

ALEXANDER KAUERTZ, HANS E. FISCHER, JÜRGEN MAYER, ELKE SUMFLETH UND MAIK WALPUSKI

## Standardbezogene Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe I

Modeling competence according to standards for science education in secondary schools

### Zusammenfassung

In diesem Artikel werden die theoretischen fachdidaktischen Grundlagen beschrieben, die zu einer Operationalisierung der Standards mit dem Ziel einer modellbasierten und fairen Testung zu Evaluationszwecken beitragen. Dazu werden zunächst einige bildungstheoretische Aspekte sowie Grundannahmen der Operationalisierung dargestellt. Im Anschluss wird das Kompetenzmodell vorgestellt, das der Evaluation der Bildungsstandards in den Naturwissenschaften zugrunde liegt, die im Auftrag der KMK 2012 durchgeführt werden soll. Ferner werden die Aufgabenentwicklung und die geplante empirische Überprüfung der Konstruktvalidität des vorgeschlagenen Kompetenzmodells erläutert.

Schlüsselwörter: Kompetenzmodelle, Bildungsstandards, Testentwicklung, Naturwissenschaften, Sekundarstufe

### Abstract

In this article we describe the fundamental science education theories and ideas that are needed to operationalize national education standards. The end goal is to develop a test that consistently and fairly evaluates these standards and the school system as a whole. The applied educational theories, basic assumptions of this operationalization and the competence model used for the test are all explained in the following. Item development and the planned empirical study to show the test's construct validity will also be illustrated.

Keywords: competence models, educational standards, test development, science, secondary education

---

## 1 Einleitung

In zahlreichen Ländern werden durch Standards und Benchmarks die Leistungsanforderungen an Schülerinnen und Schülern zu bestimmten Zeiten in ihrer Schullaufbahn definiert. Damit verbunden ist die Erwartung, Schulsystementwicklung durch eine an Lernergebnissen orientierte Prozesskontrolle besser steuern zu können. Während internationale Schulsystemvergleiche wie etwa das *Programme for International Student Assessment* (PISA) (OECD, 1999) auf systemische Bedingungen (Stundentafeln, soziale Verteilungen) als Prädiktoren für die Ergebnisse orientiert sind, sind Standards und die damit

verbundenen Tests auf die Wirkung von Unterricht fokussiert. Durch die Bildungsstandards für den Mittleren Schulabschluss sind 2003 für das Ende der Sekundarstufe I in Deutschland leistungsbezogene Standards für die Fächer Deutsch, erste Fremdsprache und Mathematik, sowie 2004 auch für die Naturwissenschaften Biologie, Chemie und Physik von der Kultusministerkonferenz beschlossen worden (KMK, 2005a, b, c). Anders als in der Schweiz (vgl. Harms, Labudde, 2007) sind diese Standards in Deutschland in kurzer Zeit und ohne breiten Diskurs von einer Expertengruppe entwickelt und politisch ohne nennenswerte Widerstände betroffener Verbände verabschiedet worden.

Seit 2005 sind die Standards für die Schulen verbindlich. Auch wenn sie noch diskutiert und weiterentwickelt werden müssen, stellen sie derzeit die gültige normative Grundlage schulischer Arbeit dar. Mit ihrer standardisierenden und normierenden Funktion – insbesondere durch die damit verbundenen Tests – sind sie unverzichtbare Grundlage einer evidenzbasierten Schulpolitik. Sie sollen Lehrkräften, Eltern, Schülerinnen und Schülern sowie der Schulverwaltung und Schulaufsicht Transparenz über die Ziele ihres Handelns und Verbindlichkeit für ihr Handeln bieten und eine faire Einschätzung der Erfolge des Systems auf Schul- und Individualebene erlauben.

Ausgehend von diesem beschrieben Ist-Zustand wird im Folgenden präzisiert, welche Ziele mit den Bildungsstandards verfolgt werden, um daraus eine Operationalisierung von Kompetenz abzuleiten, die für die Entwicklung eines nationalen Tests geeignet ist. Mit diesem Test werden die Bildungsstandards im Projekt „Evaluation der Standards in den naturwissenschaftlichen Fächern der Sekundarstufe I (ESNaS)“ normiert und es wird eine Synthese aus normativen und empirischen Ansätzen ermöglicht. Der Schwerpunkt der Darstellung liegt auf der theoretischen Beschreibung des verwendeten Kompetenzmodells. Anhand der Ergebnisse einer Prä-Pilotierung wird die noch ausstehende empirische Untersuchung skizziert.

Das hier beschriebene Verfahren basiert – im Gegensatz zu zahlreichen anderen Testverfahren – auf theoretisch begründeten a-priori-Annahmen zur Aufgabenschwierigkeit. In den anderen Fächern und bei internationalen large-scale-assessments (z. B. PISA) wird üblicherweise ein Aufgabenpool entwickelt, erprobt und anhand der statistischen Kennwerte ausgewählt. Im Nachhinein werden die Aufgaben auf gemeinsame Merkmale untersucht und dadurch bestimmte Fähigkeiten postuliert, über die die Bearbeiter zum erfolgreichen Lösen verfügen müssen. Bei der Evaluation der Bildungsstandards in den naturwissenschaftlichen Fächern wird dagegen zunächst ein Kompetenzmodell be-

schrieben, in das sich alle Aufgaben einordnen lassen. Sie können so bestimmten Ausprägungen von Kompetenz im Vorhinein zugeordnet werden. Anhand der empirischen Verteilung der Lösungshäufigkeiten der Aufgaben in einem Test kann dann das Modell validiert und die Verteilung auf die verschiedenen Kompetenzausprägungen untersucht werden. Durch die zu Grunde liegende Systematik lassen sich Unterschiede in der untersuchten Population valide beschreiben und dadurch zielgerichtete Entwicklungsmöglichkeiten aufzeigen.

## 2 Ziele der Bildungsstandards

Als Bestandteil der Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring (KMK/IQB, 2006 bzw. Walpuski et al., 2008) verfolgen die deutschen Bildungsstandards mehrere Ziele. Ihr Kerngedanke aber ist die Sicherung des Ergebnisses von Schule bis zum Mittleren Schulabschluss und die Verbesserung der Qualität von Unterricht. Dazu werden in den KMK-Standards die fachspezifischen Ziele des Unterrichts in Form von Kompetenzen und Regelstandards formuliert – es erfolgt also eine Festlegung, was deutsche Schülerinnen und Schüler im Normalfall zum Zeitpunkt des Mittleren Schulabschlusses können sollen. Mindeststandards, die alle Lernenden erreichen sollen, werden nicht formuliert, ebenso erfolgt keine Festlegung oder Beschreibung von Lernsituationen im Sinne von Opportunity to Learn-Standards. Es wird lediglich eine innerfachliche (vertikale) Vernetzung als zentrales Qualitätsmerkmal fachlichen Unterrichts genannt, die durch die Formulierung von Basiskonzepten in den Bildungsstandards explizit angeregt werden soll (KMK, 2005a, b, c).

Die Kompetenzen, die in den Bildungsstandards beschrieben sind, gehen auf die Definition von Kompetenz nach Weinert (2001) zurück. Demnach sind Kompetenzen „die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und

Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (Weinert, 2001, S. 27). Es ist davon auszugehen, dass Kompetenzen domänenspezifisch zu betrachten sind (vgl. Klieme et al., 2003), wobei die Abgrenzung verschiedener Domänen unklar ist, was sich etwa bei der Diskussion des Fächerkanons für allgemeinbildende Schulen zeigt (z.B. Naturwissenschaft vs. Physik, Chemie, Biologie). Die Bildungsstandards für die Naturwissenschaften sind disziplinspezifisch formuliert, sie stimmen jedoch in der grundsätzlichen Struktur und in zahlreichen allgemeinen Formulierungen überein. Dadurch wird die Nähe der naturwissenschaftlichen Fächer deutlich gemacht, ohne ihren eigenständigen Charakter aufzugeben.

Innerhalb der Bildungsstandards wird Naturwissenschaftliche Kompetenz in vier Kompetenzbereiche unterteilt: Fachwissen, Erkenntnisgewinnung, Bewerten und Kommunikation. Diese Unterteilung entspringt einer normativ pragmatischen bzw. politikgeleiteten Perspektive. Demnach soll Naturwissenschaftlicher Unterricht neben der ausschließlichen Vermittlung und Anwendung von Fachinhalten auch Handlungsaspekte vermitteln. Dazu gehören das Wissen zum Experimentieren und Modellieren, zu dem auch das Wissen *über* die Naturwissenschaften (z. B. ihre geschichtliche Entwicklung und der Charakter naturwissenschaftlicher Theorien) gehört, sowie Kommunikation und Bewertung im Sinne einer kritischen Haltung mündiger Bürger. Um dies hervorzuheben, stehen diesen drei Anforderungen an Unterricht gleichwertig neben dem Kompetenzbereich Fachwissen. Es ist noch zu prüfen, ob eine empirische Unterscheidbarkeit der Kompetenzbereiche gegeben ist, da Fachwissen auch für die anderen drei Kompetenzbereiche benötigt wird. Für die Kompetenzbereiche *Fachwissen* und *Erkenntnisgewinnung* gibt es zumindest

erste Hinweise, dass eine Unterscheidung empirisch haltbar ist (Klos, Henke, Kieren, Walpuski, & Sumfleth, 2008, Mayer, Grube & Möller 2008). Es ist jedoch empirisch nicht geklärt, ob die Kompetenzbereiche *Erkenntnisgewinnung*, *Kommunikation* und *Bewertung*, möglicherweise im Gegensatz zum Kompetenzbereich *Fachwissen*, naturwissenschaftenübergreifend modelliert werden können.

