

Einstellungen und Vorstellungen von Lehrpersonen zum Kompetenzbereich Bewertung der Nationalen Bildungsstandards

Zusammenfassung

Mit der Einführung der Nationalen Bildungsstandards durch die KMK 2004 hat die Diskussion um die Förderung von „Bewertungskompetenz“ in den naturwissenschaftlichen Fächern an Aktualität gewonnen. Für eine erfolgreiche Implementation der Kompetenzorientierung in der Schule ist die fundierte Kenntnis der Vorstellungen von Lehrkräften von zentralem Interesse. Die vorliegende Studie erhebt die Einstellungen und Vorstellungen von Lehrpersonen bezüglich des Kompetenzbereichs Bewertung in einem explorativen und rekonstruktiven Design auf Basis von 37 Interviews mit Biologie-, Chemie-, Physik- und Politiklehrkräften. Zu Beginn des Artikels wird ein Überblick über die nationale und internationale Diskussion um Bewertungskompetenz skizziert, um die vorliegende Studie einordnen zu können. Im Ergebnisteil des Artikels werden ausgewählte Ergebnisse dargestellt. Es können vier verschiedene Vorstellungen der Lehrkräfte von Bewertung identifiziert werden. Es wird eine Typologie der Lehrkräfte entwickelt, auf deren Basis Hinweise für die Gestaltung von Fortbildungsangeboten gegeben werden können. Die Typologie zeigt eine schwach ausgeprägte Spezifik der Zugehörigkeit der Lehrkräfte zu Unterrichtsfächern und Schulformen.

Schlagwörter: Bewertungskompetenz, Lehrervorstellungen, Implementation

Teachers' beliefs about their students' ability to make judgments

In 2004 common national standards for science education have been released. There are three separated documents for biology, chemistry, and physics teaching in secondary school which all focus on the development of a competency enabling students to make well-founded judgments and decisions. A prerequisite for a successful implementation of the standards is a thorough knowledge of science teachers' beliefs about judgement and decision-making. The article is presenting in the beginning an overview on the national standards and the discussion in science education about judgement and decision-making. The study presented here is based on 37 interviews conducted with teachers teaching physics, chemistry, biology or politics. The results of the study cover a range of four beliefs about judgement and decision-making. It was possible to develop a typology of teachers covering four different types. The typology allows insights into teachers' beliefs and which kind of problems they might face regarding the facilitation of this competency. The study indicates a weak relation between the typology and the subject the teachers do teach.

Keywords: judgement and decision-making, competence, teachers' beliefs, implementation

1. Ausgangslage

Mit der Einführung der Nationalen Bildungsstandards 2004 durch die KMK wurde für den naturwissenschaftlichen Unterricht die Forderung nach der Förderung von Bewertungskompetenz bei Schülerinnen und Schülern bis zum mittleren Schulabschluss in Deutschland stärker in die Diskussion gebracht als bisher. International wird diese

Diskussion bereits seit einigen Jahren unter verschiedenen Schlagworten geführt: naturwissenschaftliche Grundbildung bzw. „scientific literacy“ (z. B. Bybee 1997; Roth/Lee 2004; Sadler 2004; Linder et al. 2011), „science for citizenship“ mit besonderem Augenmerk auf die Ermöglichung von Laien-Experten-Dialogen (z. B. Cross/Price 1999; Jenkins 1999; Kolstoe 2001; Collins/Evand 2007) und „decision making in socio scientific issues“ (z. B. Ratcliffe 1997; Zeidler 2003; Höttecke et al. 2010; Sadler 2011).

Die Forderungen an Naturwissenschaftsunterricht bezüglich der Förderung von Bewertungskompetenz sind vielfältig. So soll die naturwissenschaftliche Bildung ein Verständnis für die Welt eröffnen und zum mündigen Bürger erziehen, der an gesellschaftlichen Diskursen teil hat, abgewogene Entscheidungen auf der Basis naturwissenschaftlichen Wissens und unter Berücksichtigung ethischer Wertvorstellungen fällt und mit Hilfe naturwissenschaftlicher Grundbildung in gesellschaftlichen Entscheidungen begründet Position bezieht.

In den Bildungsstandards der KMK werden für die drei Naturwissenschaften Biologie, Chemie und Physik je vier Kompetenzbereiche - Fachwissen, Erkenntnisgewinnung, Kommunikation und Bewertung - formuliert, die in der Darstellung der KMK gleich gewichtet sind. Damit erhalten auch die prozessbezogenen Kompetenzen, neben dem Fachwissen, eine zentrale Stellung für die Ausrichtung des Naturwissenschaftsunterrichts. Eine Analyse der Bildungsstandards der drei Fächer für den Kompetenzbereich Bewertung ergibt aber, dass sich das Verständnis von Bewertungskompetenz für die Fächer unterscheidet. Während in den Standards für Biologie die Explikation und Benennung von Werten und Normen als Ziel formuliert wird (vgl. KMK 2005a), findet sich in Physik diese Forderung nicht. Vielmehr wird in den Bildungsstandards für das Fach Physik so etwas wie Zukunftsorientierung verlangt, wenn davon die Rede ist, dass Grenzen naturwissenschaftlichen Wissens aufgezeigt, Risiken bewertet und Auswirkungen abgeschätzt werden sollen (vgl. KMK 2005c). Auch die Generierung alternativer Vorschläge zielt auf die Fähigkeit ab, Zukunft mit zu gestalten. In den Bildungsstandards für das Fach Chemie wird insbesondere auf Anwendungsbezüge fokussiert (vgl. KMK 2005b). Dabei handelt es sich sowohl um Anwendungsbereiche und -beispiele als auch um Transferaspekte. Insbesondere die Diskussion von lebenswelt- und gesellschaftsrelevanten Fragestellungen und Aussagen wird für das Fach Chemie im Kompetenzbereich Bewertung thematisiert.

Die Bildungsstandards senden aber auch je Fach den Fachlehrkräften keine klaren Botschaften, was unter Bewertungskompetenz genau zu verstehen sei. In der Übersicht zu den Kompetenzbereichen der Physik heißt es z. B., Bewertung meine „Physikalische Sachverhalte in verschiedenen Kontexten erkennen und bewerten“ (KMK 2005c: 7) zu können. Der Aspekt, der Nutzung physikalischer Sachverhalte in gesellschaftlichen Problemkontexten, wird erst in den Erläuterungen (ebd.: 10), der Beschreibung der Standards (ebd.: 12) und der Festlegung von Anforderungsbereichen (ebd.: 14) genannt. Die vordergründige und unspezifische Botschaft fokussiert einen rein innerfachlichen Bewerten-Begriff, wie er bereits kritisiert worden ist (Schecker/Höttecke 2007), der auch in den Aufgabenbeispielen der Bildungsstandards Physik bestätigt wird.

Damit geben die Bildungsstandards für die Förderung von Bewertungskompetenz im naturwissenschaftlichen Unterricht bereits eine recht heterogene Ausgangslage vor. Es ist nicht geklärt, ob Bewertung über die naturwissenschaftlichen Domänen hinweg als einheitliche Kompetenz aufgefasst oder in domänenspezifische Bewertungskompetenzen untergliedert werden muss. Diese sowohl normativ als auch empirisch zu klärende Frage stellt ein Desiderat fachdidaktischer Forschungs- und Entwicklungsarbeit dar.

In der Politikdidaktik stellt sich die Sache anders dar. In den Anforderungen an Bildungsstandards, die von der Gesellschaft für politische Jugend- und Erwachsenenbildung (GPJE) 2004 auf Anfrage der KMK hin formuliert wurden, ist zentraler Gegenstand der dort als politische Urteilsfähigkeit bezeichneten Kompetenz die Erziehung zur Mündigkeit und die Ermöglichung von Partizipation an gesellschaftlichen Diskursen. Dabei gilt die Fähigkeit zur politischen Teilhabe als bedeutsames Bildungsziel, durch das Individuen gestärkt werden und der gesellschaftliche Fortbestand der Demokratie gesichert werden soll. In der Politikdidaktik hat die Förderung der Fähigkeit zum Urteilen und Entscheiden eine lange Tradition (z. B. Sutor 1971; Bundeszentrale für politische Bildung 1997), sodass davon ausgegangen werden kann, dass dieser Kompetenzbereich in weit größerem Maße zum Selbstverständnis des Faches Politik gehört als dies in den naturwissenschaftlichen Fächern erwartbar ist.

Mit dem durch die Bildungsstandards zwar uneinheitlich aber normativ gesetzten Ziel der Förderung von Bewertungskompetenz ergeben sich Anforderungen an die Fachlehrkräfte in den Naturwissenschaften. Aus der Forschung ist bekannt, dass die Vorstellungen von Lehrerinnen und Lehrern bei der Gestaltung von Lehr-Lern-Umgebungen handlungsleitend sind (Müller/Duit 2004). Darüber hinaus wissen wir, dass eine mangelnde Passung zwischen subjektiven Theorien der Lehrpersonen und Unterrichtsmethoden, die von ihnen erwartet werden, Schwierigkeiten bei der Umsetzung im unterrichtlichen Handeln mit sich bringen, wie Watson, Swain und McRobbie (1999) am Beispiel offen gestalteter Experimentierumgebungen zeigen konnten. In diesem Sinne stellen die Vorstellungen und Einstellungen der Lehrkräfte zum Kompetenzbereich Bewertung eine wichtige Bedingung für gelingende Implementation dar (Haney/Czerniak/Lumpe 1996). Schröder und Fischler (2002) formulieren die Erkenntnis, dass Unterricht nur dann nachhaltig positiv beeinflusst werden kann, wenn die Vorstellungen von Lehrerinnen und Lehrern erweitert werden. Für erfolgreiche Implementationen curricularer Neuorientierungen bildet die Kenntnis von Lehrervor- und -einstellungen also eine wichtige Basis. Denn um das Unterrichtshandeln von Lehrpersonen erfolgreich verändern zu können, bedarf es einer fundierten Kenntnis ihrer Vorstellungen (Fischler 2001). Die Frage nach den Lehrerperspektiven gewinnt ihre Bedeutsamkeit also aus dem Implementationsproblem des Kompetenzbereichs Bewertung. Generell gelten curriculare Innovationen, die das System Schule top-down verändern sollen, als dahingehend problematisch, dass sie ein hohes Maß an Vertrauen in die bruchlose Umsetzbarkeit einer Innovation voraussetzen. Lehrkräfte dagegen interpretieren die Innovation vor dem Hintergrund ihrer eigenen biographisch erworbenen Erfahrungen und (fach-)kulturell überlieferten Einstellungen und Vorstellungen. Sie adaptieren die Innovation beeinflusst

von ihrem je vorhandenen kognitiven und affektiven Hintergrund (z. B. Snyder et al. 1992; Geijssel et al. 2001).

Die der Implementation voraus liegenden Einstellungen und Vorstellungen über Bedeutung und Wirksamkeit der Innovation sind die personenbezogenen Variablen der Implementation. Lehrpersonen sind „gatekeeper“ curricularer Innovationen (Barab/Luehmann 2003), sodass deren Akzeptanz für eine Innovation über den Erfolg mitentscheidet (Lee/Witz 2009). Dabei ist uns klar, dass eine generelle Akzeptanz einer Innovation bestenfalls eine notwendige Bedingung erfolgreicher Implementation bedeutet (Witz/Lee 2009).

2. Theoretischer Hintergrund

Im Rahmen dieser Untersuchung wird Bewertungskompetenz als die Fähigkeit verstanden, in sachlich, ethisch und emotional komplexen Problemsituationen begründet zu urteilen und zu entscheiden, indem Handlungsoptionen erkannt, generiert und abgewogen werden. Tatsächlich müsste der Begriff Urteilen dem Begriff Bewerten vorgezogen werden. Wir werden den Begriff des Bewertens aber beibehalten, um den begrifflichen Anschluss an die Bildungsstandards zu wahren und sehen eine weitgehende Übereinstimmung der Bedeutungen der beiden Begriffe.

Ein Leitziel der Entwicklung von Bewertungskompetenz ist die Befähigung zur Partizipation an gesellschaftlichen Diskursen. Beispiele sind das Urteilen über Handlungsoptionen, um dem Klimawandel zu begegnen, oder das Urteilen über die Verwendung von Mobiltelefonen, die „Elektrosmog“ und die damit verbundenen Gefahren erhöhen können. In den Fachdidaktiken wurde wiederholt darauf hingewiesen, dass Bewertungsprobleme zumindest im Sinne von „socio-scientific issues“ sich oft durch wissenschaftliche Unsicherheit auszeichnen (Kolstoe 2000; Albe 2008b). Es liegt z. B. unterschiedliche wissenschaftliche Evidenz über den Klimawandel und seinen Verlauf oder über die Auswirkungen von „Elektrosmog“ auf Mensch und Gesundheit vor. Ein Urteil über den wissenschaftlichen Tatbestand ist daher ohne weiteres gar nicht möglich. Vielmehr ist ein Wissen über Wissenschaft selbst notwendig, um zu erkennen, dass wissenschaftliche Unsicherheit oder wissenschaftliche Kontroversen keine Ausnahmezustände sind, sondern zumindest im Bereich aktueller und noch nicht kanonisierter Forschung den Normalfall abgeben.

Menthe unterscheidet vier Themenbereiche, in denen naturwissenschaftliches Urteilen in der Literatur dargelegt wird: a) ökologische Bewertungskompetenz, Umweltbildung, b) moralische Urteilsfähigkeit, c) Alltagsentscheidungen und d) naturwissenschaftliche Grundbildung als Voraussetzung demokratischer Teilhabe („mündiger Bürger“, „citizenship“) (vgl. Menthe 2006: 10). Inhaltlicher und fachlicher Rahmen der zur Rede stehenden Interviewstudie stellt der Klimawandel als komplexes ökologisches Problemfeld dar. Dabei ist die „Besonderheit von ökologischen Fragen [...], dass sie zugleich einen naturwissenschaftlichen wie einen ethisch-politischen Zugang bedingen

und daher eng mit Fragen des Urteilens und Bewertens verknüpft sind“ (Menthe/Parchmann 2006: 80).

