

Hostenbach, Julia; Walpuski, Maik

Untersuchung der Einflussfaktoren auf die Bewertungskompetenz im Fach Chemie

Zeitschrift für Didaktik der Naturwissenschaften : ZfDN 19 (2013), S. 129-157



Quellenangabe/ Reference:

Hostenbach, Julia; Walpuski, Maik: Untersuchung der Einflussfaktoren auf die Bewertungskompetenz im Fach Chemie - In: Zeitschrift für Didaktik der Naturwissenschaften : ZfDN 19 (2013), S. 129-157 - URN: urn:nbn:de:0111-pedocs-317201 - DOI: 10.25656/01:31720

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-317201>

<https://doi.org/10.25656/01:31720>

in Kooperation mit / in cooperation with:



IPN

Leibniz-Institut für die Pädagogik der
Naturwissenschaften und Mathematik

<https://www.leibniz-ipn.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Informationszentrum (IZ) Bildung

E-Mail: pedocs@dipf.de

Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

JULIA HOSTENBACH UND MAIK WALPUSKI

Untersuchung der Einflussfaktoren auf die Bewertungskompetenz im Fach Chemie

Factors influencing evaluation and decision making in chemistry education

ZUSAMMENFASSUNG

Bewertungskompetenz stellt neben *Fachwissen*, *Erkenntnisgewinnung* und *Kommunikation* einen gleichberechtigten Kompetenzbereich in den Bildungsstandards dar. Die Schülerkompetenz in diesem Bereich soll, wie auch in den anderen Kompetenzbereichen (Kauertz, Fischer, Mayer, Sumfleth & Walpuski, 2010), mit Hilfe eines Kompetenzstrukturmodells messbar gemacht werden. Nachdem in einem vorangegangenen Artikel (Hostenbach et al., 2011) die Operationalisierung und Messung der Bewertungskompetenz im Projekt ESNaS beschrieben wurde, beschäftigt sich dieser Artikel mit der Überprüfung eines Modells bezüglich des Wirkgefüges von Einflussfaktoren auf die Bewertungskompetenz. Dazu werden die Bewertungskompetenz und mögliche Einflussfaktoren auf diese mit paper-pencil-Tests erhoben und die Zusammenhänge mit Korrelations- und Regressionsanalysen untersucht. Die Ergebnisse zeigen, dass die außerfachliche Anwendung von Bewertungsstrategien, der Umgang mit Fachwissen und die Einschätzung von Datenqualität die stärksten Prädiktoren für die Bewertungskompetenz im Fach Chemie sind. Darüber hinaus konnte gezeigt werden, dass Aufgaben zur Bewertungskompetenz in den Fächern Biologie, Chemie und in alltäglichen Situationen am besten durch ein dreidimensionales Raschmodell abgebildet werden.

Schlagwörter: Standards, Bewertungskompetenz, Kompetenzmodell, Evaluation, Einflussfaktoren

ABSTRACT

The *evaluation and judgement competence* represents one area of competence in the educational standards besides *content knowledge*, *acquisition of knowledge* and *communication*. Students' competencies in this area of competence are to be assessed similar to the other areas of competence (content knowledge and acquisition of knowledge Kauertz et al., 2010) by using a model of competence. As a preceding article (Hostenbach et al., 2011) described the operationalisation of the evaluation and judgement competence in the ESNaS-project, the model of influential aspects on the evaluation and judgement competence is explained in this article. For this purpose the influential aspects and the evaluation and judgement competence in chemistry are investigated by using paper-pencil tests. Relations between students' parameters are calculated by using correlations and regressions. The results show that evaluation and judgement competence is influenced by the application of decision making strategies, the application of content

knowledge and the estimation of data quality. The test items from biology, chemistry and everyday life fit best to a three-dimensional Rasch model.

Keywords: standards, evaluation and judgement competence, model of competence, evaluation, influential aspects

Einleitung

Ausgangspunkt für die vorliegende Studie sind die Bildungsstandards für die naturwissenschaftlichen Fächer in der Sekundarstufe I. Die Bildungsstandards definieren vier verschiedene Kompetenzbereiche über die drei naturwissenschaftlichen Fächer hinweg, die in einer bundesweiten Studie evaluiert werden. Innerhalb des Projekts ESNaS (Evaluation der Standards in den Naturwissenschaften für die Sekundarstufe I), in welchem bereits Aufgaben zur Überprüfung der Kompetenzbereiche *Fachwissen* und *Erkenntnisgewinnung* entwickelt und pilotiert wurden, erfolgt im Moment die Pilotierung der Aufgaben zum Kompetenzbereich *Bewertung*, die in einem Ländervergleich 2018 eingesetzt werden sollen.

Der vorliegende Beitrag stellt die Ergebnisse einer Studie vor, die aufbauend auf das ESNaS Kompetenzstrukturmodell für die Bewertungskompetenz durchgeführt wurde. Ziel der Studie ist es, Einflussfaktoren auf die Bewertungskompetenz empirisch abgesichert benennen zu können. Die Operationalisierung der Bewertungskompetenz wird zu diesem Zweck analog zum ESNaS-Projekt vorgenommen. Die

theoretischen Grundlagen und die detaillierte Herleitung der Operationalisierung der Bewertungskompetenz im Rahmen des ESNaS-Projekts wurden in dieser Zeitschrift bereits explizit vorgestellt (Hostenbach et al., 2011) und werden daher hier nur zusammenfassend erläutert.

Bewertungskompetenz

Im Rahmen der Bildungsstandards sollen kompetente Schülerinnen und Schüler unter dem Aspekt des Kompetenzbereiches *Bewertung* „chemische Sachverhalte in verschiedenen Kontexten erkennen und bewerten“ können (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2005a, S. 7). Die Standards für die naturwissenschaftlichen Fächer beschreiben, dass Schülerinnen und Schüler sich durch das Erwerben der Bewertungskompetenz am gesellschaftlichen Diskurs beteiligen und dabei verschiedene Perspektiven übernehmen können. Schülerinnen und Schüler sollen ihre Wertentscheidungen in Probleme einfließen lassen sowie Entscheidungen sachgerecht, selbstbestimmt und verantwortungsbewusst treffen können. Im Rahmen der Bewertungskompetenz sollen durch Fachkennt-

nisse Probleme erfasst werden und dabei in Problemlösesituationen durch Abwägen eine Entscheidung getroffen werden. Die vorliegende Studie legt die Definition der Bildungsstandards zugrunde. Dabei werden die Vorgaben aus den NBS für Biologie, Chemie und Physik zusammengeführt und um einzelne Aspekte ergänzt. Im Detail ist das zugrundeliegende Verständnis von Bewertungskompetenz bei Hostenbach et al. (2011) erläutert. Die Schülerinnen und Schülern sollen folglich Probleme in realen Situationen erfassen, Interessenkonflikte ausmachen, mögliche Lösungen erwägen sowie deren Konsequenzen diskutieren. Bei der Betrachtung gesellschaftsrelevanter Themen aus unterschiedlichen Perspektiven soll gezeigt werden, dass Problemlösungen von Wertentscheidungen abhängig sind (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2005a). Dazu ist es nötig, dass sie den Sachverhalt erfassen, Bewertungskriterien anlegen, Handlungsmöglichkeiten aufzeigen und Perspektiven wechseln können (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 2005b). In der einschlägigen Forschungsliteratur lassen sich weitere Konzepte zur Beschreibung der Bewertungskompetenz finden, die zusätzliche Faktoren, wie z. B. das Wahrnehmen und Bewusstmachen eigener Einstellungen und das Generieren und Reflektieren von Sachinformationen (Eggert & Bögeholz, 2006) beinhalten. Die verschiedenen Ansätze unterscheiden sich im Wesentlichen darin, wie weit der Begriff der Bewertungskompetenz gefasst

wird. In der vorliegenden Studie werden alle Faktoren, die über das Anwenden von Kriterien, das Benennen und Auswählen von Handlungsmöglichkeiten und das Abschätzen von Folgen hinausgehen als Einflussfaktoren auf die Bewertungskompetenz betrachtet. Ziel ist es, auf diesem Weg aufzuklären, welche Faktoren sich auf die Bewertung auswirken.

Bei qualifizierten und expliziten Bewertungen im Sinne der NBS handelt es sich immer um Entscheidungen, bei denen ein Entscheidungsprozess zur Urteilsfindung durchlaufen wurde. Spontane oder implizite Entscheidungen wurden in dieser Studie daher nicht untersucht, auch wenn im täglichen Leben ein nicht zu vernachlässigender Anteil der Entscheidungen spontan getroffen wird (Patronis et. al, 1999). Die vorliegende Studie lässt sich an die internationale Forschung zu Socioscientific Issues anbinden, da hier vergleichbare Themengebiete und Probleme behandelt werden. Das Wissen, welches für Bewertungsaufgaben benötigt wird, unterscheidet sich in deklaratives und prozedurales Wissen. Zum einen muss die Situation durch das deklarative Wissen verstanden werden, zum anderen muss prozedurales Wissen angewendet werden, um zu einer Lösung der Bewertungsaufgabe zu kommen. Bei dem verwendeten Wissen handelt es sich sowohl um Fachwissen, als auch um Wissen über Bewertungen.

Die Bewertungskompetenz wurde vor allem qualitativ bereits in verschiedenen Forschungsprojekten untersucht. Diese Forschungsprojekte (u.a. Eggert & Bögeholz, 2006; Reitschert, Langlet, Hößle, Mittelsten Scheid & Schlüter, 2007), in

denen die Bewertungskompetenz bereits spezifisch für Teilbereiche ausdifferenziert wurde und innerhalb der Teilbereiche auch spezielle Definitionen für die Bewertungskompetenz vorliegen, beziehen sich vor allem auf das Fach Biologie. Auch quantitative Ansätze stammen im Wesentlichen aus der Biologie (Eggert & Bögeholz, 2010). Allerdings gibt es bis jetzt noch keinen Konsens über eine gemeinsame fachübergreifende, detaillierte Definition der Bewertungskompetenz (Hostenbach et al., 2011). Im Schwerpunkt dieser Studie sollen die Einflussfaktoren stehen, die Entscheidungen in Bewertungssituationen beeinflussen.

Entscheidungen

Ein wichtiger Teil der Bewertungskompetenz ist das Treffen von Entscheidungen, die grob durch drei Merkmale charakterisiert werden können. Entscheidungen umfassen eine Entscheidungssituation, den Entscheidungsprozess und Entscheidungsregeln. Entscheidungssituationen enthalten immer den Grundzustand, in der eine Entscheidung getroffen werden muss. Dabei sind ein Gegenstand (Objekt), ein Bewerter (Subjekt) und die Beziehung zwischen dem Bewerter und dem Gegenstand an einer Entscheidung beteiligt (Poschmann, Riebenstahl & Schmidt-Kallert, 1998). Der Bewerter selbst wird immer noch durch ein persönliches Wertesystem beeinflusst, so dass externe Faktoren in Entscheidungssituationen eine Rolle spielen und daher Entscheidungen subjektiv getroffen werden (Beach & Mitchell, 1978; Eisenführ, Langer & We-

ber, 2010; Jungermann, Pfister & Fischer, 2005).