In Anlehnung an die Einheitlichen Prüfungsanforderungen Abiturprüfung (KMK, 2004) werden in den Bildungsstandards zudem drei Anforderungsbereiche unterschieden, die explizit keinen Stufencharakter haben. Sie orientieren sich an der klassischen, z. B. von Bloom (1965) beschriebenen Unterscheidung zwischen *wiedergeben*, *anwenden* und *transferieren*.

### 3 Operationalisierung von Kompetenz

Im Unterricht der naturwissenschaftlichen Fächer an Schulen werden Kompetenzen meist durch das erfolgreiche Bearbeiten von Aufgaben beobachtbar (vgl. Fischer & Draxler, 2006). Zudem lassen sich Kompetenzen über Testaufgaben effizient, in der Regel in schriftlichen Tests, erfassen. Dabei wird üblicherweise – so auch im Projekt ES-NaS – Kompetenz auf kognitive Aktivitäten eingeschränkt und für Unterricht wesentliche Aspekte wie Motivation, Interaktion oder praktische Tätigkeiten ausgeklammert. Allerdings werden Interesse und Motivation in Bezug auf die untersuchten Fächer und den Unterricht im Rahmen von begleitenden Erhebungen erfasst. Die kognitive, fachbezogene Kompetenz wird am Mittelwert einer repräsentativen Stichprobe ausgerichtet, um Standards entsprechend zu justieren (Normierung). So konstruierte schriftliche large-scale-assessments können dazu beitragen, Problemstellen im Bildungssystem zu identifizieren. Aussagen zur Qualität von Lerngelegenheiten und Lernaufgaben können dadurch aber nur eingeschränkt gemacht werden, und die erfassten Personenmerk-

male, repräsentiert durch die Lösungshäufigkeit einzelner Aufgaben, stellen nur einen Ausschnitt individueller Kompetenz dar.

Kompetenzen können über Aufgabenmerkmale beschrieben werden. Unterscheiden sich im einfachsten Fall zwei Gruppen von Aufgaben in nur einem Merkmal und eine Person löst im Idealfall alle Aufgaben der einen aber keine Aufgabe der anderen Gruppe, so verfügt die Person offenbar über eine Fähigkeit, die unmittelbar mit diesem Unterscheidungsmerkmal der Aufgaben zusammenhängt. Für large-scale-assessments empfiehlt es sich, lösungsbezogene Merkmale, die unabhängig vom Bearbeiter beschrieben werden können, als Operationalisierung von Kompetenz zu nutzen und die Aufgaben so zu konstruieren, dass die Merkmale eindeutig in ihnen repräsentiert sind.

Die entsprechend der Kompetenzdefinition ausgewählten Aufgabenmerkmale werden in einem Kompetenzmodell systematisiert und beschrieben (vgl. Schecker & Parchmann, 2006). Kompetenzmodelle können als Koordinatensystem veranschaulicht werden, wobei die Achsen durch die Aufgabenmerkmale und die Unterteilung der Achsen durch die verschiedenen Ausprägungen des jeweiligen Merkmals gebildet werden. Jede Aufgabe kann hinsichtlich jedes der im Modell berücksichtigten Merkmale analysiert

und dann entsprechend ihrer Ausprägungen auf jeder Achse in das Modell eingeordnet werden. Für die Entwicklung eines Tests zur Überprüfung des Kompetenzmodells ist aus messtheoretischer Sicht dabei zu fordern, dass *jede* Aufgabe auf *jeder* Achse eindeutig *einer* Ausprägung zugeordnet ist (eindeutige Modelleinordnung) und es zu jeder Kombination von Ausprägungen verschiedener Merkmale ausreichend viele Aufgaben gibt (Modellsättigung).

#### 4 Kompetenzmodell

Die Bildungsstandards basieren implizit auf einem normativen Kompetenzstrukturmodell (vgl. Schecker & Parchmann, 2006, S. 48), das wie in Abbildung 1 gezeigt in einem zweidimensionalen Koordinatensystem dargestellt werden kann (KMK, 2005a, b, c). Die vier Kompetenzbereiche bilden die erste Achse, die eine eher inhaltliche Komponente repräsentiert, die drei Anforderungsbereiche bilden eine nominal skalierte Achse, auf der die kognitive Anforderung der Aufgabe beschrieben ist. Eine Aufgabe könnte z. B. auf der zweiten Achse die Ausprägung *Anforderungsbereich I* haben und auf der ersten die Ausprägung *Fachwissen* und würde somit die Teilkompetenz *Fachwissen wiedergeben* erfordern.

Anforderungsbereich III				
Anforderungsbereich II				
Anforderungsbereich I				
	Fachwissen	Erkenntnisgewinnung	Kommunikation	Bewertung

Abb. 1 Kompetenzmodell der Bildungsstandards.

Das den Bildungsstandards zugrunde liegende Kompetenzmodell ist für Testzwecke unzureichend, weil zum Teil mehrere Merkmale gleichzeitig variiert werden, die als schwierigkeiterzeugend angenommen werden können. So heißt es beispielsweise für Physik im Anforderungsniveau I: „Fakten und einfache physikalische Sachverhalte reproduzieren.“ (KMK, 2005c, S. 13) und im Anforderungsniveau III: „Wissen auf teilweise unbekannte Kontexte anwenden, geeignete Sachverhalte auswählen.“ (KMK, 2005c, S.13). Ähnliche Formulierungen lassen sich für alle Fächer finden. Dabei sind „Fakten“ und „Wissen“ mit Blick auf den damit verbundenen Umfang nicht trennscharfe Beschreibungen; es ist zwar anzunehmen, dass Wissen umfangreicher gemeint ist als Fakten, offen bleibt aber in welcher Art und Weise. Aufgaben, die Wissen erfordern, sollten damit schwieriger sein als solche, die Fakten erfordern, die Frage der Quantität bleibt aber ebenfalls offen. Parallel dazu wird bei Niveau I *Reproduzieren* erwartet, bei Niveau III *Anwenden* auf eine unbekannte Situation (Transfer), was ebenfalls einen Unterschied in der Schwierigkeit entsprechender Aufgaben hervorruft. Scheitert ein Schüler oder eine Schülerin an einer Aufgabe zu einem bestimmten Niveau, ist nicht erkennbar, an welchem der beiden schwierigkeiterzeugenden Merkmale er oder sie gescheitert ist. Zudem erlauben die Anforderungsniveaus ausdrücklich keine Graduierung von Kompetenz, was aber für eine differenzierte Evaluation der Standards nach Kompetenzniveaus notwendig erscheint. Aus diesem Grund wurden die in den Bildungsstandards nur implizit enthaltenen Graduierungen in ein Kompetenzmodell überführt, das es erlaubt, die angenommenen schwierigkeitsbestimmenden Merkmale für Testzwecke gezielt systematisch zu variieren.

Das im Folgenden beschriebene Kompetenzmodell dient dazu, das implizite Kompetenzmodell der Bildungsstandards so auszuschärfen, dass bundesweit, mittels schriftlicher Tests, eine Evaluation und Normierung der Bildungsstandards in den natur-

wissenschaftlichen Fächern für den Mittleren Schulabschluss vorgenommen werden kann. Das Kompetenzmodell wird aus den drei Aufgabenmerkmalen *Komplexität*, *kognitive Prozesse* und *Kompetenzbereich* gebildet und entspricht damit dem Aufbau des von Neumann, Kauertz und Fischer (2007) vorgeschlagenen Modells. Basierend auf dieser Grundstruktur werden die einzelnen Aufgabenmerkmale ausgeschärft und für weitere Kompetenzbereiche (Erkenntnisgewinnung) und Fächer (Biologie und Chemie) entwickelt. Dieses Aufgabenmerkmale werden im Folgenden konkreter beschrieben sowie ihre Auswahl im Sinne des von Schecker und Parchmann (2006) vorgeschlagenen iterativen Verfahrens durch vorliegende Forschungsarbeiten und den Bezug zu den Bildungsstandards gerechtfertigt.

#### 4.1 Kompetenzbereiche

In den Bildungsstandards werden vier Kompetenzbereiche vorgeschlagen, die modellkonform operationalisiert werden müssen: Fachwissen, Erkenntnisgewinnung, Bewertung und Kommunikation.

**Fachwissen:** Die drei Naturwissenschaften haben im Laufe der Zeit Theorien und Beschreibungen zu Objekten und Prozessen in der Natur entwickelt. Naturwissenschaftlich kompetente Personen können nachvollziehen, wie diese Theorien zur Lösung konkreter Probleme eingesetzt werden und sie können selbst relevante Prinzipien, Modelle und Konzepte anwenden. In den Bildungsstandards ist dies durch den Kompetenzbereich *Fachwissen* abgebildet.

**Erkenntnisgewinnung:** Neben diesem Satz an Theorien zeichnet sich die jeweilige Naturwissenschaft durch eine Reihe von Regeln und Prinzipien zur Entwicklung, Anwendung und Prüfung dieser Theorien aus. Im Kompetenzbereich *Erkenntnisgewinnung* sind diese Regeln und Prinzipien in den Bildungsstandards berücksichtigt.

**Bewertung:** Die Fähigkeit, naturwissenschaftliche Informationen adäquat in Entscheidungsprozesse einzubinden und Entscheidungen bezüglich ihrer Abhängigkeit von naturwissenschaftlichen Informationen zu reflektieren, wird in den Bildungsstandards durch den Kompetenzbereich *Bewertung* berücksichtigt.

