn: Verlag Julius Klinkhardt.
- P. Labbude (Ed.) (2007). *Bildungsstandards am Gymnasium – Korsett oder Katalysator?* Bern: hep.

- Laird, J. E., Rosenbloom, P. S. & Newell, A. (1986). Chunking in soas: the anatomy of general learning mechanisms. *Machine Learning*, 1, 11-46.
- Liu, X. & Lesniak (2006). Progression in children's understanding of the matter concept from elementary to high school. *Journal of Research in Science Teaching*, 43(3), 320-347.
- Mandl, H. & Fischer, F. (Eds.) (2000). *Wissen sichtbar machen. Mapping-Techniken für das Wissensmanagement in Lern- und Kooperationsprozessen*. Göttingen: Hogrefe.
- Morrison, C. A. & Fitzpatrick, S. J. (1992). *Direct and indirect equating: A comparison of four methods using the Rasch model*. Measurement and Evaluation Center: The University of Texas at Austin. ERIC Document Reproduction Services No. ED375152.
- Mund, H. (Eds.) (1993). *Naturwissenschaftlich denken: Lehrerinformation, Technische Anleitung, Arbeitsmaterialien für Schüler*. Dt. Ausg.; Englisch Original: Adey, P., Shayer, M. & Yates, C., Thinking Science. Aachen: Verlag der Augustinus Buchhandlung.
- Mulford, D. R. (1996). *An Inventory for Measuring College Students' Level of Misconceptions in First Semester Chemistry*. Ph.D. Dissertation, Purdue University.
- Neumann, K., Kauertz, A., Lau, A., Notarp, H. & Fischer, H. E. (2007). Die Modellierung physikalischer Kompetenz und ihrer Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 103-123.
- Novak, J. D., Cañas, A. J. (2008). *The Theory Underlying Concept Maps and How to Construct and Use Them*. Technical Report IHMC CmapTools. <http://cmap.ihmc.us>.
- OECD (2006). *PISA assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: OECD.
- Oelkers, J. (2007). Bildungsstandards am Gymnasium – Ein neues Problem? In P. Labbude (Ed.), *Bildungsstandards am Gymnasium – Korsett oder Katalysator?* (27-36). Bern: hep.
- Osborn, J. & Dillon, J. (2008). *Science Education in Europe: Critical Reflections. A Report to the Nuffield Foundation*. Verfügbar über <http://www.kcl.ac.uk/content/1/c6/01/32/03/SciEdinEuropeReportFinal2.pdf> (24.09.2009).
- Ostermeier, C. (2004). *Kooperative Qualitätsentwicklung in Schulnetzwerken. Eine empirische Studie am Beispiel des BLK-Modellversuchsprogramms „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“ (SINUS)*. Münster: Waxmann.
- Parchmann, I. (2008). Bildungsstandards und Kompetenzmodelle – Katalysatoren für fachdidaktische Forschung, Lehrerbildung und Unterrichtsentwicklung?. In D. Höttercke (Eds.), *Kompetenzen, Kompetenzmodelle, Kompetenzentwicklung*. Tagungsband der Gesellschaft für Didaktik der Chemie und Physik (pp. 5-13), Münster: Lit.
- Parchmann, I. & Freienberg, J. (2009). Bildungsstandards, Kerncurricula und Lehrerbildung – ein Gesamtkonzept zur Unterrichtsentwicklung? *Praxis der Naturwissenschaften*, 58(2), 11-17.
- Perkins, D. & Ritchhart, R. (2004). When is good thinking?. In D. Y. Dai & R. J. Sternberg, *Motivation, Emotion, and Cognition: Integrative Perspectives on Intellectual Functioning and Development* (pp. 351-384). Mahwah, NJ: Lawrence Erlbaum Associates.
- Peuckert, J. & Fischler, H. (2000). Concept Maps als Diagnose- und Auswertungsinstrument in einer Studie zur Stabilität und Ausprägung von Schülervorstellungen. In H. Fischler & J. Peuckert (Eds.) (1999), *Concept Mapping in fachdidaktischen Forschungsprojekten der Physik und Chemie* (pp. 91-116). Berlin: Logos Verlag.
- Piaget, J. (1966). *Psychologie der Intelligenz*. Zürich: Rascher.
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E. & Pekrun, R. (Eds.) (2007). *PISA 2006 – Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M. et al. (Eds.) (2004), *PISA 2003: Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.
- Prenzel, M., Häußler, P., Rost, J. & Senkbeil, M. (2002). Der PISA-Naturwissenschaftstest: Lassen sich Aufgabenschwierigkeiten vorhersagen? *Unterrichtswissenschaft*, 30(2), 120-134.
- Prenzel, M., Rost, J., Senkbeil, M., Häußler, P. & Klopp, A. (2001). Naturwissenschaftliche Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert et al. (Eds.), *PISA 2000 – Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 191-248). Opladen: Leske + Budrich.
- QCA, Department for Education and Employment (DfEE) - Qualifications and Curriculum Authority (QCA) (2000). *Science: The National Curriculum for England: Key Stages 1-4*.
- Rasch, G. (1960). *Probabilistic models for some intelligence attainment tests*. Chicago: The University of Chicago Press.