Im Rahmen naturwissenschaftlicher fachdidaktischer Forschung sind insbesondere in der Biologiedidaktik Versuche der Operationalisierung von Bewertungskompetenz unternommen worden. Das Göttinger Modell der Bewertungskompetenz (Eggert/Bögeholz 2006) strukturiert für den Bereich der nachhaltigen Entwicklung¹ Teilkompetenzen von Bewertungskompetenz. Dabei ist die Struktur eng angelehnt an das Metamodell der Entscheidungsfindung nach Betsch und Haberstroh (vgl. Betsch 2005), in dem zwischen der präselektionalen, der selektionalen und der postselektionalen Phase unterschieden wird. Da in der postselektionalen Phase auf die Handlungsintention abgehoben wird, von Kompetenz aber in erster Linie die kognitiven Dimensionen in der Operationalisierung berücksichtigt werden, bezieht das Göttinger Modell die ersten beiden Phasen in das Modell mit ein. Die vier Teilkompetenzen des Göttinger Modells umfassen das „Kennen und Verstehen von Nachhaltiger Entwicklung“, das „Kennen und Verstehen von Werten und Normen“, das „Generieren und Reflektieren von Sachinformationen“ und das „Bewerten, Entscheiden und Reflektieren“ (Eggert/Bögeholz 2006: 189). Dabei können die ersten beiden Teildimensionen als Inhaltsbereich verstanden werden, während die beiden nachgenannten Teilkompetenzen die Performanz des Bewertens betreffen. Weiterhin unterscheiden die Autorinnen mit Bezug auf Modelle der Entscheidungspsychologie (Jungermann/Pfister/Fischer 2005) zwischen kompensatorischen und non-kompensatorischen Vorgehensweisen. Das Abwägen von Informationen beschreibt eine kompensatorische Strategie. Non-kompensatorische Entscheidungsstrategien sind dagegen dadurch gekennzeichnet, dass bei Verletzen eines bestimmten qualitativen oder Überschreiten eines bestimmten quantitativen Schwellwertes eine oder mehrere Handlungsoptionen ausgeschlossen werden. In den empirischen Untersuchungen mit diesem Modell wird darauf geachtet, die Inhaltsgebiete und Fragestellungen so zu wählen, dass die kompensatorischen Strategien angewendet werden können, um den Prozess des Abwägens in den Entscheidungsprozess zu integrieren.

Das Modell von Eggert und Bögeholz fokussiert die kognitiven und rationalen Aspekte von Bewertungskompetenz und deren modell- und evidenzbasierte Strukturierung. Dabei bleiben volitionale und affektive Aspekte trotz ihrer großen Bedeutsamkeit für die Bereitschaft, eine Bewertung in abwägender, argumentierender und sachkundiger Weise vorzunehmen, noch unberücksichtigt. Die Rolle intuitiver ad-hoc-artiger Urteilsstrategien wird nicht berücksichtigt. Wir gehen davon aus, dass die Bereitschaft zum rationalen Urteilen in Kontexten gering ist, die stark emotional aufgeladen sind oder die Interessen des Subjekts eng berühren (z. B. Haidt 2001; Sadler/Zeidler 2005). Die Rolle nicht bewusster Vor-Urteile und Bewerten-Routinen, wie sie sich im Begriff der Intuition (Gebhard 2007) niederschlagen, ist bislang noch nicht ausreichend erforscht, gilt aber als relevant (Sadler/Zeidler 2005; Menthe 2006). Die Rolle von Fachwissen

1 „Nachhaltige Entwicklung ist eine Entwicklung, die die Lebensqualität der gegenwärtigen Generation sichert und gleichzeitig zukünftigen Generationen die Wahlmöglichkeit zur Gestaltung ihres Lebens erhält.“ (Bericht der Brundtland-Kommission 1987)

für Bewertungsprozesse in Kontexten, die die Naturwissenschaften betreffen, kann auf Basis bisheriger Studien als gering eingeschätzt werden (Lewis/Leach 2006). Empirische naturwissenschaftliche Evidenz wird dagegen – wenn überhaupt – dazu genutzt, eine eigene Meinung post-hoc argumentativ zu rechtfertigen (Simmons/Zeidler 2003; Sadler/Donnelly 2006).

Ethische Aspekte werden im Bereich moralischer Urteilsfindung relevant, wie sie in den Bildungsstandards Biologie genannt werden². Höhle und Bayrhuber (2006) formulieren ein Phasenmodell aus sechs Schritten, die gleichsam Charakteristika moralischer Urteilsfindung darstellen: 1.) Definieren eines geschilderten Dilemmas, 2.) Aufzählen möglicher Handlungsoptionen, 3.) Aufzählen ethischer Werte, welche die Handlungsoption impliziert, 4.) Unterscheiden zwischen deontologischen und konsequentialistischer Argumentationsweise (Systematisieren von Werten), 5.) begründete Urteilsfällung und Diskussion andersartiger Urteile und 6.) Aufzählen von Konsequenzen, die das eigene und das andersartige Urteil implizieren.

In der Literatur wurde darauf hingewiesen, dass Urteilen weit weniger rational gesteuert ist, als die hier diskutierten Modelle zunächst nahe legen. Die Forschung zeigt, dass Menschen generell kaum entlang elaborierter Bewertungsstrategien urteilen und entscheiden, sondern zumeist intuitive Entscheidungen fällen, die, wenn überhaupt, post-hoc begründet werden (vgl. Haidt 2001). Bei der Begründung von Urteilen sind persönliche Motive vorrangig (Bell/Lederman 2003; Albe 2008a). Es bedarf offenbar strukturierter Lerngelegenheiten, intuitive Bewertungsstrategien an sich selbst zu erkennen, auf sie – wie Gebhard (2007) fordert – zu reflektieren und durch einen kontra-intuitiven, sachlich begründeten Bewertungsprozess zu ergänzen. Wir gehen dabei nicht davon aus, dass es Ziel von Unterricht sein sollte, intuitive Bewertungsprozesse gleichsam auszumerzen. Die Bewertung „aus dem Bauch heraus“ ist kognitiv entlastend und macht Menschen damit in Alltagssituationen handlungsfähig. Komplexe Probleme erfordern jedoch auch komplexe Bewertungsstrategien, über die die Schülerinnen und Schüler gegebenenfalls verfügen sollten (Höttecke et al. 2010). Eine Funktion von Bewertungskompetenz förderndem Unterricht besteht dann darin, dass Schülerinnen und Schüler eigene Bewertungs-Strukturen erkennen, reflektieren und zu mehr begründeten, argumentativ abgesicherten Strukturen entwickeln können. In diesem Kontext sind Modelle wie das von Höhle und Bayrhuber (2006) u. E. sinnvoll.

Wie sich Bewertungskompetenz bei Schülerinnen und Schülern fördern lässt, wie scientific literacy und demokratische Teilhabe als Bildungsziele erreicht werden können (z. B. Kortland 2001; Menthe 2006) oder wie sich Bewertungskompetenz bei Schülerinnen und Schülern in Niveaustufen darstellen und messen lässt (Reitschert/Höhle 2007; Mittelsten Scheid/Höhle 2008; Eggert/Bögeholz 2006), wurde und wird an verschiedenen Stellen untersucht. Über welche Vorstellungen und Einstellungen Lehrpersonen

2 Z. B. in der Formulierung: „Schülerinnen und Schüler setzen beim systematischen Bewerten von Handlungsmöglichkeiten diese mit ethischen Werten in Beziehung. [...] Auf dieser Basis entwickeln und vertreten sie unter Berücksichtigung individueller und gesellschaftlich verhandelbarer Werte einen eigenen Standpunkt. Durch die ethische Bewertung wird die naturwissenschaftliche Perspektive im engeren Sinne ergänzt.“ (KMK, 2005a: 12)

insbesondere der naturwissenschaftlichen Fächer zum Kompetenzbereich Bewertung verfügen, ist bisher ungeklärt. In der vorzustellenden Studie untersuchen wir daher Aspekte der Lehrerperspektiven auf den Kompetenzbereich Bewertung im Rahmen eines qualitativen Designs.

3. Design der Studie

Das unterschiedliche Begriffsverständnis des Kompetenzbereichs Bewertung in den Bildungsstandards über die drei naturwissenschaftlichen Fächer hinweg führt zu der Annahme, dass dieser auch von den Lehrkräften der Naturwissenschaften nicht einheitlich verstanden wird. Um dieser Annahme nachzugehen, sollen jene Vorstellungen, die Lehrerinnen und Lehrer zum Kompetenzbereich Bewertung der Nationalen Bildungsstandards artikulieren, rekonstruiert werden. Zusätzlich stellt sich die Frage, ob es sich um kohärente Einstellungen und Vorstellungen innerhalb der jeweiligen Fächer handelt. Die Hauptforschungsfrage³ lautet:

Welche (unterschiedlichen) Vorstellungen und Einstellungen äußern Lehrerinnen und Lehrer verschiedener Fachkulturen (Physik, Chemie, Biologie, Politik) zum Kompetenzbereich „Bewertung“ der Nationalen Bildungsstandards und welche generellen pädagogisch-didaktischen Orientierungen liegen diesen Äußerungen zugrunde?

Um diese Frage aufzuklären, wurde ein exploratives, qualitatives, die Perspektiven der Befragten rekonstruierendes Verfahren gewählt. Rekonstruierende Verfahren sind im doppelten Sinne rekonstruktiv. In der Annahme, dass Wirklichkeit konstruiert wird, wird der Lebenswelt, dem Relevanzsystem (vgl. Bohnsack 2003) der Befragten viel Raum gegeben, ihr Handeln und ihre Vorstellungen darzulegen, Bedeutungen zu generieren und deutlich zu machen. Bohnsack formuliert dazu: „Hierzu ist eben Voraussetzung, dass denjenigen, die Gegenstand der Forschung sind, Gelegenheit gegeben wird, ihre Konstruktion und ihr kommunikatives Regelsystem zu entfalten“ (ebd.: 24). In zweiter Bedeutung heißt rekonstruktiv: den Forschungsprozess permanent rekonstruierend und überprüfend, abgleichend mit der eigenen erkenntnistheoretischen und methodologischen Praxis zu gestalten.

Im Verständnis der rekonstruierenden Forschung wird davon ausgegangen, dass der Beobachtungsgegenstand der oder dem Forscher/in in der bloßen Beobachtung unzugänglich ist – erst die Interpretation der Konstruktion bzw. die Deutung des Handelns erschließt das Handeln selbst (vgl.: 23).

Für die Untersuchung wurde ein Interviewleitfaden entwickelt und mit vier Lehrkräften unterschiedlicher Fächer pilotiert. Dabei wurden insbesondere Fragen der Ver-

3 Die Formulierung der Hauptforschungsfrage orientiert sich an den Bildungsstandards für die naturwissenschaftlichen Fächer. Für das Fach Politik wird in den Anforderungen an nationale Bildungsstandards der Kompetenzbereich mit „Politischer Urteilsfähigkeit“ bezeichnet und gilt hier als eingeschlossen.

ständigkeit, der erwünschten thematischen Ausrichtung und der schlüssigen Reihenfolge der Interviewfragen beachtet und in Teilen revidiert.

Um dem Relevanzsystem der Lehrkräfte genügend Entfaltungsraum zu geben, den inhaltlichen Schwerpunkt aber auf die Vorstellungen und Einstellungen zu Bewertungskompetenz zu fokussieren, haben wir uns für ein halbstrukturiertes, problemzentriertes, leitfadengestütztes Interview entschieden. Dabei werden die Lehrkräfte aufgefordert, zu ausgewählten Aspekten ihre Überlegungen und Gedanken ausführlich darzustellen.

Die Interviewfragen sind so gestaltet, dass sie zu Beginn des Interviews möglichst offen gestellt werden, um den Interviewten die „Entfaltung ihres Relevanzsystems“ bezogen auf die Frage zu ermöglichen. Dies stellt nach Bohnsack (2003) eine zentrale Qualitätskontrolle der Erhebung dar, weil die Interviewten selbst ihre Deutung der Frage und damit ihr Verständnis bzw. Unverständnis auslegen. Über Nachfragen und die Konkretisierung der nachfolgenden Fragen wird das Interview im weiteren Verlauf stärker fokussiert und strukturiert, um möglichst alle forschungsrelevanten Aspekte im Interview zu berücksichtigen. Die erste offen gestellte Interviewfrage⁴ lautet:

Wir sind an Ihrer Meinung interessiert – was halten Sie persönlich von dem Kompetenzbereich „Bewertung“ der Nationalen Bildungsstandards?

In dieser ersten Frage soll den Lehrkräften die Möglichkeit gegeben werden, erst einmal ganz allgemein ihre Einschätzung und Kenntnis, ihr Verständnis und ihre Assoziationen zum Thema zu erläutern. Um die thematische Einordnung von „Bewertung“ als Kompetenzbereich der Bildungsstandards zu gewährleisten, wurde den Interviewten eine Übersichtsgrafik aus den Bildungsstandards als weiterer Stimulus präsentiert. Die Frage nach einer „vorläufigen“ Definition von Bewertungskompetenz stellt dann im weiteren Verlauf des Interviews eine weniger offene Frage dar, fokussiert aber speziell auf die Darlegung der Vorstellungen zum Kompetenzbereich Bewertung.

Des Weiteren werden Lehrerinnen und Lehrer im Interview gebeten, Unterrichtsbeispiele, in denen sie Bewertungskompetenz bei Schülerinnen und Schülern gefördert haben, zu erläutern. Darüber hinaus beziehen sich Interviewfragen auf die durch die Lehrerinnen und Lehrer eingeschätzte Fachzugehörigkeit von Bewertungskompetenz und auf die Fähigkeiten, die Schülerinnen und Schüler brauchen, um Bewertungen vornehmen zu können. Insgesamt werden 14 Fragen im Interview zum Thema Bewertungskompetenz gestellt.

Mit Menthe und Parchmann (2006: 80) gehen wir davon aus, dass ein ökologischer Bezugsrahmen wie „Klimawandel“ besonders geeignet ist, die Notwendigkeit des Urteilens zu evozieren. Die Untersuchung wurde u. a. mit Lehrkräften durchgeführt, die an einem Unterrichtsentwicklungsprojekt⁵ teilgenommen haben. Der Interviewleitfaden

4 Die erste Interviewfrage ist hier in der Formulierung für die Naturwissenschaftslehrkräfte aufgeführt. Für die Politiklehrkräfte wurde die Frage sprachlich an die Formulierungen (Anforderungen an nationale Bildungsstandards, Politische Urteilsfähigkeit) angepasst.

5 „Der Klimawandel vor Gericht – Ein Projekt zur Förderung ökologischer Urteilsfähigkeit bei Jugendlichen“ (vgl. Eilks et al. 2011b; c)

besteht aus zwei Teilen, wobei nur der zweite Teil der Datenerhebung für die hier vorliegende Studie dient. Der erste Teil des Interviews dient einer weiteren Studie⁶ der AG Biologiedidaktik der Universität Oldenburg zum „pedagogical content knowledge“ der Lehrpersonen bezogen auf Unterricht über Klimawandel und Treibhauseffekt. Entsprechend wurde der gesamte Interviewleitfaden entlang dieser beiden Forschungsarbeiten strukturiert. Die Interviewführung wechselte ungefähr in der Interviewmitte. Der hier relevante zweite Teil des Interviews wurde mit einer Ausnahme von der Erst-Autorin geführt.