**Kommunikation:** Die naturwissenschaftstypische Weise Sprache, Symbole sowie Kommunikations- und Darstellungsarten zu nutzen, wird in den Bildungsstandards durch den Kompetenzbereich *Kommunikation* berücksichtigt.

Diese Kompetenzbereiche bieten verschiedene Blickwinkel an, unter denen die Inhalte der naturwissenschaftlichen Fächer betrachtet werden können. Da Theorien ohne das Wissen um ihre Rechtfertigung und Experimente oder Symbole ohne inhaltlichen Bezug keine angemessene naturwissenschaftliche Bedeutung haben, kann es sich bei den Kompetenzbereichen nicht um inhaltlich getrennte Bereiche handeln, die zusammengesetzt wie ein Puzzle die Naturwissenschaften ergeben. Versteht man Kompetenzbereiche vielmehr als unterschiedliche Blickwinkel auf die jeweiligen Fachinhalte, lassen sie sich klar hinsichtlich ihrer Intention unterscheiden. So kann zu einer biologischen, chemischen oder physikalischen Theorie ihre Anwendung auf ein konkretes Problem betrachtet werden oder die zu ihrem Beleg durchgeführten Experimente, die für ihre Beschreibung verwendeten Symbole und Begriffe und die Berücksichtigung der von ihr erklärten Daten bei Entscheidungsprozessen. Jeweils andere, auch fachtypische Aspekte der Theorie werden dabei ins Zentrum der Aufmerksamkeit treten. Für die Testentwicklung ist es erforderlich, diese verschiedenen Blickwinkel zu trennen und einzeln zu erfassen. Um die Aufgabenentwicklung in den einzelnen Bereichen und die Abgrenzung der verschiedenen Blickwinkel zu erleichtern, werden die Kompetenzbereiche durch Teilbereiche und diese

wiederum durch Aspekte näher beschrieben. Auf der Ebene der Teilbereiche wird erwartet, dass diese sich im Leistungstest statistisch unterscheiden lassen.

Für den Kompetenzbereich *Umgang mit Fachwissen* werden die durch die Bildungsstandards vorgeschlagenen Basiskonzepte als Teilbereiche genutzt. Diese sind für jedes Fach spezifisch genannt. Mit ihnen werden Phänomene und Inhalte aus der Perspektive des jeweiligen Fachs theoretisch beschrieben und strukturiert, um den Erwerb eines grundlegenden, vernetzten Wissens zu erleichtern (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2005a; 2005b; 2005c, S. 7). Weitere Ziele der Basiskonzepte sind eine „Eingrenzung und auf Schule und Unterricht bezogene Sichtweise der jeweiligen wissenschaftlichen Konzepte“ (Demuth, Ralle & Parchmann, 2005, S. 57). Für die Physik konnte in einer Studie von Kauertz (2008) gezeigt werden, dass Schülerinnen und Schüler Aufgaben zu den verschiedenen Basiskonzepten offenbar unabhängig beantworten, d. h. die Basiskonzepte erfordern unterschiedliche Teilkompetenzen. Im Fach Chemie wird das Basiskonzept *Energetische Betrachtungen bei Stoffumwandlungen* nicht als einzelner Teilbereich übernommen, alle Aufgaben mit energetischen Betrachtungen aber unter diesem Blickwinkel ein zweites Mal kodiert, da dieses Basiskonzept inhaltlich eng mit den anderen Basiskonzepten verbunden ist. Es ist daher anzunehmen, dass dieses Basiskonzept statistisch nicht von den anderen zu unterscheiden ist. In Biologie und Physik werden alle Basiskonzepte übernommen.

Der Kompetenzbereich *Erkenntnisgewinnung* wird durch die Teilbereiche *Naturwissenschaftliche Untersuchungen*, *Naturwissenschaftliche Modellbildung* und *Wissenschaftstheoretische Reflexion* strukturiert. Dieser Strukturierung liegen die in der internationalen Diskussion zu Nature of Science (vgl. Lederman, 2007, McCormas & Olson, 1998) berücksichtigten Charakteristika der Naturwissenschaften zugrunde (AAAS,

1993; National Research Council, 1996). Naturwissenschaften zeichnen sich dabei zum einen durch experimentelle Methoden aus und stellen andererseits einen kognitiv modellierenden Zugang zur Welt dar (Baumert, 2002). Die wissenschaftstheoretische Reflexion bereitet dabei die wissenschaftspropädeutische Arbeit in der Oberstufe vor, vor allem aber ist sie Teil der von Bybee beschriebenen scientific literacy (Bybee, 2000), da sie notwendige Voraussetzung dafür ist, die Rolle der Naturwissenschaften für und in unserer Gesellschaft verstehen zu können. Sie trägt somit zu einem „adäquaten Bild der Naturwissenschaften“ bei, das in den Bildungsstandards als ein zu erreichendes Ziel der Sekundarstufe formuliert ist (KMK, 2005a).

Die Operationalisierung des Aspekts *Naturwissenschaftliche Untersuchungen* als schriftlicher Test erscheint zunächst problematisch, da damit weder manuelle Handlungen noch ein exploratives Herangehen als Teil realer Forschung erfasst werden kann. Bislang ist jedoch unklar, inwieweit zielführende manuelle Handlungen beim Experimentieren von den gleichzeitig ablaufenden kognitiven Prozessen zur Strukturierung der Handlung unabhängig sind. Es gibt zwei wesentliche Gründe gegen so genannte Performance Test als large-scale-assessments, in denen Schülerinnen und Schüler ihre Kompetenz an realen Experimenten zeigen sollen. Performance Tests zeigen ein uneinheitliches Bild über die Zusammenhänge zu schriftlichen Tests. Offenbar werden zur erfolgreichen Bearbeitung einer Aufgabe verschiedene Fähigkeiten benötigt und diese können von Aufgabe zu Aufgabe stark variieren. Somit ist die von der Kombination der Prädiktoren *Person* und *Aufgabe* erzeugte Varianz größer als die von den jeweiligen Prädiktoren einzeln erzeugte (Shavelson, Baxter & Gao, 1993). Außerdem sind Performance Tests extrem aufwändig und wenig reliabel, da sie nur über eine Beurteilung der Experimentalsituationen ausgewertet werden können. Durch die verringerte Messgenauigkeit ist ihre

Aussagekraft für large-scale-assessments, in denen die Datenfülle eines umfangreichen Ratings kaum ökonomisch erhoben werden kann, deshalb weniger valide.

Der Teilbereich *Naturwissenschaftliche Untersuchungsmethoden* bezieht sich deshalb, bedingt durch die Einschränkung des Testverfahrens, auf die (kognitiv orientierte) Strukturierung der experimentellen Abläufe. Die Operationalisierung erfolgt durch die Einteilung der Untersuchung in die Schritte Fragestellung, Hypothese, Untersuchungsdesign, Datenauswertung. Diese Merkmale naturwissenschaftlicher Untersuchungen konnten für die Biologie als vier eigenständige Teilkompetenzen bei Schülerinnen und Schülern identifiziert werden (Grube, Möller & Mayer, 2007). Für die Chemie wurde dies bisher empirisch nicht gefunden, auch wenn die Annahme theoretisch sinnvoll erscheint (vgl. Klos et al., 2008), für die Physik wurden diese Teilkompetenzen bisher nicht umfassend untersucht. Unabhängig davon erscheint die Verwendung dieser Teilkompetenzen im Sinne von *Aspekten* bei der Aufgabenkonstruktion hilfreich, um möglichst vielfältige Aufgaben zu erhalten, die verschiedene Merkmale naturwissenschaftlicher Untersuchungen abbilden.

Der Teilbereich *Naturwissenschaftliche Modellbildung* ist operationalisiert durch die Anwendung von Modellen sowie die Kenntnis ihrer Grenzen und Funktionen. Der Begriff „Modell“ wird dabei jedoch in den naturwissenschaftlichen Fächern unterschiedlich verwendet. Während in der Biologie vielfach mit gegenständlichen Struktur- und Funktionsmodellen (z. B. Membranmodell) oder dynamischen Modellen (Populations-, Ökosystem, Evolutionsmodelle) gearbeitet wird, werden in der Physik und Chemie vorwiegend abstrakte oder mathematische Modelle verwendet (z. B. Atommodell, Teilchenmodelle) (vgl. Beerenwinkel & Parchmann, 2008; Mikelskis-Seifert & Fischler, 2003).

Der Teilbereich *Wissenschaftstheoretische Reflexion* beschreibt schwerpunktmäßig

eine Meta-Ebene. Wesentliche Aspekte, die in den Aufgaben zu diesem Bereich berücksichtigt werden, beziehen sich auf den Kanon von McComas und Olson (1998) (vgl. auch Osborne et al., 2003 oder Lederman, 2007): Verfolgen einer naturwissenschaftlichen Fragestellung, Nutzung vielfältiger Methoden, Mannigfaltigkeit der Interpretation von Daten, Subjektivität naturwissenschaftlichen Wissens, Daten und Interpretation als Grundlage naturwissenschaftlichen Wissens, Vorläufigkeit naturwissenschaftlichen Wissens (vgl. Zilker, Holliday et al., 2010). Darin werden Merkmale der naturwissenschaftlichen Theorien (z. B. Vorläufigkeit des Wissens), der Erkenntniswege (z. B. subjektive Einflüsse) und die Rolle der Gesellschaft (z. B. die Rolle der wissenschaftlichen Institutionen) aufgeführt, die in den meisten wissenschaftstheoretischen Auffassungen unstrittig sind (vgl. Kremer, Urhahne & Mayer 2009).

