

Ein indikatorenbasiertes Verfahren zur Einstufung von Testaufgaben in ein Kompetenzmodell

Heike Theyßen, Marita Schmidt, Erik Einhaus & Horst Schecker

Institut für Didaktik der Naturwissenschaften, Abteilung Physikdidaktik, Universität Bremen
(Eingegangen: 06.09.2006; Angenommen: 01.12.2006)

Kurzfassung

Die Entwicklung von Bildungsstandards und Kompetenzmodellen erfordert eine valide Einschätzung der Anforderungen, die Schüler bei Bearbeitung von Testaufgaben bewältigen müssen. Das hier vorgestellte Verfahren der Erfassung der Kompetenzanforderungen verwendet zwei zentrale Dimensionen — Prozesse und Ausprägungen — des Bremen-Oldenburger Kompetenzmodells: (Schecker & Parchmann 2006). Die Prozesse korrespondieren eng mit den vier Kompetenzbereichen der nationalen Bildungsstandards (KMK 2004a).

Es ergibt sich eine zweidimensionale Kompetenzmatrix, deren Zellen durch Zellindikatoren operationalisiert werden. Mit Hilfe eines webbasierten Systems schätzen Experten die Anforderungen vorgegebener Beispielaufgaben anhand der Zellindikatoren ein. Dabei kommt ein zweistufiges Rating-Verfahren zur Anwendung.

Das Verfahren wurde mit einer Gruppe von Experten erprobt und dabei die Übereinstimmung der Expertenratings a) unter einander und b) mit einer modellkonformen Mustereinstufung bestimmt. Die Ergebnisse der Studie dienen der weiteren Präzisierung der Zelloperationalisierungen. Übergeordnetes Ziel ist ein explizit regelbasiertes Verfahren für die Einordnung von Testitems in das Kompetenzmodell. Das Instrument ermöglicht eine modellbezogene Entwicklung von Testinventaren.

1. Testentwicklung zur modellbasierten Kompetenzdiagnostik

Die nationalen Bildungsstandards (NBS; KMK 2004a) und daran anschließende fachdidaktische Kompetenzmodelle (Schecker & Parchmann 2006) beschreiben mehrdimensionale Strukturen fachbezogener Kompetenz in Physik. Diese Strukturmodelle sollen als Grundlage für eine nach Kompetenzen differenzierte Förderung und Analyse von Schülerfähigkeiten genutzt werden. Gleichzeitig sind die Lehrkräfte aufgefordert, ihre Schülerinnen und Schüler beim Erwerb dieser Kompetenzen zu unterstützen.

„Den Lehrerinnen und Lehrern geben Bildungsstandards ein Referenzsystem für ihr professionelles Handeln. Die Kompetenzanforderungen einzulösen, so gut dies unter den Ausgangsbedingungen der Schülerinnen und Schüler und der Situation in den Schulen möglich ist, ist der Auftrag der Schulen.“ (Klieme 2003, S.19)

Für die gezielte Förderung der in den Bildungsstandards formulierten Kompetenzanforderungen sind speziell darauf abgestimmte Lernaufgaben notwendig. Ebenso müssen für eine modellbasierte Kompetenzdiagnostik geeignete Testinstrumente entwickelt werden. Sowohl im Bereich der Lernaufgaben als auch im Bereich der Testinstrumente stehen bislang nur wenige Beispiele zur Verfügung.

Der hier vorgestellte Zugang zur Konstruktion von Testinventaren berücksichtigt — neben modellbezogenen Neuentwicklungen von Items — die Verwendung der großen Anzahl bereits in der Praxis erprobter Aufgaben: vorhandene Aufgaben werden, ggf. in überarbeiteter Form, in das zugrunde liegende, präskriptive Kompetenzmodell eingeordnet. Die Güte der Diagnostik spezifischer Kompetenzen auf Basis solcher Aufgaben hängt entscheidend von der Objektivität und Reliabilität dieser Einordnung ab.

1.1. Das Bremen-Oldenburger Kompetenzmodell

Die NBS (KMK 2004a) beschreiben die Kompetenzen, die zum Abschluss der 10. Jahrgangsstufe bei den Schülerinnen und Schülern vorliegen sollten. Diese Beschreibung erfolgt in Form eines präskriptiven, zunächst zweidimensional angelegten Strukturmodells. Eine Dimension wird durch vier Kompetenzbereiche aufgespannt: Fachwissen, Erkenntnisgewinnung, Kommunikation und Bewertung. Die zweite Dimension wird durch drei Anforderungsbereiche aufgespannt, die aufeinander aufbauen. Der Anforderungsbereich I umfasst Reproduktion bzw. Nachvollziehen, in den Anforderungsbereichen II und III kommen Anwendung, Auswahl und Transfer von z.B. Fachwissen, Fachmethoden und Darstellungsformen hinzu. Für den Kompetenzbereich „Fachwissen“ lauten die Beschreibungen der Anforderungsbereiche beispielsweise „Wissen wiedergeben“, „Wissen anwenden“ und „Wissen transferieren“.

und verknüpfen“. Im Kompetenzbereich „Fachwissen“ werden neben den Anforderungsbereichen vier Basiskonzepte unterschieden: Materie, Wechselwirkung, System und Energie. Die Basiskonzepte stellen eine inhaltliche Gliederung der fachlichen Inhalte dar und können daher als dritte Dimension des Kompetenzmodells aufgefasst werden. Auch jede Aufgabe zur Erkenntnisgewinnung, Kommunikation oder Bewertung, muss sich inhaltlich einem oder mehreren der vier Basiskonzepte zuordnen lassen.

An den Universitäten Bremen und Oldenburg wurde ein Kompetenzmodell entwickelt, das auf dem oben skizzierten Modell der Bildungsstandards aufbaut und dieses um zwei zusätzliche Dimensionen erweitert. Eine ausführliche Beschreibung und Begründung des Modells findet sich in (Schecker & Parchmann 2006).

Die Dimension „Prozess“ entspricht im Wesentlichen den Kompetenzbereichen der Bildungsstandards. Sie unterscheidet zwischen „Fachwissen nutzen“, „Erkenntnisse gewinnen“, „Kommunizieren“ und „Bewerten“.

Die Dimension „Ausprägung“ unterscheidet, in Anlehnung an die Anforderungsbereiche der Bildungsstandards zwischen „fachlich nominell/reproduktiv“, „aktiv anwendend“ und „konzeptuell vertieft“ und erweitert diese um die Ausprägung „lebensweltlich“.

In der Dimension „Inhaltsbereich/Basiskonzept“ wird, wie in den Basiskonzepten der Bildungsstandards, die inhaltliche Dimension von Kompetenz berücksichtigt. Während die Bildungsstandards mit nur vier Basiskonzepten die gesamte Schulphysik inhaltlich sehr grob gliedern, ist bei den Inhaltsbereichen eine feinere Differenzierung vorgesehen. Die Anzahl der Komponenten in dieser Dimension ist nicht vorab festgelegt.

In der Dimension „Kontext“ werden verschiedene Kontexte unterschieden, in denen eine Kompetenz benötigt bzw. eingesetzt wird. Innerfachliche Kontexte werden hier abgegrenzt zu professionellen und persönlich/gesellschaftlichen Kontexten. In schulischen Lehr-Lernsituationen lassen sich nach den gleichen Kategorien Inhaltskontexte unterscheiden, in die Aufgabenstellungen eingebettet sind.

Die Dimension „Kognitive Anforderung“ erfasst vier der insgesamt sieben Faktoren, durch die in der nationalen Ergänzung zu PISA 2003 die Schwierigkeit der Aufgaben beschrieben wurde: „Konvergentes Denken“, „Divergentes Denken“, „Umgang mit mentalen Modellen“ und „Umgang mit Zahlen“ (Rost et al. 2005).

