

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

Modellierung der Bewertungskompetenz in den Naturwissenschaften zur Evaluation der Nationalen Bildungsstandards

Modeling the evaluation and judgement competence in science to
evaluate national educational standards

ZUSAMMENFASSUNG

Für die Überprüfung der Nationalen Bildungsstandards werden fachdidaktisch begründete und empirisch haltbare Modelle zur Beschreibung von Schülerkompetenzen bzw. zur Entwicklung von Testaufgaben benötigt. Nachdem in einem vorangegangenen Artikel (Kauertz et al., 2010) die fachdidaktischen Überlegungen beschrieben wurden, die zu einer Operationalisierung der Standards im Kompetenzbereich Fachwissen zum Tragen kommen, beschäftigt sich dieser Artikel mit der Modellierung und Operationalisierung des Kompetenzbereiches Bewerten. Dazu werden theoretische und empirische Vorarbeiten aus den naturwissenschaftlichen Fächern zusammengetragen und in ein Kompetenzstrukturmodell übertragen.

Schlüsselwörter: Bewertungskompetenz, Bildungsstandards, Evaluation, Kompetenzmodell, Testentwicklung

ABSTRACT

For the evaluation of the national educational standards models are needed which are capable to describe students' competencies theory driven. These models have to be tested empirically and are needed to develop test items. The theoretical underpinnings for the area of competence "content knowledge" were described in a former article (Kauertz et al., 2010). The article deals with the operationalization of the "evaluation and judgement competence". Theoretical assumptions and empirical studies from the different subjects are considered for the development of a model of competence.

Keywords: evaluation and judgement competence, National educational standards, evaluation, competence model, test development

1 Einleitung

In den durch die Kultusministerkonferenz festgelegten Bildungsstandards für den Mittleren Schulabschluss (KMK, 2005a, 2005b, 2005c) werden, im Sinne der Definition von Weinert, Kompetenzen für die naturwissenschaftlichen Fächer festgelegt, die die Schüler und Schülerinnen erreicht haben sollen. Neben *Fachwissen* und *Erkenntnisgewinnung* gehören auch die Kompetenzbereiche *Kommunikation* und *Bewertung* zu diesen Standards. Im Projekt *Evaluation der Standards in den Naturwissenschaften für die Sekundarstufe I* (ESNaS) werden auf Grundlage fachdidaktischer Arbeiten ein Kompetenzmodell für die drei naturwissenschaftlichen Fächer und dazu passende modellbasierte Leistungstests entwickelt. Die Tests operationalisieren dieses Modell, um das Erreichen der Standards zu überprüfen und diese mittel- bis langfristig weiterzuentwickeln. Die Evaluation der Standards erfolgt in Kooperation des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB) mit Fachdidaktikerinnen und Fachdidaktikern der Fächer Biologie, Chemie und Physik und mit Lehrkräften aus verschiedenen Bundesländern (vgl. Walpuski et al. 2008). Für die Kompetenzbereiche *Fachwissen* und *Erkenntnisgewinnung* ist die Pilotierung der Tests bereits abgeschlossen. Die Normierung findet 2011 statt und ein Ländervergleich 2012. Für die Kompetenzbereiche *Kommunikation* und *Bewertung* wird der Ländervergleich voraussichtlich erst 2018 durchgeführt. In diesem Artikel wird ein Modell für den Kompetenzbereich Bewertung theo-

retisch begründet und eine Ausdifferenzierung der Bewertungskompetenz in Teilkompetenzen und Kompetenzstufen vorgeschlagen. Grundlagen hierfür bilden allgemeine Entscheidungstheorien und fachdidaktische Vorarbeiten verschiedener Fächer.

2 Bewertungskompetenz

2.1 Definitionen für Bewertungskompetenz

Bewertungskompetenz ist nach den Nationalen Bildungsstandards (NBS) die Fähigkeit, biologische/chemische/physikalische Sachverhalte in verschiedenen Kontexten erkennen und bewerten zu können (KMK, 2005a, 2005b, 2005c). Zudem sollen sich Schülerinnen und Schüler am gesellschaftlichen Diskurs beteiligen, dabei verschiedene Perspektiven übernehmen, ihre Wertentscheidungen in Probleme einfließen lassen sowie Entscheidungen sachgerecht, selbstbestimmt und verantwortungsbewusst treffen können (siehe Tab. 1).

Für die in den deutschen Standards beschriebene Bewertungskompetenz lässt sich in der internationalen Literatur kein deckungsgleiches Konstrukt finden. Lediglich die Schweiz beschreibt ein vergleichbares Konstrukt *Einschätzen und Beurteilen* (Labudde, 2007; Labudde & Adamina, 2008; Metzger & Labudde, 2007). Im angloamerikanischen Raum dagegen repräsentieren *socioscientific issues* (SSI) bzw. *socio scientific decision making* ein zur Bewertungskompetenz

Tab. 1: Standards für Bewertungskompetenz (KMK, 2005a, 2005b, 2005c)

BIOLOGIE	CHEMIE	PHYSIK
Die Schülerinnen und Schüler...	Die Schülerinnen und Schüler...	Die Schülerinnen und Schüler...
B 1 unterscheiden zwischen beschreibenden (naturwissenschaftlichen) und normativen (ethischen) Aussagen,	B 1 stellen Anwendungsbereiche und Berufsfelder dar, in denen chemische Kenntnisse bedeutsam sind,	B1 zeigen an einfachen Beispielen die Chancen und Grenzen physikalischer Sichtweisen bei inner- und außerfachlichen Kontexten auf,
B 2 beurteilen verschiedene Maßnahmen und Verhaltensweisen zur Erhaltung der eigenen Gesundheit und zur sozialen Verantwortung,	B 2 erkennen Fragestellungen, die einen engen Bezug zu anderen Unterrichtsfächern aufweisen und zeigen diese Bezüge auf,	B 2 vergleichen und bewerten alternative technische Lösungen auch unter Berücksichtigung physikalischer, ökonomischer, sozialer und ökologischer Aspekte,
B 3 beschreiben und beurteilen Erkenntnisse und Methoden in ausgewählten aktuellen Bezügen wie zu Medizin, Biotechnik und Gentechnik, und zwar unter Berücksichtigung gesellschaftlich verhandelbarer Werte,	B 3 nutzen fachtypische und vernetzte Kenntnisse und Fertigkeiten, um lebenspraktisch bedeutsame Zusammenhänge zu erschließen,	B 3 nutzen physikalisches Wissen zum Bewerten von Risiken und Sicherheitsmaßnahmen bei Experimenten, im Alltag und bei modernen Technologien,
B 4 beschreiben und beurteilen die Haltung von Heim- und Nutztieren,	B 4 entwickeln aktuelle, lebensweltbezogene Fragestellungen, die unter Nutzung fachwissenschaftlicher Erkenntnisse der Chemie beantwortet werden können,	B 4 benennen Auswirkungen physikalischer Erkenntnisse in historischen und gesellschaftlichen Zusammenhängen.
B 5 beschreiben und beurteilen die Auswirkungen menschlicher Eingriffe in einem Ökosystem,	B 5 diskutieren und bewerten gesellschaftsrelevante Aussagen aus unterschiedlichen Perspektiven	
B 6 bewerten die Beeinflussung globaler Kreisläufe und Stoffströme unter dem Aspekt der nachhaltigen Entwicklung,	B 6 binden chemische Sachverhalte in Problemzusammenhänge ein, entwickeln Lösungsstrategien und wenden diese an.	
B 7 erörtern Handlungsoptionen einer umwelt- und naturverträglichen Teilhabe im Sinne der Nachhaltigkeit.		

in vielen Bereichen ähnliches Konstrukt, das jedoch auch Aspekte der Kommunikations- und Argumentationsfähigkeit umfasst. SSI sind komplexe, offene und umstrittene Probleme, bei denen mehrere Lösungen denkbar sind (Sadler, 2004) und werden als Kontext für die Förderung von Scientific Literacy gesehen (Chang Rundgren & Rundgren, 2010). Sie thematisieren auf Basis naturwissenschaftlichen Wissens und gesellschaftlicher Bedürfnisse alltagsnahe Themen, Probleme und Dilemmata, wie z. B. Klimawandel, Energiearten, Globalisierung und genetisch veränderte Nahrungsmittel.

Durch das Verwenden von SSI soll naturwissenschaftliches Lernen durch das Anknüpfen an das tägliche Leben für die Schülerinnen und Schüler relevanter werden (Sadler & Zeidler, 2004). Nach Sadler et al. (2007) umfasst das Bearbeiten von SSI immer vier verschiedene Merkmale: *complexity* (Komplexität des konstruierten Problems), *multiple perspectives* (verschiedene Lösungen in Abhängigkeit von der Beurteilungsperspektive), *ongoing inquiry* (Entscheidungen sind nicht endgültig, da sich die Datenlage ändern kann) und *skepticism* (Quellen müssen hinterfragt werden).

Den Verlauf der Entscheidungsfindung beschreiben Lee und Grace (2010) ähnlich wie Ratcliffe (1997) wie folgt: Zunächst müssen verschiedene Perspektiven für die Betrachtung eines Problems gesucht werden, eine einseitige Betrachtung des Problems ist zu verhindern. Danach müssen Handlungsoptionen gefunden und verglichen werden, bevor die Entscheidung begründet und kommuniziert wird.

Bewertungskompetenz ist bislang nur in wenigen Projekten für eine empirische Prüfung operationalisiert worden. Die vorliegenden Arbeiten werden im Folgenden kurz vorgestellt, um ableiten zu können, inwiefern einzelne Aspekte für die Evaluation der Nationalen Bildungsstandards adaptiert werden können. Die vor allem in der Biologie entwickelten Modelle sind für dieses Ziel hilfreich, aber nicht umfassend genug, da sie auf Teilaspekte und Teilthemen fokussieren und erwartungsgemäß die Bildungsstandards Biologie aufgreifen.