Der Entscheidungsprozess kann nach Kortland (2003) als ein Kreisprozess ausgedrückt werden, in dem ein Problem identifiziert wird und Kriterien und Alternativen generiert werden, die dann in der Evaluation untereinander abgewogen werden können, um eine abschließende Entscheidung zu treffen. Daran schließt sich die Implementation der Entscheidung und das Identifizieren eines neuen Problems an.

Entscheidungen können nach verschiedenen Arten unterschieden werden. Zum einen werden implizite (Entscheidungsprozess ist nicht sichtbar, Ergebnis sichtbar) und explizite (sowohl Entscheidungsprozess als auch Ergebnis sichtbar) Entscheidungen unterschieden, zum anderen werden auch verschiedene Entscheidungsregeln (u. a. non-kompensatorisch – kompensatorisch) unterschieden (Abelson & Levi, 1985; Hogarth, 1980; Jungermann et al., 2005), die sowohl einzeln als auch verbunden für eine Entscheidung herangezogen werden können.

Die dargestellten Charakteristika von Entscheidungen sind auch in Aufgaben der vorliegenden Studie, die die Bewertungskompetenz messen, berücksichtigt. Die Entscheidungssituation wird den Schülerinnen und Schülern durch eine alltägliche, eine chemiebezogene oder biologiebezogene Situation vorgegeben. Der Entscheidungsprozess, den die Schülerinnen und Schüler zur Beantwortung der Aufgabe verfolgen sollen wird dabei durch ihr eigenes Wertesystem, also durch persönliche Faktoren, beeinflusst. Während des Entscheidungsprozesses sollen dann

Entscheidungsregeln angewendet werden. Schülerinnen und Schüler müssen also über Strategiewissen verfügen, um Entscheidungsregeln anzuwenden.

Kompetenzstrukturmodell

Um die Bewertungskompetenz zu messen, wird ein Kompetenzstrukturmodell benötigt, welches die Verbindung zwischen den in den Bildungsstandards angegebenen Kompetenzanforderungen und den konkreten Testaufgaben darstellt (Bernholt, Parchmann & Commons, 2009). Die Aufgaben für die Operationalisierung der Bewertungskompetenz im Fach Chemie und im Fach Biologie sowie Aufgaben für die Anwendung von Bewertungsstrategien im außerfachlichen Kontext wurden aufbauend auf dem Kompetenzstrukturmodell entwickelt, welches im Projekt ESNaS für die Bewertungskompetenz verwendet wurde (Hostenbach et al., 2011).

Das Modell baut auf bereits bestehenden Kompetenzstrukturmodellen für die Bewertungskompetenz aus der Biologie sowie auf dem Kompetenzstrukturmodell für die Operationalisierung der Kompetenzbereiche *Fachwissen* und *Erkenntnisgewinnung* aus dem Projekt ESNaS (Kauertz et al., 2010; Walpuski et al., 2010) auf. Die zugrunde liegenden Kompetenzstrukturmodelle im Fach Biologie sind das Göttinger Modell der Bewertungskompetenz (Eggert & Bögeholz, 2006) und das Modell zur ethischen Urteilsbildung (Reitschert et al., 2007).

Das finale Kompetenzstrukturmodell (Abbildung 1) für die Operationalisierung der Bewertungskompetenz in der vorliegenden Studie beinhaltet drei Dimensionen (*kognitive Prozesse*, *Komplexität* und *Kompetenzteilbereiche*), wobei die Dimensionen *kognitive Prozesse* und *Komplexität* als schwierigkeiterzeugend angesehen werden.

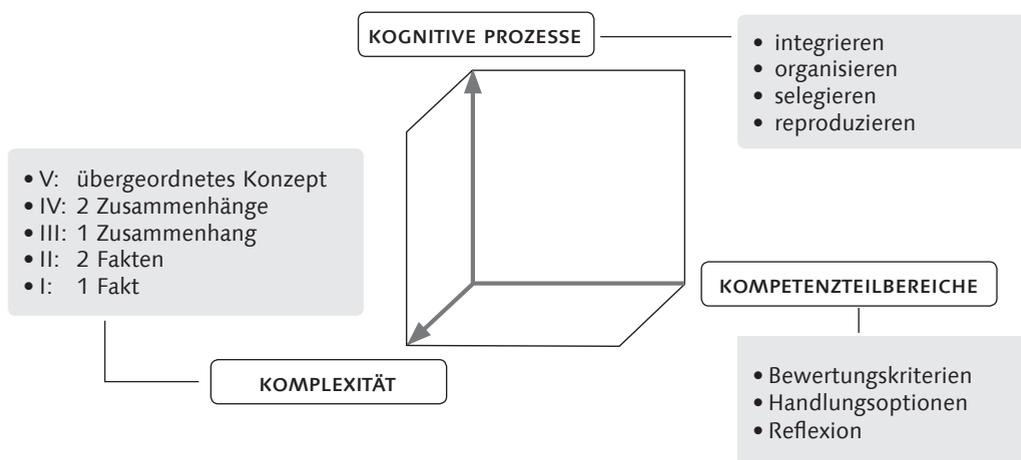


Abb. 1: Dreidimensionales Kompetenzstrukturmodell für die Operationalisierung der Bewertungskompetenz (Hostenbach et al., 2011).

Externe Einflussfaktoren

Wie bereits beschrieben, werden die Bewertungskompetenz und die im Rahmen von Bewertungsprozessen getroffenen Entscheidungen von externen Faktoren beeinflusst. Dabei handelt es sich sowohl um Faktoren, die das Wertesystem eines Bewerbers abbilden (Ajzen & Fishbein, 1969), als auch um Kenntnisse des Bewerbers (Sadler & Zeidler, 2005). Während der Entscheidungsfindung können Werte und Normen herangezogen werden, die die Entscheidung des Bewerbers beeinflussen können (Barkmann & Bögeholz, 2003,

Große & Bögeholz, 2003). Zu Werten gehören z. B. Gleichheit, Gerechtigkeit und Wohlergehen. Normen beinhalten Regelungen in der Gesellschaft oder in gesellschaftlichen Gruppen, dazu gehören auch moralische Normen, die individuelle und gesellschaftliche Vorstellungen sind. Die in dieser Studie untersuchten Faktoren können in drei Gruppen unterschieden werden (Abbildung 2), die im Folgenden genauer dargestellt werden. Die Faktoren betreffen die fachliche Seite, persönlich-individuelle Faktoren und überfachliche Faktoren.

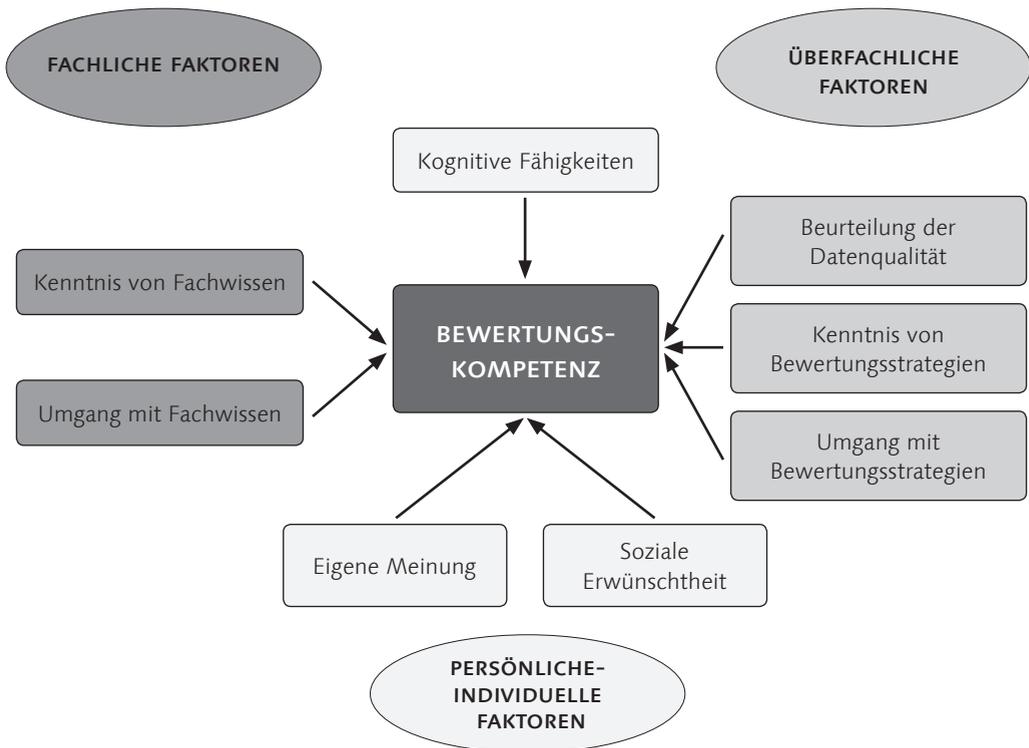


Abb. 2: Mögliche Einflussfaktoren auf die Bewertungskompetenz.

Fachliche Einflussfaktoren

Eine Seite der Einflussfaktoren stellen fachliche Faktoren dar. Dieses liegt nahe, da eine Bewertung immer vor dem Hintergrund des entsprechenden Faches erfolgt. Daher wurde die Bewertungskompetenz auch für jedes Fach einzeln in den Bildungsstandards definiert. Da in einer anderen Studie bereits herausgefunden wurde, dass das fachliche Vorwissen mit der Fähigkeit in Aufgaben bereitgestelltes Fachwissen anzuwenden korreliert (Walpuski, Ropohl & Sumfleth, 2011), ist auch anzunehmen, dass die Bewertungskompetenz verbessert wird, wenn das benötigte Fachwissen bereits als Vorwissen bekannt ist. Für die Bearbeitung von SSI konnten Patronis, Potari & Spiliotopoulou (1999) einen positiven Zusammenhang zum Fachwissen zeigen. Der Einfluss von Fachwissen auf das Treffen von Entscheidungen konnte allerdings bislang noch nicht eindeutig geklärt werden (Bell & Lederman, 2003), da es auch widersprüchliche Ergebnisse gibt. Relevantes Fachwissen sollte jedoch in Entscheidungen mit einbezogen werden (Bögeholz & Barkmann, 2005), um eine rationale und verantwortungsbewusste Entscheidung zu treffen (Höbke, 2001). Bereits in den PISA-Studien wurde davon ausgegangen, dass naturwissenschaftliches Wissen direkt auf persönliche Entscheidungen wirkt (Bybee & McCrae, 2011). In anderen Studien wurde zudem bereits ein Einfluss von Fachwissen auf die Qualität von Argumentationen gezeigt (von Aufschnaiter, Erduran, Osborne & Simon, 2008). Auch in Studien, in denen die Bearbeitung von

Aufgaben in Form von Socioscientific Issues untersucht wurde, konnte ein Zusammenhang der Bearbeitung der Aufgaben, der Nennung von Begründungen und des Argumentierens zum Fachwissen festgestellt werden (Klosterman & Sadler, 2010; Lewis & Leach, 2006; Patronis, Potari & Spiliotopoulou, 1999).