1.2. Entwicklung modellbasierter Testinstrumente

Testinstrumente zur empirischen Überprüfung der Struktur des oben beschriebenen Kompetenzmodells können aufgrund der Komplexität des Modells jeweils nur Teilbereiche abdecken. Zur Erfassung von Kompetenzstrukturen kann beispielsweise ein In-

haltsbereich ausgewählt und Kontext sowie kognitive Anforderung als Kovariaten erfasst werden. Für eine systematische Variation der Aufgabenanforderungen bleiben dann die Dimensionen „Prozess“ und „Ausprägung“. Sie spannen, wie in Abbildung 1 dargestellt, zweidimensionale Matrix mit 16 Zellen auf. In dieser Projektion entspricht das Kompetenzmodell bis auf die zusätzliche Ausprägung „lebensweltlich“ weitgehend dem Modell der nationalen Bildungsstandards.

Prozess Ausprägung	Fachwissen nutzen	Erkennt- nisse gewinnen	Kommuni- zieren	Bewerten
lebens- weltlich				
nominell/ reproduktiv				
aktiv anwendend				
konzeptuell vertieft				

Abbildung 1: Zweidimensionale Matrix zur Charakterisierung von Aufgaben bei gegebenem Inhaltsbereich (nach Schecker & Parchmann, 2006)

An der Universität Bremen werden für die Inhaltsbereiche „Wärmelehre/Thermodynamik“ (Einhaus & Schecker 2006a) und „Energie“ (Schmidt & Schecker 2006) Testinstrumente für die modellbasierte Kompetenzdiagnostik in den Dimensionen „Prozess“ und „Ausprägungen“ entwickelt. Die Anforderungen, die die Lösung einer Aufgabe stellt, werden zunächst einer oder mehreren Zellen der zweidimensionalen Kompetenzmatrix (Abb. 1) zugeordnet. Diese Einstufung wird durch Expertenratings unterstützt. Das Verfahren wird im folgenden Abschnitt dargestellt. In einem weiteren empirischen Teil werden mit den Aufgaben Schülerleistungen erhoben und anhand der modellbasiert vorgenommenen Einstufungen Kompetenzstrukturen diagnostiziert. Gleichzeitig werden die Testergebnisse dazu genutzt, die Übereinstimmung zwischen empirisch ermittelten Strukturen und den vom Modell vorhergesagten Strukturen zu überprüfen, d.h. um das theoretische Modell empirisch zu überprüfen.

2. Notwendigkeit eines indikatorenbasierten Verfahrens

Ein wesentlicher Schritt in dem oben beschriebenen Verfahren der Testentwicklung und der Überprüfung des Kompetenzmodells ist die Charakterisierung der Aufgabenantworten durch Einordnung in die Zellen der Kompetenzmatrix. Hierzu sollen Erfahrungen von Unterrichtsexperten genutzt werden. Ebenso müssen Lehrkräfte, die spezifische Kompetenzen gezielt fördern bzw. diagnostizieren wollen, die Anforderungen der von ihnen verwendeten Lern- bzw. Testaufgaben in das zugrunde liegende Kompetenzmodell einordnen können.

Erfahrungen aus der Praxis zeigen jedoch, dass dies eine hohe Anforderung darstellt. Einer der Verfasser hat in den letzten zwei Jahren das Aufgabenrating im Rahmen mehrerer Fortbildungsveranstaltungen zur Aufgabenkultur im Kontext der Nationalen Bildungsstandards aufgenommen. Teilnehmer waren erfahrene Lehrkräfte und Setkoordinatoren in den Modellversuchen SINUS-Transfer sowie Lehrkräfte und Fachdidaktiker, die im BMBF-Programm "Chemie im Kontext" mitarbeiteten.

Nach einer Einführung in die Struktur der Bildungsstandards und einer Erläuterung der Kompetenzbereiche erhielten die Teilnehmer eine oder zwei Aufgaben aus den Nationalen Bildungsstandards zur Einordnung in die Kompetenzmatrix. Den Teilnehmern lagen die dafür relevanten Beschreibungen der Kompetenzbereiche und -standards sowie der Anforderungsbereiche aus den NBS schriftlich vor. Für jedes Item gab es eine farbcodierte Antwortkarte, auf der die Zelle angegeben werden sollte, der nach Meinung der Unterrichtsexperten das Item in der Kompetenzmatrix der NBS hauptsächlich zuzuordnen ist. Die Teilnehmer besprachen diese Frage paarweise.

Die Karten wurden anschließend auf einem Meta-plan mit dem Kompetenzmodell geheftet. In dem Graustufenbild in Abbildung 2 sind exemplarisch zur Verdeutlichung die Antwortkarten schwarz bzw. grau mit schwarzem Rand markiert, die sich auf die Items 1 bzw. 3 der Aufgabe „Schilddrüse“ aus den NBS für das Fach Physik beziehen (KMK 2004a, S. 17 ff.; vgl. Anhang). Gleichzeitig sind die in den NBS abgedruckten Musterzuordnungen durch entsprechende Rahmen um Zellen der Matrix markiert. Die Aufgabe thematisiert eine Schilddrüsenuntersuchung mit radioaktiven Markersubstanzen. Die dicken schwarzen Hervorhebungen beziehen sich auf Item 1, bei dem in einer Tabelle gegebene Eigenschaften von Substanzen für eine medizinische Untersuchung von Bedeutung sind (z.B. Halbwertszeit und elektrische Leitfähigkeit). Die Aufgabe bei den grau markierten Kärtchen lautete: „Diskutieren Sie Vorteile und Gefahren einer Untersuchung, bei der radioaktive Substanzen eingesetzt werden“.

Abbildung 2 zeigt anhand der Ergebnisse auf einer Fachtagung zu Bildungsstandards im Rahmen des Projekts "Chemie im Kontext" wie stark die Antworten der Teilnehmer streuen und wie begrenzt die Übereinstimmung mit den Mustereinstufungen ist. Bei mehreren Workshops mit Physiklehrkräften aus dem Modellversuch SINUS-Transfer zeigte sich die gleiche Heterogenität in den Ratings.

Untersuchungen zur Einschätzung von Items des TIMSS III-Tests zur voruniversitären Physik (Klieme et al. 2000) haben gezeigt, dass die Übereinstimmung verschiedener Rater bei der Einschätzung von Aufgaben anhand von Merkmalen selbst dann gering ist, wenn die Merkmale kleinschrittig aufgeschlüsselt werden. Hohe Übereinstimmung wird bei Merkmalen erzielt, die direkt am Aufgabentext ab-

lesbar oder intuitiv verständlich sind. Geringe Übereinstimmungen erzielen prozessbezogene Merkmale. Zur Verbesserung der Übereinstimmung wurden von Einhaus (Einhaus, in Vorbereitung) die Merkmale ausführlich schriftlich erläutert. Diese Erläuterungen wurden jedoch bei der Einschätzung der Merkmale nicht erkennbar genutzt.

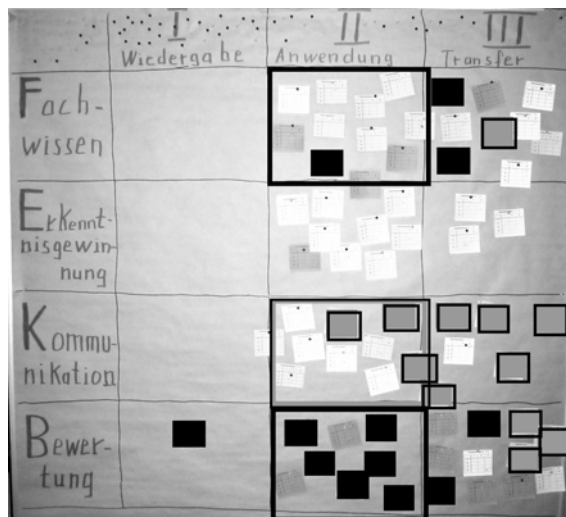


Abbildung 2: Zuordnung von Items zu Zellen der Kompetenzmatrix der nationalen Bildungsstandards; Erläuterungen im Text.

Ein großer Teil der Zellen der Kompetenzmatrix (vgl. Abb. 1) bezieht sich auf Prozesse während der Aufgabenlösung. Zu diesen Zellen sind umfangreiche Erläuterungen, Abgrenzungen und Fallbeispiele notwendig. Sie eignen sich aufgrund der oben genannten Ergebnisse nicht als Merkmale, anhand derer die Experten eine direkte Einordnung der Aufgabenantworten vornehmen könnten.