Die Definition von Bögeholz (2007, S. 209) bezieht sich auf den Kontext der nachhaltigen Entwicklung und bezeichnet Bewertungskompetenz als Fähigkeit, „sich in komplexen Problemsituationen begründet und systematisch bei unterschiedlichen Handlungsoptionen zu entscheiden“. In Bezug auf den ökologischen Kontext bezeichnet Bewertungskompetenz nach Bögeholz (2000, S. 1) die „Fähigkeit ökologische Sachverhalte systematisch auf relevante Normen und Werthaltungen zu beziehen“ (siehe auch Große & Bögeholz, 2003, S.104; Barkmann & Bögeholz, 2003, S.50). Damit umfasst Bewertungskompetenz allgemeiner betrachtet sowohl das Treffen begründeter Entscheidungen als auch das Beurteilen anhand bekannter oder vorgegebener Normen und Werthaltungen. Diese beiden Aspekte fassen Mittelsten Scheid und Hößle (2008, S. 88) zusammen, wenn sie Bewertungskompetenz als die Fähigkeit bezeichnen, die „ethische Relevanz naturwissenschaftlicher Themen wahrzunehmen, damit verbundene Werte zu erkennen und ab-

zuwägen sowie ein reflektiertes und begründetes Urteil zu fällen“. Mayer, Harms, Hammann, Bayrhuber & Kattmann (2004, S. 172) greifen diese Zusammenhänge auf und thematisieren die unterschiedliche Gewichtung von Begründungskriterien. Bewertungskompetenz wird hier beschrieben als „Fähigkeit, Kriterien heranzuziehen und zu gewichten, um deskriptives Wissen über den zu beurteilenden Sachverhalt mit individuellen oder gesellschaftlichen Wertsetzungen in transparenter Weise zu verknüpfen“. Rost (2002, S. 8) definiert Bewertungskompetenz im Kontext der nachhaltigen Entwicklung als „Fähigkeit, Entscheidungs- und Handlungsalternativen gegeneinander abzuwägen, sich dabei der involvierten Wertvorstellungen bewusst zu werden und hypothetische oder tatsächliche Entscheidungen aufgrund einer persönlichen Gewichtung vorzunehmen“. In einer weiteren, fächerübergreifenden Definition wird die Bewertung als ein Kennzeichen von Mündigkeit gesehen, und „ermöglicht in gesellschaftlichen Kontroversen (z. B. über den Klimawandel) reflektierte Teilhabe an Diskursen und Entscheidungen“ (Eilks, Feierabend, Hößle, Höttecke & Menthe et al., 2011, S. 7). Dabei soll vor allem die Meinungsbildung gefördert werden, so dass Schüler selber kritisch reflektierte und argumentativ fundierte Meinungen bilden können. Auch nach Schecker und Höttecke (2007) beinhaltet Bewertungskompetenz nicht nur das fachliche Bewerten biologischer, physikalischer oder chemischer Sachverhalte, sondern ein Abwägen verschiedener Kriterien, wie es zum Beispiel in Diskussionen über Vorteile und Gefah-

ren von Experimenten deutlich wird. Bei allen Ansätzen gehört zur Bewertungskompetenz das Erkennen und Abwägen bestimmter Bewertungskriterien, die für die Entscheidungsfindung herangezogen werden müssen. Daraus folgt, dass Schülerinnen und Schüler unter anderem dazu befähigt werden sollen, aktiv und begründet Entscheidungen zu treffen, aber auch bereits getroffene Entscheidungen oder Entscheidungen anderer kritisch zu reflektieren. Diese Fähigkeit ist zum Beispiel auch bei der naturwissenschaftlichen Beurteilung verschiedener Alltagsprodukte von Bedeutung.

2.2 Merkmale von Entscheidungsprozessen

Bewertungen sind in der Regel mit einer (zumindest hypothetischen) Entscheidungssituation verbunden. In solchen Entscheidungssituationen können Bewertungsprozesse durch spezifische Merkmale charakterisiert werden. Im Folgenden werden verschiedene Modelle zur Beschreibung von Entscheidungssituationen und -prozessen sowie entsprechende Einflussfaktoren vorgestellt.

Entscheidungssituationen sind nach Poschmann, Riebenstahl und Schmidt-Kallert (1998) durch drei Grundelemente gekennzeichnet: 1. den Bewertungsgegenstand (Objekt), 2. den Bewerter (Subjekt) und 3. deren Beziehung zueinander. Der Bewerter wird durch seine Grundhaltung bestimmt, die wiederum von seinen persönlichen Präferenzen, Zielen und anderen Abhängigkeiten bedingt wird (siehe Abb. 1).

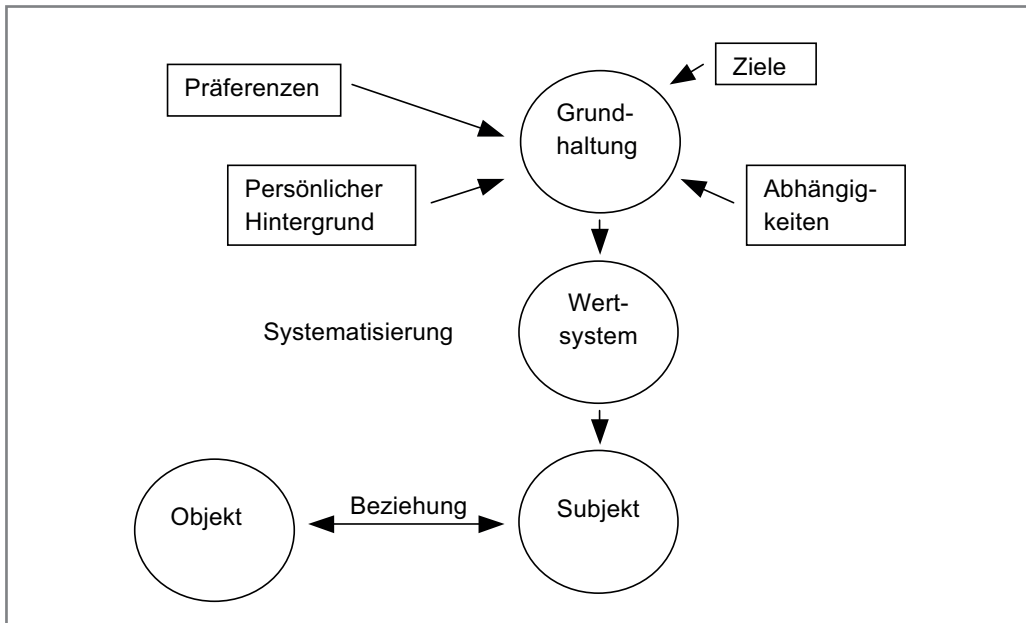


Abb. 1: Bewertungsvorgang nach Poschmann et al. (1998).

Bei einer Entscheidung hat der Bewerter mindestens zwei Optionen, zwischen denen er sich entscheiden kann oder muss. Diese Optionen können etwa Objekte, Handlungen, Regeln oder Strategien sein. Die jeweils relevanten Optionen unterscheiden sich in bestimmten Merkmalen oder Attributen. Sind die möglichen Optionen bereits vorgegeben, muss der Bewerter ausschließlich die Information verarbeiten, um eine Entscheidung zu treffen. Sind die Optionen nicht explizit aufgeführt, müssen sie erst vom Bewerter gesucht bzw. entwickelt werden. Der Verlauf individueller Entscheidungsprozesse ist bislang allerdings wenig erforscht (Jungermann, Pfister & Fischer, 2005).

Nach Kortland (2003) sind Entscheidungen durch einen Prozess gekennzeichnet, bei dem zuerst ein Problem identifiziert

wird, dann Kriterien erkannt, Alternativen generiert und mit den Kriterien abgeglichen werden (Menthe, 2006). Abschließend kann eine sinnvolle Entscheidung getroffen werden (siehe Abb. 2). Das Abgleichen von Kriterien und Generieren von Alternativen beinhaltet eine Anwendung von Entscheidungsstrategien. Das Modell beschreibt nach Menthe (2006) einen allgemeinen Ablauf der Entscheidungsfindung, der auf allen Stufen zusätzlich von individuellen Werten abhängig ist.

Jungermann et al. (2005) gehen bei der Darstellung von Entscheidungsprozessen von folgenden Annahmen aus:

1. Es können *kompensatorische* und *non-kompensatorische* Entscheidungsstrategien unterschieden werden. Bei einer

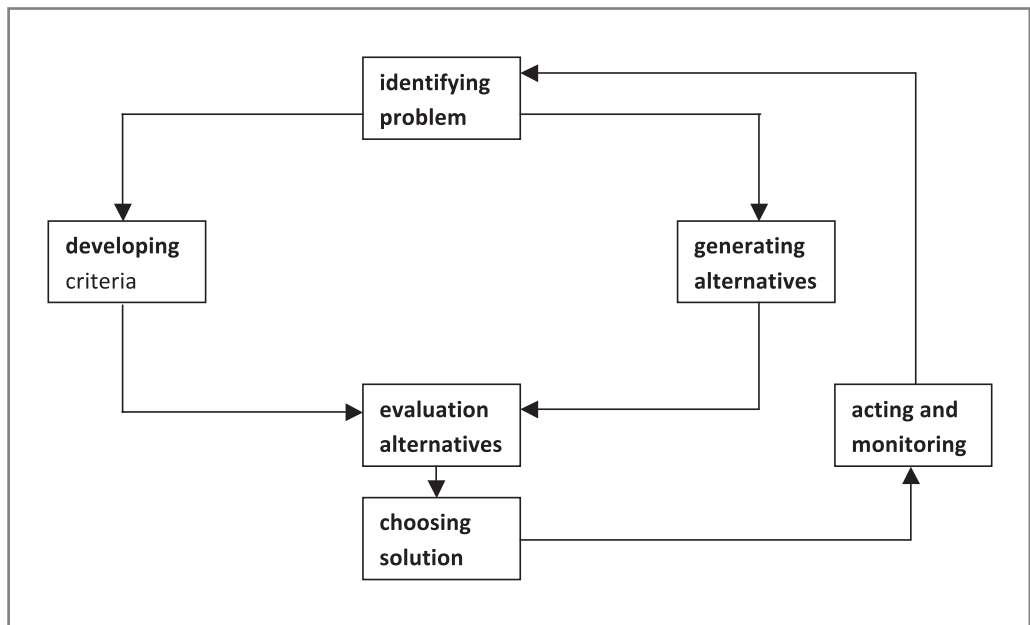


Abb. 2: Prozess der Entscheidungsfindung nach Kortland (2003).

non-kompensatorischen Entscheidung wird eine Option nicht gewählt, wenn sie in einem oder in mehreren Kriterien bestimmte Schwellenwerte (cut-off) nicht erreicht bzw. wenn sie diese Kriterien erst gar nicht besitzt. Bei einer cut-off Entscheidung sind die weiteren Kriterien der Optionen nicht relevant (Hogarth, 1980, Jungermann et al., 2005). Bei Verwendung kompensatorischer Entscheidungsregeln kann dagegen ein schlecht erfülltes Kriterium einer Option durch ein besonders gut erfülltes anderes Kriterium der gleichen Option ausgeglichen werden (trade-off). Es findet also ein Abwägen der Kriterien möglicher unterschiedlicher Optionen statt (Jungermann et al., 2005; Bögeholz, Höhle, Langlet, Sander & Schlüter, 2004). Kompensatorische

und non-kompensatorische Entscheidungsstrategien können in Entscheidungsprozessen kombiniert werden. Hierbei ist die Reihenfolge von cut-off und trade-off nicht festgelegt, wenn auch die Verwendung eines cut-offs nach einem trade-off nicht ideal ist. Bei dieser Reihenfolge wird ein höherer Aufwand betrieben als nötig ist, denn es werden alle Optionen einem trade-off unterzogen, obwohl anschließend anhand eines Kriteriums ein cut-off getroffen wird. Somit müssten eigentlich nur die Optionen beim trade-off berücksichtigt werden, die auch das cut-off-Kriterium erfüllen.

Ein Beispiel für die Anwendung einer kompensatorischen nach der Anwendung einer non-kompensatorischen Ent-

scheidungsstrategie ist die Auswahl von Chemikalien für Schülerversuche. Hierbei ist die Zulassung der Chemikalie für Schülerversuche im Chemieunterricht anhand der Gefährdungspotentiale ein cut-off-Kriterium, während anschließend Preis und bestellbare Packungsgrößen gegeneinander abgewogen werden müssen (trade-off). Wird die Reihenfolge der Kriterien vertauscht, kann dieselbe Entscheidung getroffen werden, allerdings in der Regel mit mehr Aufwand.