Die Gesamtlänge der Interviews betrug 30 bis 90 Minuten. Die Interviews wurden audiographiert, anschließend vollständig transkribiert (624 Transkriptseiten) und mit Hilfe der Software Maxqda analysiert. Nach der Transkription wurden die Texte in zweifachem Durchgang von den Interviewerinnen überarbeitet. Da die Transkription nicht von den Forscherinnen selbst durchgeführt wurde, ist die Überarbeitung wichtiger Bestandteil der Qualitätssicherung. Über das Nachhören der Interviews durch die Forscherinnen konnten noch 80% der im Transkript als unverständlich gekennzeichneten Stellen ergänzt werden. Dies lässt sich dadurch erklären, dass die Forscherinnen die Interviews selbst durchgeführt haben und über die Zusammenarbeit in Lehrerarbeitsgruppen die Personen bereits kennengelernt hatten.

4. Datenerhebung und Sample

Das Sample setzt sich aus Lehrpersonen der Unterrichtsfächer Biologie, Chemie, Physik und Politik zusammen. Die Politiklehrergruppe dient hier als Referenzgruppe. Da politische Urteilsfähigkeit seit langem fester curricularer Bestandteil ist, kann es als eine plausible Vorannahme gelten, dass die Einstellungen und Vorstellungen von Politiklehrkräften im Vergleich mit Lehrpersonen der Naturwissenschaften differenziert und elaboriert sind und sich in ihren im Interview artikulierten Einstellungen und Vorstellungen niederschlagen. Die Zusammensetzung des Samples folgte der Erwartung, Heterogenität und Varianz durch Berücksichtigung einer Vergleichsgruppe zu maximieren (Lamnek 2010). Ein theoretisches Sampling im Sinne eines während der Datenanalyse bis zur theoretischen Sättigung erweiterten Samples konnte aus forschungspragmatischen Gründen nicht erfolgen. Das Sample stand also zu Beginn der Untersuchung weitgehend fest.

Die Lehrpersonen aller vier Unterrichtsfächer unterschieden sich weiterhin durch die Teilnahme oder Nicht-Teilnahme an einem Projekt namens „Der Klimawandel vor Gericht“. Im Rahmen des Projekts wurden Lehr-Lern-Szenarien und -Materialien zur Förderung von Bewertungskompetenz im jeweiligen Unterrichtsfach entwickelt (Eilks et al. 2011b; c). Die am Projekt teilnehmenden Lehrpersonen arbeiteten zum Untersuchungszeitpunkt ca. ein Jahr lang in Arbeitsgruppen aus Lehrkräften und Fachdidaktikerinnen und -didaktikern. Die Arbeitsgruppen arbeiteten während etwa monatlich stattfindender Sitzungen gemäß eines symbiotischen Entwicklungsansatzes (vgl. Grä-

6 Vgl. Oelgeklaus, voraussichtlich 2012

sel/Parchmann 2004). In iterativen Prozessen wurde Material zur Förderung von Bewertungskompetenz im Kontext Klimawandel und Treibhauseffekt entwickelt, erprobt und überarbeitet. Die in den Arbeitsgruppen erarbeiteten Unterrichtsmaterialien integrieren neben der Erarbeitung fachlicher Grundlagen für die Förderung und Schulung von Bewertungskompetenz Plan- bzw. Rollenspiele (Höttecke et al. 2009a; Höttecke et al. 2009b; Höttecke/Mrochen 2010; Eilks et al. 2011a). Mit der Unterscheidung des Samples nach teilnehmenden und nicht-teilnehmenden Lehrpersonen ging die Erwartung einher, dass die teilnehmenden Lehrpersonen sich im Hinblick auf Einstellungen und Vorstellungen von anderen Lehrpersonen unterscheiden und sich Fortbildungseffekte niederschlagen würden. Letztlich sollte die Zusammensetzung die Heterogenität und Varianz im Sample abermals steigern. Die Verteilung geht aus Tab. 1 hervor.

Tabelle 1: Probandenverteilung

Nges=37	Biologie	Chemie	Physik	Politik
Teilnahme am Entwicklungsprojekt	5	5	5	5
keine Teilnahme am Entwicklungsprojekt	5	5	5	2

Die Lehrkräfte, die sich für eine Teilnahme am Projekt entschieden hatten, differieren über die Fächer hinweg in ihren personalen und soziokulturellen Merkmalen (Schulform, Unterrichtsfächer, Geschlecht, Alter, Berufserfahrung). Die Lehrkräfte ohne Projektteilnahme wurden unter Berücksichtigung dieser personalen und soziokulturellen Merkmale so zusammengesetzt, dass die Gruppen der Lehrkräfte über die verschiedenen Fächer hinweg möglichst ähnlich strukturiert waren.

Die Variabilität in der Länge der Interviews (30 bis 90 Minuten) lässt sich darüber erklären, dass im Interview von Lehrkräften aus der Projektzusammenarbeit deutlich auf Erfahrungen und Ereignisse im Zusammenhang mit der Projektarbeit eingegangen wurde, auf die Lehrkräfte außerhalb des Projektzusammenhangs nicht zugreifen konnten.

Die Interviews fanden im Milieu der Befragten zumeist in den jeweiligen Schulen am Rande des Unterrichts statt. So konnte eine authentische, wenn auch nicht immer ruhige, Atmosphäre hergestellt (Lamnek 2010) und die Bereitschaft der Lehrpersonen zur Teilnahme am Interview erhöht werden.

5. Methodisches Vorgehen

Die Auswertung der Interviews erfolgte nach der inhaltsanalytischen Methode des thematischen Kodierens nach Hopf (vgl. Kuckartz 2007). Hopf beschreibt folgende vier Auswertungsschritte des Verfahrens: 1.) Entwickeln der Auswertungskategorien, 2.) Kodieren des Materials, 3.) Erstellen von Fallübersichten, 4.) Vertiefende Analyse von ausgewählten Fällen (vgl. ebd.: 5).

Die Auswertungskategorien werden in einem induktiv-deduktiven Verfahren erarbeitet. Bei der Entwicklung des Interviewleitfadens werden deduktiv die theoriegeleiteten Hauptkategorien festgelegt und anschließend am Datenmaterial präzisiert. Ausgehend von der Hauptforschungsfrage werden Kategorien zu den Bereichen *Einstellungen*, *Vorstellungen* und *Umsetzung* entwickelt. Dabei bezeichnen *Einstellungen* eine affektive Komponente, *Vorstellungen* eine kognitive Komponente bzw. deklaratives Wissen über Bewertungs-Probleme, -Prozesse und -Kompetenz. Mit der Kategorie *Umsetzung* werden von den Lehrpersonen artikulierte Unterrichtserfahrungen und -beispiele zur Förderung von Bewertungskompetenz kodiert, die mittelbar weitere Rückschlüsse auf Einstellungen und Vorstellungen zulassen. Die Ausdifferenzierung des Kategoriensystems in Unterkategorien erfolgt am Datenmaterial selbst und gilt daher als induktiv. Eine scharfe Trennung zwischen deduktivem und induktivem Vorgehen ist allerdings nicht möglich.

Das Kategoriensystem wurde durch die erste Analyse von vier Interviews (je Unterrichtsfach ein Interview) entwickelt. In einem verschränkten Prozess der Ausdifferenzierung und Reduktion wurde das Kategoriensystem mehrfach überarbeitet. Nach Analyse von ca. 25% des kodierten Datenmaterials hatte sich ein Kategoriensystem ergeben, das mit mehr als 250 Kategorien sowohl in der Handhabbarkeit als auch in der Validierung an seine Grenzen stieß. Zusammenfassende Analyseschritte und die Akzentuierung zentraler Dimensionen führten schließlich zu einem Kategoriensystem aus 67 Kategorien. Dieses wurde letztlich für die Auswertung des gesamten Datenmaterials verwendet. Für die Darstellung der Ergebnisse werden für einzelne Unterkategorien insbesondere jene unter der Hauptkategorie *Vorstellungen* qualitative Ausprägungen identifiziert. Sie werden als Ergebnisse des Rekonstruktionsprozesses im Ergebnisteil dieses Artikels dargelegt.

Für alle Forschung gilt es die Gütekriterien einzuhalten und die eigene Forschung an diesen auszurichten. Dabei gelten als Gütekriterien Reliabilität, Validität und Objektivität (vgl. Bortz/Döring 2006; Lamnek 2010). Die Reliabilität wurde mit einem Interrater-Verfahren abgesichert. Dabei wurden insgesamt sieben von 37 Interviews (ca. 19%) unabhängig und doppelt kodiert. Die Erstautorin hat alle sieben Interviews kodiert. Drei Interviews wurden von der zweiten Interviewerin kodiert und vier weitere Interviews wurden von einer weiteren Person kodiert. Anschließend wurden die Kodierungen jeweils mit denen der Erstautorin verglichen, Übereinstimmungen wurden gezählt. Bortz und Döring (2006: 276) schlagen einen Übereinstimmungskoeffizienten vor, der schließlich mit Hilfe von Cohens Kappa (vgl. ebd.: 276f.) um die geschätzte zufällige Übereinstimmung korrigiert wird. Die Rater-Übereinstimmung berechnet sich als Quotient aus der Summe der übereinstimmenden Kodierungen und der Gesamtzahl kodierter Objekte. Für die Berechnung von Cohens Kappa wird die geschätzte zufällige Übereinstimmung errechnet und mit berücksichtigt.

Im Vergleich mit der Kodierung der Erstautorin ergaben sich die folgenden, jeweils zufriedenstellenden, Kappa Werte:

- a) Bei 289 Gesamtobjekten in 3 Interviews haben Kodiererin 1 und 2 204 Übereinstimmungen gehabt. Daraus errechnet sich ein Mittelwert über die 3 Interviews von $\kappa_1=0,70$.
- b) Bei 288 Gesamtobjekten in 4 Interviews mit der zweiten Co-Kodiererin waren 246 Übereinstimmungen zu verzeichnen. Der sich hier ergebende Mittelwert über die 4 Interviews ergibt ein Kappa von $\kappa_2=0,84$.

Bei κ -Werten ab 0,60 spricht man von einer guten Übereinstimmung (vgl. ebd., S. 277). Damit sind beide Werte als gute Übereinstimmungen zu bezeichnen.

Um die Validität der Ergebnisse abzusichern wurden Teile der Interviews (ca. 20%) von der Erstautorin und zwei weiteren Kodierern ausgewertet bzw. argumentativ validiert (Forschertriangulation, vgl. Lamnek 2010). Eine kommunikative Validierung der Interpretationen mit den Probanden wurde im Vorhinein ausgeschlossen, da auch sozial unerwünschte oder widersprüchliche Deutungen, die im Rahmen kommunikativer Validierung von den Probanden abgelehnt oder verneint werden könnten, berücksichtigt werden sollen. Die argumentative Validierung wurde in zwei Schritten sichergestellt:

Argumentative Validierung I: Es wurden bereits von der Erstautorin kodierte Teile des Datenmaterials diskutiert. Die Kodierungen wurden diskursiv auf ihre Angemessenheit hin untersucht. Dies wurde mit einer unabhängigen Person durchgeführt, die bereits im Interviewverfahren an der Sicherung der Ergebnisse, nicht aber an der Erhebung der Daten, beteiligt war. Diese Person wies eine fundierte Kenntnis in der Kodierung von Interviews auf. Für diesen Validierungsschritt wurden etwa 20% der Interviews ausgewählt.

Argumentative Validierung II: In zwei verschiedenen Forschergruppen wurden Auszüge aus Interviews (ca. 10% des Gesamtmaterials aus sechs verschiedenen Interviews) in Expertengruppen bearbeitet. Die Gruppen bestanden in einem Fall aus Forscherkolleginnen und -kollegen des Instituts und in einem anderen Fall aus einem interdisziplinären Arbeitskreis. Es waren jeweils sechs bis acht Experten/innen beteiligt. Die Experten/innen haben das ausgewählte Datenmaterial zunächst unabhängig von einander kodiert. Anschließend wurden die Kodierungen verglichen und in ihrer Angemessenheit und ihren inhaltlichen Dimensionen am konkreten Datenmaterial diskutiert bis Einigung über ein gemeinsames Verständnis der Interviews und der zugehörigen Kodierung erzielt werden konnte.

Lamnek (2010: 158) hebt hervor, dass der Objektivitäts-Begriff der qualitativen Sozialforschung soziale Realität interpretativ erschleie und Intersubjektivität durch Interpretation in Gruppen diskursiv vorgenommen werden könne. Der diskursive Charakter beider Validierungsschritte sichert damit zugleich die intersubjektive Geltung und damit die Objektivität der qualitativen Analyse.

Nach der Entwicklung eines Kategoriensystems und der Validierung einer Auswahl von Kodierungen können für die Beantwortung der Forschungsfrage, welche Vorstellungen Lehrerinnen und Lehrer zum Kompetenzbereich Bewertung artikulieren, erste Ergebnisse formuliert werden. In der vorliegenden Studie wurden dazu Vorstellungen extrahiert und eine Typisierung der interviewten Lehrkräfte entwickelt, die Aufschluss

über Einstellungen und Vorstellungen von Lehrpersonen bringen sollte. Die Darstellung einer Typologie bietet die Möglichkeit, über die Individuen hinaus Zusammenhänge und Merkmale im Sample zu identifizieren.

Der Typenbegriff ist seit jeher eine wichtige Vokabel der empirischen Sozialforschung. Es macht jedoch insbesondere bei der Bildung des Samples einen Unterschied, ob mit Idealtypen, Realtypen oder Prototypen gearbeitet werden soll. Als Grundstruktur der Typologie machen Kelle und Kluge (2010) und Kluge (1999) theoretisch bedeutsame Merkmalskombinationen und jeweilige Merkmalsausprägungen aus. Eine empirisch begründete Typenbildung (Kluge 2000: 4) basiert danach auf einer Verschränkung theoretischer Vorannahmen und empirischer Analysen. Dieses Vorgehen liegt der vorliegenden Untersuchung zugrunde.