3. Das indikatorenbasierte Einordnungsschema

3.1. Grundsätzlicher Aufbau

Da eine pauschale Zuordnung von Aufgabenanforderungen zu einer Zelle des Kompetenzmodells zu nachweisbar schlechten Übereinstimmungen führt, wird die direkte Einordnung durch ein mehrstufiges Verfahren ersetzt.

Die Zellbeschreibungen werden in Form von Zellindikatoren operationalisiert. Jede Zelle wird nun durch einen oder mehrere Zellindikatoren charakterisiert. Diese beschreiben konkrete Anforderungen der Aufgabenantwort, z.B. „Fakten, Zusammenhänge oder Begriffe erinnern“ oder „Eigenständige Erstellung einer adressatengemäßen Darstellungsform“. Zu den 16 Zellen der Kompetenzmatrix wurden auf diese Weise über 42 verschiedene Zellindikatoren formuliert. Ausgangspunkt für die Formulierung und Zuordnung der Zellindikatoren waren die Kompetenzstandards der NBS (KMK 2004a), die Anforderungsbereiche der Einheitlichen Prüfungsanforderungen in der Abiturprüfung Physik (KMK 2004b) sowie die Analyse einer großen Zahl von

Beispielaufgaben. Unterrichtsexperten müssen nun die Aufgabenantwort nicht direkt in die 16-zellige Kompetenzmatrix einordnen, sondern lediglich bestimmte Zellindikatoren zuordnen. Über die Zellindikatoren ist die Aufgabenantwort dann eindeutig einer oder mehreren Zellen der Kompetenzmatrix zugeordnet.

Wegen der großen Anzahl und inhaltlichen Breite der Zellindikatoren würde es jedoch einen großen Zeitaufwand erfordern, für jede Aufgabenantwort die Gültigkeit jedes einzelnen Zellindikators einzuschätzen. Deshalb wurden die Zellindikatoren zu Strängen gruppiert. Die Stränge fassen Gruppen von Zellindikatoren zusammen, die den gleichen Prozess beschreiben, d.h. in die gleiche Spalte der Kompetenzmatrix (vgl. Abb. 1) fallen. Der Strang ist also

eindeutig einem Prozess zugeordnet, umgekehrt hat jedoch in der Regel jeder Prozess mehrere Stränge. Für den Prozess „Kommunikation“ werden z.B. die Zellindikatoren zu einem Strang gruppiert, die sich auf eine adressatengerechte Darstellung beziehen (z.B. die adressatengerechte Auswahl aus vorgegebenen Darstellungsformen) und die Zellindikatoren, die sich auf spezielle Anforderungen an die sachgerechte Darstellung beziehen (z.B. die Nutzung fachtypischer Darstellungsformen oder längerer Fachtexte). Innerhalb eines Stranges differenzieren die Zellindikatoren zwischen den verschiedenen Ausprägungen (d.h. Zeilen der Kompetenzmatrix). Abbildung 3 gibt einen Überblick über den Aufbau des Einordnungsschemas.

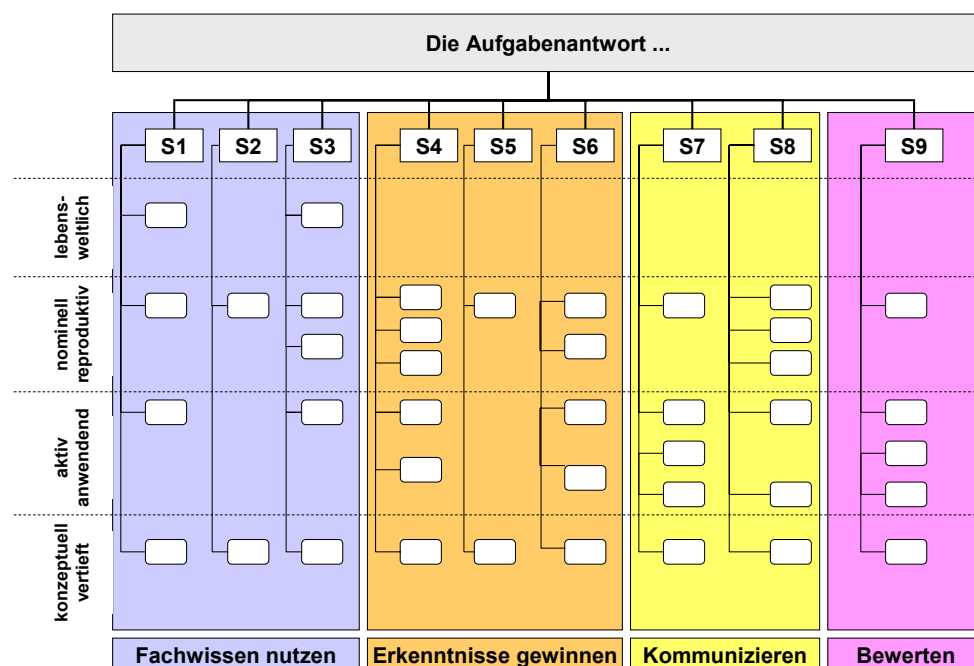


Abbildung 3: Schematische Darstellung der Gruppierung der Zellindikatoren zu Strängen und der Zuordnung zu Prozessen (vertikal) und Ausprägungen (horizontal).

Jeder der insgesamt 9 Stränge besitzt einen Strangindikator, der sozusagen den „Eingang“ zu einem Strang darstellt (Rechecke S1 bis S9 in Abb. 3).

Ein Experte wird nach dem Lesen der Aufgabe zunächst aufgefordert, zu der Aufgabenantwort einen oder mehrere Strangindikatoren auszuwählen (z.B. S6: „Umgang mit Formeln und Zahlen“). Hierbei sollte eine Fokussierung auf die wesentlichen Anforderungen der Aufgabenantwort stattfinden.

In einem zweiten Schritt sind für jeden gewählten Strangindikator nur noch die zugehörigen Zellindikatoren zu betrachten und einzuschätzen. Gleichzeitig wird die Bedeutung einzelner Zellindikatoren durch den übergeordneten Strangindikator zusätzlich illustriert. Der Experte ist nun aufgefordert, zu jedem Strang einen (ggf. mehrere) Zellindikator zur Konkretisierung der durch den Strang grob umrissenen Anforderung auszuwählen (z.B. „Termumfor-

mungen vornehmen“ als Konkretisierung von „Umgang mit Formeln und Zahlen“).

In den folgenden Abschnitten werden die den einzelnen Prozessen zugeordneten Strangindikatoren und Zellindikatoren an Beispielen erläutert.

3.2. Fachwissen nutzen

Im Prozess „Fachwissen nutzen“ lauten die Strangindikatoren „Die Aufgabenantwort ...

(S1) ... erfordert Kenntnisse bestimmter physikalischer Fakten oder Zusammenhänge (die nicht in der Aufgabenstellung enthalten sind).“

(S2) ... beinhaltet die Erklärung einer physikalischen Situation.“

(S3) ... ist Ergebnis einer Rechnung.“

Hier werden mit Benennung, Erklärung und Berechnungen drei Aspekte der Nutzung von Fachwissen unterschieden, die sich gegenseitig nicht ausschlie-

ben. Nur wenige Aufgaben sind so konstruiert, dass der Strangindikator (S1) nicht zutrifft, sondern alle zur Bearbeitung notwendigen Sachinformationen in der Aufgabe angegeben sind. Die veröffentlichten PISA-Aufgaben (OECD 2003) und analog konzipierte Aufgabensammlungen (z.B. Petri & Einhaus 2005) liefern Beispiele für derartige Aufgaben, zeigen jedoch auch, wie umfangreich der Aufgabentext dadurch in der Regel wird.

Einfach ist es hingegen, Aufgaben zu konstruieren, in denen der Strangindikator (S1) die wesentlichen Aufgabenanforderungen bereits vollständig erfasst. Aufgabe 1 ist ein typisches Beispiel hierfür. Es ist lediglich die Siedetemperatur von Wasser zu erinnern. Innerhalb des Strangs (S1) wird daher der folgende Zellindikator ausgewählt:

(Z6) Fakten, Zusammenhänge oder Begriffe erinnern.