Persönlich-individuelle Einflussfaktoren

Intelligenz wird als allgemeiner Einflussfaktor auf die Bewertungskompetenz gesehen, da kognitive Fähigkeiten hohe Zusammenhänge zu naturwissenschaftlichen Kompetenzen zeigen (Steinkamp & Maehr, 1983) und grundsätzlich ein guter Prädiktor für Schulleistungen sind (Prenzel, Walter & Frey, 2007).

In verschiedenen Studien wurde bereits herausgefunden, dass persönliche Werte einen Einfluss auf das Treffen von Entscheidungen haben (Ajzen & Fishbein, 1969, Barkmann & Bögeholz, 2003, Große & Bögeholz, 2003, Kolstø, 2006). Hierbei zeigte sich, dass Schülerinnen und Schüler unterschiedlich Entscheidungen treffen, je nachdem, wie sie das potentielle Risiko einschätzen. Weiterhin zeigte sich, dass die Aspekte Unsicherheit, Angst und Vorläufigkeit der Argumente eine Entscheidung beeinflussen (Kolstø, 2006). Ein weiterer möglicher Einflussfaktor ist die soziale Erwünschtheit, die in der vorliegenden Studie untersucht wurde. Im Prozess der Entscheidung wird der Bewerter immer von seinem eigenen Wertesystem beeinflusst, in dem die eigene

Meinung und persönliche Einstellungen eine große Rolle für die Entscheidung spielen (Chang Rundgren & Rundgren, 2010). Die Absicht, eine Entscheidung zu treffen, kann nach Ajzen und Fishbein (1969) mit der Einstellung der Person bezüglich der Handlung vorhergesagt werden. Hierbei ist davon auszugehen, dass Menschen in einer Art und Weise auf gesellschaftskritische Fragen antworten, in der sie in den Augen anderer gut dastehen. Dabei werden die Konzepte *Selbsttäuschung* und *Fremdtäuschung* unterschieden (Musch, Brockhaus & Bröder, 2002). Durch das Antworten in Form von sozial erwünschten Antworten wird der Eindruck anderer Menschen in eine bestimmte Richtung gesteuert und kontrolliert, so dass Äußerungen mit sozialen Normen und Werten der Gesellschaft in Einklang gebracht werden können (Janikisz & Moosbrugger, 2008). Durch sozial erwünschte Antworten kann eine Verzerrung des Testergebnisses erfolgen, so dass dieses auch als ein Einflussfaktor auf die Bewertungskompetenz gesehen werden kann. Besonders bei Aufgaben, die zur Erfassung der Bewertungskompetenz eingesetzt werden, werden gesellschaftskritische Themen behandelt. Hierbei sollen Entscheidungen z. B. bezogen auf ökologische Themengebiete erfolgen, bei der es Entscheidungen gibt, die von der Gesellschaft eher akzeptiert werden als andere.

Die Aufgaben zur Bewertungskompetenz beinhalten häufig Umweltthemen, so dass ein umweltbewusstes Handeln für die Entscheidungen vorausgesetzt wird. Werden umweltbewusste Entscheidungen und ein umweltbewusstes Handeln erwartet, muss auch immer eine dementsprechende Ein-

stellung vorhanden sein (Kaiser, Wölfing & Fuhrer, 1999). Somit wird davon ausgegangen, dass die persönliche Einstellung bezogen auf Umweltthemen einen Einfluss auf die Bewertungskompetenz haben kann. Da es auch gegensätzliche Befunde gibt (z. B. Rubitzko & Girwidz, 2009), soll dieser Aspekt näher untersucht werden. Zu den Einflussfaktoren der persönlich-individuellen Seite zählen dementsprechend die allgemeine Intelligenz, die soziale Erwünschtheit und die eigene Meinung.

Überfachliche Einflussfaktoren

In Entscheidungssituationen wird der Anwendung von Bewertungsstrategien eine große Rolle zugeschrieben. Hierbei geht es darum, dass Schülerinnen und Schüler generell über Bewertungsstrategien verfügen und sie auch in außerfachlichen Kontexten anwenden können. Das bedeutet, dass Kriterien gefunden werden können und Entscheidungen basierend auf Entscheidungsstrategien getroffen werden können. Eine Studie aus der Biologiedidaktik zeigt bereits, dass Schülerinnen und Schüler elaborierte Entscheidungsstrategien in einer Alltagssituation nicht auf Situationen im Bereich der nachhaltigen Entwicklung übertragen können (Eggert, Schatz & Bögeholz, 2007). Dieser Effekt zeigt sich z. B. auch bei dem Konzept Problemlösen. PISA 2003 zeigt, dass die Schülerinnen und Schüler in Deutschland eine relative Stärke im Bereich der fächerübergreifenden Problemlösekompetenz haben, bei der ein deutlicher Bezug zum außerunterrichtlichen Leben der Schüle-

rinnen und Schüler gegeben ist (Leutner, Klieme, Meyer & Wirth, 2005). Die naturwissenschaftliche Kompetenz und die Problemlösekompetenz korrelieren im internationalen Vergleich relativ hoch miteinander (Stawitz, 2010). Gleichzeitig liegen die Fähigkeiten deutscher Schülerinnen und Schüler im Bereich der naturwissenschaftlichen Kompetenz im internationalen Vergleich nur im Mittelfeld (Leutner et al., 2005; Prenzel, Baumert & Blum, 2005). Weiterhin wurde bereits herausgefunden, dass die Fähigkeit von Schülerinnen und Schülern, Entscheidungen zu treffen, verbessert wird, wenn sie über metakognitive Strategien verfügen (Zeidler, Walker, Ackett & Simmons, 2002). Wenn Schülerinnen und Schüler über die notwendigen Strategien verfügen, können sie z. B. durch die Verwendung eines cut-offs die Anzahl der Optionen verringern oder die Optionen im trade-off gegeneinander abwägen (Eggert & Bögeholz, 2010). Entscheidungen sind besonders dann verlässlich, wenn sie basierend auf vertrauenswürdigen und hochwertigen Informationen getroffen werden (Gräfe & Maaß, 2008). Nach Sadler, Barab & Scott (2007) umfasst auch die Bearbeitung von Socioscientific Issues neben den Merkmalen: *complexity* (Komplexität des konstruierten Problems) und *multiple perspectives* (verschiedene Lösungen in Abhängigkeit von der Beurteilungsperspektive) auch die Merkmale: *ongoing inquiry* (Entscheidungen sind nicht endgültig, da sich die Datenlage ändern kann) und *skepticism* (Quellen müssen hinterfragt werden). Dafür ist die Fähigkeit zur Beurteilung von Datenqualität nötig, was ein Teil der Fä-

higkeit ist, kritisch zu denken. Somit kann davon ausgegangen werden, dass auch die Beurteilung von Datenqualität ein Einflussfaktor auf die Bewertungskompetenz darstellt (Chang Rundgren & Rundgren, 2010). Die Beurteilung von Datenqualität bedeutet also, dass Schülerinnen und Schüler überprüfen können, ob vorliegende Beobachtungen und Daten repräsentativ sind (Astleitner, Brünken & Zander, 2002). Zu den überfachlichen Einflussfaktoren gehören demnach die Kenntnis und die Anwendung von Bewertungsstrategien sowie die Einschätzung von Datenqualität.

Ziele der Untersuchung

Es wurde bislang nicht untersucht, inwiefern die Bewertungskompetenz abhängig vom Fachgebiet ist bzw. inwieweit die Bewertungskompetenz bei Schülerinnen und Schülern von weiteren Faktoren beeinflusst wird. Dementsprechend verfolgt die vorliegende Studie in einem ersten Ziel, mit Hilfe eines Kompetenzstrukturmodells, welches für die Überprüfung der Bildungsstandards für das Projekt ESNaS ausdifferenziert wurde (Hostenbach et al., 2011) Testaufgaben zur Bewertungskompetenz zu entwickeln. Die Aufgaben operationalisieren Bewertungskompetenz im Fach Chemie, im Fach Biologie und im außerfachlichen Kontext, um zu überprüfen, ob die Bewertungskompetenz im Fach Chemie von der Bewertungskompetenz im Fach Biologie empirisch getrennt werden kann. Obwohl die Bewertungskompetenz für jedes Fach einzeln in den Bildungsstandards operationalisiert wurde

und demnach als fachspezifisch anzusehen ist, könnte es auch möglich sein, dass es sich um eine fachübergreifende Kompetenz handelt. In diesem Fall reichen metakognitive Strategien aus, um in jedem Kontext adäquat eine Entscheidung zu treffen.

In einem zweiten Schritt wird das Ziel verfolgt, ein Modell zu entwickeln und zu prüfen, welches darstellt, ob und in welcher Weise die Bewertungskompetenz von verschiedenen angenommenen Aspekten beeinflusst wird. Mit verschiedenen Tests wird dafür das in Abbildung 2 beschriebene, hypothetische Modell geprüft. Aus den Zielen ergeben sich folgende zwei Forschungsfragen:

- Lässt sich – basierend auf dem vorliegenden Kompetenzstrukturmodell – eine fachspezifische Bewertungskompetenz in der Chemie von der in anderen Fächern (z. B. Biologie) und von der im außerfachlichen Kontext abgrenzen?

- Wie wird die Bewertungskompetenz der Schülerinnen und Schüler durch externe (fachliche, überfachliche, individuelle) Faktoren beeinflusst?

Design

An der Studie waren insgesamt 784 Schülerinnen und Schüler aus 29 Klassen an 16 unterschiedlichen nordrhein-westfälischen Gymnasien beteiligt. Aufgrund des Umfangs der Tests wurde die Studie an zwei verschiedenen Tagen durchgeführt, die Testzeit betrug an beiden Tagen 90 Minuten. Alle Testinstrumente wurden von etwa gleich vielen Schülerinnen und Schülern bearbeitet. In der folgenden Tabelle (Tabelle 1) sind die jeweiligen Tests für den jeweiligen Testtag aufgeführt, die im Folgenden noch weiter erläutert werden. Der endgültige und reduzierte Datensatz setzt sich aus 45.3% Jungen und aus 49.7% Mädchen zusammen. 4.9% haben keine

Tab. 1: Auflistung der verwendeten Testinstrumente

1. TESTTAG	2. TESTTAG
Kenntnis von Fachwissen Chemie	Kognitiver Fähigkeitstest (Heller & Perleth, 2000)
Umgang mit Fachwissen Chemie	Kenntnis von Bewertungsstrategien
Chemie Bewertung	Anwendung von Bewertungsstrategien
Kenntnis von Fachwissen Biologie	Einstellungen zu Umweltthemen (Frey et al., 2009)
Umgang mit Fachwissen Biologie	Einschätzung von Datenqualität
Biologie Bewertung	
Soziale Erwünschtheit (Paulhus, 1998)	

Angaben zum Geschlecht gemacht. Das mittlere Alter der Schülerinnen und Schüler beträgt $M = 15.29$ Jahre ($SD = .681$).