Während es für die Kompetenzbereiche *Fachwissen* und *Erkenntnisgewinnung* Erfahrung mit large-scale-assessments gibt, fehlen diese Erfahrungen in den Bereichen *Kommunikation* und *Bewertung* weitgehend. Vor allem der fachspezifische Anteil in diesen beiden Bereichen ist schwierig zu erfassen, weil hier große Überlappungen auch mit dem Anwenden von Fachwissen vermutet werden. Für den Bereich *Bewertung* gibt es einige Aufgabenbeispiele aus den PISA-Erhebungen, deren fachlicher Anteil jedoch unklar ist. Für den Bereich der Biologie gibt Bögeholz ein Modell an, mit dem *Bewertung* (begrenzt auf den Aspekt *nachhaltige Entwicklung*) fachspezifisch beschrieben werden kann (Eggert & Bögeholz, 2006) und das aufgrund der geringen Variablenanzahl für large-scale-assessments geeignet scheint. Für den Bereich *Kommunikation* gibt es ein solches Modell bislang nicht. Die Operationalisierung der beiden Bereiche ist daher noch nicht abgeschlossen sondern Bestandteil der 2. Projektphase bis 2011.

## 4.2 Komplexität

Der Begriff Komplexität charakterisiert die Struktur von Systemen (Wunsch, 1985; Nicolis & Prigogine, 1987) und ist von Schwierigkeit und Kompliziertheit abzugrenzen. In Fachdidaktik und Psychologie wird Komplexität genutzt, um inhaltliche und kognitive Strukturen zu beschreiben. Fischer (1994) definiert Komplexität für Physikaufgaben als Anzahl funktionaler Relationen zwischen Elementen gleicher Qualität und beschreibt darüber die Veränderung kognitiver Systeme. Eine Beziehung zur Schwierigkeit von Aufgaben wird zunächst nicht hergestellt. Commons et al. (2007) beziehen den Begriff Komplexität auf die Aufgabenlösung, um eine hierarchische Stufung zu entwickeln. Sweller (1994) charakterisiert die kognitive Belastung durch den Inhalt beim Lernen über die Anzahl und Intensität der Verknüpfung von Inhaltselementen (*Element Interactivity*). Fischer et al. (2007) beurteilen die vertikale Vernetzung im naturwissenschaftlichen Unterricht anhand der Komplexität der durch Lehrer und Schüler erarbeiteten Sachstruktur. In allen Beispielen geht eine höhere Komplexität mit einem höheren Anspruch an den Bearbeiter oder mit einer umfangreicheren kognitiven Aktivität einher.

Kauertz und Fischer (2006) schlagen ausgehend von diesen Arbeiten eine Beschreibung von Kompetenz in sechs Stufen vor. In einer empirischen Studie konnte gezeigt werden, dass jede Komplexitätsstufe mit einer Verteilung der Schwierigkeit auf dieser Stufe einhergeht und sich die Verteilungen der Aufgabenschwierigkeit von Stufe zu Stufe zu einer größeren mittleren Schwierigkeit hin verschieben (Kauertz, 2008). Die Ergebnisse legen jedoch eine Reduzierung auf fünf Stufen und Präzisierung der Stufen nahe. Diese fünf Stufen für das Aufgabenmerkmal *Komplexität* sind:

- I. Fakt
- II. Zwei Fakten
- III. Zusammenhang



- IV. Zwei Zusammenhänge  
V. Übergeordnetes Konzept

Die Komplexitätsstufen basieren dabei auf den zwei Prinzipien *Anzahl an Elementen* und (funktionale) *Verknüpfungen zwischen Elementen*. Das einfachste Element ist dabei *Ein Fakt*. Es umfasst die Bezeichnung einer (biologischen, chemischen, physikalischen) Größe, Eigenschaft oder Variable. Auf der zweiten Stufe sind zwei unverbundene Fakten eingeordnet, z. B. Eigenschaften desselben Objekts oder die Unterteilung in abhängige und unabhängige Variablen einer Untersuchung. Auf der dritten Stufe wird eine funktionale Beziehung (Funktionen, Relationen, korrelative, kausale und Mengenbeziehungen) zwischen Fakten hergestellt. Daraus entsteht ein neues Element, ein Zusammenhang. Auf dieser Stufe sind im Chemie- und Physikunterricht je-desto-Beziehungen typisch bzw. die Verknüpfung von unabhängiger und abhängigen Variable bei Untersuchungen. Auf der vierten Stufe werden diese Zusammenhänge verknüpft, um dadurch beispielsweise einen Kreisprozess zu beschreiben. Werden diese Zusammenhänge so generalisiert beschrieben, dass sie von der konkreten Situation unabhängig anwendbar sind, handelt es sich um die fünfte Komplexitätsstufe *übergeordnetes Konzept*. Bei der Erkenntnisgewinnung ist dies etwa eine Diskussion der Notwendigkeit von Variablenkontrollstrategien.

Die so definierte Komplexität ist mit mittlerer Inferenz in Aufgaben beurteilbar und, basierend auf den beschriebenen Untersuchungen, als Stufenmerkmal geeignet, Kompetenz zu operationalisieren. Damit differenziert und operationalisiert es den in den Anforderungsbereichen der Bildungsstandards berücksichtigten Aspekt der Komplexität. Durch das im hier vorliegenden Modell beschriebene Aufgabenmerkmal *Komplexität* können diese Anforderungen präzise und systematische beschrieben und dadurch in Aufgaben umgesetzt werden. Da es sich bei der *Komplexität* außerdem um ein gestuftes Aufgabenmerkmal handelt, können dadurch

unterschiedliche Kompetenzniveaus beschrieben werden.

### 4.3 Kognitive Prozesse

Durch das Konstrukt *kognitive Prozesse* soll beschrieben werden, welche Denkvorgänge bei der Bearbeitung durch die Aufgabe angeregt werden. Grundsätzlich ist bei konkreten erfolgreichen Aufgabenlösungen kaum nachzuweisen, ob ein bestimmter Denkvorgang stattgefunden hat. Außerdem ist in der Literatur uneinheitlich beschrieben, welche Denkprozesse für das Lösen einer Aufgabe überhaupt benötigt werden. Weit verbreitet ist, die Beziehung zum Vorwissen in drei Abstufungen zu beschreiben, wie es auch in den Anforderungsbereichen der Bildungsstandards gemacht wird: Der Bearbeiter muss Vorwissen wiedergeben, anwenden oder transferieren. Bei bekanntem Vorwissen der Schülerinnen und Schüler lässt sich die Aufgabenschwierigkeit dadurch recht gut vorhersagen (Leutner et al., 2008). Diese Klassifikation, basierend auf der Taxonomie von Bloom (1965), kann in Stichproben, die deutlich verschiedenes Vorwissen haben können, aber nicht sinnvoll angewandt werden. Zur Entwicklung eines bundesweiten Kompetenztests muss eine alternative Beschreibung gefunden werden, da ein gemeinsamer verbindlicher Inhaltskanon über die verschiedenen Schulformen und Bundesländer hinweg fehlt.

Bei der nationalen Erweiterung der PISA-Studie führt die Unterscheidung von sieben verschiedenen, fachübergreifenden Prozessen wie *Graphen interpretieren* und *konvergentes Denken* (Senkbeil et al., 2005) zu einer aus fachdidaktischer Sicht wenig plausiblen Auswahl notwendiger Fähigkeiten. Es besteht also ein erhebliches Validitätsproblem.

In schriftlichen Tests müssen textliche oder (statische) visuelle Informationen im Aufgabenstamm aufgegriffen und dann kognitiv zur richtigen Lösung verarbeitet werden. Daher kommen als mögliche, generelle Pro-

zesse Informationsverarbeitungsstrategien in Betracht, wie sie in der Psychologie etwa von Weinstein und Meyer (1986) beschrieben wurden. Die Prozesse des *Reproduzierens* (Identifizieren von Information), *Selegierens* (Auswählen von Information), *Organisierens* (Strukturieren von Information) und *Integrierens* (Einbinden von Information in die Wissensbasis) sind dabei kumulativ, d.h. sie werden nacheinander durchlaufen und bauen aufeinander auf (vgl. Wittrock, 1989). Erst nach dem letzten Schritt kann ein zuvor unbekannter Text als „verstanden“ gelten. Sie sind zudem prinzipiell unabhängig davon, was die Person bereits weiß. Sie werden durch entsprechendes Vorwissen jedoch erleichtert (vgl. Weidenmann, 1993).

Da sich diese generellen Prozesse auf die Informationsverarbeitung beziehen, müssen sie für die Erfassung kognitiver Prozesse bei der Bearbeitung von naturwissenschaftlichen Testaufgaben operationalisiert werden. Dazu wurden drei Kriterien entwickelt, die sich auf das Verhältnis von Informationsangebot in der Aufgabenstellung zur erwarteten Information in der Lösung beziehen. Jeder Informationsverarbeitungsprozess ist durch eine spezifische Ausprägung der drei Kriterien definiert (vgl. Tab. 1). Das *erste Kriterium* kennzeichnet die Identität der Information: Alle Informationen, die die Lösung erfordert, sind bereits im Aufgabenstamm enthalten. Das *zweite Kriterium* setzt voraus, dass das erste nicht erfüllt ist, und lautet:

Zwischen den Informationen im Aufgabenstamm muss wenigstens ein nicht genannter Zusammenhang hergestellt werden, um die geforderte Information für die Lösung zu erhalten. Das *dritte Kriterium* setzt voraus, dass das zweite Kriterium erfüllt ist, und betrifft die Gesamtstruktur der gegebenen und gesuchten Information: Die in der Aufgabenstellung beschriebene Situation ist nicht gleich der in der Lösung zu beschreibenden Situation. Mit Situation ist dabei entweder das Themengebiet des Faches (Wellen in Akustik und Optik), das beschriebene Beispiel (Vergrößerung der Blendenöffnung um 1 mm), der lebensweltliche Kontext (Zentripetalkraft beim Karussell und beim Hammerwurf) oder das Abstraktionsniveau (Kraft ist proportional zur Masse, Vergleich der Gewichtskraft des halbvollen und vollen Eimers) gemeint.

