

Evaluation von Kompetenzstruktur und –niveaus zum Beobachten, Vergleichen, Ordnen und Experimentieren

Nicole Wellnitz & Jürgen Mayer

Nicole.Wellnitz@didaktik.bio.uni-giessen.de

Justus-Liebig-Universität Gießen, Institut für Biologiedidaktik
Karl-Glöckner-Str. 21 C, 35394 Gießen

Zusammenfassung

Zur Qualitätssteigerung des naturwissenschaftlichen Unterrichts wurden Bildungsstandards formuliert, die künftig im Rahmen des KMK-Bildungsmonitorings regelmäßig evaluiert werden (KMK 2006). Die naturwissenschaftliche Erkenntnisgewinnung stellt dabei einen der zu evaluierenden Kompetenzbereiche dar. Wissenschaftliche Grundlage der Evaluation sind Kompetenzmodelle, die die Beschreibung und Analyse von Struktur und Niveaus der domänenspezifischen Kompetenzen erlauben.

Im vorliegend beschriebenen Forschungsprojekt soll der Teilbereich „Naturwissenschaftliche Untersuchungen“ innerhalb der Erkenntnisgewinnung modelliert und evaluiert werden. Die Erhebungen werden im Rahmen des Kooperationsprojektes „Evaluation der Standards in den Naturwissenschaften für die Sekundarstufe I (ESNaS)“ durchgeführt (vgl. WALPUSKI et al. 2008). Dabei soll die Binnenstruktur beim Beobachten, Vergleichen, Ordnen und Experimentieren hinsichtlich ihrer spezifisch hypothetisch-deduktiven Erkenntnislogik modelliert werden. Ausgehend von einem theoriebasierten Kompetenzstrukturmodell zur Erkenntnisgewinnung (MAYER 2007), wird die bestätigte Binnenstruktur aus vier Teilkompetenzen, Fragestellung, Hypothese, Untersuchungsdesign und Datenauswertung zum Experimentieren (GRUBE et al. 2007, MÖLLER et al. 2007) anhand zweier schwierigkeiterzeugender Merkmale überprüft und auf weitere Untersuchungsmethoden übertragen. Vermutete Kompetenzunterschiede der Untersuchungsmethoden werden graduiert und evaluiert.

Abstract

In order to improve the quality of science instruction at German schools (grade 5-10) educational standards were established in 2004 (KMK 2005). These standards will be evaluated on a regular basis through national assessments (KMK 2006). One area of the intended evaluation is scientific inquiry. The scientific bases of the evaluation are competence models which allow to describe and analyse the structure as well as the levels of competence in specific domains.

The aim of this PhD project is to create and evaluate subsequently a model for the below mentioned subareas of scientific inquiry. The according research is conducted within the cooperation project "Evaluation Standards in Science for Secondary School (ESNaS)"

(WALPUSKI et al. 2008). Within the project the inner structure of various inquiry methods as observing, comparing, arranging and experimenting are modelled regarding their specific hypothetical-deductive approach. Based on a theory-grounded structural competence model for scientific inquiry (MAYER 2007) the verified inner structure of four central skills of scientific inquiry: formulating questions, generating hypotheses, planning of investigation, and interpreting data (GRUBE et al. 2007, MÖLLER et al. 2007) is screened by two difficulty-generating criteria and assigned to further research methods. Assumed competence differences of the various research methods are graduated and evaluated.

1 Einleitung

Das mäßige Abschneiden deutscher Schülerinnen und Schüler in internationalen Schulleistungstests wie TIMMS und PISA hat in Deutschland einen Paradigmenwechsel im Bildungssystem ausgelöst, da die bis dahin übliche Inputsteuerung durch inhaltliche Vorgaben nicht zu den erhofften Ergebnissen geführt hat. Als eine Reaktion auf die unbefriedigenden Befunde hat die Kultusministerkonferenz einheitliche Bildungsstandards als normative Vorgaben für den Mittleren Schulabschluss für das Fach Biologie formuliert (KMK 2005). Die Bildungsstandards greifen allgemeine Bildungsziele auf und benennen fachbezogene Kompetenzen, über die Schülerinnen und Schüler am Ende ihrer Schullaufbahn verfügen sollten (Outputsteuerung).

Eine dauerhafte Qualitätssicherung im naturwissenschaftlichen Unterricht ist nur dann möglich, wenn die Operationalisierung der kompetenzorientierten Bildungsstandards auf der Basis von Kompetenzmodellen gelingt. Während die Bildungsstandards präskriptive Kompetenzen benennen, besteht die Aufgabe fachdidaktischer Forschung darin, diese mittels theorie- und evidenzbasierter Modelle zu fundieren (KLIEME et al. 2003, KLIEME & LEUTNER 2006). Für eine differenzierte Leistungsmessung und Lernstandsdiagnostik müssen geeignete Testinstrumente entwickelt und validiert werden, um die Kompetenzen der Schülerinnen und Schülern erfassen, aufgreifen und systematisch weiter entwickeln zu können.

Kompetenzen naturwissenschaftlicher Erkenntnisgewinnung sind ein wesentliches Element naturwissenschaftlicher Grundbildung („scientific literacy“) (AAAS 1993, BYBEE 2002). Dem entsprechend werden sie in den nationalen Bildungsstandards mit dem Kompetenzbereich „Erkenntnisgewinnung“ umfangreich berücksichtigt (KMK 2005). International werden sie unter den Begriffen „scientific inquiry“ und „nature of science“ subsumiert (NRC 1996, GOTT & DUGGAN 1998, DfES & QCA 1999, LEDERMANN 2006, 2007). Zentrale Aspekte der Erkenntnisgewinnung sind die Formulierung von naturwissen-

schaftlichen Fragestellungen und Hypothesen, die Planung und Durchführung sowie die Auswertung einer naturwissenschaftlichen Untersuchung (NRC 1996, DFES & QCA 1999, KLAHR 2000, GOTT & DUGGAN 2001, GRUBE et al. 2007). Diese Aspekte werden im Rahmen des Projektes „Evaluation der Standards in den Naturwissenschaften für die Sekundarstufe I (ESNaS)“ unter dem Kompetenzteilbereich „Naturwissenschaftliche Untersuchungen“ (scientific inquiry) zusammengefasst (vgl. WALPUSKI et al. 2008).