Dieser Zellindikator fällt in die Ausprägung „fachlich nominell/reproduktiv“. Die Aufgabe erhält demnach eine eindeutige Einstufung in die zweidimensionale Kompetenzmatrix.

Aufgabe 1

Auf dem Herd steht ein Topf ohne Deckel mit 2 l Wasser. Nach einer kurzen Zeit fängt das Wasser an zu kochen. Wie hoch ist die Temperatur des Wassers jetzt ungefähr?

- A. 90 °C
- B. 100 °C
- C. 110 °C
- D. 120 °C

Aufgabe 2 ist ein Beispiel für eine doppelte Einstufung im Prozess „Fachwissen nutzen“. Hier trifft sowohl der Strangindikator (S1) als auch (S2) zu. Es sind Fakten über das Sieden von Wasser und den Phasenübergang flüssig-gasförmig zu erinnern, so dass im Strang (S1) der gleiche Zellindikator (Z6) und somit die gleiche Ausprägung wie bei Aufgabe 1 zugewiesen wird. Die Aufgabenantwort beinhaltet hier jedoch zusätzlich die Erklärung einer (physikalischen) Situation (S2). Die erinnerten Fakten sind anzuwenden auf die beschriebene Situation. Als Zellindikatoren zum Strang (S2) stehen zur Auswahl:

- (Z1) Erklärung mit lebensweltlichen Konzepten.
- (Z2) Sich an eine physikalische Erklärung erinnern.
- (Z3) Fakten, Zusammenhänge oder Begriffe zur Erklärung einer Situation nutzen.
- (Z4) Erklärung einer Situation aus übergeordnetem Modell ableiten.

Bei Aufgabe 2 wird der Zellindikator (Z3) zugewiesen, was innerhalb des gleichen Prozesses zu einer weiteren Einstufung in der Ausprägung „aktiv anwendend“ führt.

Bei der Unterscheidung zwischen der Reproduktion von Fakten, Zusammenhängen, Erklärungen etc. und deren Ableitung aus übergeordneten Modellen oder

Konzepten spielen Annahmen über die Unterrichtsvalidität eine große Rolle. Die Erwartungen, die ein Entwickler mit einer Aufgabe verbindet (z.B. „die Fakten sollten im Unterricht vorher behandelt worden sein“), müssen nicht mit der unterrichtlichen Praxis übereinstimmen. So unterliegt die Zuordnung einer Aufgabe zu einer Kompetenzausprägung u.a. der Vermutung, ob der zugrunde liegende Sachverhalt im Unterricht in der Regel vorher behandelt wird oder ob die Lösung nur durch Ableitung aus einem übergeordneten Modell zu erhalten ist. Diese Einschätzung beruht auf der eigenen unterrichtlichen Praxis bzw. auf Mutmaßungen über die gängige Praxis.

Aufgabe 2

Ein Koch stellt zwei Töpfe, jeder mit Kartoffeln und Wasser gefüllt, auf einen Herd und wartet, bis das Wasser kocht. Nachdem das Wasser in beiden Töpfen kocht, stellt er die Temperatur der einen Herdplatte so niedrig ein, dass das Wasser gerade am Kochen gehalten wird, die Temperatur der anderen Herdplatte lässt er unverändert. Er glaubt, dass die Kartoffeln im Topf auf der heißeren Herdplatte schneller gar sind. Seine Freundin ist anderer Meinung. Sie behauptet, dass die Kochzeit für die Kartoffeln in beiden Töpfen gleich ist.

Was glaubst du, wer von beiden Recht hat? Begründe deine Antwort.

Bei dem Beispiel aus Aufgabe 1 ist eine Ableitung der Siedetemperatur nicht möglich. Sie muss erinnert werden. Bei Aufgabe 2 ist es grundsätzlich möglich, dass die Fragestellung bereits im Unterricht behandelt wurde und daher die Erklärung nur noch erinnert werden muss. In diesem Beispiel wird mit der Zuweisung des Zellindikators (Z3) unterstellt, dass das sehr unwahrscheinlich ist. Die Tatsache, dass derartige Annahmen über die Unterrichtsvalidität bei vielen Aufgaben eine Rolle spielen, ist ein wesentlicher Grund dafür, Experten mit Unterrichtserfahrung in die Einstufung einzubeziehen.

Aufgaben, bei denen Einstufungen im Prozess „Fachwissen nutzen“ mit Einstufungen in anderen Prozessen kombiniert sind, finden sich in den folgenden Abschnitten.

3.3. Erkenntnisse gewinnen

Auch im Prozess „Erkenntnisse gewinnen“ wird zwischen drei Strängen unterschieden: „Die Aufgabenantwort ...

- (S4) ... benötigt Umgang mit Daten oder einem experimentellen bzw. technischen Sachverhalt.“
- (S5) ... benötigt Umgang mit Modellen und Analogien.“
- (S6) ... benötigt Umgang mit Formeln und Zahlen.“

Strang (S4) umfasst den großen Bereich des (gedanklichen) Umgangs mit Experimenten oder realen

Anwendungssituationen. Hierzu gehören auch die Auswertung und Interpretation gegebener Daten. Kompetenzen in der Planung und Durchführung von Experimenten lassen sich mit herkömmlichen Testaufgaben nur sehr begrenzt erfassen. Hier kann nicht die eigentliche Handlungskompetenz erhoben werden, sondern lediglich das „Wissen über...“. Auch dies ist ein wesentlicher Bestandteil von Strang (S4). Die Stränge (S5) und (S6) erweitern das Spektrum der Gewinnung von Erkenntnissen um den großen Bereich der Nutzung abstrakter Darstellungen. Sie unterscheiden hierbei zwischen Analogien und bildhaften Modellen (z.B. Atommodelle, Schaltpläne) einerseits und Formeln als mathematischen Modellen andererseits. Eine präzise Trennung des Strangs (S6) von Strang (S3) zum Prozess „Fachwissen nutzen“ wird erst anhand der zugehörigen Zellindikatoren möglich. Während Strang (S3) auf die Nutzung von Fachwissen bei der Auswahl und Anwendung einer Formel beschreibt, beziehen sich die Zellindikatoren aus Strang (S6) auf die Methodik, wie z.B. Termumformungen und den Umgang mit Maßeinheiten.

Aufgabe 3 gibt ein Beispiel für eine Einstufung in den umfangreichen Strang (S4), und zwar speziell die Analyse und Interpretation vorgegebener experimenteller Daten. Zwei Zellindikatoren im Strang (S4) betreffen diesen Aspekt:

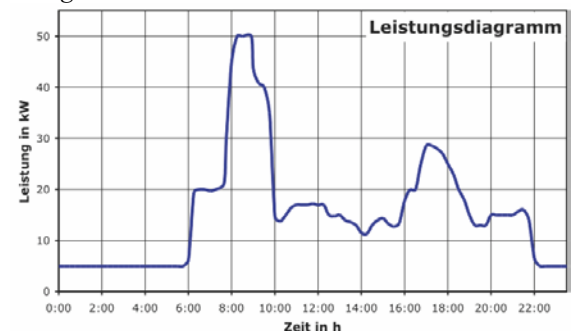
(Z16) Auswertungsmethoden auf gegebene Daten anwenden.

(Z18) Auswertungsmethoden begründet auswählen. Hier wird (Z18) zugewiesen, weil eine geeignete Auswertungsmethode, in der Regel das Auszählen von Kästchen unter der Kurve, selbst zu wählen und zu beschreiben ist. Diese Zuweisung führt zu einer Einstufung in die Ausprägung „aktiv anwendend“, während Zellindikator (Z16) lediglich als reproduktiv eingestuft ist.

Da Faktenwissen über den Zusammenhang zwischen Energie, Zeit und Leistung für die Wahl der Auswertungsmethode notwendig ist, erhält die Aufgabe eine zusätzliche Einstufung im Prozess „Fachwissen nutzen“ (Zellindikator (Z6), vgl. 2.3).