Testinstrumente

Bewertungskompetenz Chemie/Biologie: Für die Messung der Bewertungskompetenz wurde sowohl ein Test mit chemiespezifischen Aufgaben als auch ein Test mit biologiespezifischen Aufgaben entwickelt. Zur Testentwicklung für die Tests zur Messung der Bewertungskompetenz wurden Lehrpläne der Sekundarstufe I der Fächer Chemie und Biologie sowie verschiedene Schulbücher für diese Fächer herangezogen. Somit konnte die inhaltliche Validität der Aufgaben sichergestellt werden. Die Aufgaben wurden auf Grundlage des in ESNaS verwendeten Aufgabenkonstruktionsmanuals zur Bewertung standardisiert konstruiert und überprüft (Köller et al., 2009). Die Kriteriumsvalidität konnte nicht geprüft werden, da kein geeignetes Maß für ein entsprechendes Außenkriterium zur Verfügung steht. Idealerweise müssten die Probanden nach der Testung realistischen Entscheidungssituationen ausgesetzt werden, um zu prüfen, ob die vorhergesagte Fähigkeit mit dem tatsächlichen Verhalten korreliert. Nach Bortz und Döring (2002) kann in diesem Fall hilfsweise auf die Analyse von differential item functioning (DIF) zurückgegriffen werden. Liegt DIF vor, kann die Lösungswahrscheinlichkeit einer Aufgabe nicht vollständig durch die Personenfähigkeit und die Itemschwierigkeit erklärt werden. In dem vorliegenden Fall wurde die Stich-

probe nach dem Geschlecht in zwei Gruppen aufgeteilt. In diesem Fall zeigen 25 von 97 Aufgaben DIF. Davon sind 14 Aufgaben für männliche, elf Aufgaben für weibliche Probanden signifikant schwieriger. Allerdings zeigen lediglich drei der Aufgaben einen DIF Kontrast von über einem Logit. Ferner zeigen sich keine Auffälligkeiten hinsichtlich einzelner Themengebiete oder bestimmter Aufgabenmerkmale. Generell sind die geringen Zahlen kein Grund für einen Ausschluss der Items aus dem Itempool, zudem wird kein Geschlecht systematisch bevorteilt oder benachteiligt. Der Test *Chemie Bewertung* beinhaltet insgesamt 60 Aufgaben z. B. zu den Themengebieten Treibhauseffekt, Bioethanol oder Dünger, die auf zehn Testhefte mit jeweils zwölf Aufgaben in Form eines balanced incomplete block design verteilt wurden. Dafür wurden 60 Items auf zehn Blöcke (B) mit jeweils sechs Items verteilt. Anschließend wurden die Blöcke auf die Testhefte verteilt, so dass jedes Testheft (TH) insgesamt zwölf Aufgaben, bestehend aus zwei Blöcken beinhaltet (Abbildung 3). Der Test *Biologie Bewertung* enthält 30 Aufgaben, die wiederum auf zehn Testhefte mit jeweils sechs Aufgaben verteilt wurden. Beispiele für die Themengebiete aus dem Test *Biologie Bewertung* sind Sucht, Gentechnik und Massentierhaltung.

Soziale Erwünschtheit: Für die Erfassung der sozialen Erwünschtheit wurde die deutschsprachige Version des Balanced Inventory of Desirable Responding (BIDR) verwendet (Musch et al., 2002; Paulhus, 1998). Dieser Test erfasst sowohl das Konzept *Selbsttäuschung* als auch das

Block	Testheft									
	TH1	TH2	TH3	TH4	TH5	TH6	TH7	TH8	TH9	TH10
B1	X									X
B2	X	X								
B3		X	X							
B4			X	X						
B5				X	X					
B6					X	X				
B7						X	X			
B8							X	X		
B9								X	X	
B10									X	X

Abb. 3: Verteilung der Items (Chemie Bewertung) im balanced incomplete block Design.

Konzept *Fremdtäuschung*. Für die Auswertung werden alle Antworten der Probanden für die jeweilige Skala gezählt. Allerdings können in die Auswertung nur die Probanden aufgenommen werden, die alle Items des Tests beantwortet haben, da ansonsten die jeweilige Skala systematisch unterschätzt wird.

Einstellungen zu Umweltthemen: Zur Erfassung der Einstellungen zu Umweltthemen wurden drei Skalen des Fragebogens verwendet, der in PISA 2006 zur Erfassung der Einstellungen in Bezug auf Umweltthemen eingesetzt wurde (Frey et al., 2009). Dabei handelt es sich um die Skalen *Wahrnehmung von Umweltproblemen*, *Optimismus in Bezug auf Umweltprobleme* und *Einstellungen zu Umweltmaßnahmen*. *Kognitive Fähigkeiten:* Zur Erfassung der Intelligenz wurden drei Skalen des kognitiven Fähigkeitstests eingesetzt (Heller & Perleth, 2000). Dabei wurde jeweils eine Skala zur Messung verbaler, quantitativer und figuraler Fähigkeiten verwendet.

Fachwissen: Ein Fachwissenstest wurde sowohl für die Chemie als auch für die Bio-

logie konstruiert. Sowohl in Chemie als auch in Biologie ist der Fachwissenstest untergliedert in den Test *Kenntnis von Fachwissen* und in den Test *Umgang mit Fachwissen*. Abbildung 4 enthält eine Beispielaufgabe aus dem Test Chemie Bewertung sowie die Aufgaben des Tests *Kenntnis von Fachwissen Chemie* und *Umgang mit Fachwissen Chemie*. Der Test *Kenntnis von Fachwissen* beinhaltet die Abfrage reinen Faktenwissens. Im Test *Umgang mit Fachwissen* wird im Aufgabenstamm Fachinformationen angegeben, mit denen die Probanden umgehen müssen, um die richtige Aufgabenlösung zu erreichen. Das bedeutet, alle notwendigen Informationen werden vorgegeben, der Schwerpunkt liegt dann auf dem korrekten Umgang mit diesen Fachinformationen (Kauertz et al., 2010; Ropohl, 2010). Durch die Untergliederung des Tests in *Kenntnis* und *Umgang mit Fachwissen* kann sowohl deklaratives Fachwissen der Schüler (*Kenntnis von Fachwissen*) als auch prozedurales fachspezifisches Wissen (*Umgang mit Fachwissen*) erhoben werden (Ropohl, 2010). Die Aufgaben zur

Ein Forscher in einer Firma möchte ein neues Kühlpack herstellen. Dafür benötigt er Salz und Wasser. Das Lösen eines Salzes in Wasser ist ein Vorgang, der entweder exotherm oder endotherm ist. Um sich für ein Salz zu entscheiden, betrachtet der Forscher die Lösungsenergie (Kilojoule je Gramm Salz) und den Preis. Das Kühlpack soll so kalt und so günstig wie möglich werden.

Die Lösungsenergie kann

- negativ sein: die Lösung wird warm, der Lösungsvorgang ist exotherm
- null sein: die Temperatur der Lösung bleibt gleich
- positiv sein: die Lösung kühlt sich ab, der Lösungsvorgang ist endotherm

Zum Beispiel erwärmen sich 50 mL Wasser bei Zugabe von 5 g Natriumiodid um ca. 3°C

Salz	Lösungsenergie in kJ/g	Preis in €/100 g
Calciumsulfat	- 0,13	3,63
Ammoniumchlorid	+0,27	2,73
Natriumchlorid	+0,06	1,97
Natriumiodid	-0,05	19,30
Calciumfluorid	+0,23	0,79

Aufgabe 1:

cBK1Zo01

Für welches Salz sollte sich der Forscher entscheiden, wenn er die Lösungsenergie und den Preis der Salze berücksichtigt? Begründe genau!

Um den Treibhauseffekt zu verringern werden Autos entwickelt, die zum Beispiel aus alternativen Energiequellen erzeugten Wasserstoff als Treibstoff benötigen. Bei der Verbrennung von Wasserstoff wird viel Wärme frei und das Verbrennungsprodukt Wasser ist umweltfreundlich. Dagegen entsteht bei der Verbrennung von Benzin viel Kohlenstoffdioxid, welches zur Erderwärmung beiträgt. Nachteil der Autos, die mit Wasserstoff betrieben werden, ist allerdings eine geringe Reichweite.

Aufgabe 1:

cBT1Zo03

Entscheide, ob du eher Benzin oder Wasserstoff als Treibstoff nutzen würdest. Und Begründe deine Entscheidung mit den Kriterien *Umweltschutz* und *Reichweite* genau.

Abb. 4: Beispielaufgaben Chemie Bewertung.

Erhebung des Fachwissens sind in Form von Multiple-Choice Antwortformaten und in Form von Kurzantworten entwickelt worden. Die Tests zum Fachwissen behandeln genau die Themenbereiche, die auch in den Aufgaben zur Bewertung (Biologie/Chemie) untersucht wurden. Die

Schülerinnen und Schüler bekamen dabei genau die passenden Fachwissensaufgaben zu den Bewertungsaufgaben, die sie bearbeiten sollten. Der Test in Chemie (Abbildung 5) beinhaltete vier Aufgaben für den Teil *Kenntnis von Fachwissen*, und vier Aufgaben für den Teil *Umgang mit Fach*

Umgang mit Fachwissen:

Besitzt ein Stoff, wie zum Beispiel Calciumfluorid, eine positive Lösungsenergie, ist der Lösungsvorgang endotherm und die Lösung kühlt sich ab. Besitzt ein Stoff, wie zum Beispiel Calciumsulfat, eine negative Lösungsenergie, ist der Lösungsvorgang exotherm und die Lösung wird warm.

Wie verändert sich die Temperatur von Wasser, wenn man Calciumfluorid löst?

- Die Temperatur der Lösung verändert sinkt, da Calciumfluorid eine negative Lösungsenergie besitzt und somit endotherm reagiert.
- Die Temperatur der Lösung steigt, da Calciumfluorid eine negative Lösungsenergie besitzt und somit exotherm reagiert.
- Die Temperatur der Lösung sinkt, da Calciumfluorid eine positive Lösungsenergie besitzt und somit endotherm reagiert.
- Die Temperatur der Lösung steigt, da Calciumfluorid eine positive Lösungsenergie besitzt und somit exotherm reagiert.

Kenntnis von Fachwissen:

Die Lösungsenergie eines Stoffes sagt aus, ...

- ... ob sich das Volumen einer Lösung ändert, wenn man einen Stoff darin löst.
- ... ob sich die Temperatur der Lösung ändert, wenn man einen Stoff darin löst.
- ... ob sich die Siedetemperatur einer Lösung ändert, wenn man einen Stoff darin löst.
- ... ob sich die Masse einer Lösung ändert, wenn man einen Stoff darin löst.

Abb. 5: Beispielaufgaben Fachwissen.

wissen. Die Biologie Fachwissenstests beinhalten jeweils zwei Aufgaben.

Bewertungsstrategien: Der Test zu Bewertungsstrategien teilt sich wiederum in

einen Test *Kenntnis von Bewertungsstrategien* und einen Test *außerfachliche Anwendung von Bewertungsstrategien* auf (siehe Abbildung 6). Zum einen müssen

Oliver möchte sich ein neues Handy kaufen. Dafür hat er mehrere zur Auswahl. Entscheidend für ihn sind die Merkmale Internetzugang, Touchscreen und Preis.