2. Es gibt Entscheidungen, bei denen die im Prozess angewendeten Strategien *implizit* oder *explizit* sind. Bei einer impliziten Entscheidung ist nur das Endergebnis der Entscheidung sichtbar. Bei der expliziten Entscheidung wird die Option, die für den Betrachter differenziert dargestellt ist, anhand verschiedener Kriterien (evtl. partiell) bewertet. Sowohl implizite als auch explizite Entscheidungen können kompensatorisch oder non-kompensatorisch getroffen werden.

Im Prozessmodell zur allgemeinen Entscheidungsfindung nach Betsch u.a. wird die Entscheidungsfindung durch eine präsektionale, sektionale und postsektionale Phase dargestellt (Betsch & Haberstroh, 2005; Betsch, Haberstroh & Höhle, 2002). In der *präselektionalen* Phase wird ein Problem identifiziert und weitere Informationen zu diesem Problem werden gesucht. Zudem werden Kriterien identifiziert, die zur Lösung des Problems hilfreich sind. In der anschließenden *selektionalen* Phase findet der eigentliche Bewertungsprozess

statt, bei der die einzelnen Optionen auf Basis von Kriterien gegeneinander abgewogen werden, so dass eine abschließende Entscheidung getroffen werden kann. Die *postselektionale* Phase beinhaltet die Implementation des Verhaltens unter Berücksichtigung volitionaler Kriterien.

In Bewertungssituationen müssen also entweder Entscheidungen getroffen werden oder bereits gefällte Entscheidungen nachträglich bewertet werden. Dazu können die oben genannten Strategien herangezogen und die beschriebenen Prozesse durchlaufen werden. Testaufgaben zum Bewerten ohne eine Entscheidungsstrategie (implizite Strategien) werden in ES-NaS nicht eingesetzt, da die Kodierung der Entscheidungsprozesse qualitative Zusatzinstrumente (wie z. B. nachgeschaltete Interviews) erfordert, die im Large-Scale-Assessment nicht eingesetzt werden können.

3 Modelle der Bewertungskompetenz

Vor dem Hintergrund der beschriebenen theoretischen Konzepte von Bewertung sind Kompetenzmodelle als Basis für die Entwicklung von Aufgaben zur Kompetenzmessung entstanden (Hamann, 2004; Schecker & Parchmann, 2006).

3.1 Göttinger Modell der Bewertungskompetenz

Das Göttinger Modell der Bewertungskompetenz spezifiziert den Teilbereich

der *nachhaltigen Entwicklung* und baut auf den Konstruktionsprinzipien theoretisch und empirisch begründeter Modelle (Kompetenzmodelle) auf (vgl. Bögeholz, 2007, S. 214). Dabei wurde das Metamodell der Entscheidungsfindung (Betsch & Haberstroh, 2005), das im Rahmen von SEPUP (Science Education for Public Understanding Program) verwendete Modell zum decision making (Wilson & Sloane, 2000), das Modell zu Scientific Literacy (Bybee, 1997) und das Modell zur Erkenntnisgewinnung durch Experimentieren (Hamman, 2004) verwendet. Grundlage für das Göttinger Modell der Bewertungskompetenz ist die Definition der Bewertungskompetenz als „Fähigkeit, sich in komplexen Problemsituationen begründet und systematisch bei unterschiedlichen Handlungsoptionen zu entscheiden“ (Bögeholz, 2007, S. 209). Entscheidungen werden als „Reflexion(en) der eigenen Werte und Normen sowie der Werte und Normen anderer im Hinblick auf das Leitbild der nachhaltigen Entwicklung“ beschrieben und in dieser Form von den Schülerinnen und Schülern in Aufgaben verlangt (Eggert & Bögeholz, 2006, S. 183). In diesem Modell werden vier zentrale Teilkompetenzen für Bewertungskompetenz formuliert:

- a) Generieren und Reflektieren von Sachinformationen
- b) Bewerten, Entscheiden und Reflektieren
- c) Kennen und Verstehen von Werten und Normen
- d) Kennen und Verstehen von nachhaltiger Entwicklung

Die von Bögeholz benannten Teilkompetenzen lassen sich in das Prozessmodell der Entscheidung nach Betsch und Haberstroh (2005) einordnen. Dabei bezieht sich die Teilkompetenz a) des Göttinger Modells auf die präselektionale Phase. Hier werden mögliche Optionen zum Fällen einer Entscheidung gesucht. Hierzu können auch Datenerhebungen und Untersuchungen geplant, durchgeführt und aufbereitet werden (Eggert & Bögeholz, 2006). Die Teilkompetenz b) beschreibt zentrale Punkte der selektionalen Phase. Hierunter fällt auch der Prozess der Entscheidung im engeren Sinne, bei dem mögliche Optionen miteinander verglichen und untereinander abgewogen werden. Für eine systematische Entscheidung ist die Kenntnis von kompensatorischen und non-kompensatorischen Entscheidungsstrategien nötig. Die zwei weiteren Teilkompetenzen c) und d) beinhalten den Umgang mit Werten und Normen im Kontext von nachhaltiger Entwicklung. Die Ausdifferenzierung dieser beiden Teilkompetenzen ist im Göttinger Modell zunächst biologiespezifisch. Im Prozessmodell der Entscheidungsfindung von Betsch und Haberstroh (2005) werden diese beiden Teilkompetenzen jedoch nicht als Teil des Bewertungsprozesses gesehen, sondern bilden den Wissenshintergrund für eine Entscheidung.

Die Teilkompetenzen werden im Modell nach Bögeholz durch Kompetenzniveaus graduiert. Beispielhaft soll hier die Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ mit den dazugehörigen Kompetenzniveaus dargestellt werden

(siehe Tabelle 2). Die Stufung der Kompetenzniveaus erfolgt unter anderem anhand der Anzahl und Qualität der verwendeten Kriterien und Entscheidungsstrategien (Bögeholz, 2007). Für diese Stufung konnte mittels Item-Response-Theorie eine empirische Eignung für Large-Scale-Assessments gezeigt werden (Eggert, 2008; Eggert & Bögeholz, 2010).

3.2 Modell zur ethischen Urteilskompetenz

Im Modell der Arbeitsgruppe Hößle (Mittelsten Scheid, 2008; Reitschert & Hößle, 2007) wird Bewertungskompetenz als kognitive Reflexions- und Urteilskompetenz verstanden (Reitschert, 2009; Mandry, 2005; Dietrich, 2002). Nach Reitschert

Tab. 2: Kompetenzniveaus (nach Eggert & Bögeholz, 2006)

NIVEAU	BESCHREIBUNG
	Schüler(innen)...
1	<ul style="list-style-type: none"> • bewerten und entscheiden intuitiv bzw. rechtfertigend ohne Anwendung einer Entscheidungsstrategie • wählen eine Option auf der Basis von Alltagsvorstellungen aus und / oder berücksichtigen dabei maximal 1 Kriterium
2	<ul style="list-style-type: none"> • bewerten und entscheiden unter Berücksichtigung von mindestens zwei relevanten Kriterien • vergleichen gegebene Optionen teilweise im Hinblick auf die Kriterien und dokumentieren ihren Entscheidungsprozess unvollständig • entscheiden v.a. non-kompensatorisch
3	<ul style="list-style-type: none"> • bewerten und entscheiden unter Berücksichtigung von mindestens drei relevanten Kriterien • vergleichen gegebene Optionen vollständig im Hinblick auf die Kriterien und dokumentieren vollständig • entscheiden non-kompensatorisch und / oder kompensatorisch • reflektieren zentrale normative Entscheidungen im Bewertungsprozess
4	<ul style="list-style-type: none"> • bewerten und entscheiden unter Berücksichtigung von mindestens drei relevanten Kriterien • vergleichen gegebene Optionen vollständig im Hinblick auf die Kriterien und dokumentieren vollständig • entscheiden v.a. kompensatorisch • reflektieren zentrale normative Entscheidungen im Bewertungsprozess und können die Grenzen in der Anwendung von Entscheidungsstrategien erkennen

und Hößle (2007, S. 126) steht die „Befähigung zu einem sensibilisierten Zugang zum moralischen Problemgehalt moderner Biotechnologien und medizinischen Fortschritts sowie die Fähigkeit zu einer bewussten, reflektierten, kritisch hinterfragenden und argumentativ fundierten Urteilsfähigkeit“ im Vordergrund. Das Modell wurde unter Vergleich bestehender didaktischer Modelle zur Urteilsbildung aus der Biologie- und Philosophiedidaktik und unter Hinzuziehen philosophischer Grundfertigkeiten konzipiert (Reitschert 2009). Es beschreibt acht Teilkompetenzen (Reitschert, Langlet, Hößle, Mittelsten Scheid & Schlüter, 2007; Reitschert & Hößle, 2007):

- Wahrnehmen und Bewusstmachen der eigenen Einstellung
- Wahrnehmen und Bewusstmachen moralischer Relevanz
- Beurteilen
- Folgenreflexion
- Urteilen
- Ethisches Basiswissen
- Perspektivenwechsel
- Argumentieren

Perspektivenwechsel und Argumentieren werden als grundlegende Fertigkeiten gesehen, die von den anderen Teilkompetenzen nicht zu trennen sind. In Tabelle 3 wird die allgemeine Ausdifferenzierung der Niveaus ethischer Urteilskompetenz angegeben. Weiterhin gibt es für jede einzelne Teilkompetenz eine weitere Spezifizierung der Niveaue Konkretisierung (Reitschert et al., 2007). Im Rahmen qualitativer Untersuchungen (Mittelsten Scheid & Hößle, 2008; Reitschert, 2009) wurden die Teilkompetenzen des Oldenburger Modells zur Bewertungskompetenz in Form von Niveaustufungen weiter ausdifferenziert. Die empirische Ausdifferenzierung erfolgte anhand einzelner bioethischer Themen (u.a. PID), so dass die Ergebnisse nicht ohne weiteres für andere Themen und Kontexte verallgemeinert werden können.

In den NBS ist die Teilkompetenz *Argumentation* dem Kompetenzbereich *Kommunikation* zugeordnet, allerdings spielt das Finden von Gründen für Entscheidungssituationen und damit für die Bewertungskompetenz eine wichtige Rolle. Teilaspekte des Argumentierens werden

Tab. 3: Niveaue Konkretisierung (aus Reitschert et al., 2007)

NIVEAU	NIVEAUE KONKRETISIERUNG
A (Ausgangsvorstellung)	Nicht-Kennen / Nicht-Erkennen / Nicht-Anwenden
I	Kennen / Erkennen / Unreflektierte Anwendung
II	Einfache (Erst-)Begründung / Anwendung
III	Komplexere Begründung / Verstehen der Anwendung
	a) Begründung der Begründung
	b) Erweiterung des Kontextes

daher in beiden Kompetenzbereichen operationalisiert. Auch die Perspektivenübernahme ist den beiden Kompetenzbereichen *Kommunikation* und *Bewertung* zuzuordnen (KMK, 2005a, 2005b, 2005c).