Für die Entwicklung der Typisierung werden Fallübersichten erstellt (vgl. Hopf 1982) und Einzelfälle mit Hilfe von Personenvignetten in ihrer Charakteristik analysiert (vgl. Hopf 1993). Die Fälle werden anhand relevanter Merkmale bzw. Kategorien dimensionalisiert. Hierbei werden die Kombinationsmöglichkeiten der Merkmale dargestellt und theoretisch und am empirischen Material auf ihr mögliches Auftreten hin untersucht. Bei mehrdimensionalen Typisierungen wird der Prozess mehrfach wiederholt und weitere Kategorien über die Dimensionalisierungen hinweg miteinander kombiniert.

„Eine *Typologie* ist das Ergebnis eines Gruppierungsprozesses, bei dem ein Objektbereich anhand eines oder mehrerer Merkmale in Gruppen bzw. Typen eingeteilt wird [...], so dass sich die Elemente innerhalb eines Typus möglichst ähnlich sind (*interne Homogenität* auf der ‚Ebene des Typus‘) und sich die Typen voneinander möglichst stark unterscheiden (*externe Heterogenität* auf der ‚Ebene der Typologie‘) [...]“ (Kelle/Kluge 2010: 85; Hervorhebungen im Original).

Die Erarbeitung von Typen geschieht schließlich über Fallvergleich und Fallkontrastierung. Dabei sollen die Ähnlichkeiten und Unterschiede möglichst sowohl auf der Ebene der Einzelfälle als auch für eine große Anzahl von Fällen herausgearbeitet werden (vgl. Kelle/Kluge 2010; Gerhardt 1986). Damit stellt die komparative Analyse bzw. die vergleichende Kontrastierung den wichtigsten Schritt der Typenbildung dar.

Kelle und Kluge beschreiben vier Stufen, die zur Typenbildung führen. Dabei werden die ersten drei in einem rekursiven Verfahren solange wiederholt, bis alle gewünschten Merkmale miteinander kombiniert, die Typen sich je untereinander möglichst ähnlich, aber voneinander möglichst verschieden sind. Folgende Stufen umfasst das Modell (vgl. ebd.: 92f):

1. Erarbeitung relevanter Vergleichsdimensionen,
2. Gruppierung der Fälle und Analyse empirischer Regelmäßigkeiten,
3. Analyse inhaltlicher Sinnzusammenhänge und Typenbildung,
4. Charakterisierung der gebildeten Typen.

Die Verteilung und Kombination von Kategorien und Merkmalen macht es möglich, eine Typisierung zu entwickeln, die die Lehrkräfte nicht mehr auf Fall- sondern auf

Typenebene beschreibt. Es handelt sich bei der Typisierung also um eine Kombinationsreihe von Merkmalen, die in der vorliegenden Studie anhand der Kodier-Kategorien vorgenommen wird. Mit Kuckartz (vgl. Kelle/Kluge 2010: 105) werden sogenannte „idealtypische Konstrukte“ erstellt. Die wichtigsten Charakterzüge und Besonderheiten des jeweiligen Typus werden in einem Fokus zusammengefasst und entlang jeweils zentraler Merkmale charakterisiert.

Die Kategorien, die die Grundlage der Typenbildung darstellen, werden als *Vorstellungen*, *Einstellungen*, *Umsetzung*, *Rolle des Sachwissens*, *Fachzugehörigkeit* und *Voraussetzungen* bezeichnet. Dabei beinhalten die benannten Kategorien die im Folgenden inhaltlich ausgeführten Aspekte. Die identifizierten Vorstellungen von Bewertungskompetenz gehen mit unterschiedlichen Einstellungen der Lehrkräfte einher und können miteinander kombiniert werden. Auch unterscheiden sich die Lehrkräfte in ihrer Einschätzung, in welchem Verhältnis die Vermittlung von Fachwissen zu anderen Aspekten von Bewertungskompetenz steht. Ein weiteres Merkmal stellt dar, ob und in welchem Maße Lehrerinnen und Lehrer über Unterrichtsbeispiele zur Förderung von Bewertungskompetenz verfügen und wie sie diese darlegen. Zudem differieren die Einschätzungen, ob Bewertungskompetenz als Teil ihrer je eigenen Fachkultur zu verstehen ist oder eher in die Zuständigkeit anderer Unterrichtsfächer fällt. Sie haben weiterhin ein unterschiedliches Bild davon, welche Voraussetzungen Schülerinnen und Schüler für die Kompetenzentwicklung im Bereich Bewerten mitbringen müssen.

Die Generalisierbarkeit der entwickelten Typologie ergibt sich nicht aus der Angabe einer quantitativen Verteilung innerhalb einer repräsentativen Stichprobe, wie dies mit quantitativen Daten ermöglicht werden könnte. Insofern ist der Anspruch auf Verallgemeinerbarkeit in der qualitativen Forschung eingeschränkt gegeben bzw. folgt einem anderen Verständnis von Generalisierbarkeit (vgl. Lamnek 2010). Der Anspruch auf Verallgemeinerbarkeit ergibt sich durch die transparente und wohlbegründete Zusammenstellung des Samples einerseits, durch methodische Kontrolle andererseits und die Plausibilität der Ergebnisse im Lichte bekannten theoretischen und empirischen Wissens. Das Ergebnis der vorliegenden Studie sind daher Merkmale und Merkmalskombinationen, die für das Sample als typisch identifiziert werden konnten, und nicht ihre Häufigkeitsverteilung.

6. Ergebnisse

Im folgenden Abschnitt werden Ergebnisse der Studie in exemplarischer Beschränkung dargestellt. Der Fokus liegt zunächst auf der Darstellung von Vorstellungen der Probanden. Diese werden entlang der im Auswertungsverfahren entwickelten Kategorien und Unterkategorien beschrieben. Im zweiten Teil wird die entwickelte Typologie an ausgewählten Beispielen dargestellt.

6.1 Vorstellungen

Im Zuge des oben beschriebenen iterativen Analyseprozesses konnten vier Vorstellungsklassen identifiziert werden. Wir unterscheiden die Vorstellungsklassen *innerfachliche* Bewerten-Vorstellungen, *Bewerten-ist-Anwenden* Vorstellungen und *Bewerten-ist-Urteilen* Vorstellungen. Weiterhin lassen sich Vorstellungen identifizieren, die sich nicht in das Bedeutungsspektrum der Nationalen Bildungsstandards und auch nicht in die im Theorie-Teil dieses Artikels skizzierten Bewerten-Vorstellung einordnen lassen. Diese Vorstellungen werden als *andere* Vorstellungen bezeichnet.

6.1.1 Innerfachliche Bewerten-Vorstellungen

Als *innerfachliche* Bewerten-Vorstellungen werden Aussagen erfasst, in denen Lehrpersonen die Bewertung von Graphen, Tabellen, Messergebnissen oder Stoffeigenschaften benennen. Darüber hinaus wird unter diese Klasse subsumiert, die Lösungen von Aufgabenstellungen auf ihre Sinnhaftigkeit bezogen auf die Fragestellung einzuschätzen. Aber auch das Heranziehen und Einschätzen unterschiedlicher Quellen für die Lösung einer Aufgabe bzw. die Recherche in Aufgabenteilen wird unter eine *innerfachliche* Bewerten-Vorstellung gefasst. Innerfachlich meint in diesem Zusammenhang, dass die Dimensionen einer Bewertung sich allein aus fachlichen Maßstäben ergeben, Kriterien einer Bewertung als dem Fach inhärent verstanden werden und die Fragestellungen sich auf einen begrenzten inhaltlich-fachlichen Aspekt beziehen. Diese Bewerten-Vorstellung ist insbesondere in den Bildungsstandards für das Fach Physik angelegt, wird und wurde dahingehend aber auch bereits kritisiert, sofern sich Bewerten auf innerfachliche Aspekte beschränkt (vgl. Schecker/Höttecke 2007).

Einige Aspekte dieser Vorstellung, wie z. B. das Heranziehen und Einschätzen unterschiedlicher Quellen, finden sich in den Bildungsstandards jedoch nicht im Kompetenzbereich Bewertung, sondern sind dort als Standard für den Bereich Kommunikation (z. B. K1-, K2-Chemie, K4-Biologie, K3-Physik) oder den Bereich Erkenntnisgewinnung (E2-Physik) formuliert. Darüber hinaus wird auch die Beurteilung der Gültigkeit von Ergebnissen als Standard für den Kompetenzbereich Erkenntnisgewinnung formuliert (E10-Physik, E6-Chemie).

Als eine Vorstellung von Bewertungskompetenz sind *innerfachliche* Bewerten-Vorstellungen unter den Lehrkräften durchaus verbreitet. Es lassen sich bei 70% der interviewten Lehrkräfte – wenn auch z. T. in Kombination mit weiteren Vorstellungen – *innerfachliche* Bewerten-Vorstellungen identifizieren. Diese bieten insbesondere bei Physik- und Chemie-Lehrkräften oftmals einen ersten Assoziations- und Deutungsrahmen.

Ein von uns interviewter Physik-Lehrer, Herr Rade⁷, erläutert im folgenden Zitat die fachspezifische Selbstverständlichkeit dieser Bewerten-Vorstellung.

7 Alle Namensangaben von Probanden sind aus Datenschutzgründen geändert.

„Bewerten ist ja auch etwas eigentlich sehr, sehr Physikalisches, also, das hat ja auch, äh, in seiner REINsten Form was mit Messen zu tun. Also, Bewerten heißt ja auch Vergleichen. Und das ist auch was, was also auch in meinem Physikunterricht, äh, vorkommt, dass sie erstmal überhaupt, ähm, begreifen, was Messen eigentlich ist. Und dass man also, Messen nichts anderes ist als irgendetwas vergleichen. Dass wir ja auch durch das Vergleichen auch nur unsere Welt erkennen können, [...] Und, und, das geht dann ja auch in Richtung ... Bewerten. Also mehr so vom wirklich rein naturwissenschaftlichen, äh, Standpunkt.“ (Herr Rade, Phy, Abs.198-199)⁸

Herr Rade beschreibt hier Bewerten als etwas „sehr, sehr Physikalisches“, weil es „was mit Messen zu tun“ hat. Messen ist für ihn essentieller Bestandteil von Physik. Seine Vorstellung von Bewerten betrifft den physikalischen Messprozess („reinste Form“). Messprozesse werden in der Fachliteratur als Vergleich einer Messgröße mit Maßeinheiten beschrieben. Dieses Verständnis vom Messprozess reproduziert Herr Rade hier. Ein tiefes Verständnis des physikalischen Messprozesses selbst ist Herrn Rade wichtig. Entsprechend wählt er eine pathetische Semantik („REIN“, „Welt erkennen“), um die Dignität und Bedeutung dieses Ziels hervorzuheben. Bewerten ist sehr positiv konnotiert, enthält es doch den Schlüssel zur Fähigkeit, die physikalische Welt zu erkennen. Bewerten besteht darin, die „rein naturwissenschaftliche“ Erkenntnispraxis zu durchschauen und bezieht sich metarefektiv auf die Erkenntnispraxis der Physik. Die Grenzen der Bedeutsamkeit des Faches müssen hier aber von Herrn Rade nicht durch einen überfachlichen Bewerten-Begriff erweitert werden.

Diese *innerfachliche* Bewerten-Vorstellung findet sich auch in den Nationalen Bildungsstandards (z. B. B3- und B6-Chemie, B3-Physik) und ist als ein Aspekt von Bewertungskompetenz in vielen Äußerungen der Lehrkräfte identifizierbar. Im Sinne der Bildungsstandards gilt die Vorstellung als adäquat. Im Sinne des im theoretischen Teil dieses Artikels dargelegten Verständnisses von Bewertung bzw. Urteilen greift sie jedoch zumindest als isolierte Bewerten-Vorstellung deutlich zu kurz.

6.1.2 Bewerten-ist-Anwenden Vorstellung

In der Vorstellung *Bewerten-ist-Anwenden* verbinden Lehrpersonen mit dem Bewertungsbegriff vorrangig das Erreichen einer bestimmten Handlungsbereitschaft bei Jugendlichen. Dafür werden lebensweltliche Bezüge hergestellt und anhand von Anwendungsbeispielen erläutert. In Hinsicht auf das vermittelte Fachwissen äußern die Lehrkräfte zudem die Vorstellung, dass das Anwenden des Fachwissens in Transferaufgaben Teil von Bewertungskompetenz sei. Bewertungskompetenz wird in dieser Vorstellung zu einem Konglomerat aus Ideen für unterrichtliche Strukturierungen, die den Unterricht interessant und anschaulich machen, die Relevanz fachlichen Lernens an Anwendungsbeispielen verdeutlichen und darin gleichsam als sinngebende Struktur des Unterrichts fungieren sollen.

Insbesondere lebensweltliche Bezüge können als Herstellung der Beziehung zwischen Fach und Gesellschaft verstanden werden und finden sich z. B. in den Bildungs-

⁸ Das Zitat ist ungeglättet verwendet und analysiert worden. Betonungen sind durch Versalien gekennzeichnet.

standards für das Fach Chemie, dort aber nicht im Kompetenzbereich Bewertung, sondern in den Bereichen Erkenntnisgewinnung und Kommunikation (E8- und K5-Chemie, K7-Biologie).

Die Vorstellung *Bewerten-ist-Anwenden* konnte bei den Lehrkräften in unterschiedlicher Komplexität identifiziert werden. Dies bezieht sich einerseits auf die Komplexität, mit der Lehrkräfte Bewertung thematisieren, aber auch auf die Themenauswahl generell. So können von Lehrkräften z. B. selbst innerhalb komplexer Themen nur eine eingeschränkte Anzahl an Betrachtungsebenen für das Unterrichtsgeschehen thematisiert werden. Dies kann Unterrichtssituationen, dem Lernstand der Schülerinnen und Schüler oder auch der Wahl bestimmter Lernziele geschuldet sein. Im folgenden Zitat wird deutlich, dass die Nutzung von Energiesparlampen oder Glühlampen eine bewertungsrelevante Frage sein kann, auch wenn die interviewte Physik-Lehrerin Frau Vahse die Fragestellung hier nur in ausgewählten Facetten thematisiert.