	<i>Internetzugang</i>	<i>Touchscreen</i>	<i>Preis</i>
Handy A	<i>Ja</i>	<i>Ja</i>	169 €
Handy B	<i>Nein</i>	<i>Ja</i>	155 €
Handy C	<i>Nein</i>	<i>Ja</i>	150 €

Aufgabe Kenntnis von Bewertungsstrategien: KBS 11

Marie rät Oliver zu dem Handy C. Für sie ist das wichtigste Kriterium, dass das neue Handy ein modernes Touchscreen-Handy ist. Internet benötigt sie nicht, dafür kauft sie lieber das günstigere Handy.

Welche Strategie hat Marie angewendet, um ein Handy auszuwählen?

- Sie hat sich ein Preislimit gesetzt und anschließend die Handys verglichen, die das Preislimit eingehalten haben.
- Sie hat alle Merkmale gegeneinander abgewogen und dann das beste Handy ausgesucht.
- Zuerst hat sie den Touchscreen als K.O.-Kriterium gesetzt, welches unbedingt erreicht werden muss, anschließend hat sie zwei Kriterien gegeneinander abgewogen.
- Sie hat keine offensichtliche Strategie angewendet, um eine Entscheidung zu treffen.

Aufgabe Anwendung von Bewertungsstrategien: ABS1Zo09

Welches Handy würde Oliver auswählen, wenn er sich ein Preislimit von 155 Euro setzt und anschließend das Handy auswählt, welches kleiner ist?

	Internetzugang	Touchscreen	Preis	Größe
Handy A	Ja	Ja	169 €	11 cm x 5cm
Handy B	Nein	Ja	155 €	8 cm x 3 cm
Handy C	Nein	Ja	150 €	10 cm x 4,5 cm

Abb. 6: Beispielaufgaben Bewertungsstrategien.

Schülerinnen und Schüler wissen, welche Bewertungsstrategien es gibt und wie eine elaborierte Bewertung/ Entscheidung durchgeführt wird. Zu den Bewertungsstrategien gehören kompensatorische und nicht-kompensatorische Strategien, sowie die Kombinationen aus beiden Strategien. Zum anderen beinhaltet der Test zur *außerfachlichen Anwendung von Bewertungsstrategien* Aufgaben, in denen Entscheidungen in vorgegebenen Situationen getroffen und begründet werden müssen. Die Aufgaben zur außerfachlichen Anwendung von Bewertungsstrategien wurden nach dem gleichen Aufgabenkonstruktionsmanual konstruiert, welches auch für die Tests *Chemie Bewertung* und *Biologie Bewertung* verwendet wurde. Der Unterschied zwischen den Aufgaben zur außerfachlichen Anwendung von Bewertungsstrategien und den Aufgaben zur Bewertungskompetenz in Chemie und in Biologie liegt nur in dem gegebenen Fachinhalt. Während in den Tests für die Chemie und die Biologie Themengebiete wie z. B. Gentechnik oder Treibhauseffekt verwendet wurden, beinhaltet der Test bezüglich der Anwendung und der Kenntnis von Bewertungsstrategien Aufgaben zu den Themengebieten *Kamera Kauf, Urlaub oder Handy*. Der Test *Kenntnis von Bewertungsstrategien* beinhaltet vier Aufgaben, der Test *außerfachliche Anwendung von Bewertungsstrategien* beinhaltet sieben Aufgaben.

Einschätzung von Datenqualität: Der Test zur Erfassung der Datenqualität wurde aufbauend auf ein Projekt der Deutschen Gesellschaft für Informations- und

Datenqualität (DGIQ) und den dabei entwickelten 15 Informationsqualitätsdimensionen entwickelt (Rohweder, Kasten, Malzahn, Piro & Schmid, 2008). Der Test besteht aus drei Aufgaben, bei denen Schülerinnen und Schüler zuerst zehn verschiedene Medien in Bezug auf ihre Objektivität und auf ihre fachliche Richtigkeit einschätzen sollen. In einer zweiten Aufgabe soll bei 18 Kriterien (z. B. inhaltliche Fehlerfreiheit, Glaubwürdigkeit, Bebilderung) eingeschätzt werden, ob sie die Qualität von Fachbüchern steigern. Das Gleiche soll in der dritten Aufgabe in Bezug auf Internetseiten anhand von 22 Kriterien (z. B. Vollständigkeit, Bearbeitbarkeit, Objektivität) geschehen. Die Einschätzung im gesamten Fragebogen erfolgt auf einer vierstufigen Likert-Skala, die zusätzlich eine Kategorie „weiß ich nicht“ beinhaltet.

Ergebnisse

Aufgrund der Tatsache, dass, beruhend auf dem verwendeten balanced incomplete block design, in den Tests zur Bewertung sowie in den Fachwissenstests unvollständige Datensätze existieren, wurde zur Analyse zusätzlich zur klassischen Testtheorie die probabilistische Testtheorie herangezogen. Die klassische Testtheorie wurde dabei für die Analyse der weiteren Tests herangezogen. Der kognitive Fähigkeitstest, der Test zur Einschätzung von Datenqualität, der Test zur sozialen Erwünschtheit sowie der Test zur Einstellung von Umweltthemen wurden klassisch ausgewertet, da hierbei

von allen Schülerinnen und Schülern alle Aufgaben bearbeitet wurden, während die Tests zur Bewertungskompetenz und zum Fachwissen probabilistisch ausgewertet wurden.

Mehrdimensionalität der Bewertungskompetenz

Im Rahmen der Verwendung der probabilistischen Testtheorie wurde ein Schwerpunkt besonders auf die eventuelle Mehrdimensionalität des Konstrukts Bewertungskompetenz gelegt. Zur Überprüfung der Mehrdimensionalität wurden die Aufgaben der Tests *Chemie Bewertung*, *Biologie Bewertung* und *außerfachliche Anwendung von Bewertungsstrategien* betrachtet. Die Tests zur Bewertungskompetenz wurden von 777 Schülerinnen und Schülern ausgefüllt. Aufgrund der Durchführung der Tests an zwei verschiedenen Testtagen und durch das balanced incomplete block design liegen für die Aufgaben unterschiedliche Anzahlen von Schülerantworten vor. Im Test *außerfachliche Anwendung von Bewertungsstrategien* liegen 692 Antworten für jede Aufgabe vor. Die Antworten je Aufgabe im Test *Chemie Bewertung* variieren von 142–151 ($M = 147$, $SD = 3$), die Antwortanzahlen je Aufgabe im Test *Biologie Bewertung* liegen in einer Spanne von 143–150 Antworten pro Aufgabe ($M = 147$, $SD = 2.5$). 10 % aller Schülerantworten der offenen Aufgaben wurden doppelt kodiert, die Interraterreliabilität ist dabei sehr gut ($\kappa = .86$ – 1.0). Bei der Analyse der Tests wurden anhand des Likelihoodquotienten-Tests und ver-

schiedener Informationskriterien (AIC, BIC, CAIC) verschiedene berechnete Rasch-Modelle hinsichtlich ihrer Passung auf die erhobenen Daten analysiert (Rost, 2004). Hierbei wurde verglichen, ob das Konstrukt Bewertungskompetenz ein einheitliches Konstrukt unabhängig vom Fachinhalt darstellt, oder ob die Bewertungskompetenz ein mehrdimensionales Konstrukt ist, so dass die Bewertungskompetenz in jedem Fach eine latente Dimension darstellt (Wu, Adams & Wilson, 2007). Zudem wurden die Fit-Statistiken (MNSQ, t-Wert) der Items berücksichtigt, um die Güte der einzelnen Items in den Rasch-Modellen zu beurteilen. In der vorliegenden Studie wurde als Grenzwert für den MNSQ die Spanne (.75–1.33) nach Wilson (2005) verwendet. Der t-Wert sollte zwischen -2 und +2 liegen (Bond & Fox, 2007). Da der t-Wert gerade bei sehr großen Stichproben oftmals signifikante Werte annimmt, sollten nach Wilson (2005) und Bond und Fox (2007) nur die Aufgaben als problematisch angesehen werden, die sowohl im MNSQ als auch im t-Wert nicht ausreichende Werte erreichen. Diese Richtlinie wurde auch in der vorliegenden Untersuchung realisiert. Dabei stellte sich heraus, dass ein dreidimensionales Modell, bei der die Dimensionen *außerfachliche Anwendung von Bewertungsstrategien*, *Bewertung im Fach Chemie* und *Bewertung im Fach Biologie* jeweils ein einzelnes latentes Konstrukt darstellen, am besten auf die erhobenen Daten passt. Das dreidimensionale Modell ist sowohl in der Deviance als auch in den Informationskriterien (AIC, BIC, CAIC) besser als das eindimensionale Modell

(Tabelle 2). Zudem zeigt der Likelihood-quotiententest der Deviance-Differenz eine signifikant bessere Passung des dreidimensionalen Modells auf die erhobenen Daten, so dass auch diese Kennwerte für das Vorhandensein von verschiedenen Konstrukten spricht.

Des Weiteren können für die Prüfung auf Mehrdimensionalität latente Korrelationen berechnet werden. Dabei zeigt sich eine hohe latente Korrelation zwischen *Chemie Bewertung* und *Biologie Bewertung* ($r = .885$). Zwischen *Bewertung Chemie* und der *außerfachlichen Anwendung von Bewertungsstrategien* liegt die Korrelation bei $r = .547$ und zwischen *Bewertung Biologie* und der *außerfachliche Anwendung von Bewertungsstrategien* bei $r = .672$. Die Korrelation zwischen den Variablen *Bewertung Chemie* und *Bewertung*

Biologie ist zwar hoch, allerdings wurden in anderen Studien bereits ähnliche latente Korrelationen zwischen den Konstrukten *Lesekompetenz* und *Naturwissenschaftliche Grundbildung* ($r = .87$; Artelt & Schlagmüller, 2004) sowie zwischen *analytischem Problemlösen* und der *Naturwissenschaftskompetenz* festgestellt ($r = .80$; Rindermann, 2006), die auch als getrennte Konstrukte betrachtet werden. Die folgende Abbildung (Abbildung 7) zeigt die Wright Map und damit die Verteilung der Itemparameter und der Personenparameter in den Dimensionen bei einem dreidimensionalen Rasch-Modell. Hierbei zeigt sich, dass die Personenparameter bei dem Tests außerfachliche Anwendung von Bewertungsstrategien höher liegen als bei den anderen dargestellten Tests. Die Verteilung der Personenparameter sieht in al-

Tab. 2: Vergleich der Ergebnisse der Rasch-Analysen des 1-dimensionalen und des 3-dimensionalen Modells

	1-DIMENSIONALES MODELL		3-DIMENSIONALES MODELL		
			Außerfachlich	Chemie	Biologie
<i>N</i>	777		777		
EAP/PV Reliability	.73		.47	.68	.65
Deviance	16110.96		16015.62		
Varianz	.80		.97	1.08	1.3
AIC	16306.96		16221.62		
BIC	16677.48		16611.05		
CAIC	16492.22		16416.33		