Ziel der hier vorgestellten Studie ist die empirische Überprüfung eines theoriebasierten Kompetenzstrukturmodells zum Kompetenzteilbereich „Naturwissenschaftliche Untersuchungen“, das die konzeptionelle Grundlage für die Evaluation der entsprechenden Bildungsstandards darstellt (vgl. WALPUSKI et al. 2008). Das Kompetenzstrukturmodell soll für diesen Teilbereich die spezifische innere Logik der hypothetisch-deduktiven Vorgehensweise des Erkenntnisprozesses für verschiedene Untersuchungsmethoden wie Beobachten, Vergleichen, Ordnen und Experimentieren abbilden. Für die jeweiligen Untersuchungsmethoden sollen zentrale Teilkompetenzen und qualitative Kompetenzniveaus mit Hilfe des Modells konsistent beschrieben werden. Zur Überprüfung der postulierten Kompetenzstruktur und –graduierung werden Biologieaufgaben konstruiert, die hinsichtlich zweier Schwierigkeitserzeugender Merkmale systematisch variiert werden können.

2 Theoretische Grundlagen

Ausgehend von einem theorie- und evidenzbasierten Kompetenzmodell zur Erkenntnisgewinnung aus der Biologiedidaktik (GRUBE et al. 2007, MÖLLER et al. 2007, MAYER et al. in Druck) und einem entsprechenden Modell zum Fachwissen (NEUMANN et al. 2007, KAUERTZ et al. 2008) wurde im Rahmen des ESNaS-Projektes ein mehrdimensionales Kompetenzmodell entwickelt (vgl. WALPUSKI et al. 2008). Dieses bildet die Grundlage für die Beschreibung und Erfassung der in den Standards genannten Kompetenzen zur Erkenntnisgewinnung.

2.1 Das Kompetenzstrukturmodell zur Evaluation der Bildungsstandards

Die normativen Kompetenzbeschreibungen der nationalen Bildungsstandards klassifizieren die Kompetenzen nach den Merkmalen „Anforderungsbereiche“ und „Kompetenzbereiche“. Darüber hinaus wurde für eine Graduierung der Kompetenzen auf die Anforderungsbereiche für Abituraufgaben zurückgegriffen (vgl. EPA's), „da noch keine empirisch abgesicherten Kompetenzstufenmodelle“ vorlagen (KMK 2005, S. 16). Die Anforderungsbereiche werden genutzt,

um „zunächst zur *Einschätzung der in den Aufgabenbeispielen gestellten Anforderungen*“ (KMK 2005, S. 16) zu gelangen. Die Anforderungsbereiche sind keine Ausprägungen oder Niveaus einer Kompetenz, sondern Aufgabenmerkmale, „die verschiedene Schwierigkeitsgrade innerhalb ein und derselben Kompetenz abbilden können“ (KMK 2005, S. 16).

Für die Kompetenzmodellierung wird ein theoriebasiertes Kompetenzstrukturmodell genutzt, das der Evaluation der Bildungsstandards zu Grunde liegt:

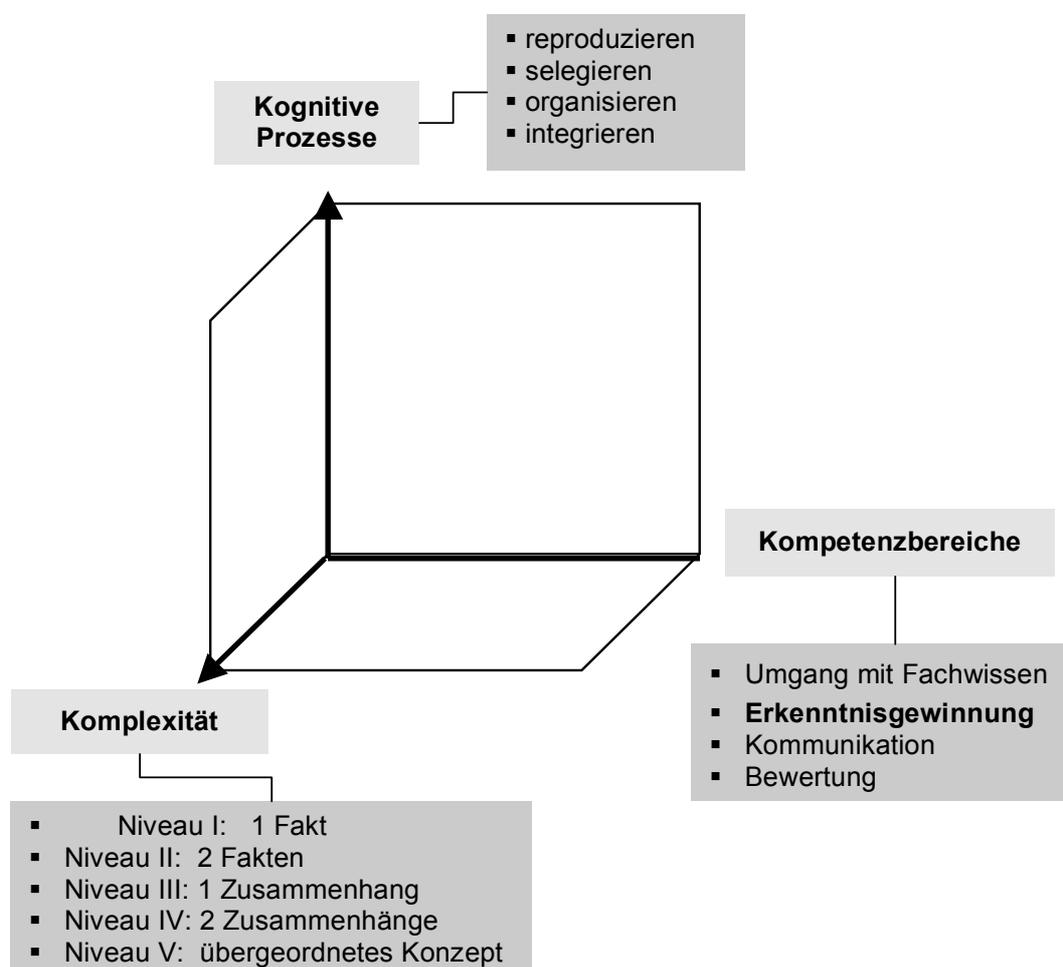


Abb. 1: Dreidimensionales Kompetenzstrukturmodell (vgl. WALPUSKI et al. 2008)

Die Anforderungsbereiche der KMK-Bildungsstandards wurden in diesem Modell in die Dimensionen „Kognitive Prozesse“ und „Komplexität“ ausdifferenziert, d.h. in zwei Merkmale, die als schwierigkeiterzeugend bekannt (NEUMANN et al. 2007, KAUERTZ et al. 2008) und bereits implizit in den Anforderungsbereichen enthalten sind. Das entwickelte Modell umfasst demnach drei

Dimensionen: Kognitive Prozesse, Komplexität und Kompetenzbereiche, wobei die kognitiven Prozesse und die Komplexität ordinal skaliert sind.