„Und, weiß ich nich, neulich, .. ähm, Energie, Thema Energie äh und Abwärme. Da .. fassen die, die, die Glühlampe an und sagen: „Ja, is ja heiß, ne?“ Und die Sp⁺ Energiesparlampe is eben nich heiß. .. Ja, und dann .. geh ich eben noch mal den Schritt und sage: „Ja, und was heißt das jetzt für uns?“ Und das is für mich Bewerten, also so offen, ne? Auf so ner kleinen Schiene, aber das, was sie dann gesehen haben im Physikunterricht und was sie, w' wo ja Physik einfach hinter steckt, dann zu übertragen, ja, okay, jetzt macht's auch wirklich Sinn, äh, ne EnerGIESparlampe zu benutzen. Und ich WEIß jetzt, warum es Sinn macht. .. So, das versuch ich unterzubringen.“ (Frau Vahse, Phy, Abs. 79)

Das Anwendungsbeispiel der Nutzung von Energiespar- oder Glühlampen stellt für Frau Vahse eine sinngebende Struktur („warum es Sinn macht“) fachlichen Lernens dar, hier über die Energieumwandlung verschiedener Leuchtmittel. Dabei denkt sie sich fachliches Lernen und Anwendungsbeispiel als auf verschiedenen Ebenen angesiedelt. Die Verschiedenheit der Ebenen Fachwissen und Bewerten/Anwenden drückt sich semantisch mehrfach aus:

- im Begriff der „Schiene“ (die Metapher impliziert einen Plural)
- einer räumlichen Topologie („wo Physik einfach hinter steckt“, „übertragen“)
- und dem Gegensatzpaar „offen“ (als Attribut von Bewertung) und der impliziten Vorstellung von Eindeutigkeit, die sich im Zitat als Faktizität klarer Schüleräußerungen („Ja, is ja heiß [...] is eben nich heiß“) ausdrückt.

Faktizität des Fachlichen und Offenheit des Bewertens stehen sich hier als geschiedene Bereiche gegenüber, deren Überwindung der unterrichtlichen Inszenierung bedarf. Die Bewertung durch die Schülerinnen und Schüler findet in einem offenen Raum statt, der sich vom Raum der Faktizität fachlichen Lernens abhebt. Zugleich wird die Offenheit der vorgestellten Bewertungs-Situation wieder eingeschränkt, da das Bewertungsergebnis (es macht Sinn, Energiesparlampen zu nutzen) feststeht und implizit Appellcharakter trägt. Sprachlich macht Frau Vahse deutlich, dass sie Bewerten mit der Bedeutung des Anwendens von Fachwissen einerseits als sehr bedeutungsvoll und dennoch zugleich als additiv zum Eigentlichen des Faches, das mit Faktizität konnotiert ist, einschätzt.

Die vorgestellte Offenheit der unterrichtlich inszenierten Bewertungs-Situation widerspricht dem Appellcharakter eines vorweggenommenen Bewertungs-Ergebnisses. Diese Widersprüchlichkeit erlebt Frau Vahse an dieser Stelle jedoch nicht offenkundig.

Die *Bewerten-ist-Anwenden* Vorstellung korrespondiert mit einigen Vorgaben der Bildungsstandards (Standards B3-Biologie, B1- und B3-Chemie, B3- und B4-Physik), aber auch mit kontextorientierten Konzepten für den naturwissenschaftlichen Unterricht, die lebensweltliche Bezüge in authentischen Kontexten für fachliches Lernen nutzen (z. B. Müller 2006).

6.1.3 Bewerten-ist-Urteilen Vorstellung

Die Vorstellung *Bewerten-ist-Urteilen* umfasst verschiedene Urteilsdimensionen und ist bei Lehrkräften in verschiedenen Ausprägungsgraden identifizierbar. Zentral an dieser Vorstellung ist die Deutung des Urteilens als Prozess des Abwägens und Argumentierens. Wichtig ist, dass es sich um den Prozess des Urteilens handelt, nicht um ein (bestimmtes) und von den Lehrkräften vorweggenommenes Ergebnis dieses Prozesses. Insbesondere das persönliche Sich-eine-Meinung-Bilden wird als Teil dieser Vorstellung gefasst. Im Rahmen dieser Vorstellung wird der Prozesscharakter des Urteilens betont und die Komplexität des Bewertens und seiner Fragestellungen werden ins unterrichtliche Handeln integriert.

Die besondere Bedeutung des Prozesshaften des Urteilens in dieser Vorstellung wird durch das Argumentieren und Abwägen durch die Schülerinnen und Schüler formuliert. Argumentieren bzw. Diskutieren von Ergebnissen gilt in den Bildungsstandards aller Fächer jedoch hauptsächlich als Teilkompetenz des Kompetenzbereichs Kommunikation (z. B. K1-, K6-Biologie, K8-Chemie, K7-Physik).

Bewerten-ist-Urteilen umfasst verschiedene Aspekte, die im Kategoriensystem Unterkategorien bilden. Dabei handelt es sich u. a. um

- Problem erkennen
- Folgen abschätzen
- Perspektiven einnehmen
- Wertedimension klären
- Hintergründe kennen und begründet argumentieren
- Handlungsalternativen generieren
- Entscheidungen anderer nachvollziehen

Die Auswahl der hier dargestellten Aspekte des Bewerten-Begriffs korrespondiert mit den empirisch begründeten Schritten nach Hölble/Bayrhuber (2006) zur Förderung von Urteilsfähigkeit und zur Erläuterung der Dimensionen eines Urteilsprozesses. Die befragten Lehrkräfte benennen diese Aspekte zur Erläuterung ihrer Vorstellung von Bewertung und betonen die Prozesshaftigkeit des Bewertens und die Vielfalt der Argumente. So erläutert Frau Salz ihren Bewertungskompetenzbegriff:

„[5sec, Salz atmet tief aus] Das heißt, für mich verschiedene .. ähhmm, sachliche Grundlagen zu kennen, erstmal. Sie ä .. sag'n wir mal abgeprüft, ein Stück, also, ähm, durchaus kritisch betrachtet, angenommen zu haben. Also für sich bewertet zu haben. Ähm, und .. ähm, akute Situationen, in denen, die ihnen dargeboten werden, wie zum Beispiel, ich denk jetzt mal einfach an eine Aufgabe, in einzelne Parameter äh zu ähmm .. nicht zu verteilen, sondern, im Grunde zu klären, wo ist meine meine, ähm eher so 'ne so 'ne emotionale Schiene, die bei mir läuft, wo sind die sachlichen Argumente, die ich anbringen kann? (l: mhmh) Und was spricht eigentlich äh wofür und wogegen? Die Dinge miteinander abzugleichen und für sich im Grunde eine Lösung daraus zu entwickeln. Welche sachlichen äh Kriterien sind so, sagen wir mal wichtig, oder äh, dass sie dringend einer ähm Änderung bedürfen, zum Beispiel. Dann kann (l: mhmh) ich auf der Sachebene was tun, oder meinewegen auch Handlungsebene, wenn es 'ne Handlung betrifft. Ähm .. und welche Dinge ähmm pff .. sind eher so meine persönliche ä .. Einschätzung, oder meine, meine äh, so was wie Empathie, oder so was? (l: mhmh) Ich kann das jetzt nicht so richtig fassen.“ (Frau Salz, Bio, Abs. 279-281)

Die Integration sachlicher und emotionaler Anteile an einer Fragestellung bildet für Frau Salz den Kern des Bewertungsprozesses. Sie erläutert den Bewertungsprozess als Offenlegen der verschiedenen Anteile, die zu einer Bewertung geführt haben. Dabei handelt es sich gewissermaßen um ein mehrstufiges Verfahren. Zu Beginn gilt es die „sachlichen Grundlagen zu kennen, [...] kritisch betrachtet, angenommen zu haben [und] für sich bewertet“ zu haben. Erst im Anschluss folgt eine Analyse der Problemstellung, in der emotionale Beteiligung und sachliche Argumente mit einander in Beziehung gesetzt und abgewogen werden. Offen bleibt an dieser Stelle, was Frau Salz mit den Änderungen der sachlichen Kriterien meint – es könnte sich gleichermaßen um die weiterführende Recherche von Sachargumenten handeln, die erneute Prüfung von Quellen auf ihren Ursprung und damit ggf. verbundene Interessensvertretungen hin oder auch die Änderung des Grundproblems und der damit verbundenen Bezugswissenschaften. Im weiteren Verlauf des Interviews wird u. a. diese Deutung nahegelegt, wenn Frau Salz als bekennende Atomenergie-Gegnerin deutlich macht, dass grundsätzlich etwas am Energieversorgungssystem geändert werden müsse, wenn das Atomenergieproblem sich nicht anders lösen lasse⁹.

Die Unsicherheit, mit der Frau Salz hier ihre Äußerung abschließt („Ich kann das jetzt nicht so richtig fassen.“), macht auch deutlich, dass es sich um keine feste definitonische Antwort handelt. Vielmehr artikuliert sie spontane Assoziationen, die kein routiniertes Unterrichtshandeln vermuten lassen. Das Prozesshafte des Urteilens und die Integration von Argumenten unterschiedlicher Qualität stellen den Kern des Bewertungsprozesses für Frau Salz dar.

Die *Bewerten-ist-Urteilen* Vorstellung korrespondiert mit einigen Teilkompetenzen der Bildungsstandards, z. B. B5-Chemie, B2-, B3-, B4-, B5-Biologie oder B2-Physik, in denen abwägendes und argumentierendes Urteilen anklingt.

9 Es ist erwähnenswert, dass die Interviews vor der Fukushima-Katastrophe in Japan im März 2011 und dem anschließenden Atomausstiegsbeschluss der Bundesregierung geführt worden sind.

6.1.4 Andere Vorstellungen

Als *andere* Vorstellungen werden jene Äußerungen der Lehrpersonen verstanden, in denen Sinngebungen artikuliert werden, die keine anschlussfähigen Deutungen an die Bildungsstandards zulassen. Lehrkräfte, deren Vorstellungen als *andere* Vorstellungen klassifiziert wurden, reagieren z. B. generell mit Unverständnis darauf, was mit Bewertungskompetenz für den naturwissenschaftlichen Unterricht überhaupt gemeint sein könne. Oder es wird deutlich, dass der Begriff Bewertungskompetenz für die betreffenden Lehrkräfte so stark z. B. mit Notengebung und Leistungsbewertung besetzt ist, dass auf weitere Bedeutungszusammenhänge nicht eingegangen wird oder werden kann. Im folgenden Zitat von Herrn Neißé kommt diese Dimension zum Tragen:

„Die Bewertung ist eins, glaub ich, wir müssen das ja alle tun. Ähm, .. versuchen, möglichst objektiv zu sein. Ich glaube, das schaffen wir NIE-mals. .. Wir sind es gehalten! Und ich glaube, jeder Lehrer auff äh, Bewertung angesprochen, wird sich immer unwohl fühlen. Hört man mir wahrscheinlich auch an. Ähm, .. es ist immer letztlich ähm, 'ne Frage, ähm, was hab ich vorgegeben im Unterricht und was wünsch ich soll rauskommen? Was wünschte ich soll rauskommen? Das ist aber ja schon wieder sehr subjektiv. Weil die ja andre Dinge und mit andern Voraussetzungen reingehen als ich. Ich kann´s ja an meinem Finger auf irgendetwas zeigen, (l:mhmm) das heißt noch lange nicht, dass sie das auch wohl sehen. (l:mhmm) Mmh, von daher ist es also schon, is' Bewertung, n' ganz und gar schwierige Geschichte. Besonders bei Gruppenarbeiten, ähm, wer hat mehr getan als der andere.“ (Herr Neißé, Phy, Abs. 150-152)

Herr Neißé formuliert hier sein Unbehagen mit der Leistungsbewertung von Schülerinnen und Schülern. Dabei beschreibt er die Schwierigkeit als Problem der Subjektivität bzw. mangelnder Objektivität. Die unterschiedliche Sichtweise und der verschiedene Verstehenshorizont von Schülerinnen und Schülern führen zu Dissonanzen darüber, was zu leisten bzw. zu wissen ist. Im weiteren Verlauf der Interviewpassage beschreibt Herr Neißé schließlich, wie er darüber in Aushandlungsprozesse mit Schülerinnen und Schülern tritt.

Wird die Ebene der Aushandlung von Kriterien, in diesem Fall für die Leistungsbewertung, betrachtet, kann auch hier ein Bewerten-Verständnis identifiziert werden, in dem Sinne, dass dieser Aushandlungsprozess sich auch auf andere Bereiche übertragen ließe. Gemeint ist es von Herrn Neißé an dieser Stelle so aber vermutlich nicht, was insbesondere mit dem Abschluss der Passage, dass Bewerten schwierig ist, „[b]esonders bei Gruppenarbeiten, ähm, wer hat mehr getan als der andere“ noch einmal deutlich zum Ausdruck kommt. In dieser Passage handelt es sich um den *anderen* Bedeutungshorizont, in diesem Fall den der Notengebung. Dieser Bedeutungshorizont ist von den Autorinnen und Autoren der Bildungsstandards nicht gemeint.

Zur Beantwortung der Forschungsfrage, welche Vorstellungen Lehrerinnen und Lehrer zum Kompetenzbereich Bewertung der Nationalen Bildungsstandards äußern, werden vier verschiedene Vorstellungen identifiziert. Dabei korrespondieren zwei der dargelegten Vorstellungen mit Bewertungskompetenzbegriffen, die so oder so ähnlich bereits an anderen Stellen (vgl. Schecker/Höttecke 2007 für den *innerfachlichen* Bewertungsbegriff, Hößle/Bayrhuber 2006 für *Bewerten-ist-Urteilen*) beschrieben

werden. Gemeinsam mit der Vorstellung *Bewerten-ist-Anwenden* befinden sie sich im Bedeutungshorizont der Bildungsstandards. Die Identifikation der vierten Vorstellung macht darauf aufmerksam, dass die Bezeichnung des Kompetenzbereichs Bewertung für die naturwissenschaftlichen Fächer insofern eine Schwierigkeit darstellt, als dass die Konfundierung des Begriffs mit Notengebung die Kommunikation über Bewertungskompetenz deutlich erschwert. Schließlich stellt das Nicht-Verstehen von Bewertungskompetenz im Sinne eines Urteilensprozesses oder die isolierte Vorstellung von Bewertung als Leistungsbewertung ein weitgehendes Missverständnis dar, das keinesfalls erwarten lässt, dass Bewertungskompetenz im Sinne argumentativ abgesicherten und abwägenden Urteilens von den Lehrpersonen gezielt gefördert wird.

6.2 Typen

Die entwickelte Typologie umfasst vier verschiedene Typen, die jeweils bezüglich ihrer Merkmalskombinationen charakterisiert werden. Alle Typen wurden mit sprechenden Namen bezeichnet, die die wesentlichen Merkmale prägnant bündeln. Aus Platzgründen wird nur der erste Typ in exemplarischer Ausführlichkeit dargestellt, während die anderen drei Typen nur in ihrer Hauptcharakteristik, dem sogenannten Fokus, vorgestellt werden.¹⁰ Die Merkmale selbst gehen aus den Kategorien des Kodierungsprozesses hervor. Alle vier Typen sind jeweils aus Kombinationen aller Merkmale entwickelt worden. Die Merkmale werden durch Kategorien abgebildet, die sich im Verlauf der Datenanalyse als besonders prägnant herausgestellt haben: die *Bewerten-Vorstellungen*, die *Bedeutung des Sachwissens*, die *Umsetzung von Bewertungskompetenz im unterrichtlichen Handeln* und die *Einstellungen*.