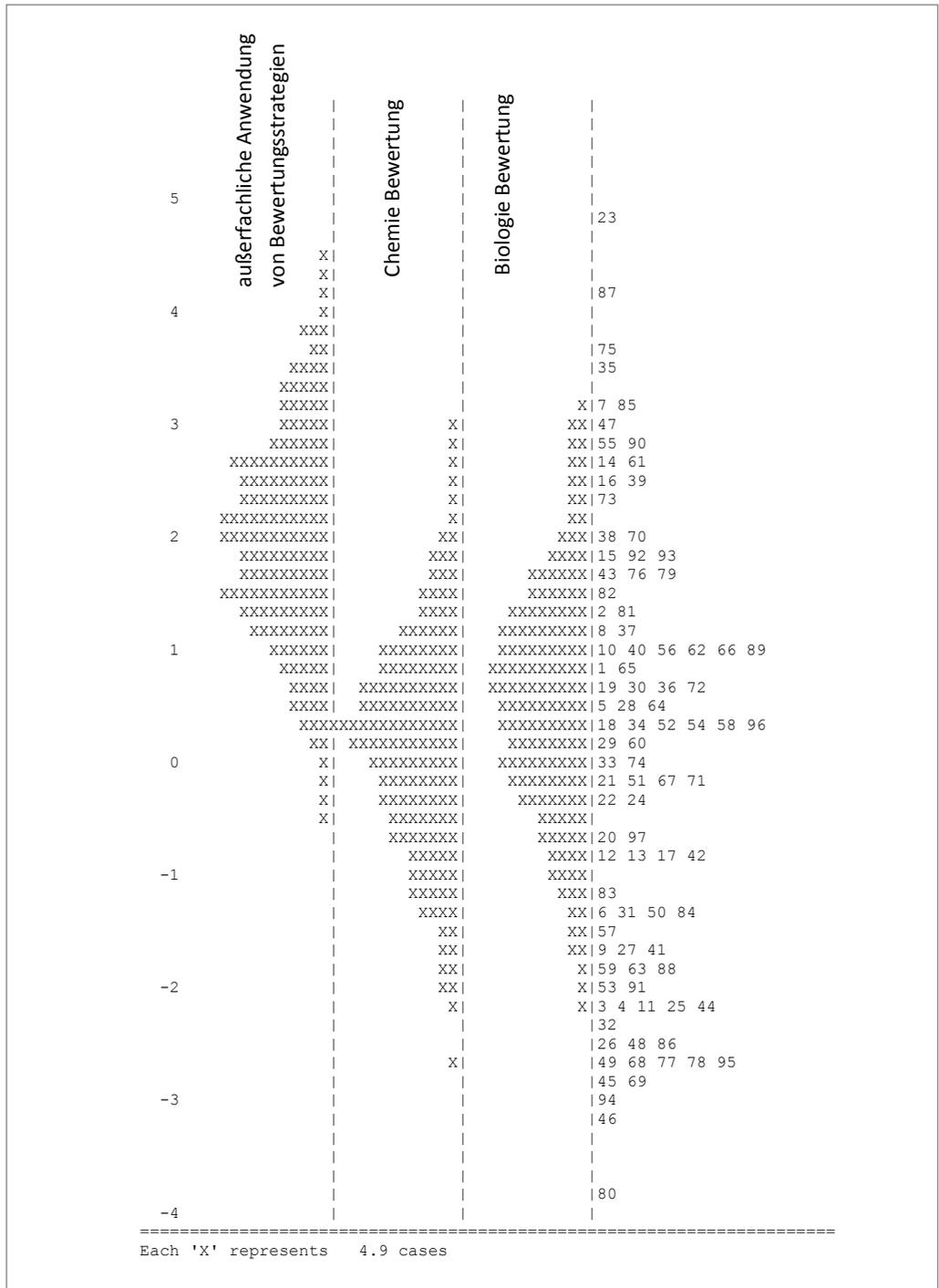


Abb. 7: Wright Map für die Verteilung der Personen- und der Itemparameter im 3-dimensionalen Modell.

len drei Tests ähnlich aus, allerdings liegen sie verschoben auf der Skala.

Somit zeigen die Deviance, der Likelihoodquotiententest der Deviance-Differenz, die Informationskriterien, sowie die latente Korrelation und auch die Verschiebung der Personenparameter insgesamt, dass die Dimensionen verschiedene Konstrukte darstellen.

Es kann also herausgestellt werden, dass es sich bei der Bewertungskompetenz um eine fachspezifische Kompetenz handelt, die jeweils vom fachlichen Kontext abhängt. Zudem kann festgestellt werden, dass Schülerinnen und Schüler die Anwendung von Bewertungsstrategien im außerfachlichen Kontext sehr gut beherrschen. Die bekannten Bewertungsstrategien können dagegen in einem fachlichen Kontext nicht in dem gleichen Maße genutzt werden.

Kennwerte der weiteren Tests

Sowie die Tests zum Fachwissen, als auch der Test zur Kenntnis von Bewertungs-

strategien wurden mit Hilfe der probabilistischen Testtheorie ausgewertet. Da die Tests aus vergleichsweise wenigen Items bestanden, wurden zur Optimierung des EAP-Schätzers Hintergrundmodelle gerechnet. Für die Skala *Kenntnis von Fachwissen* wurden die Skalen *KFT*, *PISA Einstellungen* und *Umgang mit Fachwissen* in das Hintergrundmodell aufgenommen, für die Skala *Umgang mit Fachwissen* die Skalen *KFT*, *PISA Einstellungen* und *Kenntnis von Fachwissen*. Für die Skala *Kenntnis von Bewertungsstrategien* wurde die Skala *KFT* in das Hintergrundmodell aufgenommen. Die Ergebnisse sind in der folgenden Tabelle (Tabelle 3) dargestellt.

Zwar sind die Reliabilitäten der verschiedenen Tests nicht optimal in Bezug auf die Messung einzelner Personenfähigkeiten, für Gruppenvergleiche sind sie jedoch ausreichend. Die Reliabilität für die Schätzung der Itemparameter ist hingegen sehr gut. Die in Regressionsanalysen betrachteten Itemparameter konnten im Raschmodell sehr verlässlich geschätzt werden, sodass hier keine Einschränkungen hin-

Tab. 3: Kennwerte der Tests zum Fachwissen und des Tests Kenntnis von Bewertungsstrategien

	KENNTNIS VON FACHWISSEN	UMGANG MIT FACHWISSEN	KENNTNIS VON BEWERTUNGSSTRATEGIEN
<i>N</i>	779	779	779
EAP/PV Reliability	.67	.84	.75
Item Separation Reliability	.96	.99	.98
Varianz	.42	2.47	.91

sichtlich der Interpretierbarkeit gemacht werden müssen.

Die folgende Tabelle (Tabelle 4) zeigt die Stichproben und die Reliabilitäten der Tests, die mit der klassischen Testtheorie ausgewertet wurden. Insgesamt zeigt sich hierbei eine zufriedenstellende bis gute Reliabilität. Die Reliabilitäten der Tests zur sozialen Erwünschtheit und zu Einstellungen zu Umweltthemen zeigen in einzelnen Subskalen nur ausreichende Reliabilitäten, diese sind aber noch hinreichend für Gruppenvergleiche und wurden auch bereits in anderen Studien in dieser Größenordnung herausgefunden (Musch et al., 2002).

Zusammenhangsmaße

Die geschätzten Personenparameter des dreidimensionalen Rasch-Modells wurden in einem weiteren Schritt genutzt, um

den Einfluss von externen Faktoren auf die Bewertungskompetenz zu überprüfen. Dabei wurde für jeden Test ein Personenparameter berechnet, um daraufhin anhand von Korrelations- und Regressionsanalysen herauszufinden, wie die gemessenen Fähigkeiten zusammenhängen und damit zu erklären, wie die Bewertungskompetenz von den äußeren Faktoren beeinflusst wird. Im Folgenden werden die Zusammenhangsanalysen der externen Faktoren und der Bewertungskompetenz im Fach Chemie dargestellt.

Korrelationsanalyse

Bei allen berechneten Korrelationen zwischen der Bewertung im Fach Chemie und den möglichen Einflussfaktoren auf die Bewertungskompetenz handelt es sich nur um sehr geringe bis geringe Korrelationen (Tabelle 5). Die Korrelationsanalyse zeigt, dass die *Bewertung im Fach Chemie*

Tab. 4: Kennwerte der Tests, die mit Hilfe der klassischen Testtheorie ausgewertet wurden

TEST	STICHPROBE	RELIABILITÄT
KFT	N=688	Cronbachs α = .87
Einschätzung von Datenqualität	N=684	Cronbachs α = .81
Soziale Erwünschtheit		
Skala Fremdtäuschung:	N=658	Cronbachs α = .72
Skala Selbsttäuschung:	N=682	Cronbachs α = .56
Einstellungen zu Umweltthemen		
Wahrnehmung von Umweltproblemen	N=688	Cronbachs α = .59
Optimismus in Bezug auf Umweltprobleme	N=688	Cronbachs α = .51
Einstellungen zu Umweltmaßnahmen	N=688	Cronbachs α = .67

am stärksten mit der *Bewertung im Fach Biologie*, der *Kenntnis von und dem Umgang mit Fachwissen im Fach Chemie* sowie mit der *Kenntnis von und dem Umgang mit außerfachlichen Bewertungsstrategien* korreliert. Offensichtlich ist eine ausreichende Wissensbasis – sowohl über Bewertungsstrategien als auch im Bereich des Fachwissens – notwendig, um fachbezogene Bewertungen vornehmen zu können. Eine weitere signifikante Korrelation konnte zwischen der *Einschätzung von Datenqualität* und der *Bewertungskompetenz* gefunden werden, die im negativen Bereich liegt, da die Polung im Test *Einschätzung von Datenqualität* entgegengesetzt zu der Polung des Tests *Chemie Bewertung* ist. Die Korrelation zwischen den *kognitiven Fähigkeiten* und der *Bewertungskompetenz* liegt bei $r = .265$ ($p \leq .001$). Die Korrelationen mit den anderen möglichen Einflussfaktoren liegen in einem sehr kleinen Bereich $r < .200$, so dass auf diese hier nicht weiter eingegangen wird.

Regressionsanalyse

Für die Analyse des Einflusses der möglichen externen Einflussfaktoren auf die Bewertungskompetenz im Fach Chemie wurde eine multiple lineare Regression berechnet. Dabei zeigt sich, dass der stärkste Prädiktor für die Bewertungskompetenz die *außerfachliche Anwendung von Bewertungsstrategien* ist ($\beta = .229$, $p \leq .001$). Der zweitstärkste Prädiktor ist die *Kenntnis von Fachwissen* ($\beta = .185$, $p \leq .001$). Auch die *Einschätzung von Datenqualität* ($\beta = -.168$, $p \leq .001$) kann als Prädiktor für die Be-

wertungskompetenz gesehen werden. Der Einfluss der Fähigkeit, die Datenqualität einzuschätzen zeigt einen negativen Regressionskoeffizienten aufgrund der Polung des Tests. Erreichen die Schülerinnen und Schüler einen niedrigen Wert im Test *Einschätzung von Datenqualität* zeigt dieses, dass ihre Fähigkeit in diesem Bereich höher ist. Die *Kenntnis von Bewertungsstrategien* ($\beta = .073$, $p = .54$) wird als Prädiktor für die Bewertungskompetenz gerade nicht mehr signifikant. Die kognitive Fähigkeit ist kein signifikanter Prädiktor für die Bewertungskompetenz ($\beta = .033$, *n.s.*). Als Ergebnis der Regressionsanalyse kann festgehalten werden, dass die genannten Einflussfaktoren nur knapp ein Fünftel der Varianz der Bewertungskompetenz aufklären können ($R^2 = .19$). Es zeigen sich insgesamt bei der Regressionsanalyse keine sehr hohen, direkten Einflüsse auf die Bewertung im Fach Chemie.