Diese drei Kriterien konkretisieren die Anforderungsbereiche der Bildungsstandards. Da sie gezielt bei der Aufgabenentwicklung berücksichtigt werden, ist die Zuordnung von Aufgaben zu den kognitiven Prozessen eindeutig möglich (siehe z. B. Abb. 3). Dass dies auch zu einer empirischen Unterscheidbarkeit führt lassen Ergebnisse von Mannel (Mannel, Sumfleth & Walpuski, 2008) erwarten. Entsprechend der Anforderungen an einen nationalen Test beziehen sich die Kriterien auf die vorgegebenen Informationen statt auf Vorwissen. Wiedergeben wird so verstanden, dass keine neuen Informa-

Tab. 1 Kognitive Prozesse und variierende Kriterien

Kriterien Kognitive Prozesse	Verhältnis zwischen der Menge der vorgegebenen Information und der erwarteten Information	Notwendigkeit, Zusammenhänge herzustellen	Ähnlichkeit der Situation in Aufgabenstellung und Lösung
Reproduzieren	Identisch	Nein	hoch
Selegieren	Teilmenge	Nein	hoch
Organisieren	Erweiterung	Ja	hoch
Integrieren	Erweiterung	Ja	niedrig

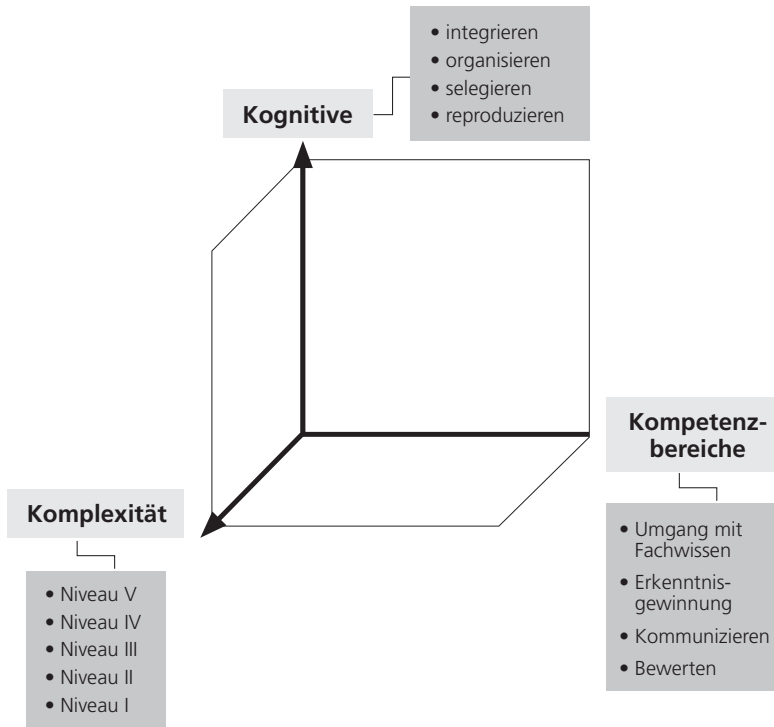


Abb. 2 Das Kompetenzmodell von ESNaS.

tionen erzeugt werden müssen, Anwenden bedeutet, dass aus vorhandener Information ein logischer Schluss gezogen werden muss und Transfer bedeutet, dass vorhandene Information in eine andere Situation überführt werden muss.

Die Informationen, auf die die Prozesse angewandt werden, sind die jeweils fachlich relevanten Informationen der Aufgabe. Für den Prozess des *Reproduzierens* wird in den Aufgaben das Wiedergeben der passenden Fachinformationen verlangt, die im Text genannt sind. Da dazu der Text Sinn erfassend gelesen werden muss, dürfte auf dieser basalen Stufe der Unterschied zum Textverstehen gering sein, obwohl sich ersten Ergebnissen zur Folge nur mittlere Korrelationen nachweisen lassen. Es ist jedoch nach Ergebnissen der PISA-Studie davon auszugehen, dass etwa 20% der Schülerinnen und Schüler mit dem Lesen solcher Fachtexte erhebliche Schwierigkeiten haben (Baumert et al., 2003). Die Kompetenz in dieser Gruppe soll durch diese basale Stufe kognitiver Prozesse

erstmalig ausreichend aufgelöst werden. Auf dieser Stufe Standards zu beschreiben, erscheint den Autoren nicht angemessen, die Stufe sollte jedoch unbedingt getestet werden, um dieser Gruppe entsprechende Fördermöglichkeiten zu eröffnen.

Die zweite Stufe kognitiver Prozesse, das *Selegieren*, ist ebenfalls stark mit Textverstehen konfundiert. Da es um das Auswählen von fachlicher Information geht, wird hier jedoch, etwa durch die Vorgabe von typischen, Alltags- und Schülervorstellungen als Distraktoren, bereits erkennbarer naturwissenschaftliche Kompetenz abgebildet. Die beiden höchsten Stufen beschreiben die typischen naturwissenschaftlichen Aufgaben, in denen Diagramme ausgewertet, Formeln genutzt, Beziehungen zwischen Realität und Modell hergestellt werden etc. Auch hier ist eine Konfundierung mit Textverstehen nicht zu vermeiden, da die Informationen lediglich textlich und (statisch) visuell vorliegen. Diese Einflussgröße wird durch entsprechende Lesekompetenztests

zu erfassen sein. Durch den kumulativen Charakter der kognitiven Prozesse lässt sich aussagen, ob die Schülerinnen und Schüler daran gescheitert sind die notwendigen Informationen zu erkennen und auszuwählen. In dem Fall können sie bereits die Aufgaben auf den Stufen *Reproduzieren* und *Selektieren* nicht lösen, andernfalls scheitern sie an der Strukturierung (*Organisieren*) und Einbettung der Informationen in die Wissensbasis (*Integrieren*). Diesen theoretischen und in Bezug auf Strategien des Textverstehens auch empirischen Ergebnissen zur Folge stellen die so beschriebenen kognitiven Prozesse ein gestuftes Aufgabenmerkmal dar und tragen somit zur Beschreibung von Kompetenzstufen bei.

#### 4.4 Zusammenfassung

Naturwissenschaftliche Kompetenz am Ende der Sekundarstufe I wird durch die Fähigkeit operationalisiert, komplexe Sachverhalte unter den in den vier *Kompetenzbereichen* beschriebenen Blickwinkeln zu bearbeiten. Die beiden Aufgabenmerkmale *Komplexität* und *kognitive Prozesse* werden als allgemeines Strukturmerkmal bzw. allgemeine Informationsverarbeitungsstrategien für naturwissenschaftliche Strukturen und Informationen konkretisiert. Die in den Bildungsstandards vorgeschlagenen vier Kompetenzbereiche werden übernommen: Fachwissen, Erkenntnisgewinnung, Bewertung und Kommunikation.

Durch die Fähigkeiten, die mit den drei Aufgabenmerkmalen *Kompetenzbereich*, *Komplexität* und *kognitive Prozesse* verbunden sind, ist ein dreidimensionales Kompetenzmodell beschrieben (siehe Abbildung 2). Es basiert im Wesentlichen auf empirischen Ergebnissen und erlaubt eine reliable und valide Einordnung jeder Aufgabe in das Modell. Durch das Modell können die in den Bildungsstandards beschriebenen Kompetenzen erfasst werden, weitere Kompetenzen können im Modell verortet werden, indem sie bezüglich des inhaltlichen Ver-

knüpfungsgrads, der Informationsverarbeitung und des Bezugs zu den naturwissenschaftlichen Fächern analysiert werden. Damit gelingt die Strukturierung konkreter, unterrichtlich orientierter Kompetenzen, wie sie in verschiedenen Bildungsplänen oder auch bei HarmoS (vgl. Labudde, 2007) verwendet werden. Die Auflistung solcher Kompetenzen ist im Allgemeinen nicht abgeschlossen und bezüglich Auswahl und Reihenfolge normativ.

Die Operationalisierung von Kompetenz durch dieses Modell ist, wie dargestellt, am normativen Bildungsziel der KMK-Bildungsstandards orientiert. Es können direkte Beziehungen zwischen den in den Bildungsstandards benannten Kompetenzen und den einzelnen Bereichen im Modell hergestellt werden. Durch die drei Dimensionen werden zentrale Merkmale der naturwissenschaftlichen Fächer und des naturwissenschaftlichen Arbeitens und Erkennens mit wesentlichen kognitiven Fähigkeiten und Fertigkeiten kombiniert. Die Kompetenzbereiche *Fachwissen*, *Erkenntnisgewinnung* und *Bewertung* sind aus Sicht der einzelnen naturwissenschaftlichen Fächer im Modell repräsentiert, wobei insbesondere die Erkenntnisgewinnung viele fächerübergreifende Elemente enthält. Eine empirische Prüfung der Konstruktvalidität des Modells wird über einen Vergleich mit Items der PISA-Studie in der Pilotierung erfolgen. Zur diskriminanten Validierung werden zusätzlich kognitive Fähigkeitstests und Lesetests eingesetzt.