6.2.1 Typ 1: Sachwissen ist das A und O

Fokus

Die Hauptaufgabe und Bedeutung des Unterrichts liegt für Lehrerinnen und Lehrer dieses Typs in der Vermittlung von Sach- und Fachwissen. Insbesondere der Naturwissenschaftsunterricht schöpft seine Bedeutung daraus, dass nicht leicht zu erlernendes Fachwissen vermittelt werden muss. Dafür bedarf es Zeit und Anstrengung. Alles, was diesem Ziel dient, kann und wird in den Unterricht integriert. Die Lehrpersonen verstehen ihren Bildungsauftrag darin, die kompetente Vermittlung von Fachwissen sicher zu stellen. Dabei kann es vonnöten sein, Schwerpunkte zu setzen und sich ggf. über bildungsadministrative Vorgaben hinwegzusetzen bzw. diese – weil irrelevant – zu ignorieren.

Dieser Typ ist in den Interviewdaten am häufigsten auszumachen. Er findet sich über alle vier Fächer hinweg in etwa der gleichen Häufigkeit und auch das Alter und die Berufserfahrung der Lehrkräfte differieren über die gesamte Breite. Auffällig ist jedoch, dass die Lehrkräfte diesen Typs fast ausschließlich am Gymnasium unterrichten. Dies

¹⁰ Eine ausführliche Darstellung wird in der Dissertation der Erstautorin vorgestellt.

korrespondiert mit der besonderen Betonung von Leistung und Wissen, das als zentrales Merkmal bei diesem Typ zu finden ist. Lehrkräfte, die am Projekt teilgenommen haben und damit im Bereich Bewertungskompetenz über 1-1,5 Jahre fortgebildet sind, finden sich gleichermaßen in diesem Typ wie Lehrkräfte, die ohne Fortbildung an der Interviewstudie teilgenommen haben.

Bewerten-Vorstellung

Bei diesem Typ verbindet sich die Überzeugung, dass die Hauptaufgabe des (Naturwissenschafts-)Unterrichts die Vermittlung von Fachwissen sei mit *innerfachlichen* Bewerten-Vorstellungen oder jenen, in denen lebensweltliche Bezüge hergestellt oder technische Anwendungen thematisiert werden. Dabei dient die Förderung von Bewertungskompetenz im Unterricht dem Ziel, das Fachwissen besser zu verstehen, tieferen Einblick zu ermöglichen oder eigene fachliche Einschätzungen vornehmen zu können. Bewerten wird als eine Art Vehikel genutzt, um Fachwissen lebendig und verständlich zu gestalten, insbesondere mit der Vorstellung *Bewerten-ist-Anwenden*. Einige Lehrkräfte befürchten, dass die Orientierung auf Bewertungskompetenz im Unterricht vom eigentlichen Geschehen ablenke und sich als hinderlich für das Lernen der Fachsystematik erweisen könne.

Die *innerfachliche* Bewerten-Vorstellung geht häufig mit der Überzeugung einher, dass es dem Fach immanent sei, Fehler- oder Ergebnisabschätzungen vorzunehmen. Inhaltlich prägt sich die Vorstellung darüber hinaus so aus, dass das Erkennen und Einordnen von Stoffen und deren Eigenschaften z. B. wichtiger Teil des notwendigen Wissens ist, aber auch das Messen und Vergleichen ein Verständnis für die Fachlogik eröffnet.

Mit komplexeren Bewerten-Vorstellungen konfrontiert äußern Lehrerinnen und Lehrer dieses Typs einerseits Interesse und bekunden die Bedeutung der Fragestellungen, es wird aber deutlich darauf hingewiesen, dass eine Beschäftigung mit ethisch, politisch und sozial aufgeladenen Themen Aufgabe anderer Unterrichtsfächer sei und nicht dem Charakter des (naturwissenschaftlichen) Unterrichts entspreche. Dabei kommt auch zum Ausdruck, dass sich insbesondere die Naturwissenschaftslehrkräfte von den Anforderungen bzw. Ansprüchen der in den Bildungsstandards formulierten Kompetenzen überfordert fühlen.

Bedeutung von Sachwissen

Dem Sachwissen kommt in den Vor- und Einstellungen dieses Typs eine besondere Bedeutung zu. Es wird als zentrale Aufgabe verstanden, Fachwissen zu vermitteln, dafür Sorge zu tragen, dass sich Schülerinnen und Schüler einen fundierten Wissensbestand erarbeiten und auf dessen Basis zu mündigen Bürgerinnen und Bürgern werden. Ob dieses gelingen kann, hängt dann weniger vom Unterricht oder der Initiierung von Urteilsprozessen ab als von der Lernbereitschaft und Intelligenz der Schülerinnen und Schüler. Schließlich wird auch benannt, dass das jeweilige Fach, weil es komplex und schwierig ist, eine besondere Aufmerksamkeit auf das Fachwissen legen müsse, um dessen Vermittlung zu ermöglichen.

Erst wenn der Bildungsauftrag in Form der Vermittlung des Fachwissens erfolgreich erfüllt worden ist, können weitere Anwendungs- und Lebensweltbezüge hergestellt werden, die das Gelernte illustrieren bzw. dessen Transfer in andere Gebiete schulen.

Umsetzung

Bewertungskompetenz wird, wie bereits in den vorangegangenen Abschnitten angedeutet, in den Unterricht integriert, wenn es der Vermittlung von Fachwissen dient. Dabei wird z. T. auch die Erarbeitung komplexer Sachverhalte als ein zentraler Teil zur Förderung von Bewertungskompetenz verstanden und ist insofern immer Bestandteil des Unterrichts. Auch in den innerfachlich geprägten Bewerten-Vorstellungen hat Bewertung einen wichtigen Stellenwert im Unterricht, wenn es z. B. um die Vermittlung von Stoffeigenschaften geht, die Struktur des Messens verdeutlicht, auf Fehler und Fehlerrechnung eingegangen oder Sachwissen richtig erläutert werden soll.

Im Sinne von auch ethisch komplexen und aufgeladenen Themen findet die Förderung von Bewertungskompetenz nur begrenzt Eingang in das Unterrichtsgeschehen. Dabei wird zwar die grundsätzliche Relevanz von Bewertungskompetenz nicht dementiert, aber die Herausforderungen des (naturwissenschaftlichen) Wissens seien eben so groß, dass für darüber hinausgehende Fragestellungen die Unterrichtszeit nicht ausreiche. Zudem wird die Ansicht vertreten, dass insbesondere eine solide fachwissenschaftliche Erarbeitung die Schülerinnen und Schüler in ihrer Bewertungskompetenz implizit fördere.

Einstellung

Lehrpersonen, die davon ausgehen, dass die Vermittlung von Fachwissen einen zentralen Beitrag für die Förderung von Bewertungskompetenz leistet bzw. konfligierende Evidenz als dem Fach innewohnend verstehen und diese auch mit Schülerinnen und Schülern thematisieren, bezeugen zunächst einmal eine sehr positive Einstellung dem Kompetenzbereich Bewertung gegenüber. Dabei bezieht sich ihre Zustimmung des Ziels, Bewertungskompetenz zu fördern, jedoch nicht auf die Form von formulierten Bildungsstandards, Output-Orientierung oder gar bildungsadministrativen Vorgaben. Diese werden eher als hinderlich und der Schulrealität nicht gerecht werdend empfunden.

Ablehnung oder zumindest Vorbehalte äußern Lehrpersonen dieses Typs dem Thematisieren von Werten und Normen, ethischen Fragestellungen und Unterrichtsthemen gegenüber, die Bewertungskompetenz in einem umfassenden Sinn fördern sollen. Die Einstellung umfasst sowohl Vorbehalte, die die prinzipielle Erreichbarkeit des Ziels, Bewertungskompetenz zu fördern, anzweifeln, und reicht bis zur expliziten Ablehnung einer „Neu- oder Umorientierung“ des Unterrichts weg von Fachinhalten und hin zu „weicheren“ und schwer überprüfbaren Zielen wie Bewertungskompetenz. Insbesondere Naturwissenschaftslehrkräfte dieses Typs artikulieren Bedenken, überfordert zu sein mit der Förderung von Bewertungskompetenz in einem umfassenden Sinn. Sie schätzen sich als für solche unterrichtsmethodischen Varianten nicht ausreichend ausgebildet ein. Die hohe Bedeutung, die naturwissenschaftlichem Fachverständnis zugewiesen wird,

begünstigt es selbst bei jüngeren Lehrkräften, „weichere“ Ziele wie die Förderung von Bewertungskompetenz für ihren Unterricht abzulehnen.

Ablehnung wird auch dahingehend geäußert, dass die Bildungsstandards in einer Top-down Vorgabe schulisches Know-how außer Acht lassen und sich in eine Kette von häufig veränderten Vorgaben einreihen, die als Schikane empfunden werden.

6.2.2 Typ 2: Bewerten ist alles

Fokus

Anders als der oben dargestellte Typ steht dieser Typ für eine radikalere Veränderung des Naturwissenschaftsunterrichts weg von der Vermittlung „trockenen, toten“ Fachwissens hin zu lebensweltrelevanten, gesellschaftlichen Fragestellungen, für deren Bearbeitung das benötigte Fachwissen erarbeitet werden soll. Im Vordergrund steht der Anspruch, Schülerinnen und Schüler für die zumeist weniger beliebten naturwissenschaftlichen Fächer zu motivieren, mit aktuellen Fragestellungen zu interessieren und zu vermitteln, dass Naturwissenschaften tatsächlich erlernbar sind und Spaß machen.

Die Bewerten-Vorstellungen des zweiten Typs umfassen ähnlich wie in Typ 1 *innerfachliche* Bewerten-Vorstellungen und die Vorstellung *Bewerten-ist-Anwenden*. Dabei ist die Konnotation und Bedeutung der Vorstellungen denen des ersten Typs jedoch diametral entgegengesetzt. Während im ersten Typ die Förderung von Bewertungskompetenz als Vehikel für die Fachwissensvermittlung genutzt wird, wird von Typ 2 Bewertungskompetenz als eigenständiges Ziel des Unterrichts verstanden. Dieses Ziel zu erreichen wird über verschiedene Wege und Angebote versucht. Dafür werden vielfältige lebensweltliche, aktuelle, politische, soziale und ethische Bezüge und Diskussionen in den Unterricht aufgenommen. Vorrangig sind diesem Typ also die *Bewerten-ist-Anwenden* Vorstellungen sowohl in der Nutzung von Anwendungsbeispielen als auch in der Herstellung lebensweltlicher Bezüge. Die *Bewerten-ist-Anwenden* Vorstellung wird von Lehrkräften dieses Typs artikuliert und nimmt für den Unterricht eine zentrale Rolle für die Ausrichtung hin zu offeneren Themen und Unterrichtsmethoden ein. Dabei wird auch in Kauf genommen, dass die Erarbeitung von Fachinhalten dafür in den Hintergrund treten kann.

Innerfachliche Bewerten-Vorstellungen „schleichen“ sich eher unbemerkt ein und haben im Interview vor allem die Funktion zu verdeutlichen, dass sich Bewertungskompetenz überall findet und fördern lässt. Sie werden damit gleichsam zur Rechtfertigung der eigenen Schwerpunktsetzung. Dabei kann es passieren, dass der Bewertungskompetenzbegriff unscharf wird, was schließlich zu einem Konglomerat verschiedenster Ansätze führt, die mit Bewertungskompetenz in Verbindung gebracht werden. Insofern kann Bewerten „alles“ sein, was den Unterricht anschaulich werden lässt, und hat in der eigenen Unterrichtsausrichtung einen besonders hohen Stellenwert, bedeutet den Lehrkräfte also „alles“.

Dieser Typ ist unter den Lehrkräften weniger oft vertreten als der vorangegangene Typ, findet sich aber außer bei den Politiklehrkräften für alle naturwissenschaftlichen

Fächer. Besonders ausgeprägt ist dieser Typ bei sehr jungen Lehrkräften, die in ihrer Ausbildung die Förderung prozessbezogener Kompetenzen als wichtigen progressiven Anspruch kennen gelernt und als solchen verinnerlicht haben.

6.2.3 Typ 3: Die Anspruchsvollen

Fokus

Die Anspruchsvollen verfolgen mit ihrem Unterricht in erster Linie allgemeinbildende Unterrichtsziele. Die Befähigung zur Teilhabe an der Gesellschaft ist zentrales Ziel, ebenso wie die Ausprägung eines kritischen Bewusstseins und der reflektierte Umgang mit v. a. ökologischen Fragen. Die Schülerinnen und Schüler sollen lernen, sich eine eigene Meinung zu bilden, diese zu begründen, darzulegen und zu vertreten. Dafür benötigen sie einen kompetenten Umgang mit dem gelernten Fachwissen. Dies kann auch bedeuten, sich gegen eine weitere fachliche Komplexitätserhöhung auszusprechen und auf Sachwissensnutzung zu verzichten, um das behandelte Fachwissen zu festigen.

Die Bewerten-Vorstellungen dieses Typs changieren zwischen der *Bewerten-ist-Anwenden* und der *Bewerten-ist-Urteilen* Vorstellung. Es handelt sich bei allen geäußerten Vorstellungen um recht differenzierte Vorstellungen, die dem Bewerten eine hohe Bedeutung und große Vielschichtigkeit einräumen. Mit der *Bewerten-ist-Anwenden* Vorstellung geht die priorisierte Nutzung lebensweltlicher Bezüge einher, um das Fachwissen zu beleben, besser einordnen zu können und ihm einen tieferen Sinn zu verleihen. Es wird damit aber auch verbunden, Schülerinnen und Schülern mit Hilfe des naturwissenschaftlichen Wissens „richtiges Handeln“ zu vermitteln. Insbesondere im ökologischen Themenfeld, dem sich Lehrpersonen diesen Typs verpflichtet fühlen, möchten sie Schülerinnen und Schülern einen Einblick in komplexe Zusammenhänge ermöglichen und ihnen an Ökologie und Nachhaltigkeit gebundene Werte vermitteln. Insgesamt lassen sich in den Äußerungen der Lehrkräfte diesen Typs besonders häufig *Bewerten-ist-Urteilen* Vorstellungen identifizieren. Auch wenn sie in ihrer Orientierung auf ökologische Themen bestimmte Werthaltungen vermitteln wollen, legen sie viel Wert auf die Prozesshaftigkeit des Bewertens. Die zentrale Aufgabe von Unterricht sehen sie entsprechend darin, dass Schülerinnen und Schüler eigene Urteilsprozesse vollziehen und Bewertungs-Strukturen erlernen, um ihr Bewerten kompetenter zu gestalten. Dabei wird von den Lehrkräften diesen Typs eine Vielzahl von Facetten des Urteils-Begriffs benannt und für wichtig erachtet.

