

THILO KLEICKMANN, ILONCA HARDY, KORNELIA MÖLLER, JUDITH POLLMEIER, STEFFEN TRÖBST UND CHRISTINA BEINBRECH

Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter: Theoretische Konzeption und Testkonstruktion

Modeling scientific competence of primary school children: Theoretical background and test construction

Zusammenfassung

Zur Beschreibung naturwissenschaftlicher Kompetenz von Schülerinnen und Schülern lagen bislang in erster Linie normativ entwickelte Ansätze vor. Zwar gibt es für den Sekundarbereich bereits einige empirisch überprüfte Modelle zur Beschreibung dieser Kompetenzen, für den Primarbereich ist die Forschungslage diesbezüglich jedoch äußerst dünn, insbesondere was die Modellierung der Entwicklung naturwissenschaftlicher Kompetenzen angeht. Ziel des Kooperations-Projektes Science-P ist daher die theoretische Modellierung und psychometrische Überprüfung der Entwicklung naturwissenschaftlicher Kompetenz in der Grundschule. Als zentrale Bereiche naturwissenschaftlicher Kompetenz werden die Bereiche „Naturwissenschaftliches Wissen“ und „Wissen über Naturwissenschaften“ untersucht. Dabei werden in beiden Bereichen die Kompetenzniveaus „naïve Vorstellungen“, „Zwischenvorstellungen“ und „wissenschaftliche Vorstellungen“ unterschieden. Das methodische Vorgehen orientiert sich an der konstrukt-basierten Instrumententwicklung nach Wilson. Ergebnisse einer Pilotstudie im Bereich „Naturwissenschaftliches Wissen“ geben erste Hinweise auf Reliabilität und Validität der entwickelten Aufgaben. Schlüsselwörter: Naturwissenschaftliche Kompetenz, Kompetenzmodelle, Grundschule, Testkonstruktion

Abstract

Up to now, scientific competence of school children had been mainly described by normative approaches. For the secondary level, some empirically validated models of scientific competence exist, but for the primary level such research is almost missing, especially concerning the development of scientific competence. Against this background, the project Science-P aims at theoretically modeling and empirically testing the development of scientific competence of primary school children. The project focuses on two main dimensions of scientific competence: “scientific knowledge” and “knowledge of the nature and methods of science”. In both dimensions the competence levels of “naïve conceptions”, “intermediate conceptions” and “scientific conceptions” are distinguished. In order to construct measures capturing scientific competence, we followed the approach of construct-based test construction suggested by Wilson. Results of a pilot study provide some evidence that reliable and valid measures had been constructed. Keywords: Scientific competence, competence models, primary school, test construction

Das Projekt Science-P ist ein im Rahmen des DFG-Schwerpunktprogrammes „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“ (SPP 1293, Klieme & Leutner, 2006) gefördertes Projekt mit dem Ziel der theoretischen Modellierung und

psychometrischen Erfassung der Entwicklung naturwissenschaftlicher Kompetenz in den beiden Kompetenzbereichen *Naturwissenschaftliches Wissen* und *Wissen über Naturwissenschaften* von der zweiten bis zur vierten Klassenstufe. Science-P ist als interdisziplinäres Projekt angelegt, in

dem Vertreter der Psychologie, der Sachunterrichtsdidaktik und der Erziehungswissenschaft zusammenarbeiten (Pollmeier et al., 2009).

In diesem Beitrag wird auf theoretische Grundlagen der Entwicklung des Kompetenzmodells für den Kompetenzbereich *Naturwissenschaftliches Wissen* und auf die Entwicklung von gruppentestfähigen Aufgaben zur Erfassung der in dem Modell beschriebenen Kompetenzen eingegangen. Außerdem werden erste Ergebnisse aus der Pilotierung der Aufgaben berichtet.

1 Kompetenzmodelle in den Naturwissenschaften

Mit Bezug auf eine Definition von Weinert (2001) können Kompetenzen zunächst allgemein als individuell verfügbare Kollektion von notwendigen Voraussetzungen für erfolgreiches Handeln in bedeutsamen Aufgabenbereichen verstanden werden. Kompetenzen gelten dabei prinzipiell als erlern- und vermittelbar und sie beziehen sich immer auf spezifische Domänen oder Situationen (Klieme, Hartig, & Rauch, 2008; Klieme & Leutner, 2006). Im Gegensatz zu der weiten, auch bspw. motivationale Voraussetzungen berücksichtigenden Kompetenz-Definition von Weinert wurde im o.g. Schwerpunktprogramm eine engere Definition zugrunde gelegt, die nur kognitive Voraussetzungen in den Blick nimmt (Klieme & Leutner, 2006). In Science-P wird ebenfalls dieser engere Kompetenzbegriff verwendet.