Die Schülerinnen und Schüler werden in den Aufgabenstellungen mit variablen Situationen konfrontiert, die kognitiv verarbeitet werden müssen. Der Schwierigkeitsgrad eines solchen Prozesses wird einerseits durch die „Komplexität“ und andererseits durch den „kognitiven Prozess“ bestimmt. Um Aufgaben unterschiedlicher Schwierigkeitsgrade zu konstruieren, werden diese Eigenschaften gezielt variiert bzw. kontrolliert (vgl. WALPUSKI et al. 2008).

Die **Dimension „Kognitive Prozesse“** *„beschreibt die zur erfolgreichen Anwendung von Kompetenz notwendigen kognitiven Informationsverarbeitungsstrategien im Sinne einer kognitiven Anforderung...“* (NEUMANN et al. 2007, S. 114). Eine Aufgabe ist umso schwieriger, je weniger die vorgegebenen Informationen mit den für die Lösung notwendigen Informationen übereinstimmen. Die Herstellung von Zusammenhängen oder der Transfer vorgegebener Informationen in einen neuen Kontext erhöhen ebenfalls die Aufgabenschwierigkeit (vgl. NEUMANN et al. 2007).

Die **Dimension „Komplexität“** beschreibt wie viele Informationen für die richtige Lösung einer Aufgabe von den Schülerinnen und Schülern verarbeitet werden müssen. Ein Ansteigen der Komplexität geht mit einer Zunahme von Lösungsschritten einher (vgl. NEUMANN et al. 2007).

Die **Dimension „Kompetenzbereiche“** beschreibt analog zu den Bildungsstandards die fachspezifischen inhaltlichen Komponenten. Unterschieden werden „Umgang mit Fachwissen“, „Erkenntnisgewinnung“, „Kommunikation“ und „Bewertung“. In allen Kompetenzbereichen wird kein deklaratives Wissen abgefragt, sondern die Kompetenz beim Umgang mit vorgegebenen Informationen (WALPUSKI et al. 2008). Während der Kompetenzbereich Fachwissen mittels Basiskonzepten bereits strukturiert vorlag, musste den Kompetenzen der Erkenntnisgewinnung zunächst eine explizite Struktur zugrunde gelegt werden.

2.2 Naturwissenschaftliche Erkenntnisgewinnung

Innerhalb des Kompetenzbereichs Erkenntnisgewinnung stellen „Naturwissenschaftliche Untersuchungen“ (scientific inquiry) – neben dem Naturwissenschaftsverständnis (nature of science) (vgl. KREMER et al. in Druck) – einen zentralen Teilbereich dar (Mayer 2007). Internationale Curricula verstehen darunter das Verständnis und die Fähigkeiten für die Durchführung des naturwissenschaftlichen Erkenntnisprozesses (AAAS 1989, NRC 1996). Aus lerntheoretischer Perspektive wird naturwissenschaftliche Erkenntnisgewinnung als

komplexer Problemlöseprozess mit spezifischen Prozeduren verstanden, in dem auf methodisches und inhaltliches Konzeptwissen zurückgegriffen wird. Die Güte der Problemlösung wird von der Qualität dieser Prozeduren sowie von Personenvariablen und Situationsvariablen bestimmt (MAYER 2007).

Naturwissenschaftliche Erkenntnisgewinnung kann auf vielgestaltige Art und Weise stattfinden. Auf der "Suche nach Wissen" kommen je nach Fragestellung verschiedene Untersuchungsmethoden und Instrumente zum Einsatz. In der Regel beginnt eine naturwissenschaftliche Untersuchung mit einem wahrnehmbaren Phänomen oder aus einer theoretischen Überlegung heraus, woraus eine naturwissenschaftliche Frage abgeleitet wird. Darauf aufbauend kann eine zu testende Hypothese formuliert und ein entsprechendes Untersuchungsdesign entworfen werden. Die gewonnenen Daten werden ausgewertet und die Ergebnisse interpretiert, um die eingangs aufgestellte Hypothese zu überprüfen und die zu Grunde liegende Theorie anzupassen.

Der Verlauf eines Forschungsprozesses erfordert demnach von den Schülerinnen und Schülern mehrere Teilkompetenzen, die in der Regel eine iterative und rekursive Abfolge von wissenschaftlichen Schritten umfassen, die zur Lösung eines Problems, einer naturwissenschaftlichen Frage, beitragen. Kompetenzmessung sollte die Binnenstruktur einer Kompetenz, d.h. diese Teilkompetenzen genauso wie die Niveaustufen explizit darstellen (KLIEME & LEUTNER 2006). Als Binnenstruktur des Experimentierens können vier eigenständige Teilkompetenzen identifiziert werden: naturwissenschaftliche Fragen formulieren, Hypothesen generieren, Untersuchungen planen sowie Daten interpretieren (GRUBE et al. 2007). Diese Teilkompetenzen bereiten den Lernenden mehr oder weniger spezifische Probleme, deren Ursachen von den verschiedenen Autoren unterschiedlich interpretiert werden (vgl. CAREY et al. 1989, GOTT & DUGGAN 1995, DUGGAN et al. 1996, KLAHR 2000, HAMMANN et al. 2007). Für jede Teilkompetenz lassen sich qualitative Kompetenzniveaus unterscheiden, die sich hinsichtlich Komplexität und qualitativer Abstufung im Sinne eines Problemlöseprozesses zuordnen lassen (MÖLLER et al. 2007).

2.3 Untersuchungsmethoden der Biologie

Ein wesentliches Element wissenschaftsmethodischen Verständnisses ist, dass Wissenschaftler unterschiedliche Untersuchungsmethoden nutzen, um zur Lösung eines naturwissenschaftlichen Problems zu gelangen (vgl. GROPPENGIEBER & KATTMANN 2006). In der Biologie nehmen vor allem das Beobachten (STAECK 1998, KALAS 2002, BREZMANN 2003, KILLERMANN et al. 2005), Vergleichen (HAMMANN 2002, WASMANN-FRAHM 2007), Ordnen (HAMMANN 2004, KRÜGER

& BURMESTER 2005, KATTMANN 2007) und Experimentieren (HAMMANN et al. 2006, MAYER 2006) einen zentralen Stellenwert ein. Die Kenntnis und Beherrschung dieser Methoden wird explizit in den Bildungsstandards im Fach Biologie gefordert (vgl. KMK 2005). Die genannten Methoden weisen hinsichtlich der zu untersuchenden Zusammenhänge jeweils eine spezifische Charakteristik auf (s. Tab. 1). Dabei werden unter dem Terminus „Untersuchungsmethoden“ nur solche verstanden, die das spezifisch hypothetisch-deduktive Vorgehen der Naturwissenschaften verlangen. Bei jeder Methodik ist der Verlauf des Forschungsprozesses immer an die Lösung eines naturwissenschaftlichen Problems durch Hypothesenbildung und deren Überprüfung gebunden.