## 5 Aufgabenkonstruktion und empirische Prüfung

Das Kompetenzmodell ist Grundlage für die Entwicklung von Testaufgaben. Gleichzeitig ist es ein Rahmen für die Entwicklung von gerade entstehenden Lernaufgaben, so dass eine Beziehung zwischen Lern- und Testaufgaben bestehen wird. Die Testaufgaben werden von Lehrkräften aus den verschiedenen Bundesländern entwickelt und von Experten und Expertinnen der Fachdidaktik des jeweili-

ligen naturwissenschaftlichen Faches, der Psychometrie und der Sprachwissenschaften bewertet. Deren Rückmeldung fließt in den Entwicklungsprozess ein und führt zu einer Überarbeitung der Aufgaben (vgl. Walpuski et al., 2008). Parallel dazu führen sowohl die Entwicklergruppe als auch die fachdidaktische Leitung Prä-Pilotierungen in kleinen Stichproben durch. Diese führen neben ersten empirischen Hinweisen auf die Modell- und Aufgabengüte zur Entwicklung von Distraktoren und Kodieranweisungen. Auf diese Weise wird eine hohe schulische Validität (im Sinne eines implementierten Lehrplans) und Aufgabenqualität sichergestellt. Die curriculare Validität (im Sinne des intendierten Lehrplans) wird in einem zusätzlichen Projekt empirisch überprüft (Notarp, Kauertz & Fischer, 2008).

Die Entwicklung der Aufgaben ist an einem Manual orientiert, das neben der Modellbeschreibung weitere Aufgabenmerkmale festlegt. So werden für die Pilotierung zunächst zwei Typen von Aufgaben unterschieden. Beim ersten Typ sind die für die Bearbeitung der Aufgaben notwendigen Fachinformationen im Aufgabenstamm vorgegeben, der Schwerpunkt liegt auf dem korrekten Umgang mit Fachinformationen. Beim zweiten Typ werden fundamentale Kenntnisse im jeweiligen Fach vorausgesetzt. Für die Aufgaben vom Typ eins sind die für die Lösung unmittelbar erforderlichen Inhalte vollständig angegeben. Die Schülerinnen und Schüler müssen also nicht auf ihr Vorwissen zurückgreifen, sondern die Inhalte zur Lösung der beschriebenen Probleme anwenden. Diese Festlegung entspricht der Auffassung von Kompetenz als Fähigkeit und Fertigkeit zur *Anwendung* fachlicher Inhalte in Problemlösesituationen, außerdem haben die im Detail uneinheitlichen Curricula der einzelnen Bundesländer so geringe inhaltliche Schnittmengen, dass kaum Inhalte als sicher bekannt vorausgesetzt werden können (vgl. Notarp, Kauertz & Fischer, 2008). Beim zweiten Typ von Aufgaben werden Inhalte, die im jeweiligen Fach als fundamental angesehen werden, zur Bearbeitung

als bekannt vorausgesetzt. Sie erfassen neben Kompetenz als Fähigkeit und Fertigkeit somit auch die Kenntnis von Inhalten. Damit wird der Auffassung Rechnung getragen, dass ohne vorhandenes Grundwissen anspruchsvolle fachliche Probleme nicht gelöst werden können. Um den Effekt der im Aufgabenstamm enthaltenen Informationen zu kontrollieren, werden Aufgaben teilweise mit und ohne diese Information pilotiert (Ropohl, Walpuski & Sumfleth, 2009).

Die inhaltliche und sprachliche Gestaltung der Aufgaben orientiert sich nicht an fachlichen Formalisierungen, sondern an der in der Sekundarstufe I üblichen Darstellung. Die Zielpopulation ist der 10. Jahrgang an allgemeinbildenden Schulen mit dem mittleren Schulabschluss. Abbildung 3 zeigt ein Aufgabenbeispiel aus dem Fach Chemie und seine Einordnung in das Kompetenzmodell (vgl. Bernholt et al., 2009; weitere Aufgabenbeispiele sind verfügbar unter [http://www.iqb.hu-berlin.de/arbbereiche/projekte?pg=p\\_34&spg=r\\_7](http://www.iqb.hu-berlin.de/arbbereiche/projekte?pg=p_34&spg=r_7)).

Besondere Bedeutung kommt der sprachlichen und graphischen Gestaltung der Aufgaben zu. Da die Aufgabenstellung über ein textliches Format vermittelt wird, muss der Text möglichst leicht zu verstehen sein, um naturwissenschaftliche Kompetenz mit den Aufgaben valide messen zu können. Dazu orientiert sich die sprachliche Gestaltung an den Textverständlichkeitskriterien nach Sumfleth und Schüttler (1994, 1995). Um die Lesezeit möglichst kurz zu halten, beschränken sich die Texte auf die unmittelbar notwendigen Angaben, was eine kontextuelle Einbettung erschwert, da Kontexte auch unwesentliche Aspekte einer Situation anführen, um anschaulich zu sein und alltagsübliche Graphiken auch nichtfachliche Gestaltungselemente besitzen können.

In den verschiedenen Phasen ihrer Entwicklung wird eine Aufgabe in kleineren Schülerstichproben bearbeitet. In den frühen Phasen werden so Distraktoren gewonnen, Kodieranweisungen angepasst und textliche Ausschärfungen vorgenommen. Sobald eine Aufgabe den gesamten Begutachtungsprozess

## Aufgabenbeispiel Chemie

### Aufgabenstamm:

Peter möchte herausfinden, ob die Geschwindigkeit einer Reaktion von der Temperatur (T) und/oder von der Konzentration (c) der Edukte (Ausgangsstoffe) abhängt. Er weiß bereits, dass der Zerteilungsgrad eines Stoffes die Reaktionsgeschwindigkeit beeinflusst.

### Itemstamm:

In einem Experiment gibt Peter Salzsäure ( $c = 0,1 \text{ mol/L}$ ) in 4 Reagenzgläser. Jedes Reagenzglas stellt er in ein anderes Wasserbad. Diese haben verschiedene Temperaturen ( $T = 10 \text{ }^\circ\text{C}$ ,  $T = 20 \text{ }^\circ\text{C}$ ,  $T = 40 \text{ }^\circ\text{C}$ ,  $T = 80 \text{ }^\circ\text{C}$ ). Er gibt in jedes Reagenzglas einen gleich großen Zinkstreifen. Er misst für jedes Reagenzglas die Zeit, bis sich der Zinkstreifen vollständig aufgelöst hat.

#### Item 1: 1 Zusammenhang selektieren:

Welche Frage kann Peter mit seinem Experiment beantworten? Kreuze an.  
Hängt die Reaktionsgeschwindigkeit...

- von der Konzentration der Säure ab?
- von der Temperatur ab?
- vom Zerteilungsgrad der Edukte ab?
- von der Temperatur und von der Konzentration der Säure ab?

#### Item 2: 1 Zusammenhänge organisieren:

Welches Experiment muss Peter wählen, um zu beweisen, dass der Zerteilungsgrad die Reaktionsgeschwindigkeit beeinflusst? Kreuze an.

	<input type="checkbox"/> Exp. 1	<input type="checkbox"/> Exp. 2	<input type="checkbox"/> Exp. 3	<input type="checkbox"/> Exp. 4
<b>Zeit</b>	Messen	Konstant halten	Messen	Messen
<b>Konzentration</b>	Beliebig	Konstant halten	Beliebig	Konstant halten
<b>Zerteilungsgrad</b>	Gezielt variieren	Beliebig	Konstant halten	Gezielt variieren

Abb. 3 Aufgabenbeispiel Chemie.

durchlaufen hat (vgl. Walpuski et al., 2008), wird sie von der fachdidaktischen Leitung einem ersten kleinen Feldtest unterzogen. Für eine Reihe von Aufgaben zu den Kompetenzbereichen Fachwissen und Erkenntnisgewinnung liegen dadurch bereits erste Ergebnisse vor. Der Umfang dieser Prä-Pilotierungen erlaubt noch keine empirisch fundierten Aussagen. Es zeigt sich aber bereits – unter Vernachlässigung der Unterscheidung

in Fächer und Kompetenzbereiche – mit Hilfe einer ANOVA ein kleiner Effekt der Dimensionen *Komplexität* ( $F = 8,73$ ;  $df = 4$ ;  $p < .001$ ;  $\eta^2 = .12$ ) und *kognitive Prozesse* ( $F = 7,08$ ;  $df = 3$ ;  $p < .001$ ;  $\eta^2 = .08$ ) auf die Aufgabenschwierigkeit (vgl. Abbildung 4 und 5). Aufgrund der Vermischung zahlreicher Varianzquellen und der daraus resultierenden Überlappung der Stufen, ist der Effekt hier eher konservativ geschätzt und für einzelne

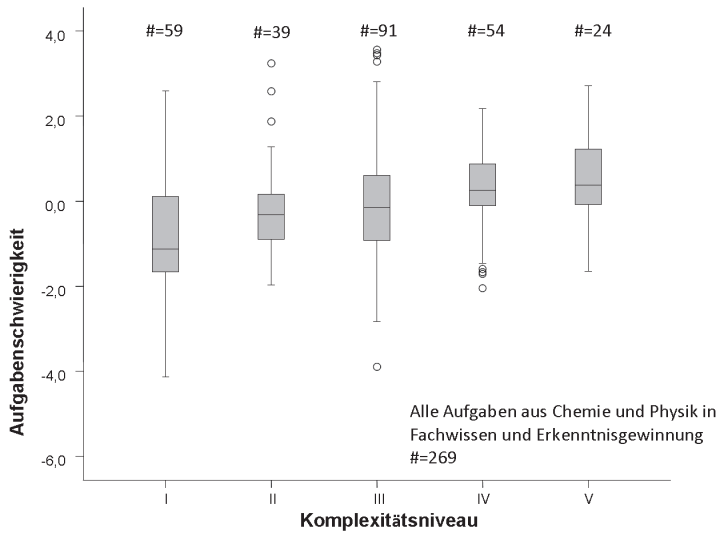


Abb. 4 Ergebnisse der Prä-Pilotierungen: Zusammenhang Komplexitätsniveau und Antwortschwierigkeit.