Beide bereits bestehenden Kompetenzmodelle erfassen die Bewertungskompetenz für die naturwissenschaftlichen Fächer nicht vollständig, da sie im Wesentlichen einzelne Aspekte von Bewertungskompetenz im Fach Biologie beschreiben. Sie bieten jedoch eine Grundlage für die Entwicklung eines Modells zur Bewertungskompetenz in den anderen naturwissenschaftlichen Fächern, das zudem eine Erfassung von Bewertungskompetenz im Large-Scale-Assessment ermöglichen soll. Die von Bögeholz (2007) verfolgte Nutzung von Qualität und Quantität von Merkmalen zur Beschreibung der Kompetenzstufen wurde in den dem Projekt ESNaS zugrunde liegenden Kompetenzstrukturmodellen für die Kompetenzbereiche Fachwissen und Erkenntnisgewinnung ebenfalls verwendet. Auf diese Art lassen sich Kompetenzstufen auch im Fachwissen (Kauertz, 2008; Kauertz et al., 2010) und der Erkenntnisgewinnung (Mayer, Grube, & Möller 2008) empirisch adäquat beschreiben.

4 Ausdifferenzierung eines Modells der Bewertungskompetenz für ESNaS

Die Operationalisierung von Kompetenz bei ESNaS erfolgt mit dem Ziel, diese in einem Large-Scale-Assessment am Ende der Sekundarstufe zu nutzen. Um außer-

dem sowohl Kompetenzbereiche als auch Fächer untereinander vergleichen zu können und dadurch Hinweise auf Gemeinsamkeiten und Entwicklungspotenziale zu erhalten, liegen der Operationalisierung folgende grundsätzliche Entscheidungen zu Grunde:

1. Die Operationalisierung berücksichtigt strukturelle Gemeinsamkeiten der drei naturwissenschaftlichen Fächer.
2. In allen Kompetenzbereichen erfolgt die Graduierungen von Kompetenz durch Komplexität und kognitive Prozesse. Grundlegend ist dabei die Annahme, dass sich in jedem Kompetenzbereich spezifische Arten von Komplexität beschreiben und mit kognitiven Prozessen der Informationsentnahme und -verarbeitung verknüpfen lassen (Kauertz et al., 2010). Ergänzend muss die inhaltspezifische Achse des Kompetenzstrukturmodells für jeden Kompetenzbereich adäquat ausdifferenziert werden.
3. Die Kompetenzbereiche sollen empirisch unabhängig voneinander sein, d.h. es wird eine möglichst geringe Korrelation zwischen den Bereichen angestrebt. Dies wird erreicht, indem Inhalte unter neuen Perspektiven und einem kompetenzbereichsspezifischen Fokus bearbeitet werden.
3. Bei der Operationalisierung wird daher versucht ein Grundmodell bestehend aus Kompetenzbereich, Komplexität und kognitiven Prozessen anhand vorhandener Forschungsergebnisse zum jeweiligen Kompetenzbereich spezifisch auszugestalten.

4.1 Das ESNaS-Kompetenzmodell

Für die Ausdifferenzierung der Kompetenzbereiche *Fachwissen* und *Erkenntnisgewinnung* wurde im Projekt ESNaS ein fachübergreifendes Kompetenzstrukturmodell für die Naturwissenschaften entwickelt (Kauertz et al., 2010). Das Modell bildet die Grundlage für die Entwicklung von Testaufgaben, mit denen in einem Large-Scale-Assessment das Erreichen der Bildungsstandards und die Standards selbst deutschlandweit evaluiert werden. Das dreidimensionale Modell (siehe Abb. 3) besteht aus den Dimensionen *Kompetenzbereich*, *Komplexität* und *kognitive Prozesse*, wobei die Dimensionen *kognitive Prozesse* und *Komplexität* als schwierigkeiterzeugende Merkmale nachgewiesen werden konnten, auch wenn sich nicht in allen Studien alle Stufen empirisch tren-

nen ließen und daher teilweise zusammengefasst wurden (Kauertz, 2008; Neumann, 2011; Ropohl, 2010). Durch Konstruktion von Aufgaben unterschiedlicher Komplexität und unterschiedlicher kognitiver Prozesse, die zur Lösung benötigt werden, entsteht ein Test, der unterschiedlich ausgeprägte Kompetenz abbildet. Je nach Kombination von Komplexitätsstufe und kognitivem Prozess sind die Testaufgaben unterschiedlich schwierig. Der gesamte Test bildet damit a priori das gesamte Spektrum der im Modell dargestellten Kompetenzen ab. Auf Basis der anschließend gewonnenen Daten wird wiederum geprüft, ob die Modellannahmen auch für diesen Kompetenzbereich haltbar sind. Für den Kompetenzbereich *Fachwissen* geben die Bildungsstandards bereits Basiskonzepte als weitere Untergliederung vor, für den Kompetenzbereich *Erkennt-*

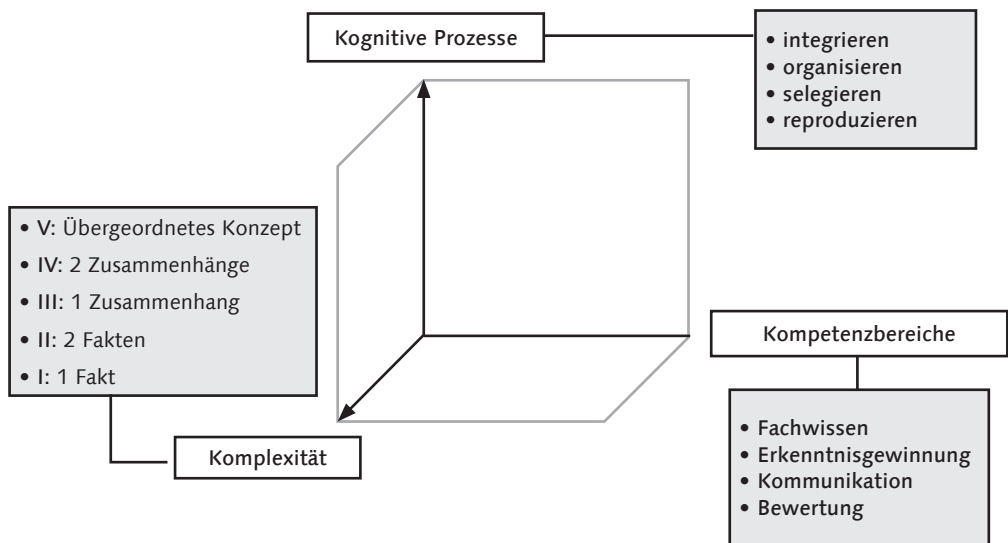


Abb. 3: ESNaS Kompetenzmodell.

nisgewinnung wurden die Kompetenzteilbereiche *Naturwissenschaftliche Untersuchungen*, *Naturwissenschaftliche Modellbildung* und *Wissenschaftstheoretische Reflexion* definiert.

Die fünfstufige Gliederung der Komplexität in 1 Fakt, 2 Fakten, 1 Zusammenhang, 2 Zusammenhänge und übergeordnetes Konzept – wobei diese Bezeichnungsweise im Kompetenzbereich Fachwissen entstanden ist - basiert auf der Anzahl an Elementen und Verknüpfungen zwischen den Elementen (Kauertz et al., 2010; Kauertz, 2008) und erlaubt eine Beschreibung unterschiedlicher Kompetenzniveaus. Ein Fakt kann allgemeiner als eine Entität bezeichnet werden, ein Zusammenhang als zwei verknüpfte Entitäten usw. Was genau unter den einzelnen Entitäten und Verknüpfungen zu verstehen ist, wird für jeden Kompetenzbereich separat ausdifferenziert.

Im Hinblick auf die Modellierung von Kompetenz sind neben den Inhalten und ihrer Komplexität kognitive Prozesse zu beschreiben, die die Anwendung von Wissen auf ein Problem oder zur Lösung einer Fragestellung spezifizieren (Anderson et al., 2001; Haladyna, 2004). Allerdings ist laut Haladyna (2004) keine der bisher verwendeten Taxonomien kognitiver Prozesse ausreichend validiert worden. Die kognitiven Prozesse des Kompetenzstrukturmodells ESNaS sind Theorien der Informationsverarbeitung entlehnt (Mayer, 2001; Weinstein & Mayer, 1986). Sie wurden für die Bearbeitung von Testaufgaben und den Umgang mit aufgabenrelevanten Informationen adaptiert (Walpuski et al., 2010) und in die vier hierarchischen Stufen reproduzieren, selektieren, organi-

sieren und integrieren untergliedert (Kauertz et al., 2010; Walpuski, Kampa, Kauertz & Wellnitz, 2008). Im ESNaS-Modell werden die kognitiven Prozesse für alle Kompetenzbereiche gleich operationalisiert, da in allen Kompetenzbereichen schriftliche Informationen der Aufgabenstellung bei der Lösung der Aufgaben herangezogen werden müssen. Zudem hat sich in den bereits durchgeführten Studien (Ropohl, 2010; Kauertz et al., 2010; Neumann, 2011) gezeigt, dass sich diese Annahme eines Schwierigkeitsanstiegs der Aufgaben vom Reproduzieren zum Integrieren bewährt hat.

4.2 Das ESNaS-Kompetenzmodell zu Bewertung

Die Bewertungskompetenz wird als dritte Komponente (neben Fachwissen und Erkenntnisgewinnung) auf der Dimension *Kompetenzbereich* abgebildet. Um möglichst differenzierte Aussagen über das Kompetenzgefüge der Stichprobe machen zu können, muss die Bewertungskompetenz möglichst unabhängig von den anderen Kompetenzbereichen operationalisiert werden. Bei der Operationalisierung wird diese Rahmenbedingung besonders beachtet. Die Teilkompetenz *Generieren und Reflektieren von Sachinformationen* des Göttinger Modells ist im ESNaS-Modell bereits im Kompetenzbereich *Erkenntnisgewinnung* verortet, etwa durch das Erkennen einer Problemstellung, die Vergegenwärtigung der Wissensbasis, das Modellieren und Reflektieren von Modellen sowie die Verknüpfung ge-

wonnener Erkenntnisse mit bereits geläufigen Konzepten, Modellen und Theorien. Das *Argumentieren* im Modell zur ethischen Urteilskompetenz (Reitschert et al., 2007) ist in den Bildungsstandards mit Ausnahme des „Begründens“ der *Kommunikation* zugeordnet. Diese Teilkompetenzen werden daher – um Doppelzuschreibungen zu vermeiden – im ESNaS-Modell nicht der Bewertungskompetenz zugeordnet.

Da das Modell die Bewertungskompetenz fachübergreifend darstellen soll, muss das Modell unabhängig von fachlichen Themen sein. Dennoch werden wichtige Aspekte des Modells von Bögeholz (2004) und des Modells von Reitschert et al. (2007) als Teilaspekte in dieses naturwissenschaftsspezifische Modell integriert. Da die Operationalisierung des Modells für ein Large-Scale-Assessment erfolgt, müssen alle Aufgaben quantitativ auswertbar sein. Es gibt also keine Möglichkeit zusätzliche qualitative Methoden einzusetzen. Dies ist eine besondere Herausforderung an die Entwicklung der Aufgaben, da in den bisherigen Studien hauptsächlich qualitative Methoden oder eine Kombination aus qualitativen und quantitativen Methoden eingesetzt wurden.