Naturwissenschaftliches **Beobachten** ist aufmerksamkeitsgeleitetes und an bestimmten Kriterien orientiertes planmäßiges Erfassen von Objektmerkmalen und deren zeitlichen Veränderungen (STURM 1974). Es dient einem bestimmten Forschungszweck bei dem nichts dem Zufall überlassen wird. In wissenschaftsmethodischer Hinsicht stellt das Beobachten eine komplexe Erkenntnismethode dar, welche sich aus dem systematischen Planen, dem aufmerksamkeitsgeleiteten Beobachten, der systematischen Beschreibung und Aufzeichnung, dem Zählen und der Hinterfragung von Kausalzusammenhängen zusammensetzt. Sie erfordert eine hypothetisch-deduktive Vorgehensweise. Die naturwissenschaftliche Beobachtung ist, wie andere Untersuchungsmethoden auch, wiederholten Prüfungen und Kontrollen hinsichtlich der Gültigkeit, Zuverlässigkeit und Genauigkeit der Daten unterworfen. Im Gegensatz zum Experimentieren werden die zu beobachtenden Merkmale oder Eigenschaften nicht künstlich isoliert oder variiert.

Eine Beobachtung kann direkt oder indirekt erfolgen (MAHNER & BUNGE 2000). Bei einer direkten Beobachtung werden keine grundlegenden Veränderungen am Erkenntnisobjekt oder deren zeitlichen Veränderungen vorgenommen, da die Erfassung bestimmter Faktoren mit den bloßen Sinnesorganen erfolgt (STURM 1974).

Moleküle, Gene oder Zellorganellen können nicht direkt beobachtet werden. Die meisten Beobachtungen sind deshalb indirekte Beobachtungen (Untersuchungen), bei denen technische Hilfsmittel (Mikroskop, Reagenzien, Sezierbesteck) zur Anwendung kommen. Die indirekte Beobachtung ist stets hypothesen- oder theorieabhängig, da ihre Vermutungen auf theoretischen Vorüberlegungen basieren (MAHNER & BUNGE, 2000). *„Ohne theoretischen Hintergrund wären ... die von indirekten Beobachtungen gelieferten Rohdaten wissenschaftlich unbegründet ... unverständlich und somit nutzlos, weil sie keinen offenkundigen Bezug zu dem beobachteten Faktum haben“* (MAHNER & BUNGE

2000, S. 70). Der Einsatz von Hilfsmitteln ermöglicht es, direkt beobachtbare Fakten wie Bilder auf einem Monitor, Töne oder Farbumschläge zu erhalten.

Tab. 1: Hypothetisch-deduktive Schritte des wissenschaftlichen Erkenntnisprozesses

Erkenntnismethoden				
Aspekte	Beobachten	Vergleichen	Ordnen	Experimentieren
Frage	Welche Merkmale, Strukturen, räumlichen Beziehungen oder zeitlichen Abläufe lassen sich beim Objekt identifizieren?	Durch welche Unterschiede, Gemeinsamkeiten oder Ähnlichkeiten in den Merkmalen zeichnen sich die Objekte aus?	Welche Objekte mit konstanten gemeinsamen oder ähnlichen Merkmalen lassen sich einer Kategorie zuordnen?	Welcher Zusammenhang besteht zwischen den Faktoren x und y?
Hypothese	Das Objekt zeigt das Merkmal X in der Ausprägung x. Das Objekt besteht aus den Teilen x, y, z.	Die Objekte A und B gleichen sich im Merkmal x und unterscheiden sich im Merkmal y.	Alle A und B weisen im Gegensatz zu C und D typische gemeinsame Merkmale auf und gehören damit der Kategorie AB an.	Die (unabhängige) Variable x hat einen Einfluss auf die (abhängige) Variable y.
Untersuchungsdesign	<ul style="list-style-type: none"> ▪ Planung, Durchführung und Datengewinnung ▪ Wahl des Erkenntnisgegenstandes und der Hilfsmittel ▪ Versuchsaufbau; Festlegung von Ort, Zeitpunkt und Zeitspanne ▪ Datenerfassung; qualitativ oder quantitativ; Identifizierung von Störgrößen 			
	<p>planmäßig</p> <ul style="list-style-type: none"> ▪ Beobachtung eines oder mehrerer Objekte ohne Ähnlichkeiten zu erfassen 	<p>kriteriengeleitet</p> <ul style="list-style-type: none"> ▪ Gegenüberstellen von mindestens zwei Objekten ▪ vergleichen 	<p>systematisch</p> <ul style="list-style-type: none"> ▪ Gegenüberstellen von mehreren Objekten ▪ sortieren 	<p>kontrolliert</p> <ul style="list-style-type: none"> ▪ Identifizierung der abhängigen (AV), unabhängigen (UV) und zu kontrollierenden (KV) Variablen
	<ul style="list-style-type: none"> ▪ Nennung der spezifischen Kriterien und ihren Ausprägungen 			
	<ul style="list-style-type: none"> ▪ Beobachtung von dynamischen Systemeigenschaften 	<ul style="list-style-type: none"> ▪ Berücksichtigung der Merkmalsvariation und Kriterienstetigkeit ▪ Innergruppen- oder Zwischengruppenvergleich 	<ul style="list-style-type: none"> ▪ Isolierung und systematische Variation ▪ Kontrolle 	
Datenauswertung	<ul style="list-style-type: none"> ▪ Dokumentation der Daten (Beschreibung, Zeichnung, Messwerte, Fotos) ▪ Aufbereitung der Daten (Tabelle, Diagramm, Berechnung) ▪ Analyse (Genauigkeit, Fehler, Störgrößen) und Interpretation der Daten 			
	<ul style="list-style-type: none"> ▪ korrelative Zusammenhänge ▪ Erklärung durch morphologisch-anatomische bzw. Form-Funktions-Zusammenhänge 	<ul style="list-style-type: none"> ▪ Prüfung der Äquivalenzrelation ▪ Phylogenetische oder ökologische Erklärung eines Phänomens (Homologien, Analogien) 	<ul style="list-style-type: none"> ▪ Ordnungssystem mit übergeordnetem Vergleichskriterium ▪ Klassifizierung: taxonomisch-systematisch ▪ Stammbaumrekonstruktion 	<ul style="list-style-type: none"> ○ kausale Zusammenhänge ○ Erklärung eines Phänomens durch seine physiologischen, ökologischen oder genetischen Ursachen
		<ul style="list-style-type: none"> ▪ Bestimmen (Prüfung, ob Merkmale im Schlüssel, mit denen des vorliegenden Objekts übereinstimmen) 		