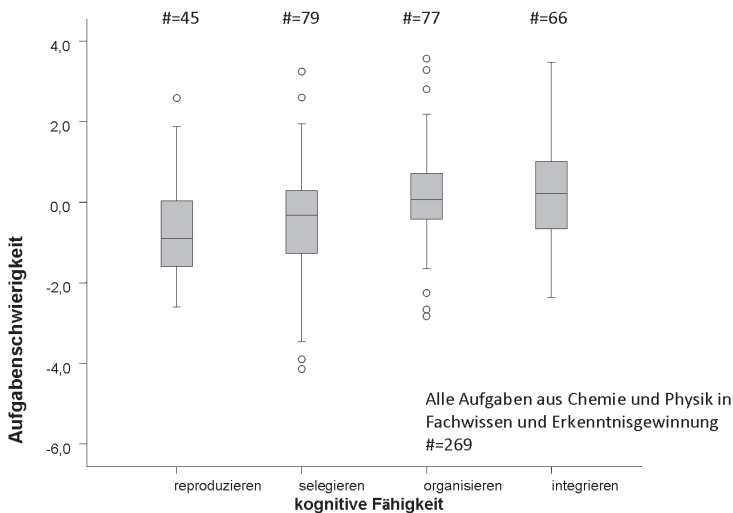


Abb. 5 Ergebnisse der Prä-Pilotierungen: Zusammenhang Kognitive Prozesse und Antwortschwierigkeit.

Fächer und Kompetenzbereiche vermutlich größer. Bislang reicht die Anzahl an Aufgaben und Probanden nicht für detailliertere Aussagen. Hier lassen sich erst durch die Auswertung der im Herbst 2009 durchgeführten Pilotierung Aussagen belegen.

In der Pilotierung werden vor allem Zusammenhänge der Modelldimensionen *Komplexität, kognitive Prozesse* und *Kompetenzbereich* mit der Aufgabenschwierigkeit untersucht. Lassen sich diese Zusammenhänge nachweisen, können die späteren Testergebnisse mit diesem Modell gedeutet werden. Unabhängig davon wird zu prüfen sein, wie hoch der Einfluss der kognitiven Fähigkeiten und der Lesekompetenz auf die so operationalisierte naturwissenschaftliche Kompetenz ist. Insbesondere im unteren Leistungsbereich ist ein hoher Zusammenhang zur Lesekompetenz zu erwarten, da diese den Erwerb und die Prüfung naturwissenschaftlicher Kompetenzen erst ermöglicht (vgl. Baumert, 2002). Erste Ergebnisse der Prä-Pilotierung deuten darauf hin, dass es im Unterschied zu PISA mit dem hier vorgestellten Test gelingen kann, auch den unteren Leistungsbereich durch die gewählte Operationalisierung im Test gut abzubilden und trennscharfe Aufgaben für diese Gruppe zu entwickeln. Diese gute Auflösung in allen Leistungsbereichen erlaubt es erst, Regelstandards zu beschreiben, die tatsächlich am mittleren Leistungsniveau orientiert sind und die darüber hinaus eine Entwicklungsperspektive für untere und mittlere Leistungsbereiche bieten. Die für 2010 geplanten Veröffentlichungen von modellkonformen Lernaufgaben und der Aufgaben-Entwicklungsmanuale sorgen zudem für Transparenz im Verfahren ohne die notwendige Geheimhaltung der Aufgaben aufgeben zu müssen.

## Literatur

- AAAS (1993). *Benchmarks for science literacy*. New York: Oxford University Press.
- Aufschnaiter, S. v. & Welzel, M. (1997). Wissensvermittlung durch Wissensentwicklung. Das Bremer Komplexitätsmodell zur quantitativen Beschreibung von Bedeutungsentwicklung und Lernen. *Zeitschrift für Didaktik der Naturwissenschaften*, 3(2), 43–58.
- Baumert, J. (2002). Deutschland im internationalen Bildungsvergleich. In: Killius, Nelson; Kluge, Jürgen & Reisch, Linda (Hg.), *Die Zukunft der Bildung*. Frankfurt: Suhrkamp.
- Baumert, J., Artelt, C., Carstensen, C.H., Sibberns, H. & Stanat, P. (2002). Untersuchungsgegenstand, Fragestellungen und technische Grundlagen der Studie. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich*. Opladen: Leske + Budrich, 11–38.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W. et al. (2003). *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*. Opladen: Leske + Budrich.
- Beerenwinkel, A.; Parchmann, I. (2008). Metadiskussion über Modelle. Historische Aspekte als Impuls. *Praxis der Naturwissenschaften Chemie*, 4/57 Jg., 13–16.
- Bernholt, S.; Walpuski, M.; Parchmann, I.; Sumfleth, E. (2009). Kompetenzentwicklung im Chemieunterricht – mit welchen Modellen lassen sich Kompetenzen und Aufgaben differenzieren? *Naturwissenschaften im Unterricht – Chemie* 20, 111/112, 78–85.
- Bloom, B. S. (1965). *Taxonomy of educational objectives I: cognitive domain*. New York: Longman Green.
- Bybee, Roger W. (2000). Teaching Science as Inquiry. In: J. Minstrell, Jim and E. H. van Zee (Eds.). *Inquiring into Inquiry Learning and Teaching in Science*. AAAS Publication 00-1S: Washington DC. pp. 20–47
- Commons, M. L., Goodheart, E. A., Pekker, A., Dawson, T. L., Draney, K., & Adams, K. M. (2007). Using Rasch scaled stage scores to validate orders of hierarchical complexity of balance beam task sequences. In E. V. Smith, Jr. & R. M. Smith (Eds.). *Rasch Measurement: Advanced and Specialized Applications* (pp. 121–147). Maple Grove, MN: JAM Press.
- Demuth, R., Ralle, B. & Parchmann, I. (2005). Basiskonzepte – eine Herausforderung an den Chemieunterricht. *Chemie Konkret*, 12, 55–60.



- Eggert, S; Böggeholz, S (2006). Göttinger Modell der Bewertungskompetenz. Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ für Gestaltungsaufgaben Nachhaltiger Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, S. 177–194.
- Fischer, H. E. & Draxler, D. (2006). Konstruktion und Bewertung von Physikaufgaben. In E. Kircher, R. Girwitz & P. Häußler (Hrsg.), *Physikdidaktik. Theorie und Praxis*. 639–655. Berlin: Springer.
- Fischer, H. E., Glemnitz, I., Kauertz, A. und Sumfleth, E. (2007). Auf Wissen aufbauen – kumulatives Lernen in Chemie und Physik. In: Kircher, Girwitz & Häußler (Hrsg.) *Physikdidaktik, Theorie und Praxis*. (S. 657–678) Berlin, Heidelberg, New York: Springer.
- Fischer, H.E. (1994). Physiklernen: Eine Herausforderung für Unterrichtsforschung. In Nachtigall, D. (Hrsg.): *Didaktik und Naturwissenschaften*, Band 3. Frankfurt: Lang.
- Fischer, H.E., Kauertz, A., Neumann, K. (2008). Standards of Science Education. In: Mikelskis-Seifert, Ringelband & Brückmann (Hrsg.) *Four Decades of Research in Science Education – from Curriculum Development to Quality Improvement*. (pp. 29–42) Münster: Waxmann.
- Gräber, W., Bolte, C. (Hrsg.) (1997). *Scientific Literacy – An International Symposium*. Kiel: Institut für Pädagogik der Naturwissenschaften.
- Grube, C., Möller, A., & Mayer, J. (2007). Dimensionen eines Kompetenzstrukturmodells zum Experimentieren. In H. Bayrhuber, U. Harms, D. Krüger, A. Sandmann, A. Unterbrunner, A. Upmeyer zu Belzen & H. Vogt (Eds.), *Ausbildung und Professionalisierung von Lehrkräften*. Internationale Tagung der Fachgruppe Biologiedidaktik im VBIO – Verband Biologie, Biowissenschaften & Biomedizin (pp. 31–34). Kassel.
- Kauertz, A. & Fischer, H. E. (2006). Assessing Students' Level of Knowledge and Analysing the Reasons for Learning Difficulties in Physics by Rasch Analysis. In: Liu, X. & Boone, J. (Eds.), *Applications of Rasch Measurement in Science Education*. (pp. 212–246). Minnesota: Maple Grove.
- Kauertz, A. (2008). *Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben*. Studien zum Physik- und Chemielernen. Berlin: Logos Verlag.
- Klafki, W. (1996): *Neue Studien zur Bildungstheorie und Didaktik. Zeitgemäße Allgemeinbildung und kritisch-konstruktive Didaktik*. 4. Auflage. Weinheim: Beltz.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. J. (2003). *Expertise Zur Entwicklung nationaler Bildungsstandards*. Bundesministerium für Bildung und Forschung (BMBF) (Hrsg.)
- Klos, S., Henke, C., Kieren, C., Walpuski, M. und Sumfleth, E. (2008): Naturwissenschaftliches Experimentieren und chemisches Fachwissen – zwei verschiedene Kompetenzen. *Zeitschrift für Pädagogik* 3, S.304–321.
- KMK [Sekretariat der Ständigen Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland] (Hrsg.) (2005c). *Bildungsstandards im Fach Physik für den mittleren Schulabschluss*. München: Luchterhand.
- KMK [Sekretariat der Ständigen Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland] (Hrsg.) (2005b). *Bildungsstandards im Fach Chemie für den mittleren Schulabschluss*. München: Luchterhand.
- KMK [Sekretariat der Ständigen Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland] (Hrsg.) (2005a). *Bildungsstandards im Fach Biologie für den mittleren Schulabschluss*. München: Luchterhand.
- KMK [Sekretariat der Ständigen Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland] (Hrsg.) (2004). *Einheitliche Prüfungsanforderungen in der Abiturprüfung Physik*. [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/1989/1989\\_12\\_01-EPA-Physik.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1989/1989_12_01-EPA-Physik.pdf) (18.06.2010)
- KMK/IQB (Hrsg.) (2006): Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring. München: Luchterhand.
- Kremer, K., Urhahne, D. & Mayer, J. (2009). Naturwissenschaftsverständnis und wissenschaftliches Denken bei Schülerinnen und Schülern der Sek. I. In U. Harms & A. Sandmann (Hrsg.), *Forschungsbeiträge der biologie-didaktischen Lehr- und Lernforschung* (pp. 29–43). Innsbruck: Studienverlag.
- Labudde, P. (2007). Schule und Unterricht harmonisieren. Bildungsstandards in der Schweiz. *Unterricht Physik* 18 (97). 40–41.
- Lederman, N. G. (2007). Nature of science: Past, present, and future. In S. K. Abell & N. G. Lederman. *Handbook of research on science education* (pp. 831–879). Mahwah: Erlbaum.
- Leutner, D., Fischer, H.E., Kauertz, A., Schabram, N. & Fleischer, J. (2008). Instruktionspsychologische und fachdidaktische Aspekte der Qualität von Lernaufgaben und Testaufgaben im Physikunterricht. In J. Thonhauser (Hrsg.) *Aufgaben als Katalysatoren von Lernprozessen* (pp. 169–181). Münster: Waxmann.