4.2.1 Teilbereiche der Bewertungskompetenz

Die Ausdifferenzierung des Kompetenzbereichs Bewertung stellt den Kern der Operationalisierung dar. Mit Hilfe der schwierigkeitsbestimmenden Dimensionen *kognitive Prozesse* und *Komplexität*

werden im Kompetenzbereich Bewerten Aufgaben von a priori definierter Schwierigkeit konstruiert. Aus den oben genannten Vorarbeiten und den Anforderungen der Bildungsstandards werden zur Beschreibung der Inhaltsdimension die folgenden Kompetenzteilbereiche abgeleitet: *Bewertungskriterien*, *Handlungsoptionen* und *Reflexion*. Die Teilbereiche bilden typische Merkmale eines Bewertungsprozesses ab, sie sind aber nicht als trennscharfe Unterscheidung im testtheoretischen Sinne zu sehen. Vielmehr soll durch die Einordnung von Aufgaben in Teilbereiche eine Vielfalt von Bewertungsprozessen abgebildet werden. Die fachlichen Themengebiete, die in Bewertungsaufgaben verwendet werden, sind fachtypische Themen der Sekundarstufe I. Das für die Bearbeitung der Aufgaben nötige Fachwissen wird im Aufgabenstamm angegeben, um den Einfluss von Vorwissen durch die Erinnerung an bereits Gelerntes möglichst gering zu halten (Eggert & Bögeholz, 2006). Auf Basis der Kompetenzteilbereiche können Aufgaben mit fachspezifischen Inhalten entstehen – um so fachspezifischen Unterschieden in den Standards Rechnung zu tragen – die dennoch durch ein gemeinsames Kompetenzstrukturmodell beschrieben werden können.

Im Teilbereich *Bewertungskriterien* werden Aufgaben entwickelt, bei denen sowohl die Situation als auch die Bewertungskriterien, die in der Aufgabe behandelt werden, im Sinne der präselektionalen Phase erschlossen werden müssen. Insbesondere bedeutet dies, relevante Fakten zur Bewertung, die Bewertungskriterien, bzw. zugehörige Werte und Normen zu

In Nordamerika lagern im Sandboden die zweitgrößten Ölvorkommen der Welt in einem Gemisch aus 83 % Sand, 4 % Wasser und 3 % Lehm, dem sogenannten „Ölsand“. Um dieses Öl zu fördern, müssen Urwälder abgeholzt und Flüsse und Seen stark verschmutzt werden. Dies wird damit gerechtfertigt, dass es zurzeit noch keinen geeigneten Ersatzstoff für das Erdöl gibt. Weiterhin wird durch den Abbau eine strukturschwache Region gefördert, viele Menschen finden Arbeit und Kanada ist unabhängig von ausländischen Erdöllieferungen.

Um Rohöl aus Ölsand zu gewinnen, muss ein großer Aufwand betrieben werden: Um ein Barrel (159 L) Rohöl zu erhalten, müssen zwei Tonnen Ölsand ausgebaggert und in Brechanlagen zerkleinert werden. Das Öl wird dann unter Zusatz von Natriumhydroxid, 950 L Frischwasser und weiteren Lösungsmitteln abgetrennt.

Ein Umweltaktivist und der Sprecher einer Ölsandfirma stehen sich in einer Diskussion gegenüber.

Welche Gründe könnten sie anbringen?

Nenne jeweils einen Grund für den Abbau von Ölsand und einen Grund dagegen.

Dafür:

Dagegen:

Abb. 4: Beispielaufgabe zum Teilbereich Bewertungskriterien.

erkennen oder anzuwenden. Aufgaben zu diesem Teilbereich sind folglich stark sachbezogen, so dass in ihnen der Fachbezug der Bewertungskompetenz besonders deutlich wird.

Im Teilbereich *Handlungsoptionen* stehen in den Aufgaben die Handlungsmöglichkeiten im Vordergrund, die sich aus der beschriebenen Situation ergeben. Der Teilbereich ähnelt der Teilkompetenz *Bewerten, Entscheiden und Reflektieren* im Göttinger Modell und verlangt das Generieren von

Handlungsoptionen, das Bewerten von Handlungsoptionen, Perspektivübernahme und die Folgenabschätzung.

Unter dem Generieren von Handlungsoptionen ist zu verstehen, dass im Sinne der präselektionalen Phase situationsbezogene Optionen genannt werden müssen. Aufgaben zur Bewertung von Handlungsoptionen sind der selektionalen Phase zuzuordnen und erfordern beispielsweise den Vergleich von Handlungsoptionen mit Hilfe von Entscheidungsstrategien.

An einigen Gebäuden findet man Solarzellen. Solarzellen wandeln auftreffende Energie des Sonnenlichtes in elektrische Energie um. Hierbei fallen keine klimaschädlichen Abgase wie Kohlenstoffdioxid und auch keine radioaktiven Abfälle an, die bei anderen Methoden zur Erzeugung von elektrischer Energie entstehen.

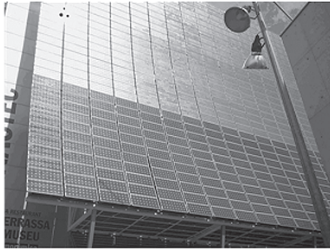


Abbildung: Fassade eines Hauses mit Solarzellen

Eine Firma möchte auf einem Dach der Wolkenstein-Schule eine Solaranlage bauen. Die Schule hat mehrere Gebäude mit verschiedenen großen Dächern. Alle Dächer sind Flachdächer.

Es soll möglichst viel elektrische Energie erzeugt werden. Andererseits möchte die Firma eine möglichst preisgünstige Lösung finden. Die Schülerinnen und Schüler einer Arbeitsgemeinschaft stellen die folgende Tabelle auf, mit deren Hilfe der beste Standort für die Solaranlage gefunden werden soll:

Gebäude	gewonnene Energie	Anschaffungskosten
Hauptgebäude	16.645 kWh/a	149.940 €
Seitenflügel	21.250 kWh/a	107.100 €
Neubau	23.800 kWh/a	166.940 €
Turnhalle	26.775 kWh/a	167.790 €

Bei ihrer Begutachtung der Dächer fällt den Kindern auf, dass an der Hauswand des Seitenflügels mehrere Schwalben ihre Nester gebaut haben und gerade brüten. Durch das Anbringen der Solarzellen würden die Tiere stark gestört werden.

Um zu einer Entscheidung zu kommen, welches Dach ausgewählt werden soll, sollen alle Kriterien berücksichtigt werden. Jedes Gebäude hat bestimmte Nachteile.

Auf welchem Gebäude sollen deiner Meinung nach die Solarzellen angebracht werden und warum?

Nenne das Gebäude und erkläre, welchen Nachteil das Gebäude hat und warum du es trotzdem auswählst.

Begründung:

<div style="border: 1px solid black; width: 95%; height: 95%; background-color: #e0e0e0;"></div>
--

tiven zu identifizieren und zu erschließen. Bei der Folgenreflexion werden Folgen von Entscheidungen für die betroffenen Personen(kreise) antizipiert. Ähnlich wie in dem Modell zur ethischen Urteilskompetenz nach Reitschert & Hößle (2007) werden für das Generieren und Bewerten von Handlungsoptionen Perspektivübernahme, Folgenreflexion, Sachlage und Wertbezug verknüpft.

Der dritte Teilbereich, die *Reflexion*, ist dadurch gekennzeichnet, dass Situationen vorgegeben werden, in denen eine Bewertung bereits vorgenommen wurde. Im Mittelpunkt der Aufgaben steht eine rückblickende Reflexion des Bewertungsprozesses – z. B. kann der Ablauf des Prozesses kritisch reflektiert werden oder die Frage bearbeitet werden, ob die den Bewertenden wichtigen Bewertungskriterien ausreichend berücksichtigt wurden. Damit repräsentiert der Teilbereich hauptsächlich die selektionale Bewertungsphase. In den Aufgaben ist es erforderlich Perspektiven und Folgen einer Entscheidung zu erschließen, diese für Entscheidungen zu berücksichtigen, zu gewichten und ggf. unter der Anwendung von Entscheidungsstrategien

Abb. 5: Beispielaufgabe zum Teilbereich Handlungsoptionen.

Perspektivübernahme bedeutet, betroffene Personen(kreise) bzw. deren Perspek-

tiven, zu gewichten und ggf. unter der Anwendung von Entscheidungsstrategien

Erdkröten gehören zu den Amphibien. Die Tierart ist geschützt. Zur Eiablage (Laichen) suchen die Erdkröten im Frühjahr stehende Gewässer, wie z. B. Teiche, auf.

In einem Feuchtgebiet am Rande einer Ortschaft ist ein neues Wohngebiet entstanden. Beim Bau der Häuser und Straßen wurden einige Laichgewässer der Kröten vernichtet.

Um das Wohngebiet gut erreichbar zu machen, wurde eine Straße gebaut. Diese Straße müssen die Erdkröten überqueren, wenn sie ihre Laichgewässer erreichen wollen. Viele Kröten werden dabei von Autos überfahren und sterben auf der Straße. In einer Versammlung der Naturschützer werden Maßnahmen zur Rettung der Erdkröten besprochen. Dabei werden mögliche Folgen dieser Maßnahmen in Bezug auf die Zahl geretteter Kröten, die Einschränkungen für Anwohner und die Höhe der Kosten dieser Maßnahmen beurteilt.

Im Protokoll wird die folgende Tabelle veröffentlicht:

Maßnahme \ Folgen	Anzahl geretteter Kröten	Einschränkungen für die Anwohner	Höhe der Kosten
Autofahrer aufmerksam machen z. B. durch Aufstellen von Straßenschildern mit Erklärungen	wenige	gering	gering
Geschwindigkeit begrenzen	wenige	teils/teils	gering
Sperrern der Straße während der Wanderungszeit der Kröten	viele	groß	gering
Bau eines Krötentunnels	viele	gering	hoch
Bau eines Zaunes entlang der Autostraße, Kröten regelmäßig einfangen und über die Straße tragen	viele	gering	hoch

Die Anwohnerin Özil ist Naturschützerin. Darum ist ihr der Schutz der Erdkröten wichtig. Gleichzeitig ist sie aber daran interessiert schnell und sicher mit dem Auto zur Arbeit zu gelangen.

Nenne eine Maßnahme aus der Tabelle, die Özil unterstützen würde. Begründe deine Wahl aus der Sicht von Özil.

<p>Maßnahme:</p>
<p>Begründung:</p>

Abb. 6: Beispielaufgabe zum Teilbereich Reflexion.

eine Abwägung vorzunehmen. Bei der Abwägung kommen verschiedene Entscheidungsstrategien zum Einsatz (z.B. kompensatorisch, non-kompensatorisch). Die Entscheidungsstrategien unterschei-

den sich u. a. darin, ob und inwieweit relevante Sachinformationen etc. berücksichtigt werden und ob diese gewichtet und abgewogen werden. In einem Teil der Aufgaben muss zudem eine Evaluation des Bewertungsprozesses und der Entscheidung durchgeführt werden.

4.2.2 Aspekte der Bewertung

In allen drei Teilbereichen kann der Fokus der Aufgaben auf *persönliche, gesellschaftliche* oder *ethische* Aspekte gelegt werden. Die Berücksichtigung dieser Aspekte bei der Aufgabenkonstruktion ermöglicht das Erfassen unterschiedlicher Perspektivübernahmen durch die Probanden. Dabei werden die Aspekte nicht als schwierigkeitsgenerierendes Merkmal angesehen, durch die Verwendung dieser Aspekte soll eine Vielfalt von Bewertungsperspektiven sichergestellt werden.