Beim kriteriengeleiteten **Vergleichen** werden mindestens zwei Objekte oder Vorgänge einander gegenübergestellt und bezüglich der Ähnlichkeiten oder Gemeinsamkeiten und Unterschiede erfasst (BREZMANN 2004). Das Vergleichen erfolgt nach ausgewählten Kriterien innerhalb einer Gruppe von Objekten oder Vorgängen (Innergruppen-Vergleich) oder zwischen verschiedenen Gruppen (Zwischengruppen-Vergleich). Welche Objekte sich ähneln, hängt nicht nur davon ab, welche Eigenschaften sie besitzen, sondern auch davon, wer den Vergleich anstellt und zu welchem Zweck. Ähnlichkeiten sind also kontextabhängig (HAMMANN 2002).

Das Vergleichen kann zum systematischen **Ordnen** führen, indem Objekte oder Vorgänge mit ähnlichen Eigenschaften oder Merkmalen in Kategorien zusammengefasst werden. Dabei spielen Ähnlichkeitserwägungen ebenso eine Rolle wie die Anwendung von Regeln (KEIL et al. 1998). Das Ordnen kann nach verschiedenen Kriterien erfolgen und entsprechend unterschiedliche Ergebnisse liefern (KRÜGER & BURMESTER 2005). Es erfordert die systematische Verwendung von unterschiedlichen Ordnungskriterien und die Vermeidung von Kriterienwechsel innerhalb eines Klassifikationssystems. Dabei ist es sinnvoll zwischen einem Kriterium (z.B. Form) und der Ausprägung eines Kriteriums (z.B. oval, rund, eckig) zu differenzieren. Ein Klassifikationssystem kann dann als kriterienstet bezeichnet werden, „*wenn für die Bildung aller Gruppen eines Klassifikationssystems ausschließlich die Ausprägungen eines einzigen Kriteriums verwendet werden*“ (HAMMANN 2004, S. 41).

Beim **Experimentieren** wird zielgerichtet in einen Ablauf eingegriffen. Es werden zur Aufdeckung von kausalen Zusammenhängen künstlich veränderte Bedingungen kontrolliert, unter denen die zu erforschende Erscheinung verläuft. Durch die systematische Isolation, Variation und Kombination einzelner Faktoren können die Abhängigkeit bestimmter Abläufe von diesen Faktoren erkannt und Gesetzmäßigkeiten abgeleitet werden. Als wichtige Faktoren müssen beim Experimentieren die abhängige Variable (AV), eine oder mehrere unabhängige Variablen (UV) und die Kontrollvariable (KV) identifiziert werden.

Die unabhängige Variable wird gezielt verändert (z.B. Licht), um deren Auswirkungen auf die abhängige Variable (z.B. Pflanzenwachstum) zu erfassen. Die unabhängige Variable ist die vermutete Einflussgröße. Die Veränderung der abhängigen Variablen wird infolge des Einflusses der unabhängigen Variablen gemessen. Die Kontrollvariable kann neben der zu untersuchenden unabhängigen Variablen ebenfalls einen Einfluss auf die abhängige Variable besitzen, jedoch steht sie nicht im Zentrum des gewählten Forschungsansatzes.

Die Kontrollvariable wird gezielt kontrolliert, um sicherzustellen, dass die Veränderung der abhängigen Variable, allein auf die Variation der unabhängigen Variable zurückzuführen ist (WATSON et al. 2000, ROBERTS 2001, ROBERTS & GOTT 2003).

3 Fragestellungen

Aus den oben dargelegten theoretischen Grundlagen lassen sich folgende Fragestellungen ableiten:

- Lässt sich die Kompetenzstruktur mit vier Teilkompetenzen (Fragestellung, Hypothese, Untersuchungsdesign, Datenauswertung) anhand des dreidimensionalen ESNaS-Modells für den Kompetenzteilbereich „Naturwissenschaftliche Untersuchungen“ bestätigen?
- Lassen sich verschiedene Untersuchungsmethoden (Beobachten, Vergleichen, Ordnen, Experimentieren) als Subkompetenzen des Modells abbilden?
- Können die Kompetenzen mittels der beiden Dimensionen „Komplexität“ und „kognitive Prozesse“ in Kompetenzniveaus graduiert werden?
- Welche Zusammenhänge zeigen sich zwischen Kompetenzen „Naturwissenschaftlicher Untersuchungen“ und dem Fachwissen (deklaratives Wissen)?

4 Methodik

Für die Entwicklung eines geeigneten Testinstruments zur Erfassung von naturwissenschaftlichen Lernerkompetenzen wird das theoriebasierte Kompetenzstrukturmodell genutzt, das der Evaluation der Bildungsstandards zu Grunde liegt (vgl. WALPUSKI et al. 2008).

4.1 Datenerhebung

Die Datenerhebung erfolgt im Rahmen des Kooperationsprojektes „Evaluation der Standards in den Naturwissenschaften für die Sekundarstufe I (ESNaS)“ (WALPUSKI et al. 2008). Darin werden Schülerinnen und Schüler der 9. und 10. Jahrgangsstufe aller drei Schulformen einem Leistungstest zum Kompetenzteilbereich „Naturwissenschaftliche Untersuchung“ unterzogen. Zunächst werden ausgewählte Aufgaben in einer Vorstudie präpilotiert (N=150). In der Hauptuntersuchung (2009) werden die überarbeiteten Aufgaben in einer bundesweiten Pilotierung (N=3000) eingesetzt.

In der Vor- und Hauptstudie werden neben der Aufgabenschwierigkeit als Beurteilungskriterien, mit einem zeitgleich eingesetzten Schülerfragebogen

Zeugnisnote (Biologie, Chemie, Deutsch, Physik), Vertrautheit mit dem Kontext und Motivation erfasst. Zur Bearbeitung eines Testheftes einschließlich des Schülerfragebogens sind ca. 35 Minuten vorgesehen. Zehn Minuten dienen der Einführung durch die Testleiter. Die erhobenen Daten aus der Vorstudie werden der Aufgabenanalyse unterzogen, um eine Entscheidungsgrundlage bei der Aufgabenauswahl für die Hauptuntersuchung zu ermöglichen.