Aufgabe 3

Dieses Leistungsdiagramm zeigt die genutzte elektrische Energie einer Schule mit Sporthalle im Verlauf eines sonnigen Winter-Tages. Wie kann man mit Hilfe des Diagramms die an diesem Tag genutzte Energie bestimmen?



3.4. Kommunizieren

Der Kompetenzbereich „Kommunikation“ der NBS umfasst im Unterschied zum analogen Kompetenzbereich der EPA neben dem adressaten- und sachbezogenen Austausch über physikalische Fragen auch den Aspekt der Erschließung und Zusammenstellung physikbezogener Informationen. Bei einer sehr weit gefassten Interpretation wäre damit fast jede Aufgabe eine Kommunikationsaufgabe, weil Informationen zur Aufgabenstellung und ggf. zum Kontext aus dem Aufgabentext zu erschließen sind (vgl. Schecker & Theyßen, im Druck).

Das Bremen-Oldenburger Kompetenzmodell legt hier eine erheblich engere Interpretation von Kommunikationsfähigkeit zugrunde. Der Aspekt des Erschließens von Informationen aus dem Aufgabentext wird nicht dem Prozess „Kommunikation“ zugeordnet, sondern gesondert betrachtet (vgl. 3.6). Voraussetzungen für eine Einstufung unter „Kommunikation“ sind Anforderungen an eine adressatengerechte Darstellung von Informationen (z.B. die Aufbereitung für den physikalischen Laien oder zur Unterstützung einer bestimmten Argumentation) oder an eine sachgerechten Darstellung, die über einen bloßen Antwortsatz deutlich hinaus geht (z.B. die Erstellung eines Energieflussdiagramms). Beide Aspekte sind durch Strangindikatoren repräsentiert:

„Die Aufgabenantwort ...

(S7) ... lässt Überlegungen über eine adressatengerechte Verständigungsebene erkennen.“

(S8) ... beinhaltet sachgerechte Darstellungsweise.“

Speziell der zweite Strang schließt noch nicht aus, dass auch alle Aufgaben, die lediglich einen kurzen (aber sachgerechten) Antwortsatz verlangen, hier eingestuft werden. Dies geschieht über die zur Verfügung stehenden Zellindikatoren zu diesem Strang. Findet sich hierin keine passende Präzisierung, entfällt auch die Einstufung in den Strang. Die einzigen Zellindikatoren in (S8), die sich auf textbasierte Darstellungen beziehen, sind „Verwendung angemessener Fachsprache in einem zusammenhängenden Text“ und „Diskursives Argumentieren zu physikalischen Sachverhalten auf angemessenem Niveau“, was beides über einen reinen Antwortsatz weit hinaus geht.

Aufgabe 4 ist ein typisches Beispiel für eine Aufgabe zur Umsetzung gegebener Sachinformationen in eine andere Darstellungsform. Die Anforderung liegt in der Erstellung des Schaubildes, wobei die Darstellungsform selbstständig und sachgerecht zu wählen ist. Der passende Zellindikator aus Strang (S8) lautet:

(Z34) Umsetzung einer selbst gewählten geeigneten Darstellungsform

Aufgabe 4

Fasse die unten aufgeführten Informationen mit einem erklärenden Schaubild zusammen:

Mechanische Energie ist der Oberbegriff für Bewegungsenergie, Spannenergie und Lageenergie.

Die Begriffe Bewegungsenergie und kinetische Energie beschreiben die gleiche Energieform.

Lageenergie und Spannenergie sind zwei Formen der potenziellen Energie.

Diese Zuweisung führt zu einer Einstufung in die Ausprägung „aktiv anwendend“ im Prozess „Kommunizieren“.

Aufgabe 5 (aus Schecker & Theyßen, im Druck) ist ein Beispiel für eine Einstufung in den Strang (S7), weil hier die adressatengerechte Darstellung im Vordergrund steht. Im Erwartungshorizont für diese Aufgabe ist enthalten, dass der Antwortbrief die Vermischung von Alltags- und Fachsprache aufgreift und die Begriffe Stromverbrauch, Energieverbrauch und Energieumwandlung differenziert erklärt - in einer für Paul und seine Freunde verständlichen Sprache. Eine solche Antwort erfordert zunächst das Erkennen des Missverständnisses und dann die Auswahl einer adressatengerechten fachlichen Tiefe. Die entsprechenden Zellindikatoren lauten:

(Z31) Fragen und Erklärungsbedarf anderer erkennen

(Z29) Auswählen einer adressatengemäßen fachlichen Tiefe

Fachwissen, das zur Klärung der Begriffe benötigt wird, führt zu zusätzlichen Einstufungen im Prozess „Fachwissen nutzen“ (Zellindikatoren (Z6) und (Z3), vgl. 3.2).

Aufgabe 5

Lisa, Paul und Simone (9.Klasse) unterhalten sich während der großen Pause.

Lisa fragt: "Im Physikunterricht heißt es immer, dass Energie nicht verbraucht, sondern umgewandelt wird. Dann gilt das auch für elektrische Energie, oder?"

Simone antwortet: "Na klar, das ist doch auch Energie!"

Lisa: "Und was ist mit Strom? Mein Vater sagt immer, ich solle Strom sparen, weil der so teuer geworden sei."

Paul (halb im Scherz): "Lass uns doch mal an das Elektrizitätswerk schreiben, dass wir aus Prinzip keine Stromrechnung mehr bezahlen. Strom wird doch gar nicht verbraucht."

Wie müsste das Energieversorgungsunternehmen antworten? Stelle Dir vor, Du bist Dipl. Ing. Gerdes aus der technischen Abteilung der "Energieversorgung Nord". Verfasse einen kurzen Antwortbrief (ca. eine halbe Seite).

3.5. Bewerten

Ähnlich wie der Prozess „Kommunizieren“ legt der Prozess „Bewerten“ eine zu weit gefasste Interpretation nahe. Ein Fehler, der sich auch bei der Kategorisierung der Beispielaufgaben in den NBS findet, ist die Zuordnung der Anforderung „Bewertung“ bei der Beurteilung eines Sachverhalts unter rein innerfachlichen Aspekten, wie der fachlichen Korrektheit einer Aussage, der Angemessenheit einer experimentellen Vorgehensweise oder der Zuverlässigkeit von Messergebnissen (vgl. Schecker & Höttecke, im Druck). Solche „Bewertungsaufgaben“ würden vielmehr Einstufungen in den Prozessen „Fachwissen nutzen“ und „Erkenntnisse gewinnen“ erhalten (z.B. im Strang (S4), vgl. 3.3). Sowohl in den NBS (KMK 2004a) als auch in den EPA (KMK 2004b) werden überfachliche Bewertungskompetenzen beschrieben, z.B. „Die Schülerinnen und Schüler vergleichen und bewerten alternative technische Lösungen auch unter Berücksichtigung physikalischer, ökonomischer, sozialer und ökologischer Aspekte“ oder „Die Prüflinge sind vertraut mit Bewertungsansätzen und sind in der Lage, persönlich, sachbezogen und kritikoffen Stellung zu beziehen.“

Auch im Bremen-Oldenburger Kompetenzmodell bezieht sich der Prozess „Bewerten“ auf eine überfachliche Bewertungskompetenz, wie die Unterscheidung zwischen physikalischen und nicht-physikalischen Fragestellungen und Perspektiven sowie deren Nutzung zur Meinungsbildung oder Entscheidungsfindung. Hierzu gehört die Beurteilung der Auswirkung physikalischer Erkenntnisse und Entwicklungen in anderen (gesellschaftlichen, technischen, sozialen, politischen,...) Kontexten. Die Unterschiede zwischen den genannten Aspekten der Bewertungskompetenz liegen in der Ausprägung. Im Einstufungsschema werden sie deshalb in einem Strang zusammengefasst:

(S9) „Die Aufgabenantwort benötigt die Verbindung oder Unterscheidung naturwissenschaftlicher und nicht-naturwissenschaftlicher Perspektiven.“

Aufgabe 6

Vier Personen diskutieren darüber, wie ein Energiespar-Haus am besten beheizt werden könnte. Unten sind diese Personen und verschiedene Aussagen dargestellt. Von wem könnte welche Aussage stammen? Zeichne je eine Linie von der Person zu der Aussage, die er/sie am wahrscheinlichsten gemacht haben könnte.