Bei der Bewertung unter persönlichen Aspekten sollen

Situationen konstruiert werden, in denen für die Schülerinnen und Schüler alltagsbezogene Entscheidungen möglich sind, (z. B. „Du möchtest ein möglichst gesundes Mineralwasser kaufen ...“). Dabei

ist zu beachten, dass eine naturwissenschaftliche Betrachtung der Entscheidungssituation möglich sein muss. Da die persönliche Einstellung der Schülerinnen und Schüler zu einem gegebenen Problem unter Umständen stark variieren kann, ist es wichtig, dass das Ziel in der Aufgabe klar definiert ist. Die Handlungsoptionen sind in der Regel nicht kontrovers sondern alternativ darzustellen (Bsp.: „Wähle ein Mineralwasser aus, das du als Sportgetränk verwenden würdest. Begründe deine Wahl mit den genannten Kriterien.“ Die Kriterien sind im Aufgabenstamm beschrieben). Die richtige Handlung ergibt sich aus der Erreichung eines Ziels (z. B. Gesundheit, Mineralstoffausgleich) mit einer optimalen Strategie. Werden verschiedene Marken von Mineralwässern angegeben, die sich im Preis und im Mineralstoffgehalt unterscheiden, kann je nach Kriterium (Preis, Mineralstoffgehalt), ein anderes Mineralwasser ausgewählt werden. Die Auswahl muss dann jeweils durch ein Abwägen der Kriterien begründet werden. Hier können aus fachlicher Sicht also verschiedene Resultate korrekt sein, je nachdem, wie die Kriterien bei der Lösung der Aufgabe gewichtet wurden. Bei der Entscheidung im Hinblick auf die Richtigkeit der Aufgabe spielt demnach die Passung zwischen Begründung und Resultat eine wesentliche Rolle.

Bei Bewertungen unter gesellschaftlichen Aspekten sollen Bewertende sich in andere Rollen hineinversetzen und dabei Interessenskonflikte identifizieren und mögliche Lösungen finden. Hierzu gehören z. B. Bewertungsprozesse unter ökologi-

schen oder ökonomischen Gesichtspunkten bzw. die Betrachtung gesellschaftlicher Folgen bestimmter Entscheidungen. Typische Aufgaben in diesem Bereich thematisieren unterschiedliche Einschätzungen aufgrund verschiedener Sichtweisen, trotz gleicher Kriterien, z. B. unterscheidet sich bei technischen oder chemischen Anlagen die Sichtweise der dort wohnenden Menschen meist von der des Betreibers. Ein weiterer charakteristischer Aufgabentyp im gesellschaftlichen Bereich ist die Bewertung der tatsächlichen oder vermuteten Auswirkungen menschlichen Handelns auf gesellschaftliches und individuelles Leben (zum Beispiel bei Eingriffen in Ökosysteme). Meist erfolgt diese Bewertung unter unterschiedlichen Perspektiven, zum Beispiel einer naturwissenschaftlichen, ökonomischen, sozialen, individuellen und ökologischen Perspektive. Zentrale Motive sind dabei das Risiko und die Chancen zukünftiger Entwicklung. Kompetente Schülerinnen und Schüler wählen die für eine vorgegebene Perspektive passenden Kriterien aus oder benennen die zu den gegebenen Kriterien passende Perspektive. Die Perspektive kann dabei auch ein Kompromiss zwischen verschiedenen bekannten Positionen sein. Die Bewertung unter ethischen Aspekten umfasst Bewertungssituationen, die mit gesellschaftlichen Werten und Normen verknüpft sind. Sie spielen in allen naturwissenschaftlichen Fächern eine Rolle, sind jedoch insbesondere für das Fach Biologie von großer Bedeutung. Beispielsweise können hier die Bereiche Tierethik, Ökologische Ethik und Medizinethik angesprochen werden. Die Be-

wertungsprozesse sind zusätzlich von der Zuschreibung eines bestimmten (ethischen) Wertes abhängig. Werte sind dabei normative Grundlagen des Handelns, mit denen Objekten Ideen oder Beziehungen zugeordnet werden. Charakteristisch für diesen Teilbereich ist die kongruente Verknüpfung von Handlung und Werten. Die Handlungsoptionen stellen sich oft als Dilemmata dar, d. h. bei jeder möglichen Handlung werden unter Umständen Werte verletzt. Es kann also oft keine „richtige“ Lösung gewählt werden, die allen Werten gerecht wird, Ziel ist es daher gut begründete Entscheidungen zu treffen. Kriterium für die Bewertung einer Handlung ist die Kongruenz von Handlung und Wert oder Norm. Werte sind dabei im Sinne der Ethik beispielsweise Gleichheit, Gerechtigkeit und Wohlergehen. Des Weiteren kann aber auch Objekten ein ästhetischer, ökologischer oder wissenschaftlicher Wert beigemessen werden. Dagegen stellen Normen Regelungen in der Gesellschaft oder in gesellschaftlichen Gruppen dar. Normen verweisen auf den ihnen zugrunde liegenden Wert. Werte stellen die hinter den Normen liegenden Begründungen dar (Beispiel: Norm: Du sollst Organe ohne Ansehen der Person Bedürftigen zuteilen. Wert: Gerechtigkeit). Moralische Normen sind individuelle und gesellschaftliche Vorstellungen, die gemeinhin in einer bestimmten Gemeinschaft als gut oder schlecht angesehen werden (Eggert & Hößle, 2006). Die von Bögeholz gewählte Leitidee „nachhaltige Entwicklung“ stellt beispielsweise eine solche Norm dar. Auch im technischen Bereich werden Entschei-

dungen durch Normen beeinflusst, indem z. B. Grenzwerte u. ä. durch verbindliche Regelungen vorgegeben werden (DIN-Normen usw.).

4.2.3 Komplexität

Die Komplexität der Entscheidungssituationen hängt im Wesentlichen von der Anzahl der zu berücksichtigenden Kriterien und den Beziehungen zwischen ihnen ab (Poschmann et al., 1998; Bögeholz et al., 2004; Eggert & Bögeholz, 2006). Die Aufgabenkomplexität kann also auch für Entscheidungssituationen von der Anzahl der für die Lösung der Aufgabe zu verarbeitenden Informationen und von den Zusammenhängen zwischen den Informationseinheiten definiert werden. Das Merkmal der Komplexität, das für die Kompetenzbereiche *Fachwissen* und *Erkenntnisgewinnung* bereits verwendet wurde, wird für den Kompetenzbereich *Bewerten* für die einzelnen Kompetenzteilbereiche ausdifferenziert. Für den Teilbereich *Bewertungskriterien* wird die Ausdifferenzierung der Komplexität im Folgenden beispielhaft dargestellt.

- Niveau I: Ein Bewertungskriterium
- Niveau II: Zwei Bewertungskriterien
- Niveau III: Ein Zusammenhang zwischen zwei Bewertungskriterien
- Niveau IV: Zwei Zusammenhänge zwischen zwei oder mehr Bewertungskriterien
- Niveau V: Übergeordnetes Konzept Anwendung/Analyse eines Bewertungsprozesses

Tab. 4: Komplexitätsniveau je Teilbereich von Bewertungskompetenz

		Bewertungskriterien	Handlungsoptionen	Reflexion
Komplexität	V	Metaverstehen, z.B.	Metaverstehen, z.B.	Metaverstehen, z.B.
		• Differenzierung von deskriptiven und normativen Aussagen	• Verständnis der Mehrperspektivität von Handlungen (inkl. Perspektivübernahme)	• Reflexion bzw. Evaluation eines Bewertungsprozesses unter Berücksichtigung von Zweck, Wirkung und Grenzen
		• Wert-Norm-Relation	• Konzept der Folgenabschätzung	• Einsicht in Begründungsstrukturen sowie Gütekriterien von Begründungen
		• Explizites Verständnis von Dilemmata	• Folgen von Handlungen auf unterschiedlichen Ebenen (Individuum, Gesellschaft)	
	IV	• Zwei Beziehungen zwischen drei (alternativen, additiven, gegensätzlichen) Bewertungskriterien oder Werten r/s/o/i	• Zwei Beziehungen von Handlungsoptionen (und ihren Folgen/) r/s/o/i	• Zwei Beziehungen von Begründung und Handlungsoption (Pro- und Kontraargumente) r/s/o/i unter Bezugnahme auf
			• in Bezug auf un- oder mittelbar betroffene Personen und in Bezug auf die Gesellschaft	• zugrunde liegende Bewertungskriterien
			• kurzfristig und langfristig	• mögliche Folgen
				• unterschiedliche Perspektiven
	III	• Eine Beziehung zwischen zwei Bewertungskriterien bzw. Sachlage und Bewertungskriterium r/s/o/i	• Eine Beziehung von zwei Handlungsoptionen (und ihren Folgen/) r/s/o/i	• Eine Beziehung von Begründung und Handlungsoption r/s/o/i unter Bezugnahme auf
			• in Bezug auf un- oder mittelbar betroffene Personen	• zugrunde liegende Bewertungskriterien
			• kurzfristig und langfristig	• mögliche Folgen
				• unterschiedliche Perspektiven
	II	• Zwei Bewertungskriterien für einen Sachverhalt r/s/o	• Zwei Handlungsoptionen für eine Situation r/s/o	• Zwei Begründungen für eine Bewertung/Alternative/Folge r/s/o
	I	• Ein Bewertungskriterium für einen Sachverhalt r/s	• Eine Handlungsoption für eine Situation r/s	• Eine Begründung/Alternative/Folge für eine Bewertung r/s

r = reproduzieren, s = selektieren, o = organisieren, i = integrieren

Analog zu der Strukturierung der Komplexität für den Teilbereich *Bewertungskriterien* werden die Teilbereiche *Handlungsoptionen* und *Reflexion* strukturiert (siehe Tabelle 3).

Vergleichbar zum Göttinger Kompetenzmodell wird die Stufung anhand der Anzahl der Kriterien und der Anwendung einer Entscheidungsstrategie vorgenommen.

Die Komplexität im ESNaS-Modell stimmt im Wesentlichen mit der Niveau-konkretisierung im Modell nach Reitschert et al. (2007) überein. So steigt zum einen die Komplexität der Niveaus in beiden Modellen von einer einfachen zu einer komplexen Begründung an. Zum anderen gleichen die Niveaustufen in der Ausdifferenzierung den Teilkompetenzen nach Reitschert et al. (2007) den Komplexitätsstufen nach dem ESNaS-Modell.

4.2.4 Kognitive Prozesse

Entscheidungssituationen erfordern, je nach Art der gegebenen Situation bzw. aufgrund der Erfahrungen des Entscheiders, einen unterschiedlichen kognitiven Aufwand (Jungermann et al., 2005, S. 18). Auch die Niveaus nach Reitschert beinhalten unterschiedliche kognitive Prozesse (Kennen, Anwenden, Verstehen), die den kognitiven Prozessen im ESNaS-Modell ähneln. Die kognitiven Prozesse beziehen sich auch in diesem Kompetenzbereich auf die Verarbeitung der in den Aufgaben vorgegebenen Information und nicht auf das Vorwissen.