4.2 Testinstrumente

Für die Datenerhebung werden Aufgaben mit ca. 100 Items zum Beobachten, Vergleichen, Ordnen und Experimentieren eingesetzt. Die Aufgaben umfassen die Teilkompetenzen „Fragestellung“, „Hypothese“, „Untersuchungsdesign“ und „Datenauswertung“. Die eingesetzten Aufgaben weisen drei verschiedene Antwortformate auf. Es werden geschlossene (multiple-choice), halboffene (Kurzantwort) und offene Antwortformate konstruiert und getestet, die sich hinsichtlich ihrer Schwierigkeit gezielt modifizieren und kontrollieren lassen (s. Abb. 2 und 3).

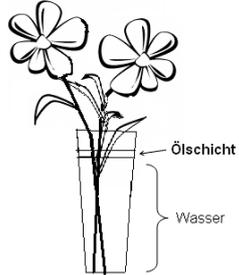
2 Zusammenhänge integrieren¹ zum Aspekt „Hypothese“ (Beobachtung)		
Aufgabenstamm: Der Wasserfernttransport bei Pflanzen funktioniert durch den Transpirationssog. Das ist jene Kraft, die durch die Wasserabgabe der Blätter einer Pflanze entsteht. Über die Spaltöffnungen der Blätter wird Wasser verdunstet und es entsteht ein Sog, der dafür sorgt, dass das Wasser aus der Wurzel durch die Pflanze nach oben gezogen wird.		
Item: Auf dem Foto siehst du Schnittblumen, die in einem Glas mit Wasser stehen. Die Ölschicht soll die Verdunstung des Wassers verhindern. Welche Vermutung (Hypothese) hast du, was nach einigen Tagen zu beobachten sein wird?		
Die Wassersäule wird ansteigen, weil das in der Pflanze gespeicherte Wasser aufgrund der Schnittfläche am Spross austreten wird.	<input type="checkbox"/>	
Die Wassersäule bleibt unverändert, weil die zähe Ölschicht auf der Wasseroberfläche einen An- oder Abstieg der Wassersäule verhindert.	<input type="checkbox"/>	
Die Wassersäule wird sinken, weil an den Blättern eine Saugkraft besteht, die das Wasser von der Schnittfläche am Spross bis zu den Blättern hoch transportiert.	<input checked="" type="checkbox"/>	
Die Wassersäule wird sinken, weil das Wasser in dem Glas langsam verdunstet.	<input type="checkbox"/>	

Abb. 2: Aufgabenbeispiel nach LISA (2004, S. 38); verändert

- 1 Beim Integrieren müssen die Informationen aus dem Aufgabenstamm von den Schülerinnen und Schülern angewendet und in eine andere Situation übertragen werden. Die richtige Antwort ist nicht im Aufgaben- oder Itemstamm gegeben (vgl. NEUMANN et al. 2007).

Die Zusammenstellung der Testhefte erfolgt nach dem Multi-Matrix-Design (vgl. PRENZEL et al. 2007) Dabei werden alle Testhefte mit systematisch variierenden Aufgaben erstellt und zufällig den Schülerinnen und Schülern zugewiesen. Aufgrund dieses Erhebungsdesigns werden von den Schülerinnen und Schülern nur Teilmengen der insgesamt eingesetzten Aufgaben bearbeitet. Alle Testhefte sind deshalb durch Ankeritems miteinander verlinkt, so dass die verschiedenen Testhefte einige identische Aufgaben enthalten. Die Aufgabenschwierigkeit kann somit über die Testhefte hinweg miteinander in Beziehung gesetzt werden (vgl. CARSTENSEN et al. 2007).

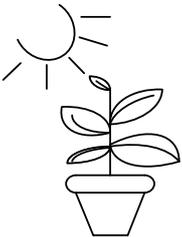
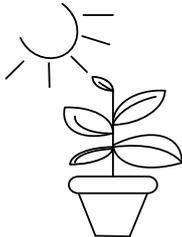
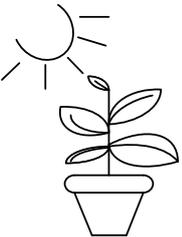
1 Fakt reproduzieren ² zum Aspekt „Hypothese“ (Experiment)		1 Fakt selektieren ³ zum Aspekt „Hypothese“ (Experiment)	
Aufgabenstamm: Deine Biologielehrerin bereitet zum Thema Pflanzenwachstum ein Experiment vor. Dazu stellt sie zwei unterschiedliche Versuchsansätze (Pflanzen) auf die Fensterbank:		Aufgabenstamm: Deine Biologielehrerin bereitet zum Thema Pflanzenwachstum ein Experiment vor. Dazu stellt sie zwei unterschiedliche Versuchsansätze (Pflanzen) auf die Fensterbank:	
<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>Sonnenlicht</p>  <p>Sand, Mineralstoffe und Wasser</p> </div> <div style="text-align: center;"> <p>Sonnenlicht</p>  <p>Sand und Wasser</p> </div> </div>		<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>Sonnenlicht</p>  <p>Sand, Mineralstoffe und Wasser</p> </div> <div style="text-align: center;"> <p>Sonnenlicht</p>  <p>Sand und Wasser</p> </div> </div>	
Item: Welche Vermutung (Hypothese) will sie untersuchen? Kreuze an!		Item: Welche Vermutung (Hypothese) will sie untersuchen? Kreuze an!	
Das Wachstum einer Pflanze ist abhängig von...		Das Wachstum einer Pflanze ist abhängig von...	
...den Mineralstoffen.	<input checked="" type="checkbox"/>	...den Mineralstoffen.	<input checked="" type="checkbox"/>
...dem Kohlenstoffdioxid.	<input type="checkbox"/>	...dem Sonnenlicht.	<input type="checkbox"/>
...der Temperatur.	<input type="checkbox"/>	...dem Sandboden.	<input type="checkbox"/>
...dem Sauerstoff.	<input type="checkbox"/>	...dem Wasser.	<input type="checkbox"/>

Abb. 3: Aufgabenbeispiele nach BAUMERT et al. (1998, S. 75); verändert

² Beim Reproduzieren ist die richtige Antwort mit der Vorgabe im Aufgabenstamm inhaltlich identisch. Die falschen Antwortoptionen sind nicht im Aufgaben- oder Itemstamm enthalten (vgl. NEUMANN et al. 2007).