Ein zukünftiger Bewohner	„Ich empfehle eine Gasheizung. Die ist zwar nicht ganz so umweltfreundlich, dafür hat sie geringe Kosten und ist bewährt.“
Eine Geldgeberin (finanziert den Bau)	„Ich bevorzuge eine Solaranlage. Sie nutzt die Energie der Sonne und benötigt weder fossile noch andere Brennstoffe.“
Eine Naturwissenschaftlerin	„Obwohl eine Holzheizung umweltfreundlich ist, würde ich sie auf keinen Fall wählen. Da muss man ja mehrmals am Tag Holz nachlegen.“
Ein Umweltschützer	„Holzheizungen arbeiten Kohlenstoffdioxid-neutral und können einen sehr hohen Wirkungsgrad erreichen. Deshalb würde ich eine Holzheizung empfehlen.“

Aufgabe 6 ist ein Beispiel für eine Bewertungsaufgabe, in der verschiedene, naturwissenschaftliche und andere (Geldgeberin, Umweltschützer, zukünftiger Bewohner) Perspektiven einander gegenüber gestellt werden. Die Sachinformationen sind in den vorgegebenen Argumentationen enthalten. Die Anforderung besteht darin, die Perspektive zu beurteilen, aus der die unterschiedlichen Argumentationen vorgebracht werden. Der hierfür zutreffende Zellindikator lautet:

(Z36) Argumente einer Position zuordnen

Der Zellindikator ist der Ausprägung „fachlich nominell/reproduktiv“ zugeordnet.

3.6. Informationen erschließen als Zusatzkodierung

Das Erschließen von Informationen aus dem Aufgabentext und Zusatzmaterial ist im Bremen-Oldenburger Kompetenzmodell aus nicht, wie in den NBS, dem Prozess „Kommunizieren“ zugeordnet (vgl. 3.4). Dennoch ist zu erwarten, dass die Art, wie die zur Bearbeitung nötigen Informationen in der Aufgabenstellung kodiert sind, Einfluss auf die Lösungswahrscheinlichkeit hat. Werden beispielsweise schon alle für die Lösung wesentlichen Größen (mit den üblichen Formelzeichen) im Aufgabentext gegeben oder sind sie einer Situationsbeschreibung oder Grafiken zu entnehmen? Sind die Angaben genau passend oder in hohem Maße redundant? Deshalb werden neben der Einordnung der Aufgabenantworten in das Kompetenzmodell zusätzlich die Anforderungen an das Erschließen von Informationen aus der Aufgabe kodiert. Dies geschieht in einem weiteren Strang, der in diesem Modell keinem der Prozesse zugeordnet ist und optional mit in das Ratingschema aufgenommen werden kann. Somit bleibt das Schema also kompatibel mit dem Modell der NBS.

4. Erprobung des Verfahrens

4.1. Zusammenstellung eines Aufgabensets

Zur Erprobung des Einstufungsverfahrens wurde ein Set von zehn Aufgaben zu den Inhaltsbereichen Wärmelehre und Energie einer Gruppe von acht Ratern vorgelegt, die nicht an der Entwicklung des Kompetenzmodells oder der Indikatoren beteiligt waren. Das Aufgabenset beinhaltet die oben vorgestellten sechs Beispielaufgaben.

Bei der Zusammenstellung der Aufgaben wurde angestrebt, möglichst viele Zellen der Kompetenzmatrix zu erfassen. Die Mustereinstufung weist bei den zehn Aufgaben insgesamt 16 verschiedene Zellindikatoren aus zehn unterschiedlichen Zellen der Kompetenzmatrix zu, Zellindikatoren aus dem Bereich „Fachwissen nutzen“ zum Teil bei mehreren Aufgaben.

4.2. Durchführung der Befragung

Für die Befragung wurde ein im Folgenden beschriebenes webbasiertes Verfahren entwickelt und eingesetzt. Es führt die Rater schrittweise durch die Einstufung der Aufgaben und verlangt anschließend einen Abgleich der eigenen Einstufung mit einer Mustereinstufung.

Die einzustufende Aufgabe wird vorgegeben, und bei offenen Antwortformaten zusätzlich ein Erwartungshorizont, auf den sich die Einstufung beziehen soll.

Im ersten Schritt sind lediglich die Strangindikatoren auszuwählen. Dies geschieht zugunsten der Übersichtlichkeit und um sicherzustellen, dass zunächst nur die Strangindikatoren und anschließend nur die jeweils zugehörigen Zellindikatoren in die Beurteilung der Anforderungen eingehen.

Anschließend wird jeder ausgewählte Strangindikator noch einmal mit allen zugehörigen Zellindikatoren vorgelegt. Die mit der Auswahl des Strangindikators getroffene Einstufung ist nun anhand der Zellindikatoren zu verfeinern. Dies geschieht nacheinander für alle gewählten Strangindikatoren.

Zum Abschluss wird einem weiteren Schritt die eigene Einstufung der Aufgabe zusammengestellt und zum Vergleich der Mustereinstufung gegenüber gestellt. Bei Zellindikatoren der Mustereinstufung, die nicht in der eigenen Einstufung enthalten sind, ist der Rater gebeten, zu entscheiden, ob

- er sich der Mustereinstufung anschließt,
- er sie nachvollziehen kann, aber sich ihr dennoch nicht anschließt oder
- er sie nicht nachvollziehen kann.

Für Zellindikatoren der eigenen Einstufung, die in der Mustereinstufung nicht auftreten, ist zu entscheiden, ob sie angesichts der Alternativen in der Mustereinstufung

- beibehalten oder
- verworfen werden.

Hier bietet sich auch die Möglichkeit, eine zu umfangreiche Einstufung mit vielen Zellindikatoren so

zu reduzieren, dass die Hauptanforderungen beschrieben werden.

4.3. Beurteilung der Übereinstimmung

Je nach Fragestellung können verschiedene Kriterien für die Beurteilung der Übereinstimmung zwischen den Ratern herangezogen werden.

Um die Eignung der Zellindikatoren zu überprüfen, ist eine Auswertung bezüglich einzelner Indikatoren notwendig, auch wenn diese auf die gleichen Zellen der Kompetenzmatrix (Abb. 1 und 3) verweisen. Um die Einstufung der Aufgabe in die Kompetenzmatrix zu überprüfen, reicht eine Auswertung bezüglich der zugewiesenen Zellen aus. Werden von unterschiedlichen Ratern unterschiedliche Zellindikatoren zugewiesen, die auf die gleiche Zelle der Matrix verweisen, läge in dieser Hinsicht dennoch eine Übereinstimmung vor. Die Betrachtung der Übereinstimmung gesamter Einstufungen ist nicht sinnvoll, weil die Zellindikatoren sehr unterschiedliche Aspekte der Aufgabenanforderung beschreiben. Es kann durchaus bezüglich einer Anforderung (z.B. an eine adressatengemäße Darstellung) große Übereinstimmung der Rater bestehen, während bezüglich eines anderen Aspekts keine Übereinstimmung erzielt wird. Beispielsweise führen bereits andere Annahmen über die Unterrichtsvalidität zu unterschiedlichen Einschätzungen darüber, ob Fakten erinnert oder aus einem übergeordneten Konzept abgeleitet werden müssen. Es werden deshalb die Übereinstimmungen jeweils für Aufgabe-Zellindikator-Kombinationen (bei Auswertung bezüglich einzelner Zellindikatoren) betrachtet. Die Auswertung nach Zellen führt zu gleichen oder höheren Übereinstimmungen und wird hier nicht im Detail vorgestellt.

Es werden zunächst alle Zellindikatoren der Mustereinstufung betrachtet. Zur Quantifizierung der Einstufungen und des Abgleichs mit der Mustereinstufung werden diese auf einer vierstufigen Skala abgebildet:

- 3 „Auswahl“: Der Zellindikator wurde vom Rater selbst ausgewählt.
- 2 „Anschließen“: Der Rater schließt sich der Zuweisung des Zellindikators an.
- 1 „Nachvollziehen“: Der Rater kann die Zuweisung des Zellindikators nachvollziehen.
- 0 „Nicht nachvollziehen“: Der Rater kann die Zuweisung des Zellindikators nicht nachvollziehen.