Beim *Reproduzieren* sind Informationen aus dem Aufgaben- bzw. Itemstamm wiederzugeben, bzw. zu nennen. Inhaltlich ist also bei geschlossenen Aufgaben die richtige Antwortmöglichkeit direkt aus der Vorgabe im Aufgaben- bzw. Itemstamm ableitbar. In Bezug auf den Kompetenzbereich *Bewertung* können zum Beispiel vorgegebene Bewertungskriterien oder Handlungsoptionen sowie Entscheidungsstrategien aus dem Stamm wiedergegeben werden.

Auf der zweiten Stufe der kognitiven Prozesse müssen die Schülerinnen und Schüler Informationen aus dem Aufgaben- bzw. Itemstamm auswählen, so dass es sich hierbei um das *Selektieren* handelt. Mögliche Aufgaben beinhalten eine Vorgabe an Bewertungskriterien, bei denen spezielle Kriterien nach verschiedenen Gesichtspunkten aus mehreren Möglichkeiten ausgesucht werden. Hierbei ist nicht im klassischen Sinn zwischen richtig und falsch zu unterscheiden sondern zu prüfen, ob die Entscheidung anhand der vorgegebenen Kriterien/Perspektive getroffen wurde.

Die dritte Stufe kognitiver Prozesse, das *Organisieren*, beinhaltet das Ordnen bzw. Strukturieren von Informationen aus dem Aufgaben- bzw. Itemstamm. Die für die richtige Antwort nötigen Informationen sind im Aufgaben- oder Itemstamm enthalten, müssen jedoch von den Schülerinnen und Schülern sinnvoll in Beziehung gesetzt oder nach bestimmten Kriterien angeordnet werden. Denkbar sind zeitliche und hierarchische Abfolgen (z.B. Bewertungsabläufe, Bewertungshierarchien, Entscheidungsstrategien), Strukturierungen nach Wichtigkeit oder Wirksamkeit einzelner Kriterien, Zuordnungen, Bildung von Kategorien u. v. m. Dabei müssen die Regeln für die Kategorisierung oder Klassifizierung im Itemstamm genannt werden. Die richtige Antwort ist immer das Ergebnis eines Organisierens nach der genannten Regel.

In dem hier vorgeschlagenen Modell wird die höchste Stufe der kognitiven Prozesse durch das *Integrieren* beschrieben. Unter anderem sollen dabei zusätzliche Kri-

terien in bestehende Informationen aus dem Aufgaben- bzw. Itemstamm eingliedert oder die Informationen aus dem Aufgaben- bzw. Itemstamm aus verschiedenen Perspektiven angewendet bzw. interpretiert werden. Zur Beantwortung der Aufgabe ist die Perspektive zu wechseln und ein weiteres Kriterium anzuwenden. In der Aufgabe und im Itemstamm muss zunächst die Situation erfasst und die dort beschriebenen inhaltlichen Zusammenhänge müssen durch Hinzuziehen eines zusätzlichen Kriteriums aus verschiedenen Perspektiven erneut bewertet werden. Verschiedene Informationen müssen also variabel und situationsbezogen verwendet werden, um zu einer korrekten Antwort zu kommen.

Die Dimensionen Komplexität und kognitive Prozesse werden in einer Matrix miteinander gekreuzt, so dass 20 verschiedene Kombinationen der beiden Dimensionen entstehen (Kauertz et al., 2010). Die Anzahl der Kombinationen wird dadurch limitiert, dass nicht jede Kombination operationalisiert werden kann. So lassen sich auf der untersten Komplexitätsstufe „ein Bewertungskriterium“ nur die kognitiven Prozesse Reproduzieren und Selektieren beschreiben, da beim Organisieren (z. B. für eine Hierarchisierung) mindestens zwei Bewertungskriterien nötig sind. Daher wird für diese Kombination auf die Konstruktion von Aufgaben verzichtet. Ähnliches gilt für „zwei Bewertungskriterien integrieren“. Erst Zusammenhänge lassen sich sinnvoll aus verschiedenen Perspektiven bewerten. Außerdem sind die Zellen „Anwendung/Analyse eines Bewertungsprozesses reproduzieren bzw. se-

legieren“ zwar theoretisch möglich, aber schwer zu konstruieren, da der Aufgabenstamm meist sehr umfangreich wird.

5 Anforderungen an die Aufgabentypen

Zu dem oben beschriebenen theoretisch hergeleiteten Modell werden Aufgaben mit verschiedenen Aufgabenformaten konstruiert, die nach Möglichkeit die gesamte Schwierigkeitsskala abbilden. Als geschlossene Antwortformate werden Multiple-Choice (single select) mit vier Antwortalternativen und einer richtigen Antwort angewendet und ein komplexes Format bei dem jede Antwortmöglichkeit mit z. B. „stimmt/stimmt nicht“ oder „ja/nein“ bewertet werden muss. Das halboffene Antwortformat beinhaltet Kurzantworten, wie z. B. bei Lückentexten und anderen Wort- oder Zeichenergänzungen. Aufgaben mit einem offenen Format, bei denen also eine erweiterte Antwort erwartet wird, sind besonders geeignet für Aufgaben auf höheren Stufen der Komplexität und der kognitiven Prozesse, weil es hier besonders schwierig ist, adäquate Distraktoren zu konstruieren, die noch eine angemessene Kürze haben. Für Antworten auf Faktenniveau, bei denen oft nur ein Begriff genannt werden muss, eignen sich aus Gründen der Auswerteökonomie MC-Aufgaben besonders. Allerdings ist in beiden Fällen zu berücksichtigen, dass eine Konfundierung von Aufgabenschwierigkeit und Aufgabenformat wahrscheinlich ist (Walpuski & Ropohl, 2011). Aus diesem Grund werden auf allen Schwierig-

keitsstufen alle Formate eingesetzt, um diese bestimmen zu können. Die Aufgaben bestehen immer aus einem kurzen Aufgabenstamm (Einführungstext), der den Schülern und Schülerinnen die Bewertungssituation vermitteln soll. Dieser sollte einen Sinn stiftenden Kontext (Häußler & Lind, 1998) beinhalten. Anschließend folgt ein Itemstamm, in dem auch die fachlichen Informationen dargestellt werden. Das benötigte Fachwissen für die Beantwortung der Fragen ist im Aufgabenstamm bzw. im Itemstamm vorgegeben, um eine größere Testfairness zwischen den verschiedenen Bundesländern und den daraus resultierenden verschiedenen Curricula zu gewährleisten und die Bewertungskompetenz möglichst unabhängig vom Kompetenzbereich *Fachwissen* zu erfassen. Um den Einfluss des Lesens möglichst gering zu halten, sind die Texte kurz und verständlich zu konstruieren. Für alle Aufgaben wird ein *scoring guide* erstellt, der insbesondere die Auswerteobjektivität bei den offenen Aufgaben sicherstellen soll. Offene Aufgaben werden zu diesem Zweck dichotomisiert und es werden Kriterien für die Entscheidung *Punkt* oder *kein Punkt* definiert. Diese Kriterien werden auf Basis von Schülerantworten aus Präpilotierungen um Beispiele für richtige und falsche Lösungen erweitert. Ziel ist es, dass 5 unabhängige Kodierer in 95 % aller Entscheidungen zu einem einstimmigen Urteil kommen. Um dieses Ziel zu erreichen, werden die *scoring guides* in der Regel in mehreren Kodiererschulungen überarbeitet.

6 Zusammenfassung und Ausblick

Basierend auf Vorarbeiten zur Bewertungskompetenz in verschiedenen Fächern und auf Ergebnissen zur Operationalisierung der Kompetenzbereiche *Fachwissen* und *Erkenntnisgewinnung* konnte ein theoretisch begründetes Kompetenzstrukturmodell entwickelt werden, das die Aufgabenschwierigkeit von Testaufgaben im Large-Scale-Assessment a priori vorhersagen soll. Gleichzeitig wird eine inhaltliche Ausdifferenzierung des Kompetenzbereiches in Kompetenzteilbereiche vorgenommen. Schülerinnen und Schüler müssen sich in den Testaufgaben mit *Bewertungskriterien*, *Handlungsoptionen* und der *Reflexion* des Bewertungsprozesses auseinandersetzen und dabei unter persönlichen, gesellschaftlichen und ethischen Aspekten Bewertungen vornehmen oder beurteilen. Das vorgestellte Modell muss sich in einem nächsten Schritt in der empirischen Prüfung bewähren, ferner ist zu klären, welchen Einfluss weitere, im Modell nicht beschriebene externe Faktoren (wie z. B. das Fachwissen) auf die Bewertungskompetenz haben können.

Literatur

- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R. et al. (2001). *A Taxonomy for Learning, Teaching and Assessing: a Revision of Bloom's Taxonomy of Educational Objectives*. New York [u.a.]: Addison Wesley Longman.
- Barkmann, J. & Bögeholz, S. (2003). Kompetent gestalten, wenn es komplexer wird. „21“ – *Bildung für eine nachhaltige Entwicklung*, 3, 49–52.
- Betsch, T. & Haberstroh, S. (2005). *Current Research on Routine Decision Making: Advances and Prospects, The Routines of Decision Making*. Mahwah, NJ: Erlbaum Associates.
- Betsch, T., Haberstroh, S., & Höhle, C. (2002). Explaining Routinized Decision Making. *Theory & Psychology*, 12(4), 453–488.
- Bögeholz, S. (2000). Vorversion des Artikels: Bögeholz (2000) Naturerfahrung: Ein Baustein der Bildung für Nachhaltige Entwicklung – Natur erleben und gestalten. *Politische Ökologie (Sonderheft 12)*, 17–18.
- Bögeholz, S. (2007). Bewertungskompetenz für systematisches Entscheiden in komplexen Gestaltungssituationen Nachhaltiger Entwicklung. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung Ein Handbuch für Lehramtsstudenten und Doktoranden* (S. 209–220). Berlin Heidelberg New York: Springer.
- Bögeholz, S., Höfle, C., Langlet, J., Sander, E., & Schlüter, K. (2004). Bewerten – Urteilen – Entscheiden im biologischen Kontext: Modelle in der Biologiedidaktik. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 89–115.
- Bybee, R. (1997). Toward an Understanding of Scientific Literacy. In W. Gräber & C. Bolte (Hrsg.), *Scientific Literacy* (S. 37–68). Kiel.
- Chang Rundgren, S.-N. & Rundgren, C.-J. (2010). SEE-SEP: From a separate to a holistic view of socioscientific issues. *Asia-Pacific Forum on Science Learning and Teaching*, 11(1), Article 2.
- Dietrich, J. (2002). Moralpädagogik. In M. Düwell, C. Hübenthal & M. H. Werner (Hrsg.), *Handbuch Ethik* (S. 527–533). Stuttgart / Weimar: Verlag J.B. Metzler.
- Eggert, S. & Bögeholz, S. (2006). Göttinger Modell der Bewertungskompetenz – Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ für Gestaltungsaufgaben Nachhaltiger Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 177–197.
- Eggert, S. & Höfle, C. (2006). Bewertungskompetenz im Biologieunterricht Ein Überblick. *Praxis der Naturwissenschaften – Biologie in der Schule*, 1/55, 1–9.
- Eilks, I., Feierabend, T., Höfle, C., Höttecke, D., Menthe, J., Mrochen, M. et al. (2011). Bewerten lernen und Klimawandel in vier Fächern. *MNU*, 64(1), 7–10.
- Große, F. & Bögeholz, S. (2003). Förderung der Bewertungskompetenz von Schülerinnen und Schülern der Sekundarstufe II am Beispiel der Streuobstwiese. *Erkenntnisweg Biologiedidaktik*, 103–119.
- Grube, C., Mayer, J. (2010). Wissenschaftsmethodische Kompetenzen in der Sekundarstufe I: eine Untersuchung zur Entwicklung des wissenschaftlichen Denkens. In: Harms, U. (Hrsg.) *Lehr- und Lernforschung in der Biologiedidaktik (Band 4), Heterogenität erfassen – individuell fördern im Biologieunterricht* (S.155–168). Innsbruck: Studien Verlag.
- Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items* (3. Auflage). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Hamann, M. (2004). Kompetenzentwicklungsmodelle. *MNU*, 57(4), 196–203.
- Härtig, H. (2010). *Sachstrukturen von Physikschulbüchern als Grundlage zur Bestimmung der Inhaltsvalidität eines Tests*. Berlin: Logos.
- Häußler, P. & Lind, G. (1998). Weiterentwicklung der Aufgabenkultur im mathematisch-naturwissenschaftlichen Unterricht. BLK-Programmförderung „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“. Erläuterungen zu Modul 1 mit Beispielen für den Physikunterricht, IPN. Verfügbar unter: www.learnline.nrw.de/angebote/sinus/zentral/grundlagen/module/modul1.doc. [25.10.2010].