- Mannel, S., Sumfleth, E. & Walpuski, M. (2008). Schwierigkeitsbestimmende Faktoren von Aufgaben zu experimentell-naturwissenschaftlichen Arbeitsweisen für den unteren Leistungsbereich. In D. Höttecke (Hrsg.), Gesellschaft für Didaktik der Chemie und Physik: Chemie- und Physikdidaktik für die Lehramtsausbildung. (pp.380–383). Berlin: Lit.
- Mayer, J., Grube, C. & Möller, A. (2008). Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung. In U. Harms & A. Sandmann (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik – Ausbildung und Professionalisierung von Lehrkräften* (3. Band) (S. 63–79). Innsbruck: Studienverlag.
- McComas, W.F., & Olson, J.K. (1998). The nature of science in international science education standard documents. In W. F. McComas (Ed.). *The nature of science in science education: Rationales and strategies* (pp. 41–52). Dordrecht: Kluwer Academic.
- Mikelskis-Seifert, S.; Fischler, H. (2003). Die Bedeutung des Denkens in Modellen bei der Entwicklung von Teilchenvorstellungen. *Zeitschrift für die Didaktik der Naturwissenschaften*, 9, 75–88.
- National Research Council (1996). *National science education standards*. Washington, DC: National Academy Press.
- 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.
- Nicolis, G. & Prigogine, I. (1987). *Die Erforschung des Komplexen. Auf dem Weg zu einem neuen Verständnis der Naturwissenschaften*. München: Piper.
- Notarp, H., Kauertz, A., Neumann, K. & Fischer, H. E. (2008). Kontenvalidität von Tests über Sachstrukturen von Physikschulbüchern. In D. Höttecke (Hrsg.) *Gesellschaft für Didaktik der Chemie und Physik: Chemie- und Physikdidaktik für die Lehramtsausbildung*. (pp.137–140), Berlin: Lit Verlag.
- OECD [ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT] (1999). *Measuring student knowledge and skills. A New Framework for Assessment*. Paris: OECD Publication Service.
- Osborne, J., Collins, S., Ratcliffe, M., Milar, R., & Duschl, R. (2003). What “ideas about science” should be taught in school science? A Delphi study of the expert community. *Journal of Research in Science Teaching*, 40, 692–720.
- Ropohl, M., Walpuski, M. & Sumfleth, E. (2009). Leistungsmessung im Kompetenzbereich Fachwissen Chemie. In D. Höttecke (Hrsg.) *Chemie- und Physikdidaktik für die Lehramtsausbildung*. (pp. 419–421). Berlin: Lit Verlag.
- Schecker, H. & Parchmann, I.(2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 45–66.
- Sumfleth, E. & Schüttler, S. (1994). Chemische Inhalte und verständliche Texte – ein Widerspruch in sich? *Fachsprache – Internationale Zeitschrift für Fachsprachenforschung, -didaktik und Terminologie*, 16 (3/4), 141–158
- Sumfleth E. & Schüttler, S. (1995). Linguistische Textverständlichkeitskriterien – Helfen sie bei der Darstellung chemischer Inhalte weiter? *Zeitschrift für Didaktik der Naturwissenschaften* 1. 55–72
- 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., Baxter, G. P., Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, 30(3), 215–232.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4 (1), 295–312.
- Walpuski, M., Kampa, N., Kauertz, A. & Wellnitz, N. (2008). Evaluation der Bildungsstandards in den Naturwissenschaften. *Der mathematische und naturwissenschaftliche Unterricht* 61, (6) 323–326.
- Weidenmann, B. (1993). Multicodierung und Multimodalität im Lernprozess. In L. J. Issing & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia*. Weinheim: Psychologie Verlags Union, 65–84.
- Weinert, F. E.(2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen*. S. 17–31. Weinheim und Basel: Beltz.
- Weinstein, C., & Mayer, R. (1986). The teaching of learning strategies. In M.C. Wittrock (Ed.), *Handbook of Research on Teaching*, 3rd Edition (pp. 315–327). New York: Macmillan.
- Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist* 24, 345–376.
- Wunsch, G. (1985). *Geschichte der Systemtheorie : dynamische Systeme und Prozesse*. München: Oldenbourg.

Zilker, I., Holliday, G. M., Fischer, H.E., Kauertz, A., Lederman, J.S., Lederman, N., G. (2010). Are historical contexts suitable for assessing students' competences in the field of nature of science and scientific inquiry? In M.F. Taşar & G. Çakmakçı (Eds.), *Contemporary science education research: international perspectives* (pp. 437–444). Ankara, Turkey: Pegem Akademi.

### Kontakt

Dr. Alexander Kauertz  
Pädagogische Hochschule Weingarten, Physik  
Kirchplatz 2  
88250 Weingarten  
[kauertz@ph-weingarten.de](mailto:kauertz@ph-weingarten.de)

### Autoreninformation

Dr. rer. nat. Alexander Kauertz ist Professor für naturwissenschaftliches Lernen mit Schwerpunkt Physik an der Pädagogischen Hochschule Weingarten. Seine Forschungsschwerpunkte sind Physikkompetenzmodellierung und -diagnose in der Sekundarstufe sowie Unterrichtsanalysen im naturwissenschaftlichen Sachunterricht der Primarstufe.

Dr. rer. nat. Hans Ernst Fischer ist Professor für Didaktik der Physik in der Fakultät für Physik der Universität Duisburg-Essen und Sprecher der DFG-Forschergruppe naturwissenschaftlicher Unterricht. Seine Forschungsschwerpunkte liegen im Bereich der fachspezifischen Unterrichtsforschung, der fachspezifischen Analysen zur Lehrerprofessionalisierung unter Berücksichtigung von Lernvoraussetzungen und Lernergebnissen.

Dr. rer. nat. Jürgen Mayer ist Professor für Didaktik der Biologie an der Universität Kassel. Seine Forschungsschwerpunkte sind die Kompetenzmodellierung im Biologieunterricht, die Entwicklung und Analyse von Lernumgebungen, insbesondere des Forschenden Lernens, sowie Lehrerprofessionalisierung.

Dr. Elke Sumfleth ist Professorin für Didaktik der Chemie und Sprecherin des DFG-Graduiertenkollegs „Naturwissenschaftlicher Unterricht“ an der Universität Duisburg-Essen. Ihre Forschungsschwerpunkte sind Empirische Lehr-Lern-Forschung in Chemie mit unterschiedlichen Schwerpunkten und Entwicklung und Evaluation von Materialien für den Chemieunterricht.

Dr. rer. nat. Maik Walpuski ist Juniorprofessor für Didaktik der Chemie an der Universität Osnabrück und Mitglied der DFG-Forschergruppe Naturwissenschaftlicher Unterricht. Seine Forschungsschwerpunkte sind die Kompetenzmodellierung und Kompetenzmessung für das Fach Chemie sowie die prozessanalytische Untersuchung des Chemieunterrichts, insbesondere in Experimentierphasen.