³ Beim Selektieren ist die richtige Antwort mit der Vorgabe im Aufgabenstamm inhaltlich identisch. Für die falschen Antwortoptionen werden Informationen aus dem Aufgabenstamm aufgegriffen, sodass die Probanden die richtige Lösung auswählen müssen (vgl. NEUMANN et al. 2007).

Bei der Zusammenstellung der Testhefte wird auf eine ausgewogene Verteilung der Aufgaben geachtet. Einerseits sollten sich die Anforderungen beziehungsweise Schwierigkeiten der Testhefte durch die Aufgabenauswahl möglichst wenig unterscheiden und andererseits sollte innerhalb eines Testheftes selbst, eine möglichst optimale Verteilung von schwierigeren und leichteren Aufgaben erfolgen. Durch die Positionierung der Aufgaben sollen Effekte minimiert werden. Somit soll verhindert werden, dass leistungsstarke Lerner gleich zu Beginn durch zu einfache Aufgaben unterfordert bzw. leistungsschwächere Lerner durch zu schwere Aufgaben überfordert werden. Bei der weiteren Bearbeitung der Testhefte finden die Schülerinnen und Schüler, in einem angemessenen Wechsel, mal schwierigere, mal leichtere Aufgaben vor.

4.3 Auswertung

Unter Verwendung probabilistischer Testtheorien (Item-Response-Theorie) werden zur Berechnung von Kompetenzwerten die erhobenen Daten vergleichend analysiert (ROST 2004). Die Auswertung erfolgt mit dem Programm ConQuest (WU et al. 1997). Somit kann eine Voraussage darüber getroffen werden, ob das Kompetenzstrukturmodell geeignet ist, wissenschaftsmethodische Kompetenzen im Fach Biologie zuverlässig und präzise abzubilden. Der Nachweis der Messgütekriterien des entwickelten Testinstruments findet anhand von Validitäts- und Reliabilitätsindikatoren statt. Die durch das Kompetenzmodell beschriebenen Dimensionen werden in reliable Skalen überführt, indem die angenommene Binnenstruktur mit Hilfe der probabilistischen Testtheorie überprüft wird.

Ausblick

Im Frühjahr 2009 wird im Rahmen des Dissertationsvorhabens eine erste Datenerhebung der Aufgaben an einer bundesweiten Stichprobe vorgenommen.

Zitierte Literatur

- AAAS, American Association for the Advancement of Science (1989): Science for all Americans. Washington, DC.
- AAAS, American Association for the Advancement of Science (1993): Benchmarks for Science Literacy. New York, Oxford University Press.
- BAUMERT, J., LEHMANN, R., LEHRKE, M. u. a. (1998): TIMMS - Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde. Leske + Budrich, Opladen.

- BREZMANN, S. (2003): Erkenntnistätigkeiten im Unterricht. Teil 1: Einleitung, das Beobachten und Beschreiben. In: PdN-BioS 7 (52), 43- 45.
- BREZMANN, S. (2004): Erkenntnistätigkeiten im Unterricht. Teil 5: Das Definieren und Vergleichen. In: PdN-BioS 4 (53), 28- 30.
- BYBEE, R. W. (2002): Scientific Literacy – Mythos oder Realität? In: W. Gräber, P. Nentwig, T. Koballa, R. Ewans [Hrsg.]: Scientific Literacy. Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung, Leske + Budrich, Opladen, 21-43.
- CAREY, S., EVANS, R., HONDA, M., JAY, E. & C. UNGER (1989): ‘An experiment is when you try it and see if it works’: a study of grade 7 students understanding of the construction of scientific knowledge. *International Journal of Science Education* 11, 514-529.
- CARSTENSEN, C. H., FREY, A., WALTER, O. & S. KNOLL (2007): Technische Grundlagen des dritten internationalen Vergleichs. In: M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun [Hrsg.]: PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie. Waxmann, Münster, 367-390.
- DfES (Department for Education and Skills) & QCA (Qualification and Curriculum Authority) (1999): Science. The National Curriculum for England. HMSO, London.
- DUGGAN, S., JOHNSON, P. & R. GOTT (1996): A Critical Point in Investigative Work: Defining Variables. *Journal of Research in Science Teaching* 33 (5), 461-474.
- GOTT, R. & S. DUGGAN (1995): *Investigative Work in Science Curriculum*. Open Univ. Press, Buckingham.
- GOTT, R. & S. DUGGAN (1998): Understanding Scientific Evidence – Why It Matters and How It Can Be Thought. In: M. Ratcliffe (1998): *ASE Guide to Secondary Science Education*. Hatfield.
- GROPENGIEßER, H. & U. KATTMANN (2006): *Fachdidaktik Biologie*. Aulis Verlag Deubner, Köln.
- GRUBE, C., MÖLLER, A. & J. MAYER (2007): Dimensionen eines Kompetenzstrukturmodells zum Experimentieren. In: H. Bayrhuber, U. Harms, D. Krüger, A. Sandmann, U. Unterbrunner, A. Upmeyer zu Belzen & H. Vogt [Hrsg.] (2007): *Ausbildung und Professionalisierung von Lehrkräften*. Internationale Tagung der Fachgruppe Biologiedidaktik im VBIO – Verband Biologie, Biowissenschaften & Biomedizin, Essen, 31-34.
- HAMMANN, M. (2002): *Kriteriengeleitetes Vergleichen im Biologieunterricht*. StudienVerlag, Innsbruck.
- HAMMANN, M. (2004): Tiere ordnen. Ein Methodentraining zum kriteriengeleiteten Vergleichen. In: R. Duit, R. Gropengießer & H. Stäudel [Hrsg.] (2004): *Naturwissenschaftliches Arbeiten-Unterricht und Material für die Fächer Biologie, Chemie und Physik*. Friedrich Verlag, 38-46, Seelze.
- HAMMANN, M. (2007): Das Scientific Discovery as Dual Search-Modell. In: D. Krüger & H. Vogt [Hrsg.]: *Theorien in der biologiedidaktischen Forschung*. Springer, Berlin/Heidelberg, 187-196.
- HAMMANN, M., PHAN, T.T.H. & H. BAYRHUBER (2006): Fehlerfrei Experimentieren. In: *MNU* 59 (5), 292-299.
- KATTMANN, U. [Hrsg.] (2007): *Ordnen und Bestimmen. Einheiten in der Vielfalt. Unterricht Biologie Kompakt*, Friedrich Verlag, Velber.
- KALAS, S. (2002): Schauen – Fragen – Beobachten – Erforschen. In: *UB* 276 (26), 4-11.
- KILLERMANN, W., HIERING, P. & B. STAROSTA (2005): *Biologieunterricht heute. Eine moderne Fachdidaktik*. Auer Verlag, Donauwörth.
- KLAHR, D. (2000): *Exploring Science. The Cognition and Development of Discovery Processes*. MIT, Cambridge.
- KAUERTZ, A. FISCHER, H. E., LAU, A. & K. NEUMANN (2008): Kompetenzmessung durch Leistungstests – Hilfe oder Druckmittel? In: *MNU*, 61 (2), 75-79.
- KEIL, F. C., SMITH, W. C., SIMONS, D. J. & D. T. LEVIN (1998): Two Dogmas of Conceptual Empiricism: Implications for Hybrid Models of the Structure of Knowledge. In: *Cognition* 65, 103-135.
- KLIEME, E. & D. LEUTNER (2006): Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. In: *Z.f.Päd.* 52 (6): 876-902.
- KLIEME, E., AVENARIUS, H., BLUM, W., DÖBRICH, P., GRUBER, H., PRENZEL, M., REISS, K., RIQUARTS, K., ROST, J., TENORTH, H.-E. & H. J. Vollmer (2003): *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Bundesministerium für Bildung und Forschung (BMBF), Bonn.
- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005): *Beschlüsse der Kultusministerkonferenz. Bildungsstandards im Fach Biolo-*