Lehrpersonen dieses Typs blicken auf einen (breiten,) oftmals umweltaktivistischen Erfahrungsschatz außerhalb der Schule zurück und haben sich später als ihre Kolleginnen und Kollegen, aber sehr bewusst für die Arbeit in der Schule entschieden. Ihre Tätigkeit vor der Arbeit als Lehrkraft war oftmals im Umweltbildungsbereich angesiedelt, für den sie sich nach wie vor besonders einsetzen. Beim Unterrichten der Schülerinnen und Schüler schätzen sie die Möglichkeiten, vermittelt über die Bildung der Schüler und Schülerinnen die Welt positiv zu verändern.

Die Lehrkräfte, die diesem Typ zugeordnet werden können, sind zumeist Biologie- oder Politiklehrkräfte. Insbesondere bei den Biologielehrkräften lässt sich eine ökologische Orientierung ausmachen. Die Politiklehrkräfte vertreten häufig einen ähnlich hohen ökologischen oder aber sozial-politischen Anspruch an die Wirksamkeit ihres Unterrichts.

6.2.4 Typ 4: Aufs Kerngeschäft beschränken

Fokus

Lehrkräfte dieses Typs unterrichten oftmals an Schulen mit problematischen Bedingungen. Die Schülerschaft ist geprägt von einem hohen Anteil an Schülerinnen und Schülern, die leistungsschwach sind oder Probleme mit der Unterrichtssprache haben. Frustrierende Erfahrungen bezogen auf die Leistungsfähigkeit und -bereitschaft der Schülerschaft, die Menge an Disziplinschwierigkeiten und das Wissen um die Perspektivlosigkeit eines Hauptschulabschlusses haben sowohl den Anspruch ihrer Fachwissensvermittlung als auch die Idee, wohlformulierte Bildungsziele zu erreichen, in ein anderes Licht gestellt als z. B. zu Beginn ihrer Lehrtätigkeit.

Die unterrichtsrelevanten Fragestellungen dieser Lehrpersonen sind weit von den bildungsadministrativen Neuausrichtungen hin zur Out-put-Orientierung entfernt. Dies erklärt z. T. die Schwierigkeit der Verständigung über den Kompetenzbereich Bewertung. Bei einigen Lehrkräften dieses Typs lassen sich in erster Linie alternative Bewertung-Vorstellungen identifizieren, die mit den curricularen Vorgaben der Bildungsstandards in keinem Zusammenhang stehen. Hierbei handelt es sich z. B. um das Missverstehen des Bewerten-Begriffs als Leistungsbeurteilung.

Eine weitere Verständigungsschwierigkeit lässt sich identifizieren, wenn jegliche Assoziationen zu Fragen nach Bewertungskompetenz fehlen. Hierbei wird deutlich, dass die Förderung von Bewertungskompetenz als weit entfernt von der eigenen unterrichtsrelevanten Wirklichkeit gesehen wird. Selbst verschiedene Nachfrage-Stimuli konnten im Interview keine Assoziation in Richtung Urteilen und Bewerten im hier gemeinten Sinn hervorrufen.

Wenn bei diesem Typ Bewerten-Vorstellungen im Sinne des hier untersuchten Kompetenzbereichs geäußert werden, handelt es sich zumeist um *innerfachliche* Bewerten-Vorstellungen (z. B. Überprüfen von Messwerten, Bestimmen der Toxizität von Stoffen). Bewertungskompetenz-Vorstellungen, die von Komplexität gekennzeichnet sind (z. B. Thematisierung von Wertvorstellungen, Meinungsbildung), werden als nicht erreichbar und unrealistisch eingeschätzt. Sie finden keinen Eingang in konkrete unterrichtsrelevante Überlegungen, während pädagogische Anforderungen als dominant und vordergründig erachtet werden. Die Lehrkräfte fühlen sich davon völlig eingenommen und konstatieren einen Mangel an Rückmeldungen ihrer Schülerinnen und Schülern, die Interesse an komplexen Sachverhalten wie z. B. ökologischen Themen erkennbar werden ließen. Die Lehrkräfte sehen sich der Herausforderung gegenüber, eine möglichst gute Lernatmosphäre herzustellen, grundlegende Fähigkeiten zu fördern und allgemeinbildende Ziele zu verfolgen.

Lehrkräfte, die diesem Typ zugeordnet wurden, finden sich in allen naturwissenschaftlichen Fächern, jedoch nicht unter den Politiklehrkräften. Keiner von ihnen unterrichtet an einem Gymnasium und sie blicken zumeist auf eine langjährige Schul- und Unterrichtserfahrung zurück.

7. Diskussion

Für die Beantwortung der Forschungsfrage nach den Vorstellungen und Einstellungen und zugrunde liegenden pädagogisch-didaktischen Überzeugungen von Lehrkräften verschiedener Fächer zum Kompetenzbereich Bewertung der Nationalen Bildungsstandards wurden in einer qualitativen Untersuchung mit halbstrukturierten, leitfadengestützten, problemzentrierten Interviews Daten erhoben. Die Analyse der Daten erfolgte in einem rekursiven, die Perspektiven der Lehrpersonen rekonstruierenden Verfahren und mittels eines in einem induktiv-deduktivem Verfahren entwickelten Kategoriensystem. Vier Vorstellungen über Bewertungskompetenz ließen sich voneinander abgrenzen und charakterisieren.

Lehrkräfte vertreten nicht eine isolierte Vorstellung von Bewertung. Stattdessen lassen sich auf Fallebene in aller Regel verschiedene Vorstellungen identifizieren. In die Konstruktion und Unterscheidung von vier Lehrer-Typen gingen daher Vorstellungen ein, die sich in verschiedenem Maße und in verschiedenen Konstellationen überlappen. Dass Bewerten-Vorstellungen nicht in Reinform beobachtbar sind, ist nicht verwunderlich. Der Bewerten-Begriff der Alltagssprache offenbart für Lehrkräfte eine große Bedeutungsvielfalt mit zahlreichen semantischen Anknüpfungspunkten, die es ermöglichen, jeweils jene Bedeutungen betont zu artikulieren, die den zugrunde liegenden pädagogisch-didaktischen Überzeugungen gemäß sind. Bewertung kann eben vieles heißen. Entsprechend kann Bewerten den Prozess der Urteilsbildung mit dem Ziel normativ erwünschter Meinungsbildung bedeuten (*Typ 3 – Die Anspruchsvollen*). Ein solcher Bewerten-Begriff kommt den im Theorie-Teil dieser Arbeit skizzierten Vorstellungen nah. Sehen Lehrkräfte ihre Aufgabe aber eher darin, fachliche Lehr-Lern-Prozesse zu initiieren (*Typ 1 – Sachwissen ist das A und O*) oder für fachliches Lernen zu motivieren (*Typ 2 – Bewerten ist alles*), dann tritt eine *Bewerten-ist-Urteilen* Vorstellung zugunsten der *innerfachlichen* Bewerten-Vorstellung und der *Bewerten-ist-Anwenden* Vorstellung in den Hintergrund. *Typ 4 (Aufs Kerngeschäft beschränken)* zeigt dagegen, dass elaborierte Vorstellungen von Bewertung gar nicht vertreten werden, sondern die neuartige curriculare Anforderung der Förderung von Bewertungskompetenz gleichsam an einem anspruchsvollen und fordernden pädagogisch-didaktischen Alltag der Lehrkräfte abprallt. Entsprechend werden jene Bedeutungsschichten von Bewertung, die die Bildungsstandards oder auch die elaborierte fachdidaktische Diskussion (s. Theorie-Teil) generiert haben, nicht wahrgenommen.

Es zeigt sich, dass die unklare und uneindeutige Botschaft der Bildungsstandards, wie sie in der Beschreibung und den Erklärungen zu den Kompetenzbereichen zu finden ist, sich bei den Lehrkräften fortsetzt. Bezogen auf die Bildungsstandards lässt sich fest-

stellen, dass Lehrpersonen in ihre Bewerten-Vorstellungen Dimensionen integrieren, die in den Bildungsstandards z. B. im Kompetenzbereich Kommunikation oder auch im Bereich Erkenntnisgewinnung verortet werden. So konnte z. B. gezeigt werden, dass für viele Lehrkräfte das Abwägen und Darlegen von Argumenten einen Teilaspekt von Bewertungskompetenz darstellt. Diese Aspekte sind in den Bildungsstandards vornehmlich unter dem Kompetenzbereich Kommunikation zugewiesen.¹¹ Damit weisen die Lehrkräfte – und das mit Recht – ein breiteres Verständnis von Bewertung auf, als es die Bildungsstandards für den gleichnamigen Kompetenzbereich nahelegen. Es erscheint uns sinnvoll und nötig, dass administrative Vorgaben, was unter Bewertung überhaupt zu verstehen sei, harmonisiert werden, um weitere Konfusion und Unklarheit zu vermeiden. Dabei könnten Aufgaben- und Unterrichtsbeispiele und didaktische Kommentare für Lehrkräfte insbesondere der naturwissenschaftlichen Fächer hilfreich sein, um diese darin zu unterstützen, in ihrem Unterricht Bewertungskompetenz zu fördern.

Die in den Bildungsstandards je nach Fach unterschiedliche Akzentuierung des Kompetenzbereichs Bewertung spiegelt sich in der Fachgruppenzugehörigkeit der Lehrkräfte nur schwach wieder. Weder die in dieser Studie identifizierten Bewerten-Vorstellungen der Lehrkräfte noch die entwickelte Typologie weisen eine ausgeprägte Fachgruppenspezifität auf. Dem Typ 3 (*Die Anspruchsvollen*) lassen sich insbesondere Biologielehrkräfte oder Lehrkräfte, die auch Biologie unterrichten, zuordnen, Typ 4 (*Aufs Kerngeschäft beschränken*) keine Politiklehrkräfte. Da Typ 4 v.a. skeptische Einstellungen gegenüber Bewertungskompetenz umfasst, schlägt sich hier das tradierte Selbstverständnis der Politiklehrkräfte als Förderer politischer Urteilsfähigkeit nieder. Lehrkräfte der „harten“ Naturwissenschaften Physik und Chemie entsprechen in der Tendenz eher jenen Typen, die die Dominanz fachlicher Lernziele vertreten. Auch wenn viele Physik- und Chemie-Lehrkräfte sogenannte *innerfachliche* Bewerten-Vorstellungen äußern, finden sich diese auch bei Biologie- und Politiklehrkräften. Ebenso verhält es sich mit den Vorstellungen *Bewerten-ist-Urteilen* und *Bewerten-ist-Anwenden*. Lehrkräfte aller Fächer artikulieren Ausprägungen, die zu einer *Bewerten-ist-Urteilen* Vorstellung gefasst wurden. Auffällig ist der Befund, dass die Zugehörigkeit der Lehrkräfte zu verschiedenen Schulformen die Nähe zu Typen erklären kann. Der Typ 1 wird eher von Gymnasiallehrkräften, der Typ 4 von Lehrkräften mit eher leistungsschwachen Schülerinnen und Schülern gestellt.

Die Typologie kann Hinweise für die Gestaltung von Lehrerfortbildungen geben. Ein Teil der interviewten Lehrkräfte hatte zum Zeitpunkt der Datenerhebung bereits eine Fortbildung zur Förderung von Bewertungskompetenz im Themenfeld Klimawandel erhalten. Es handelt sich um die am Projekt „Der Klimawandel vor Gericht“ beteiligten Lehrkräfte, die sich bereits eineinhalb Jahre mit der Entwicklung von Unterricht zur Förderung von Bewertungskompetenz befasst hatten. Im Rahmen der Veranstaltung wurde den Lehrkräften mehrfach erläutert, was in den Bildungsstandards ihres jeweiligen Faches unter Bewertung verstanden wird und welche Unterrichtskonzeptionen

11 Vergleiche die Erläuterungen zu beiden Kompetenzbereichen in den Bildungsstandards (KMK 2005a; b; c).

bereits entwickelt worden sind. Auch die so fortgebildeten Lehrkräfte lassen sich allen vier Typen der Typologie zuordnen. Eine auffällige Ausprägung der jeweiligen Typen lässt sich in dieser Gruppe nicht identifizieren. Das heißt z. B. für den Typ 4 (*Aufs Kerngeschäft beschränken*), dass selbst regelmäßige und zeitlich umfangreiche Fortbildungsangebote nicht dazu führen müssen, dass die angebotenen Bedeutungshorizonte von Bewertung in die eigenen Sinnhorizonte integriert werden.

Die entwickelte Typologie gibt Hinweise auf Probleme, die sich z. B. im Rahmen künftiger Lehrerfortbildungen zum Thema Bewertungskompetenz-Förderung einstellen könnten. Typ 1 und 2 verstehen Bewertung – wenn auch mit unterschiedlichen Schwerpunktsetzungen – als innerfachlichen Prozess. Sie müssten im besonderen Maße dafür sensibilisiert werden, dass Bewertung sich auf überfachliche Kontexte beziehen kann, die von ethischen Implikationen gekennzeichnet sind. Bewerten meint dann die begründete Urteilsbildung über Handlungsoptionen. Die Begeisterung des Typs 2 für die Förderung prozessbezogener Kompetenzen in lebensweltlichen Kontexten stellt eine gute Ausgangslage dar. Typ 3 scheint es bereits gut zu gelingen, fachlichen Anspruch und prozessbezogene Bewertungskompetenz zu integrieren. Es könnte für Typ 3 allerdings eine besondere Herausforderung darstellen, auch solche Urteile von Schülerinnen und Schülern zuzulassen und anzuerkennen, die nicht ihren persönlichen normativ aufgeladenen Erwartungen entsprechen. Fortbildungsangebote für Typ 4 müssten die spezifischen Bedingungen von Schule und Unterricht (Leistungsschwäche, mangelnde Motivation der Schülerinnen und Schüler), unter denen dieser Typ arbeitet, in besonderem Maße berücksichtigen.

Die hier vorgestellte Studie ist explorativ angelegt. Es wäre zu prüfen, ob die identifizierten Vorstellungen als auch die Typologie sich in einem quantitativen Design an einer repräsentativen Stichprobe bewähren. Weitere Forschung müsste ebenso klären, ob die Zuordnung einer Lehrkraft zu einem Typ auch Konsequenzen auf der Ebene der Unterrichtsgestaltung und – vermittelt über die Wirkungskette Lehrer-Unterricht-Schüler – Wirkungen auf Schülerseite zeitigt.