- Renkl, A. & Nückles, M. (2006). Lernstrategie der externen Visualisierung. In H. Mandl & H. F. Friedrich (Eds.), *Handbuch Lernstrategien* (pp. 135-147). Göttingen: Hogrefe.
- Renkl, A. & Stern, E. (1994). Die Bedeutung von kognitiven Eingangsvoraussetzungen und Lernaufgaben für das Lösen von einfachen und komplexen Textaufgaben. *Zeitschrift für pädagogische Psychologie*, 8, 27-39.
- Rost, J., Walter, O., Carstensen C. H., Senkbeil, M. & Prenzel, M. (2005). Der nationale Naturwissenschaftstest PISA 2003. *Der mathematische und naturwissenschaftliche Unterricht* (MNU), 58(4), 196-204.
- Ruiz-Primo, M. A. (2004). Examining Concept Maps as an Assessment Tool. *Proceedings of the First International Conference on Concept Mapping*, Pamplonā, Spain.
- Schecker, H., Fischer, H. E. & Wiesner, H. (2004). Physikunterricht in der gymnasialen Oberstufe. In H.-E. Tenorth, *Kerncurriculum Oberstufe II* (pp. 148-234). Weinheim: Beltz.
- Schecker, H. & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 45-66.
- Schmidt, H.-J. (1997). Students' Misconceptions – Looking for a Pattern. *Science Education*, 81(2), 123-135.
- Schmidt, M. (2008). Kompetenzmodellierung und -diagnostik im Themengebiet Energie der Sekundarstufe I. In: *Studien zum Physik- und Chemielernen*, Band 87, Berlin: Logos.
- Schnotz, W. (1994). *Aufbau von Wissensstrukturen*. Untersuchungen zur Kohärenzbildung beim Wissenserwerb mit Texten. Weinheim: Beltz PVU.
- Scholz, A. (2008). *Entwicklung und Einstufung von Aufgaben zum Themengebiet Säure-Base-Reaktionen*. Unveröffentlichte Bachelor-Arbeit. Universität Oldenburg.
- Senkbeil, M., Rost, J., Carstensen, C. H. & Walter, O. (2005). Der nationale Naturwissenschaftstest PISA 2003. Entwicklung und empirische Überprüfung eines zweidimensionalen Facettendesigns. *Empirische Pädagogik*, 19(2), 166-189.
- Shavelson, R. J. (2007). *A Brief History of Student Learning Assessment: How We Got Where We Are and a Proposal for Where to Go Next*. Washington: Association of American Colleges and Universities.
- Shavelson, R. J., Ruiz-Primo, M. A. & Wiley, E. W. (2005). Windows into the mind. *Higher Education*, 49, 413-430.
- Steiner, G. (2006). Lernen und Wissenserwerb. In A. Krapp & B. Weidenmann (Eds.), *Pädagogische Psychologie* (pp. 137-202). Weinheim: Beltz PVU.
- Süß, H. M. (1996). *Intelligenz, Wissen und Problemlösen*. Göttingen: Hogrefe.
- Tergan, S.-O. (1986). *Modelle der Wissensrepräsentation als Grundlage qualitativer Wissensdiagnostik*. Opladen: Westdeutscher Verlag.
- Vygotsky, L. S. (1962). *Thought and Language*. New York: Wiley.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard Business Press.
- Waddington, D., Nentwig, P. und Schanze, S. (Eds.) (2007). *Standards in Science Education*. Münster: Waxmann.
- Wienert, B. (2008). *Entwicklung und Evaluierung von Aufgaben zur Bewertungskompetenz am Beispiel alternativer Kraftstoffe*. Unveröffentlichte Bachelor-Arbeit. Universität Oldenburg.

Kontakt

Ilka Parchmann, Sascha Bernholt
IPN Kiel
Abteilung Didaktik der Chemie
Olshausenstr. 62
24098 Kiel
parchmann@ipn.uni-kiel.de
bernholt@ipn.uni-kiel.de

Michael L. Commons
Harvard Medical School
Department of Psychiatry
Beth Israel Deaconess Medical Center
234 Huron Avenue
Cambridge, MA 02138
commons@tiac.net

Autoreninformation

Sascha Bernholt hat Mathematik und Chemie für das Lehramt an Gymnasien studiert und ist Doktorand an der Universität Oldenburg. Er promoviert zum Thema Kompetenzdiagnose und -entwicklung in der Chemie. Zudem ist er wissenschaftlicher Mitarbeiter im BMBF-Projekt „Impact of Affective and Informative Feedback on Learning in Children Before and After A Reattribution Training“.

Die Forschungs- und Entwicklungsgebiete von Ilka Parchmann liegen zum einen im Bereich der Verknüpfung von Kontext basierendem, situiertem Lernen mit einer strukturierten Entwicklung von Basis Konzepten und Kompetenzen, zum anderen in der Untersuchung und Optimierung von Konzepten zur Lehreraus- und -weiterbildung. Ilka Parchmann ist Mitglied im Vorstand der Fachgruppe Chemieunterricht der Gesellschaft Deutscher Chemiker (GDCh) und in diesem Rahmen Vorsitzende der Chemistry Education Group in dem Europäischen Dachverband EChemS (European Association for Chemical and Molecular Sciences).

Michael L. Commons ist zurzeit Assistenz-Professor im *Department of Psychiatry* an der *Harvard Medical School* und Direktor am *Dare Institute*, Cambridge, MA. Seine Forschungsinteressen liegen in der quantitativen Analyse im Gebiet der Lern- und Entwicklungspsychologie. Er ist Gründer des *Journal of Adult Development* und Mitbegründer der *Society for Quantitative Analysis of Behavior* sowie der *Society for Research in Adult Development*.