Zusammenfassung und Diskussion

Basierend auf dem im Projekt ESNaS entwickelten Modell zur Erfassung der Bewertungskompetenz in den naturwissenschaftlichen Fächern wurden in diesem Projekt Testaufgaben entwickelt, um die Bewertungskompetenz von Schülerinnen und Schülern für die Fächer Chemie und Biologie zu messen. Mithilfe weiterer Testinstrumente konnte zusätzlich untersucht werden, welche weiteren Faktoren die Bewertungskompetenz beeinflussen. Die Informationen aus diesen Analysen können auch verwendet werden, um das ESNaS-Modell zu anderen Modellen, die diese

Tab. 5: bivariate Korrelationen zwischen der Bewertungskompetenz im Fach Chemie und den möglichen Einflussfaktoren

VARIABLE	BEWERTUNG CHEMIE
Bewertung Biologie Korrelation nach Pearson <i>N</i>	.481*** 648
außerfachliche Anwendung von Bewertungsstrategien Korrelation nach Pearson <i>N</i>	.294*** 648
kognitive Fähigkeiten Korrelation nach Pearson <i>N</i>	.265*** 643
Umgang mit Fachwissen Chemie Korrelation nach Pearson <i>N</i>	.218*** 643
Kenntnis von Fachwissen Chemie Korrelation nach Pearson <i>N</i>	.309*** 643
Umgang mit Fachwissen Biologie Korrelation nach Pearson <i>N</i>	.191*** 638
Kenntnis von Fachwissen Biologie Korrelation nach Pearson <i>N</i>	.189*** 645
Einschätzung von Datenqualität Korrelation nach Pearson <i>N</i>	-.267*** 637
Kenntnis von Bewertungsstrategien Korrelation nach Pearson <i>N</i>	.200*** 643
UMWELTEINSTELLUNGEN	
Wahrnehmung von Umweltproblemen Korrelation nach Spearman <i>N</i>	-.132** 644
Optimismus in Bezug auf Umweltprobleme Korrelation nach Spearman <i>N</i>	.084* 646
Einstellungen zu Umweltmaßnahmen Korrelation nach Spearman <i>N</i>	-.141** 646
SOZIALE ERWÜNSCHTHEIT	
Fremdtäuschung Korrelation nach Spearman <i>N</i>	.134** 583
Selbsttäuschung Korrelation nach Spearman <i>N</i>	-.012 601

Faktoren teilweise in die Bewertungskompetenz mit einbeziehen, in Beziehung zu setzen. Mit Hilfe von Rasch-Analysen konnte gezeigt werden, dass es sich bei der Bewertungskompetenz in Abhängigkeit von den untersuchten Fachinhalten (Chemie/Biologie/außerfachlich) um ein mehrdimensionales Konstrukt handelt. Die Ergebnisse dieser Kompetenztests zeigen, dass Schülerinnen und Schüler über eine ausgeprägte Bewertungskompetenz im außerfachlichen Kontext verfügen, die bekannten Bewertungsstrategien können dagegen in einem fachinhaltlichen Kontext nicht in dem gleichen Maße genutzt werden. Der klare Unterschied in der Schwierigkeit ist vermutlich darin begründet, dass Schülerinnen und Schüler in ihrem alltäglichen Leben fortwährend Entscheidungen treffen, während sie im unterrichtlichen Kontext selten explizite Bewertungen vornehmen. Für die Unterrichtspraxis bedeutet das, dass teilweise schon auf vorhandene Fähigkeiten im Bereich der Bewertungskompetenz zurückgegriffen werden kann, dass aber die Anwendung im fachlichen Kontext explizit geübt werden muss.

Für die weitere Analyse wurden die Personenparameter des drei-dimensionalen Rasch-Modells verwendet, um somit Zusammenhänge zwischen externen Faktoren und der Bewertungskompetenz im Fach Chemie anhand von Korrelations- und Regressionsanalysen zu untersuchen. Als Ergebnis lässt sich hier festhalten, dass der stärkste Prädiktor für die Bewertungskompetenz die *außerfachliche Anwendung von Bewertungsstrategien* ist. Dieser Befund deckt sich mit Ergebnissen von Eg-

gert (2008), die zeigen konnte, dass bei Bewertungsaufgaben im Fach Biologie nur 29,1 % der Probanden weder trade-offs noch cut-offs als Strategie anwendeten. In Kombination mit den relativ guten Testergebnissen in diesem Bereich lässt sich sagen, dass ein wichtiger Grundstein für die fachliche Bewertung in vielen Fällen bereits gelegt ist, aber dass die fachspezifische Anwendung noch deutlich vertieft werden muss. Des Weiteren stellen die *Kenntnis von Fachwissen* und die *Einschätzung von Datenqualität* weitere Prädiktoren für die Bewertungskompetenz dar. Die *kognitiven Fähigkeiten* zeigen nur einen kleinen Einfluss auf die Bewertungskompetenz in Chemie. Gar kein bzw. nur ein sehr kleiner Zusammenhang konnte zwischen der *Bewertungskompetenz* und den Aspekten *Einstellungen zu Umweltthemen* oder *soziale Erwünschtheit* gefunden werden. Einschränkend ist hier allerdings anzumerken, dass es sich bei den Testaufgaben um fiktive Bewertungssituationen handelt. Es besteht also durchaus die Möglichkeit, dass Schülerinnen und Schüler diese Situationen eher analytisch behandeln, während sie in realen Situationen stärker ihren persönlichen Interessen folgen. Um diesen Zusammenhang weiter aufzuklären, müsste das Testergebnis mit dem Entscheidungsverhalten in authentischen Entscheidungssituationen in ähnlichen Kontexten in Beziehung gesetzt werden, was in der Praxis nur schwer möglich sein dürfte. Positiv festzuhalten ist aber die geringe Störanfälligkeit des Tests für sozial erwünschte Antworten und subjektive Einstellungen.

Ein erwähnenswerter Anteil der Varianz lässt sich durch die hier eingesetzten Instrumente nicht aufklären. Dieser Anteil kann ein eigenständiges Merkmal der Bewertungskompetenz sein, aber möglicherweise auch durch Variablen wie Interesse und Motivation beeinflusst werden, die hier nicht erhoben wurden. Hier sind noch weitere Forschungsarbeiten notwendig, um präzisere Aussagen zu ermöglichen.

Zusammenfassend kann festgehalten werden, dass in Testsituationen die Schülerleistung stärker von der Kompetenz beeinflusst werden, Bewertungssituationen im Alltag zu bewältigen als von persönlichen Einstellungen und sozialer Erwünschtheit. Zusätzlich wird insbesondere auch Fachwissen zur Bewertung von Sachverhalten benötigt. Inwiefern die Testaufgaben als Prädiktoren für tatsächliches Handeln herangezogen werden können, wurde im Rahmen dieser Studie nicht untersucht, zumal die Anzahl der Bewertungsaufgaben für eine Einzeldiagnose nicht ausreichend war. Die Daten liefern Hinweise darauf, dass im Chemie und Biologieunterricht das Bewerten zu selten geübt wird, um hier ausreichende Kompetenzen zu entwickeln. Dieses Ergebnis lässt den Studien, die sich mit der Förderung der Bewertungskompetenz beschäftigen, große Bedeutung zukommen.

Danksagung

Wir danken dem Institut zur Qualitätsentwicklung im Bildungswesen (IQB) für die Förderung des Forschungsvorhabens sowie der DFG für die Unterstützung der

Forschungsarbeit im Rahmen der Forschergruppe und Graduiertenkolleg naturwissenschaftlicher Unterricht (GRK 902). Des Weiteren danken wir allen beteiligten Schülerinnen und Schülern, sowie Lehrerinnen und Lehrer für die Teilnahme an der Studie.

Literatur

- Abelson, R.P. & Levi, A. (1985). Decision Making and Decision Theory. In Lindzey, G. & Aronson, E. (Hrsg.), *Handbook of Social Psychology* (S. 231–309). New York: Eandom House.
- Ajzen, I. & Fishbein, M. (1969). The Prediction of Behavioral Intentions in a Choice Situation. *Journal of Experimental Social Psychology*(5), 400–416.
- Artelt, C. & Schlagmüller, M. (2004). Der Umgang mit literarischen Texten als Teilkompetenz im Lesen? Dimensionsanalysen und Ländervergleiche. In Schiefele, U., Artelt, C., Schneider, W. & Stanat, P. (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz* (S. 169–196). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Astleitner, H., Brünken, R. & Zander, S. (2002). Können Schüler und Lehrer kritisch denken? Lösungserfolg und -strategien bei typischen Aufgaben. *Salzburger Beiträge zur Erziehungswissenschaft*(6), 51–61.
- Aufschnaiter, C. von, Erduran, S., Osborne, J. & Simon, S. (2008). Arguing to Learn and Learning to Argue: Case Studies of How Students' Argumentation Relates to Their Scientific Knowledge. *Journal of Research in Science Teaching*, 45(1), 101–131.
- Baltes-Götz, B. (2010). Analyse von Strukturgleichungsmodellen mit Amos 18. <http://www.uni-trier.de/fileadmin/urt/doku/amos/v18/amos18.pdf> (21.3.2011).
- Barkmann, J. & Bögeholz, S. (2003). Kompetent gestalten, wenn es komplexer wird. „21“ – *Bildung für eine nachhaltige Entwicklung*, 3, 49–52.