- gie für den Mittleren Schulabschluss (Jahrgangsstufe 10). Beschluss vom 16.12.2004. Luchterhand, München.
- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2006): Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring. LinkLuchterhand, Bonn.
- KREMER, K., URHAHNE, D. & J. MAYER (in Druck): Naturwissenschaftsverständnis und wissenschaftliches Denken bei Schülerinnen und Schülern der Sek. I. In: U. Harms & A. Sandmann [Hrsg.]: *Ausbildung und Professionalisierung von Lehrkräften*. Band 3.
- KRÜGER, D. & A. BURMESTER (2005): Wie Schüler Pflanzen ordnen. In: *ZfDN* **11**, 85-102.
- LEDERMANN, N. G. (2006): "Definitions" of nature of science and scientific inquiry that guide project ICAN: A cheat sheet. <http://www.projectican.com/documents/Subjectmatterwithoutcontext-acheatsheet.pdf>; download: 03.05.2007.
- LEDERMANN, N. G. (2007): Nature of Science: Past, present, and future. In: S. K. Abell & N. G. Ledermann [Hrsg.]: *Handbook of research on science education*. Mahwah, NJ, Lawrence Erlbaum Associates, 831-879.
- LISA, Landesinstitut für Lehrerfortbildung, Lehrerweiterbildung und Unterrichtsforschung von Sachsen-Anhalt (2004): Niveaubestimmende Aufgaben für den naturwissenschaftlichen Unterricht – Schuljahrgang 6. Kultusministerium Sachsen-Anhalt, Halle, <http://www.rahmenrichtlinien.bildung-lsa.de/forum/niveau/nivnawi6.pdf>; download: 26.01.2008.
- MAHNER, M. & M. BUNGE (2000): *Philosophische Grundlagen der Biologie*. Springer, Berlin.
- MAYER, J. [Hrsg.] (2006): *Offenes Experimentieren*. Unterricht Biologie 317, Friedrich Verlag, Velber.
- MAYER, J. (2007): Erkenntnisgewinnung als wissenschaftliches Problemlösen. In: D. Krüger & H. Vogt [Hrsg.]: *Theorien in der biologiedidaktischen Forschung*. Springer, Berlin/Heidelberg, 177-186.
- MAYER, J., GRUBE, C. & A. MÖLLER (in Druck): Kompetenzmodell naturwissenschaftlicher Erkenntnisgewinnung. In: U. Harms & A. Sandmann [Hrsg.]: *Ausbildung und Professionalisierung von Lehrkräften*. Band 3.
- MÖLLER, A., GRUBE, C. & J. MAYER (2007): Kompetenzniveaus der Erkenntnisgewinnung bei Schülerinnen und Schülern der Sekundarstufe I. In: H. Bayrhuber, U. Harms, D. Krüger, A. Sandmann, U. Unterbrunner, A. Upmeyer zu Belzen & H. Vogt [Hrsg.] (2007): *Ausbildung und Professionalisierung von Lehrkräften*. Internationale Tagung der Fachgruppe Biologiedidaktik im VBIO – Verband Biologie, Biowissenschaften & Biomedizin, Essen, 55-58.
- NEUMANN, K., KAUERTZ, A., LAU, A. & H. E. FISCHER (2007): Die Modellierung physikalischer Kompetenz und ihrer Entwicklung. In: *ZfDN* **13**: 103-123.
- NRC, National Research Council (1996): *National science education standards*. National Academy Press, Washington, DC.
- PRENZEL, M., CARSTENSEN, C. H., FREY, A., DRECHSEL, B. & S. RÖNNEBECK (2007): PISA 2006 – Eine Einführung in die Studie. In: M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun [Hrsg.]: *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Waxmann, Münster, 31-59.
- ROBERTS, R. (2001): Procedural understanding in biology: the 'thinking behind the doing'. In: *Journal of Biological Education* **35** (3), 113-117.
- ROBERTS, R. & R. GOTT (2003): Assessment of biology investigations. In: *Journal of Biological Education* **37** (3), 114-121.
- STAECK, L. (1998): Praktisches Arbeiten im Biologieunterricht. Teil 2: Das Beobachten. In: *Biologie in der Schule* **47** (2), 65-67.
- ROST, J. (2004): *Lehrbuch Testtheorie – Testkonstruktion*. (2. Aufl.), Hans Huber, Bern.
- STURM, H. (1974): Beobachtung im Biologieunterricht – ein Versuch zur Begriffsklärung. In: *MNU* **27** (6), 339-344.
- WALPUSKI, M., KAMPA, N., KAUERTZ, A. & N. WELLNITZ (2008): Evaluation der Bildungsstandards in den Naturwissenschaften. In: *MNU* **61** (6), 323-326.
- WASMANN-FRAHM (2007): Erkenntnismethode als Werkzeug selbst gesteuerten Lernens. Einführung in die Wirbeltiere mit einer Methodenschulung im Vergleichen. In: *MNU* **60** (3), 162-171.
- WATSON, R., GOLDSWORTHY, A. & V. WOOD-ROBINSON (2000): Beyond the fair test. In: J. Sears & P. Sorensen [Hrsg.]: *Issues in Science Teaching*. RoutledgeFalmer, London, New York, 70-79.
- WU, M. L., ADAMS, R. J. & M. WILSON (1997): *ConQuest - Generalised item response modelling software*, Draft Release 2. Australian Council for Educational Research, Camberwell.