- Hogarth, R. M. (1980). *Judgement and choice: The psychology of decision*. A Wiley-interscience publication. Chichester: Wiley. Verfügbar unter: <http://www.gbv.de/dms/bowker/toc/9780471914792.pdf> [25.10.2010].
- Jungermann, H., Pfister, H. R., & Fischer, K. (2005). *Die Psychologie der Entscheidung*. Heidelberg, Berlin: Spektrum Akademischer Verlag.
- Kauertz, A. (2008). *Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben*. Berlin: Logos Verlag.
- Kauertz, A., Fischer, H. E., Mayer, J., Sumfleth, E., & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 135–153.
- KMK-Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2005a). *Bildungsstandards im Fach Biologie für den Mittleren Bildungsabschluss*. München, Neuwied: Luchterhand.
- KMK-Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2005b). *Bildungsstandards im Fach Chemie für den Mittleren Bildungsabschluss*. München, Neuwied: Luchterhand.
- KMK-Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2005c). *Bildungsstandards im Fach Physik für den Mittleren Bildungsabschluss*. München, Neuwied: Luchterhand.
- Kortland, J. (2003). *Designing and validating a didactical structure for a problem-posing approach to teaching decision making about the waste issue*. ESERA-Konferenz 2003, Noordwijkerhout. Verfügbar unter: http://www.phys.uu.nl/~kortland/art_esera-03.pdf [25.10.2010].
- Labudde, P. (2007). Naturwissenschaftliche Bildung: Quo vadis? In P. Labudde (Hrsg.), *Bildungsstandards am Gymnasium. Korsett oder Katalysator?* (S. 283–291). Bern: hep.
- Labudde, P. & Adamina, M. (2008). HarmoS Naturwissenschaften: Impulse für den naturwissenschaftlichen Unterricht von morgen. *Beiträge zur Lehrerbildung*, 26(3), 351–360.
- Lee, Y. C. & Grace, M. (2010). Students' reasoning processes in making decisions about an authentic, local socio-scientific issue: bat conservation. *Journal of Biological Education*, 44(4), 156–165.
- Mandry, C. (2005). Das Ethisch-philosophische Grundlagenstudium. In M. Maring (Hrsg.), *Ethisch-philosophisches Grundlagenstudium* (S. 3–14). Münster: Lit Verlag.
- Mayer, J., Grube, C., & Möller, A. (2008). Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung. In U. Harms & A. Sandmann (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik* (Band 3, S. 63–79). Innsbruck: Studien Verlag.
- Mayer, J., Harms, U., Hammann, M., Bayrhuber, H., & Kattmann, U. (2004). Kerncurriculum Biologie der gymnasialen Oberstufe. *MNU*, 57(3), 166–173.
- Mayer, R. E. (2001). A Cognitive Theory of Multimedia Learning. In R. E. Mayer (Hrsg.), *Multimedia Learning* (S. 41–62). Cambridge: Cambridge University Press.
- 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 (Univ., Diss.--Kiel, 2006.).
- Metzger, S. & Labudde, P. (2007). HarmoS Naturwissenschaften+. Bildungsstandards für die Schweiz. *Praxis der Naturwissenschaften – Physik in der Schule*, 56(6), 14–18.
- Mittelsten Scheid, N. (2008). *Niveaus von Bewertungskompetenz. Eine empirische Studie im Rahmen des Projekts „Biologie im Kontext“ (Studien zur Kontextorientierung im naturwissenschaftlichen Unterricht)*. Tönning: Der Andere Verlag
- Mittelsten Scheid, N. & Hößle, C. (2008). Bewerten im Biologieunterricht: Niveaus von Bewertungskompetenz. *Erkenntnisweg Biologiedidaktik*, 6, 87–104.
- Neumann, I. (2011). *Beyond Physics Content Knowledge. Modeling Competence Regarding Nature of Scientific Inquiry and Nature of Scientific Knowledge*. Berlin: Logos.
- Nicolis, G. & Prigogine, I. (1989). *Exploring complexity: An introduction*. New York, Freeman.

- Poschmann, C., Riebenstahl, C., & Schmidt-Kallert, E. (1998). *Umweltplanung und -bewertung*. Gotha: Justus Perthes Verlag.
- Ratcliffe, M. (1997). Pupil decision-making about socio-scientific issues within the science curriculum. *International Journal of Science Education*, 19(2), 167–182.
- Reitschert, K. & Hößle, C. (2007). Wie Schüler ethisch bewerten. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 125–142.
- Reitschert, K., Langlet, J., Hößle, C., Mittelsten Scheid, N., & Schlüter, K. (2007). Dimensionen Ethischer Urteilskompetenz. *MNU*, 60(1), 43–51.
- Reitschert, K. (2009). *Ethisches Bewerten im Biologieunterricht: Eine qualitative Untersuchung zur Strukturierung und Ausdifferenzierung von Bewertungskompetenz in bioethischen Sachverhalten bei Schülern der Sekundarstufe I*. Hamburg: Kovač.
- Ropohl, M. (2010). *Modellierung von Schülerkompetenzen im Basiskonzept Chemische Reaktion: Entwicklung und Analyse von Testaufgaben*. Berlin: Logos.
- Rost, J. (2002). *Umweltbildung – Bildung für eine nachhaltige Entwicklung: Was macht den Unterschied*. Verfügbar unter: <http://satgeomuenchen.de/geomuc/infos/umweltbildung.pdf> [25.10.2010].
- Schecker, H. & Höttecke, D. (2007). „Bewertung“ in den Bildungsstandards Physik. Aufgaben zum Kompetenzbereich „Bewertung“. *Unterricht Physik*, 18(97), 29-37.
- Schecker, H. & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 45–66.
- Sadler, T. D., Barab, S. A. & Scott, B. (2007). What Do Students Gain by Engaging in Socioscientific Inquiry. *Research in Science Education*, 37, 371–391.
- Sadler, T. D. (2004). Informal Reasoning Regarding Socio Scientific Issues: A Critical Review of Research. *Journal of Research in Science Teaching*, 41, 513–536.
- Sadler, T. D. & Zeidler, D. L. (2004). The Morality of Socioscientific Issues: Construal and Resolution of Genetic Engineering Dilemmas. *Science Education*, 88(1), 4–27.
- Walpuski, M., Kampa, N., Kauertz, A., & Wellnitz, N. (2008). Evaluation der Bildungsstandards in den Naturwissenschaften. *MNU*, 61(6), 323–326.
- Walpuski, M., Kauertz, A., Kampa, N., Fischer, H. E., Mayer, J., Sumfleth, E. et al. (2010). ESNaS – Evaluation der Standards für die Naturwissenschaften in der Sekundarstufe I. In A. Gehrmann, U. Hericks, & M. Lüders (Hrsg.), *Bildungsstandards und Kompetenzmodelle – Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht* (S. 171–184). Bad Heilbrunn: Julius Klinkhardt.
- Walpuski, M. & Ropohl, M. (2011). Einfluss des Testaufgabendesigns auf Schülerleistungen in Kompetenztests. *Naturwissenschaften im Unterricht Chemie*, 22(124/125), 82–86.
- Weinstein, C. E. & Mayer, R. E. (1986). The Teaching of Learning. In M. C. Wittrock (Hrsg.), *Handbook of research on teaching* (3. Auflage, S. 315–327). New York: MacMillan Reference Books.
- Wilson, M. & Sloane, K. (2000). From Principles to Practice: An Embedded Assessment System. *Applied Measurement in Education*, 13(2), 181–208.

KONTAKT

Prof. Dr. Maik Walpuski
Universität Duisburg-Essen
Schützenbahn 70
45127 Essen
maik.walpuski@uni-due.de

AUTORENINFORMATION

Dr. rer. nat. Julia Hostenbach
promovierte in den Arbeitsgruppen
Walpuski und Sumfleth zum Thema
„Bewertungskompetenz im Chemieunter-
richt“ an der Universität Duisburg-Essen.

Dr. rer. nat. Hans Ernst Fischer
ist Professor für Didaktik der Physik in
der Fakultät für Physik der Universität
Duisburg-Essen und Sprecher der DFG-
Forschergruppe naturwissenschaftlicher
Unterricht. Seine Forschungsschwer-
punkte liegen im Bereich der fachspe-
zifischen Unterrichtsforschung, der
fachspezifischen Analysen zur Lehrer-
professionalisierung unter Berücksich-
tigung von Lernvoraussetzungen und
Lernergebnissen.

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

Dr. rer. nat. Jürgen Mayer
ist Professor für Didaktik der Biologie
an der Universität Kassel. Seine For-
schungsschwerpunkte sind die Kompe-
tenzmodellierung im Biologieunterricht,
die Entwicklung und Analyse von Ler-
numgebungen, insbesondere des
Forschenden Lernens, sowie Lehrerpro-
fessionalisierung.

Dr. rer. nat. Elke Sumfleth
ist Professorin für Didaktik der Chemie
an der Universität Duisburg-Essen und
Sprecherin des DFG-Graduiertenkollegs
„Naturwissenschaftlicher Unterricht“ an
der Universität Duisburg-Essen. Ihr For-
schungsschwerpunkt liegt in der empi-
rischen Lehr-Lern-Forschung in Chemie.

Dr. rer. nat. Maik Walpuski
ist Professor für Didaktik der Chemie
an der Universität Duisburg-Essen und
Mitglied der DFG-Forschergruppe Na-
turwissenschaftlicher Unterricht. Seine
Forschungsschwerpunkte sind die Kom-
petenzmodellierung und Kompetenz-
messung für das Fach Chemie sowie die
prozessanalytische Untersuchung des
Chemieunterrichts, insbesondere in Ex-
perimentierphasen.