- Beach, L.R. & Mitchell, T.R. (1978). A Contingency Model for the Selection of Decision Strategies. *The Academy of Management Review*, 3(3), 439–449.
- Bell, R. L. & Lederman, N. G. (2003). Understandings of the Nature of Science and Decision Making on Science and Technology Based Issues. *Science Education*, 87, 352–377.
- Bernholt, S., Parchmann, I. & Commons, M.L. (2009). Kompetenzmodellierung zwischen Forschung und Unterrichtspraxis. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 219–245.
- Bögeholz, S. & Barkmann, J. (2005). Rational Choice and Beyond: Handlungsorientierende Kompetenzen für den Umgang mit faktischer und ethischer Komplexität. In Klee, R., Bayrhuber, H., Sandmann, A. & Vogt, H. (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik* (Band 2) (S. 211–224). Innsbruck: StudienVerlag.
- Bond, T.G. & Fox, C.M. (2007). *Applying the Rasch model fundamental measurement in the human sciences*, England: Lawrence Erlbaum Associates.
- Bybee, R. & McCrae, B. (2011). Scientific Literacy and Student Attitudes: Perspectives from PISA 2006 science. *International Journal of Science Education*, 33(1), 7–26.
- Byrne, B.M. (2001). *Structural equation modeling with AMOS. Basic concepts, applications, and programming*, Mahwah, N.J.: Lawrence Erlbaum Associates.
- 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.
- Eggert, S. (2008). Bewertungskompetenz für den Biologieunterricht – Vom Modell zur empirischen Überprüfung: Dissertation. Retrieved from <http://webdoc.sub.gwdg.de/diss/2008/eggert/eggert.pdf> [Zuletzt geöffnet: 21.04.2013].
- 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. & Bögeholz, S. (2010). Students' Use of Decision Making Strategies with regard to Socioscientific Issues – An Application of the Rasch Partial Credit Model. *Science Education*, 94, 230–258.
- Eggert, S., Schatz, J. & Bögeholz, S. (2007). Die Anwendung von Entscheidungsstrategien bei Schüler(inne)n und Studierenden. In Bayrhuber, H., Harms, U., Krüger, D., Sandmann, A., Unterbruner, U., Upmeyer Belzen, A. zu & Vogt, H. (Hrsg.), *Ausbildung und Professionalisierung von Lehrkräften* (S. 23–26).
- Eisenführ, F., Langer, T. & Weber, M. (2010). *Rationales Entscheiden*, Berlin [u. a.]: Springer.
- Frey, A., Taskinen, P., Schütte, K., Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E. & Pekrun, R. (Hrsg.) (2009). *PISA-2006-Skalenhandbuch. Dokumentation der Erhebungsinstrumente*, Münster, New York, NY, München, Berlin: Waxmann.
- Gräfe, G. & Maaß, C. (2008). Bedeutung der Informationsqualität bei Kaufentscheidungen im Internet. In Hildebrand, K. (Hrsg.), *Daten- und Informationsqualität. Auf dem Weg zur Information excellence* (S. 172–197). Wiesbaden: Vieweg + Teubner.
- Große, F. & Bögeholz, S. (2003). Förderung der Bewertungskompetenz von Schülerinnen und Schülern der Sekundarstufe II am Beispiel der Streuobstwiese. *Erkenntnisweg Biologiedidaktik*, 103–119.
- Heller, K.A. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen Revision (KFT 4-12+R)*, Göttingen: Hogrefe.
- Hogarth, R.M. (1980). *Judgement and choice. The psychology of decision*, Chichester: Wiley.
- Höbke, C. (2001). Ethische Dimensionen der Gentechnik im Unterricht Teil 1: Risikoanalyse am Beispiel des gentechnischen Produktionsverfahrens von Humaninsulin. *Praxis der Naturwissenschaften – Biologie in der Schule*, 5/50, 31–35.

- Hostenbach, J., Fischer, H.E., Kauertz, A., Mayer, J., Sumfleth, E. & Walpuski, M. (2011). Modellierung der Bewertungskompetenz in den Naturwissenschaften zur Evaluation der Nationalen Bildungsstandards. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 261–288.
- Jankisz, E. & Moosbrugger, H. (2008). Planung und Entwicklung von psychologischen Tests und Fragebogen. In Moosbrugger, H. (Hrsg.), *Testtheorie und Fragebogenkonstruktion. Mit 43 Tabellen* (S. 27–72). Heidelberg: Springer.
- Jungermann, H., Pfister, H.R. & Fischer, K. (2005). *Die Psychologie der Entscheidung*, Heidelberg, Berlin: Spektrum Akademischer Verlag 2.
- Kaiser, F.G., Wölfling, S. & Fuhrer, U. (1999). Environmental Attitude and Ecological Behaviour. *Journal of Environmental Psychology*(19), 1–19.
- 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.
- Klosterman, M.L. & Sadler, T.D. (2010). Multi-level Assessment of Scientific Content Knowledge Gains Associated with Socioscientific Issues-based Instruction. *International Journal of Science Education*, 32(8), 1017–1043.
- Köller, O., Fischer, H.E., Mayer, J., Sumfleth, E., Göbel, J., Härtig, H., Hartmann, S., Katzenbach, M., Kauertz, A., Kremer, K., Neumann, I., Walpuski, M. & Wellnitz, N. (2009). *Evaluation der Standards in den Fächern Biologie, Chemie und Physik für die Sekundarstufe I (ESNaS). Band 6: Aufgabenkonstruktionsanleitung für die Aufgabenentwicklung in den Fächern Biologie, Chemie und Physik für den Kompetenzbereich „Bewertung“*.
- Kolstø, S. D. (2006). Patterns in Students' Argumentation Confronted with a Risk-focused Socio-scientific Issue. *International Journal of Science Education*, 28(14). 1689–1716.
- Kortland, J. (2003). Designing and validating a didactical structure for a problem-posing approach to teaching decision making about the waste issue. http://www.phys.uu.nl/~kortland/art_esera-03.pdf.
- Leutner, D., Klieme, E., Meyer, K. & Wirth, J. (2005). Die Problemlösekompetenz in den Ländern der Bundesrepublik Deutschland. In Prenzel, M., Baumert, J. & Blum, W. (Hrsg.), *PISA 2003. Der zweite Vergleich der Länder in Deutschland – was wissen und können Jugendliche?* (S. 125–146). Münster [u.a.]: Waxmann.
- Lewis, J. & Leach, J. (2006). Discussion of Socioscientific Issues: The role of science knowledge. *International Journal of Science Education*, 28(11), 1267–1287.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse*, München: Beltz Psychologie Verlags Union.
- Musch, J., Brockhaus, R. & Bröder, A. (2002). Ein Inventar zur Erfassung von zwei Faktoren sozialer Erwünschtheit. *Diagnostica*, 48(3), 121–129.
- Patronis, T., Potari, D. & Spiliotopoulou, V. (1999). Students' argumentation in decision-making on a socio-scientific issue: implications for teaching. *International Journal of Science Education*, 21(7), 745–754.
- Paulhus, D.L. (1998). *The Balanced Inventory of Desirable Responding*, Toronto/Buffalo: Multi-Health Systems.
- Poschmann, C., Riebenstahl, C. & Schmidt-Kallert, E. (1998). *Umweltplanung und -bewertung*, Gotha: Justus Perthes Verlag.
- Prenzel, M., Baumert, J. & Blum, W. (Hrsg.) (2005). *PISA 2003. Der zweite Vergleich der Länder in Deutschland – was wissen und können Jugendliche?*, Münster [u.a.]: Waxmann.
- Prenzel, M., Walter, O. & Frey, A. (2007). PISA misst Kompetenzen. Eine Replik auf Rindermann (2006): Was messen internationale Schulleistungsstudien? *Psychologische Rundschau*, 58, 128–136.
- Reitschert, K., Langlet, J., Hößle, C., Mittelsten Scheid, N. & Schlüter, K. (2007). Dimensionen Ethischer Urteilskompetenz. *MNU*, 60/1, 43–51.

- Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? Schulleistungen, Schülerfähigkeiten, kognitive Fähigkeiten, Wissen oder allgemeine Intelligenz. *Psychologische Rundschau*, 57(2), 69–86.
- Rohweder, J.P., Kasten, G., Malzahn, D., Piro, A. & Schmid, J. (2008). Informationsqualität – Definitionen, Dimensionen und Begriffe. In Hildebrand, K. (Hrsg.), *Daten- und Informationsqualität. Auf dem Weg zur Information excellence* (S. 25–45). Wiesbaden: Vieweg + Teubner.
- Ropohl, M. (2010). *Modellierung von Schülerkompetenzen im Basiskonzept Chemische Reaktion. Entwicklung und Analyse von Testaufgaben*, Berlin: Logos.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*, Bern: Huber.
- Rubitzko, T., Girwidz, R. (2009). Wissen, Einstellung und Handlungsbereitschaft bei Umweltproblemen im interkulturellen Vergleich. In D. Höttecke (Hrsg.), *Chemie- und Physikdidaktik für die Lehrerbildung* (S.71–73). Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Schwäbisch Gmünd 2008. Münster: LIT-Verlag.
- 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. & Zeidler, D. L. (2005). The Significance of Content Knowledge for Informal Reasoning Regarding Socioscientific Issues: Applying Genetics Knowledge to Genetic Engineering Issues. *Science Education*, 85, 71–93.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005a). *Bildungsstandards im Fach Chemie für den Mittleren Bildungsabschluss. Beschluss vom 16.12.2004*, München, Neuwied: Luchterhand.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005b). *Bildungsstandards im Fach Biologie für den Mittleren Bildungsabschluss. Beschluss vom 16.12.2004*, München, Neuwied: Luchterhand.
- Stawitz, H. (2010). *Auswirkung unterschiedlicher Aufgabenprofile auf die Schülerleistung. Vergleich von Naturwissenschafts- und Problemlöseaufgaben der PISA 2003-Studie*, Berlin: Logos.
- Steinkamp, M.W. & Maehr, M.L. (1983). Affect, Ability, and Science Achievement: A Quantitative Synthesis of Correlational Research. *American Educational Research Association*(53), 369–396.
- Walpuski, M., Kauertz, A., Fischer, H.E., Kampa, N., Mayer, J., Sumfleth, E. & Wellnitz, N. (2010). ESNas – Evaluation der Standards für die Naturwissenschaften in der Sekundarstufe I. In Gehrman, A. (Hrsg.), *Bildungsstandards und Kompetenzmodelle. Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht*. Bad Heilbrunn: Klinkhardt.
- Walpuski, M., Ropohl, M. & Sumfleth, E. (2011). Students' knowledge about chemical reactions – development and analysis of standard-based test items. *Chemistry Education Research and Practice*(12), 174–183.
- Wilson, M. (2005). *Constructing measures. An item response modeling approach*, Mahwah, N.J: Lawrence Erlbaum Associates.
- Wu, M., Adams, R.J. & Wilson, M.R. (2007). *ACER ConQuest version 2.0: generalised item response modelling software*, Camberwell, Victoria: ACER Press.
- Zeidler, D.L., Walker, K.A., Ackett, W.A. & Simmons, M.L. (2002). Tangled Up in Views: Beliefs in the Nature of Science and Responses to Socioscientific Dilemmas. *Science Education*, 86(3), 343–367.

KONTAKT

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

AUTORENINFORMATION

Dr. rer. nat. Maik Walpuski ist Professor für Didaktik der Chemie an der Universität Duisburg-Essen. Nach dem Studium der Fächer Chemie und Deutsch für das Lehramt an Gymnasien an der Universität Duisburg-Essen promovierte er 2006 im Rahmen der DFG-Forschergruppe Naturwissenschaftlicher Unterricht und erhielt für seine Dissertation den Universitätspreis. 2007 schloss er das 2. Staatsexamen ab und wurde 2009 als Juniorprofessor an die Universität Osnabrück berufen. Von dort wechselte er 2011 auf eine Professur an die Leuphana Universität Lüneburg bevor er an die Universität Duisburg-Essen zurückkehrte. Seine Forschungsinteressen gelten der Kompetenzmodellierung und -messung im Fach Chemie sowie der Prozessanalyse von Experimentierphasen.

Dr. rer. nat. Julia Hostenbach promovierte zum Thema „Bewertungskompetenz im Chemieunterricht“ an der Universität Duisburg-Essen und arbeitet derzeit als Lehrerin an einem Gymnasium in Nordrhein-Westfalen.