Die Skalenwerte zu jeder Aufgabe-Zellindikator-Kombination werden in einem Ratingvektor mit acht Einträgen (für die acht Rater) zusammengefasst. Ein Ratingvektor (3,3,2,0,1,3,3,2) besagt, dass die Rater 1, 2, 6 und 7 den Zellindikator vor Einsicht in die Mustereinstufung selbst zugewiesen haben, die Rater 3 und 8 sich der Mustereinstufung bezüglich dieses Indikators anschließen, Rater 5 sie nachvollziehen kann und Rater 4 nicht.

Als Maß für die Einigkeit der Rater wird der von Einhaus und Schecker (2006b) konstruierte Einigkeitskoeffizient η verwendet. Der Koeffizient berücksichtigt zwei Möglichkeiten der Nichtübereinstimmung von Ratern:

1. Die Anzahl der paarweise nicht übereinstimmenden Einstufungen pro Aufgabe und Indikator.
2. Den Abstand (auf der Skala von 0 bis 3) zwischen den Einstufungen pro Aufgabe und Indikator als Grad der Unterschiedlichkeit.

Einzelne „Ausreißer“ (alleinige Minima bzw. Maxima im Zeilenvektor) werden bei der Berechnung von η regelbasiert gestrichen, sofern sich dadurch ein höherer Wert ergibt.

Die genaue Konstruktionsvorschrift ist in (Einhaus, in Vorbereitung) angegeben. Der Koeffizient nimmt Werte zwischen 0 und 1 an. Bei einer vierstufigen Skala wird bei Werten ab 0,4 von einer Übereinstimmung der Rater gesprochen und der im Zeilenvektor vorherrschende Skalenwert zugewiesen. Übereinstimmung bedeutet hier, dass die Rater mehrheitlich diese Einstufung selbst getroffen haben bzw. sich mehrheitlich anschließen konnten usw.

Die Bevorzugung von η gegenüber der Varianz als Maß für die Einigkeit mehrerer Rater erklärt das folgende Beispiel. Für die Ratingvektoren (0,0,1,1,2,2,3,3) und (0,1,3,3,3,3,3,3) erhält man praktisch die gleiche Varianz (ca. 1,4). Dennoch würde man im zweiten Fall viel eher als im ersten Fall von einer Einigkeit sprechen. Der Koeffizient η quantifiziert diese intuitive Einschätzung durch Werte von 0,09 und 0,41. Der zweite Fall würde also gerade noch als Einigkeit gewertet, der erste nicht.

4.4. Ergebnisse

Tabelle 1 zeigt für alle Aufgabe-Zellindikator-Kombinationen zu den hier vorgestellten Beispielaufgaben die Ratingvektoren, die η -Werte, die im Falle einer Übereinstimmung zugewiesenen Skalenwerte. Darüber hinaus ist angegeben, wie viel Prozent der Rater bereits selbst den betreffenden Zellindikator auswählen (Auswahl (3) in %) und, kumuliert, wie viel Prozent sich der Zuweisung des Indikators zumindest anschließen können (Auswahl und Anschließen (2 und 3) in %).

Die Tabelle zeigt, dass bei der überwiegenden Zahl der Zellindikatoren eine sehr hohe Übereinstimmung der Rater mit der Mustereinstufung erzielt werden kann. In 7 der 12 Fälle können sich die Rater der Mustereinstufung zu 100% anschließen. In 3 Fällen wird die Mustereinstufung bereits vor dem Abgleich von 100% der Rater gewählt.

Eine besonders deutliche Übereinstimmung wird bei den Zellindikatoren aus dem Bereich „Fachwissen nutzen“ erzielt (Z6, Z3 und Z9), insbesondere beim Erinnern von Fakten (Z6). Die geringsten Übereinstimmungen erzielen Zellindikatoren aus dem Bereich Kommunizieren, z.B. (Z29) bei Beispielaufgabe 5.

Beispiel aufgabe	Zellindikator	Ratingvektor (sortiert)	Auswahl (3) in Prozent	Auswahl und Anschließen (3 und 2) in %	η	Übereinstimmung bei ...
1	Z6	(3,3,3,3,3,3,3,3)	100	100	1,00	Auswahl (3)
2	Z3	(1,2,3,3,3,3,3,3)	75	87,5	0,54	Auswahl (3)
	Z6	(3,3,3,3,3,3,3,3)	100	100	1,00	Auswahl (3)
	Z38	(0,1,1,2,2,2,2,3)	12,5	62,5	0,49	Anschließen (2)
3	Z6	(2,2,3,3,3,3,3,3)	75	100	0,73	Auswahl (3)
	Z18	(2,2,2,2,2,2,3,3)	25	100	0,57	Anschließen (2)
4	Z34	(2,2,3,3,3,3,3,3)	75	100	0,73	Auswahl (3)
5	Z3	(1,1,2,2,2,2,3,3)	25	75	0,36	-
	Z6	(2,2,3,3,3,3,3,3)	75	100	0,73	Auswahl (3)
	Z29	(1,1,2,2,3,3,3,3)	50	75	0,34	-
	Z31	(1,2,3,3,3,3,3,3)	75	87,5	0,73	Auswahl (3)
6	Z36	(3,3,3,3,3,3,3,3)	100	100	1,00	Auswahl (3)

Tabelle 1: Auswertung des Ratings der Beispielaufgaben durch acht Experten

Die weiteren Aufgaben, die hier nicht vorgestellt sind, liefern ähnliche Ergebnisse. Die Vermutung liegt nahe, dass dies auf die bislang vorherrschenden Aufgabenformen zurückzuführen ist, in denen es vornehmlich um die Reproduktion und Anwendung von Fachwissen geht. Anforderungen im Bereich „Kommunikation“, insbesondere, wenn sie das Eingehen auf spezielle Adressaten betreffen, sind für Physikaufgaben bislang eher ungewöhnlich und daher schwieriger zu beurteilen. Anforderungen an eine fachgerechte Darstellung (Aufgabe 4) werden hingegen übereinstimmender beurteilt.

Die Berechnung des Einigkeitskoeffizienten η ergibt für 10 der 12 Fälle eine hinreichende Übereinstimmung der Rater. In 8 Fällen wählen die Rater übereinstimmend den Zellindikator aus (3), in 2 Fällen schließen sie sich übereinstimmend der Mustereinstufung an (2). Ein Vergleich mit den berechneten Prozentanteilen für „Auswahl“ und „Anschließen“ sowie mit den Ratingvektoren zeigt, dass η die erwartete Quantifizierung der intuitiven Einschätzung der Übereinstimmung liefert.

In einigen Fällen ergeben sich noch deutlich höhere Übereinstimmungen, wenn man statt der einzelnen Zellindikatoren die Zellen betrachtet, die ggf. mehrere Indikatoren enthalten. Dies gibt Hinweise darauf, dass die Indikatoren für bestimmte Zellen noch nicht trennscharf formuliert sind und ggf. zusammengefasst werden können.

Einen weiteren Hinweis auf die Eignung des Einstufungsverfahrens liefert die Analyse der von den Ratern zusätzlich zur Mustereinstufung zugewiesenen Zellindikatoren. Hierbei werden nur die Zuweisungen betrachtet, die im zusammenfassenden Überblick beibehalten werden (vgl. 4.2). Bei einigen Aufgaben ergibt sich ein breites Spektrum an zusätzlichen Nennungen, wobei die Nennungshäufigkeit jedoch nur in wenigen Fällen über 25% (2 von 8 Ratern) hinausgeht. Die betroffenen Aufgaben sind primär Aufgaben zu den Prozessen „Kommunikation“ und „Bewerten“. Von den oben aufgeführten Beispielaufgaben treten nur bei Aufgabe 5 Zusatz-

nennungen mit Nennungshäufigkeiten über 25% auf. Diese betreffen ausschließlich Zellindikatoren der Prozesse „Kommunizieren“ und „Bewerten“. Im Einzelfall geben Zusatznennungen mit hoher Nennungshäufigkeit ($\geq 50\%$) Anlass, die Mustereinschätzung zu erweitern bzw. die Anforderungen der Aufgabe in dem genannten Punkt zu reduzieren.