Autorenangaben

Maria Mrochen
Universität Bremen
Institut für Didaktik der Naturwissenschaften
Abt. Physikdidaktik
Fachbereich 1, Physik/Elektrotechnik
E-Mail: mariamro@uni-bremen.de

Dietmar Höttecke
Universität Hamburg
Fakultät EPB, Fachbereich Erziehungswissenschaft
Physikdidaktik
E-Mail: dietmar.hoettecke@uni-hamburg.de

Literatur

- Albe, V. (2008a): Students' positions and considerations of scientific evidence about a controversial socioscientific issue. In: *Science & Education*, 17, S. 805-827.
- Albe, V. (2008b): When Scientific Knowledge, Daily Life Experience, Epistemological and Social Considerations Intersect: Students' Argumentation in Group Discussions on a Socioscientific Issue. In: *Research in Science Education*, 38, S. 67-90.
- Barab, S. A./Luehmann, A. L. (2003): Building Sustainable Science Curriculum: Acknowledging and Accommodating Local Adaptation. In: *Science Education*, 87, 4, S. 454-467.
- Bell, R./Ledermann, N. G. (2003): Understandings of the nature of science and decision making on science and technology based issues. In: *Science Education*, 87, 4, S. 352-377.
- Betsch, T. (2005): Wie beeinflussen Routinen das Entscheidungsverhalten? In: *Psychologische Rundschau*, 56, 4, S. 261-270.
- Bohnsack, R. (2003): *Rekonstruktive Sozialforschung. Einführung in qualitative Methoden*. Opladen: Leske + Budrich.
- Brundtland-Bericht (1987): General Assembly. Development and International Economic Co-Operation: Environment. Report of the World Commission on Environment and Development. Note by the Secretary-General. Bericht der Brundtland-Kommission 1987. http://www.bne-portal.de/coremedia/generator/unesco/de/02__UN-Dekade_20BNE/01__Was_20ist_20BNE/Brundtland-Bericht.html [Zugriff: 3.2.2011]
- Bundeszentrale für politische Bildung (1997): *Politische Urteilsbildung. Aufgabe und Wege für den Politikunterricht*. Schriftenreihe, Band 344. Schwalbach/Ts.: Wochenschau-Verlag.
- Bybee, R. (1997): *Achieving scientific literacy: from purposes to practices*. Portsmouth: Heilmann.
- Collins, H. M./Evand, R. (2007): *Rethinking Expertise*. Chicago: The University of Chicago Press.
- Cross, R. T./Price, R. F. (1999): The social responsibility of science and the public understanding of science. *IJSE*, 21, 7, S. 775-785.
- Eggert, S./Bögeholz, S. (2006): Göttinger Modell der Bewertungskompetenz - Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ für Gestaltungsaufgaben Nachhaltiger Entwicklung. *ZfDN*, 12, S. 177-197.
- Eilks, I. et. al. (2011a, im Druck): *Der Klimawandel vor Gericht - Unterrichtsmaterialien für den Fachunterricht oder fächerübergreifende Projekte*. Köln: Aulis.
- Eilks, I., et. al. (2011b): Bewerten Lernen und der Klimawandel in vier Fächern - Einblicke in das Projekt „Der Klimawandel vor Gericht“ (Teil 1). *MNU*, 64, 1, S. 7-11.
- Eilks, I., et. al. (2011c): Bewerten Lernen und der Klimawandel in vier Fächern – Einblicke in das Projekt „Der Klimawandel vor Gericht“ (Teil 2). *MNU* 64, 2, S. 71-76.
- Fischler, H. (2001): Verfahren zur Erfassung von Lehrer-Vorstellungen zum Lehren und Lernen in den Naturwissenschaften. *ZfDN*, 7, S. 105-120.
- Gebhard, U. (2007): Intuitive Vorstellungen bei Denk- und Lernprozessen: Der Ansatz „Alltagsphantasien“. In: Krüger, D./Vogt, H. (Hrsg.): *Theorien in der didaktischen Forschung. Ein Handbuch für Lehramtsstudierende und Doktoranden*. Berlin: Springer-Verlag.
- Geijsel, F./Sleegers, P./Berg, R. v. d./Lelchtermans, G. (2001): Conditions Fostering the Implementation of Large-Scale Innovation Programs in Schools. Teachers' Perspectives. In: *Educational Administration Quarterly*, 37, 1, S. 130-166.
- Gesellschaft für Politikdidaktik und politische Jugend- und Erwachsenenbildung (2004): *Anforderungen an Nationale Bildungsstandards für den Fachunterricht in der Politischen Bildung in Schulen*. Ein Entwurf. Schwalbach/Ts.: Wochenschau Verlag.
- Gräsel, C./Parchmann, I. (2004): Implementationsforschung - oder: der steinige Weg, Unterricht zu verändern. In: *Unterrichtswissenschaft*, 32, 3, S. 196-214.

- Haidt, J. (2001): The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. In: *Psychological Review*, 108, 4, S. 814-834.
- Haney, J. J./Czeraniak, C. M./Lumpe, A. T. (1996): Teacher Beliefs and Intentions Regarding the Implementation of Science Education Reform Strands. In: *Journal of Research in Science Teaching*, 33, 9, S. 971-993.
- Hopf, C. (1982): Norm und Interpretation. Einige methodische und theoretische Probleme der Erhebung und Analyse subjektiver Interpretationen in qualitativen Untersuchungen. In: *Zeitschrift für Soziologie*, 11, 3, S. 307-329.
- Hopf, C. (1993): Rechtsextremismus und Beziehungserfahrungen. *Zeitschrift für Soziologie*, 22, 6, S. 449-463.
- Höbke, C./Bayrhuber, H. (2006): Sechs Schritte moralischer Urteilsfindung. Aktuelle Beispiele aus der Bioethikdebatte. In: *Praxis der Naturwissenschaften-Biologie*, 55, 4, S. 1-6.
- Höttecke, D. et. al. (2009a): BNE in der Sekundarstufe - Planspiele als Lernansatz mit Alltags- und Ich-Bezug, Bewertungskompetenz und Bildung für nachhaltige Entwicklung in der Sekundarstufe. In: Brickwedde, F./Bittner, A. (Hrsg.): *Kindheit und Jugend im Wandel! Umweltbildung im Wandel?* Berlin: Erich-Schmidt-Verlag, S. 61-69.
- Höttecke, D./Maiseyken, V./Rethfeld, J./Mrochen, M. (2009b): Den Treibhauseffekt verstehen. *Naturwissenschaften im Unterricht Physik*. S. 111-112, S. 24-37.
- Höttecke, D. et. al. (2010): Judgment and Decision-Making about Socio-Scientific Issues: A Fundament for a Cross-Faculty Approach towards Learning about Climate Change. In: Eilks, I./Ralle, B. (Hrsg.): *Contemporary science education- implications from science education research about orientations, strategies and assessment. A collection of invited papers inspired by the 20th Symposium on Chemical and Science Education held at the University of Bremen, 27.19 May 2010.* Shaker-Verlag, S. 179-191.
- Höttecke, D./Mrochen, M. (2010): Bewerten Lernen im Treibhaus. Physikalisches Wissen beim Bewerten und Entscheiden nutzen. In: *Praxis der Naturwissenschaften – Physik in der Schule*, 59, 2, S. 26-35.
- Jenkins, E. W. (1999): School science, citizenship and the public understanding of science. *IJSE*, 21, 7, S. 703-710.
- Jungermann, H./Pfister, H.-R./Fischer, K. (2005): *Die Psychologie der Entscheidung. Eine Einführung.* München: Spektrum-Verlag.
- Kelle, U./Kluge, S. (2010): *Vom Einzelfall zum Typus. Fallvergleich und Fallkontrastierung in der qualitativen Sozialforschung.* Wiesbaden: VS Verlag.
- Kluge, S. (1999): *Empirisch begründete Typenbildung. Zur Konstruktion von Typen und Typologien in der qualitativen Sozialforschung.* Opladen: Leske + Budrich.
- Kluge, S. (2000): *Empirisch begründete Typenbildung in der qualitativen Sozialforschung* [14 Absätze]. *Forum Qualitative Sozialforschung /Forum: Qualitative Social Research*, 1(1), Art. 14, <http://nbn-resolving.de/urn:nbn:de:0114-fqs0001145>. [Zugriff: 7/2008]
- KMK (2005a): *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss. Beschluss vom 16.12.2004.* Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.). München: Luchterhand.
- KMK (2005b): *Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss. Beschluss vom 16.12.2004.* Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.). München: Luchterhand.
- KMK (2005c): *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss. Beschluss vom 16.12.2004.* Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.). München: Luchterhand.
- Kolstoe, S. D. (2000): Consensus Projects: Teaching Science for Citizenship. In: *IJSE*, 22, 6, S. 645- 664.

- Kolstoe, S. D. (2001): Scientific literacy for citizenship: Tools for dealing with the science dimension of controversial socioscientific issues. *Science Education*, 85, 3, S. 291-310.
- Kortland, K. (2001): A problem-posing approach to teaching for scientific literacy: The case of decision-making about packaging waste. In: Jong, O. d./Savelsbergh, E. R./Alblas, A. (Hrsg.): *Teaching for Scientific literacy: Context, Competency, and Curriculum*. CD-B Series, 38. Utrecht: CD-B Press, S. 87-98.
- Kuckartz, U. (2007): *Einführung in die computergestützte Analyse qualitativer Daten*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lamnek, S. (2010): *Qualitative Sozialforschung*. Weinheim: Beltz-Verlag.
- Lee, H./Witz, K. G. (2009): Science Teachers' Inspirations for Teaching Socio-scientific Issues: Disconnection with reform efforts. In: *IJSE*, 31, 7, S. 931-960.
- Lewis, J./Leach, J. (2006): Discussion of Socio-scientific Issues: The role of science knowledge. In: *IJSE*, 28, 11, S. 1267-1287.
- Linder, C. et. al. (Hrsg.) (2011): *Exploring the Landscape of Scientific Literacy*. New York: Routledge.
- Menthe, J. (2006): Urteilen im Chemieunterricht. Eine empirische Untersuchung über den Einfluss des Chemieunterrichts auf das Urteilen von Lernenden in Alltagsfragen. Tönning: Der Andere Verlag.
- Menthe, J./Parchmann, I. (2006): Trink- oder Mineralwasser? Bewerten - ein Kinderspiel? In: *Naturwissenschaften im Unterricht Chemie*, 17, 94/95, S. 80-84.
- Mittelsten Scheid, N./Höbke, C. (2008): Wie Schüler unter Verwendung syllogistischer Elemente argumentieren. Eine empirische Studie zu Niveaus von Argumentation im naturwissenschaftlichen Unterricht. In: *ZfDN*, 14, S. 145-165.
- Müller, C./Duit, R. (2004): Funktionen des Experiments: Vorstellungen von Lehrern und Unterrichtsrealität. In: Pitton, A. (Hrsg.): *Chemie- und Physikdidaktische Forschung und naturwissenschaftliche Bildung*. Münster: LIT-Verlag, S. 33-35.
- Müller, R. (2006): Kontextorientierung und Alltagsbezug. In: Mikelskis, H. (Hrsg.): *Physik Didaktik. Praxishandbuch für die Sekundarstufe I und II*. Berlin: Cornelsen Scriptor, S. 102-119.
- Ratcliffe, M. (1997): Pupil decision-making about socio-scientific issues within the science curriculum. In: *IJSE*, 19, 2, S. 167-182.
- Reitschert, K./Höbke, C. (2007): Wie Schüler ethisch bewerten. Eine qualitative Untersuchung zur Strukturierung und Ausdifferenzierung von Bewertungskompetenz in bioethischen Sachverhalten bei Schülern der Sek. I. In: *ZfDN*, 13, S. 125-143.
- Roth, W.-M./Lee, S. (2004): Science education as/for participation in the community. *Science Education*, 88, 2, S. 263-291.
- Sadler, T. D. (2004): Moral and ethical dimensions of socioscientific decision making as integral components of scientific literacy. In: *The Science Educator*, 13, S. 39-48.
- Sadler, T. D./Zeidler, D. L. (2005): Patterns of informal reasoning in the context of socioscientific decision making. In: *Journal of Research in Science Teaching*, 42, 1, S. 112-138.
- Sadler, T. D. /Donnelly, L. A. (2006): Socioscientific Argumentation: The effects of content knowledge and morality. In: *IJSE*, 28, 12, S. 1463-1488.
- Sadler, T. D. (2011): Socio-scientific Issues as Contexts for Learning and Practice in Science Education. In: Höttecke, D. (Hrsg.): *Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie*. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Potsdam 2010. Münster: LIT-Verlag, S. 6-16.
- Schecker, H./Höttecke, D. (2007): „Bewertung“ in den Bildungsstandards Physik. Aufgaben zum Kompetenzbereich „Bewertung“. In: *Unterricht Physik*, 97, 18, S. 29-36.
- Simmons, M. L./Zeidler, D. L. (2003): Beliefs in the Nature of Science and Responses to Socioscientific Issues. In: Zeidler, D. L. (Hrsg.): *The Role of Moral Reasoning on Socio-*

- scientific Issues and Discourse in Science Education. Dordrecht: Kluwer Academic Publishers.
- Snyder, J./Bolin, F./Zumwalt, K. (1992): Curriculum Implementation. In: Jackson, P. W. (Hrsg.): Handbook of Research on Curriculum. New York: Macmillan, S. 402-435.
- Sutor, B. (1971): Didaktik des politischen Unterrichts: eine Theorie der politischen Bildung. Paderborn: Schöningh.
- Tesch, M./Duit, R. (2004): Experimentieren im Physikunterricht - Ergebnisse einer Videostudie. In: ZfDN, 10, S. 51-69.
- Watson, J.R./Swain, J.R.L./McRobbie, C. (1999): The interaction between teaching styles and pupil autonomy in practical science investigations – a case study. In: Leach, J. & Paulsen, A. C. (Hrsg.): Practical Work in Science Education – Recent Research Studies. Dordrecht: Kluwer & Roskilde: Roskilde University Press, S. 148-159.
- Witz, K. G./Lee, H. (2009): Science as an ideal: teachers' orientations to science and science education reform. In: Journal of Curriculum Studies, 41, 3, S. 409-431.
- Zeidler, D. L. (Hrsg.) (2003): The Role of Moral Reasoning on Socioscientific Issues and Discourse in Science Education. Dordrecht: Kluwer Academic Publishers.