5. Zusammenfassung und Ausblick

Zur Operationalisierung des Bremen-Oldenburger Kompetenzmodells in den Dimensionen „Prozesse“ und „Ausprägungen“ wurden Zellindikatoren formuliert, die die Basis eines zweistufigen Einstufungsschemas für Aufgabenanforderungen darstellen.

Mit einer webbasierten Umsetzung dieses Systems können Rater Aufgaben einstufen und ihre Einstufung mit einer Mustereinstufung abgleichen.

Das System dient primär der konsensfähigen Charakterisierung von Aufgaben hinsichtlich ihrer Kompetanzanforderungen. Zusätzlich liefern die Ergebnisse Hinweise für die Auswahl, Formulierung und Zellzuweisung der Indikatoren und dienen damit der Weiterentwicklung der Operationalisierung des Kompetenzmodells.

Erste Erprobungen zeigen eine befriedigende Übereinstimmung der Rater untereinander und mit der nachträglich präsentierten Mustereinschätzung. Dies gilt insbesondere bezüglich der Einstufung in die Zellen der zweidimensionalen Kompetenzmatrix, und damit erfüllt das System seinen primären Zweck.

Die grundsätzliche Tendenz, dass die Rater mehr Stränge und Zellindikatoren auswählen als in der Mustereinstufung enthalten sind, wird darauf zurückgeführt, dass sie nicht auf die Hauptanforderungen fokussieren. Daher wird beim weiteren Einsatz verstärkt darauf hinzuweisen sein, dass die Einstufung auf die für die Bearbeitung wesentlichen Anforderungen beschränkt werden soll.

Besonders hohe Übereinstimmung in der Zuweisung erzielen Zellindikatoren aus dem Prozess „Fachwissen nutzen“, besonders geringe solche aus den Pro-

zessen „Kommunizieren“ und „Bewerten“. Zusammenhänge mit der vorherrschenden Aufgabekultur können anhand der vorliegenden Datenbasis nur vermutet werden. Diese Vermutungen können jedoch Ausgangshypothesen für weitere Untersuchungen mit Hilfe des vorgestellten Systems sein.

6. Literatur

Einhaus & Schecker 2006a EINHAUS, E.; SCHECKER, H.: Item-Merkmale im Expertenrating. In: PITTON, A. (Hrsg.) *Lehren und Lernen mit Neuen Medien. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung der GDCP in Paderborn 2005*, Münster: LIT-Verlag, 2006, S. 111-113

Einhaus & Schecker 2006b EINHAUS, E.; SCHECKER, H.: Erstellung eines Testinventars für den Themenbereich Wärmelehre/Thermodynamik. In: NORDMEIER, V.; OBERLÄNDER A. (Hrsg.): *Didaktik der Physik. Beiträge zur Frühjahrstagung Kassel 2006*. Berlin: Lehmanns-Media, 2006

Einhaus, in Vorbereitung EINHAUS, E.: Schülerkompetenzen im Bereich Wärmelehre — Entwicklung eines modellbezogenen Tests. Dissertation, Universität Bremen (in Vorbereitung)

Klieme 2003 KLIEME, E.: *Zur Entwicklung nationaler Bildungsstandards*. Bonn: BMBF, 2003

KMK 2004a SEKRETARIAT DER STÄNDIGEN KONFERENZ DER KULTUSMINISTER DER LÄNDER IN DER BUNDESREPUBLIK DEUTSCHLAND (Hrsg.): *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss*. München: Luchterhand, 2004

KMK 2004b SEKRETARIAT DER STÄNDIGEN KONFERENZ DER KULTUSMINISTER DER LÄNDER IN DER BUNDESREPUBLIK DEUTSCHLAND (Hrsg.): *Ein-*

heitliche Prüfungsanforderungen in der Abiturprüfung Physik. Beschluss der Kultusministerkonferenz vom 01.12.1989 i.d.F. vom 05.02.2004

OECD 2003 *PISA 2003 Beispielaufgaben aus dem Naturwissenschaftstest*. OECD Programme for International Student Assessment, 2003

Petri & Einhaus 2005 PETRI, J.; EINHAUS, E.: *Wege zu einer neuen Aufgabekultur. Beispiele aus dem Bereich der naturwissenschaftlichen Grundbildung. Band 1 und 2*. Bremen: Senator für Bildung und Wissenschaft / Universität Bremen, Institut für Didaktik der Naturwissenschaften, 2005

Rost et al. 2005 ROST, J.; WALTER, O.; CARSTENSEN, C. H.; SENKBEIL, M.; PRENZEL, M.: Der nationale Naturwissenschaftstest PISA 2003. In: *Der mathematische und naturwissenschaftliche Unterricht* (2005), 58(4), S. 196-204

Schecker & Parchmann 2006 SCHECKER, H.; PARCHMANN, I.: Modellierung naturwissenschaftlicher Kompetenz. In: *Zeitschrift für Didaktik der Naturwissenschaften* (2006), 12, S. 45-66

Schecker & Höttecke, in Druck SCHECKER, H.; HÖTTECKE, D.: Aufgaben zum Kompetenzbereich Bewerten. Erscheint in: *Unterricht Physik*

Schecker & Theyßen, in Druck SCHECKER, H.; THEYßEN, H.: Aufgaben zum Kompetenzbereich Kommunikation. Erscheint in: *Unterricht Physik*

Schmidt & Schecker 2006 SCHMIDT, M.; SCHECKER, H.: Kompetenzmodellierung im Themenbereich Energie - Entwicklung eines Testinventars. In: PITTON, A. (Hrsg.): *Lehren und Lernen mit Neuen Medien. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung der GDCP in Paderborn 2005*, Münster: Lit Verlag (2006), S. 108 - 110

Anhang

Beispielaufgabe „Schilddrüse aus den nationalen Bildungsstandards (KMK, 2004a, S. 17 ff.)

Schilddrüse

Zur Untersuchung einer Schilddrüse soll eine geeignete radioaktive Substanz (als sogenannter Marker) ausgewählt werden. Diese Substanz wird in einer Verbindung mit anderen Stoffen vom Patienten eingenommen und verteilt sich durch Stoffwechselprozesse im Körper. Mit einer besonderen Kamera wird nach einigen Stunden die Stärke der Strahlung, die von der Substanz ausgeht, für jeden Punkt der Schilddrüse aufgenommen und daraus ein Bild berechnet. Auf diesem Bild sind Veränderungen erkennbar.

S 1. Entscheiden Sie jeweils, ob die in den Tabellen aufgeführten Eigenschaften für eine medizinische Nutzung von Bedeutung sind.

Eigenschaften von Substanzen (Marker) allgemein

giftig		grün		reflektierend		elektrisch leitend	
ja	nein	ja	nein	ja	nein	ja	nein

Halbwertszeit		ausscheidbar		Teilchendurchmesser		nachweisbar	
ja	nein	ja	nein	ja	nein	ja	nein

S 2. Welche der angegebenen Substanzen A, B, C, D ist für die beschriebene Untersuchung geeignet? Begründen Sie ihre Entscheidung auf der Basis der folgenden Tabelle.

Strahlungseigenschaften von Substanzen (Marker)

Substanz	Strahlungsart	mittlere Reichweite		Halbwertszeit *
		in Luft	in Gewebe	
A	α	3,8 cm	0,1 mm	4 Stunden
B	β	5,5 m	2,5 cm	6 Stunden
C	β	6,7 m	4,2 cm	25 Jahre
D	γ	viele m	einige m	mehrere Stunden

* Die Halbwertszeit gibt die Zeit an, in der die ursprüngliche Strahlungsintensität einer Substanz auf die Hälfte abgesunken ist.

S 3. Diskutieren Sie Vorteile und Gefahren einer Untersuchung, bei der radioaktive Substanzen eingesetzt werden.