Für den Bereich Naturwissenschaften besteht in theoretischer Hinsicht weitgehender Konsens darin, dass kognitive Voraussetzungen naturwissenschaftlicher Kompetenz verschiedene Komponenten umfassen (Norris & Phillips, 2003). Duit, Häußler und Prenzel (2001) unterscheiden zwischen einer inhaltlich/konzeptuellen Komponente, einer Komponente der naturwissenschaftlichen Untersuchungsmethoden und Denkweisen, einer Komponente des Wissen-

schaftsverständnisses (Nature of Science) und einer Komponente des gesellschaftlichen Bezugs.

Bisher lagen insbesondere normativ-konzeptionell entwickelte Ansätze zur Beschreibung naturwissenschaftlicher Kompetenzen vor (vgl. auch American Association for the Advancement of Science 1993; National Science Education Standards, 1996; Qualifications and Curriculum Authority, 2000; Gesellschaft für die Didaktik des Sachunterrichts, 2002) oder es wurden im Rahmen von internationalen Schulleistungstudien Modelle mit post-hoc ermittelten Kompetenzniveaus entwickelt, wie bspw. bei PISA 2006 oder IGLU-E 2003 (Prenzel et al., 2007; Bos et al., 2003). Wegen der nachträglichen Zuordnung von Items zu Kompetenzniveaus und wegen der mangelnden Übereinstimmung von Kompetenzniveaus und Aufgabenschwierigkeiten werden die post-hoc konstruierten Modelle als unzureichend kritisiert (Klieme, Avenarius, Blum, Döbrich, Gruber, Prenzel et al., 2003).

Der Bereich der Grundschule wird in bisherigen Arbeiten nur selten berücksichtigt (siehe jedoch Hammann, 2004 sowie das schweizerische Projekt Harnos: Labudde, 2008), und zur Dimensionalität naturwissenschaftlicher Kompetenz liegen kaum empirische Ergebnisse vor. Angesichts der Forderung nach einer systematischen und modellbasierten Generierung von Testitems, die für die empirische Prüfung von Kompetenzmodellen erforderlich sind, einerseits (Helmke & Hosenfeld, 2004; Klieme et al., 2003) und der Probleme bei der Operationalisierung und Überprüfung komplexer Kompetenzmodelle andererseits liegt es nahe, sich bei der Entwicklung von Kompetenzmodellen zunächst auf die Modellierung von ausgewählten Bereichen naturwissenschaftlicher Kompetenz zu beschränken.

Im Rahmen von Science-P werden die Bereiche *Naturwissenschaftliches Wissen* (d.h. inhaltlich/konzeptuelles Wissen) und *Wissen über die Naturwissenschaften* (d.h. Wissen über naturwissenschaftliche Methoden und Wissenschaftsverständnis) untersucht.

Diese Bereiche finden sich nicht nur in der aktuellen Beschreibung der in PISA 2006 erfassten naturwissenschaftlichen Kompetenz (Prenzel et al., 2007) und in Ansätzen zu Scientific Literacy (Norris & Phillips, 2003) wieder, sondern sie stellen auch zentrale Zielbereiche naturwissenschaftlichen Lernens in der Grundschule dar. Dies spiegelt sich in unterschiedlichen für den Grundschulbereich formulierten Standards bzw. Curricula wider (Bos et al., 2003; National Science Education Standards, 1996; Qualifications and Curriculum Authority, 2000; Gesellschaft für die Didaktik des Sachunterrichts, 2002). Inwieweit die beiden theoretisch angenommenen Bereiche *Naturwissenschaftliches Wissen* und *Wissen über die Naturwissenschaften* distinkte, empirisch unterscheidbare Dimensionen darstellen und wie eng sie in Beziehung stehen, ist bisher erst in Ansätzen untersucht. Studien für das Sekundarschulalter weisen zwar auf funktionale Zusammenhänge zwischen den beiden Bereichen hin (z.B. Stathopoulou & Vosniadou, 2007), für das Grundschulalter gibt es aber nahezu keine entsprechenden Studien (siehe aber Grygier 2008; auch Sodian, Jonek-Schauer & Kircher, 2006). Im Folgenden fokussiert der Beitrag, wie in der Einleitung beschrieben, auf die theoretisch postulierte Dimension des *Naturwissenschaftlichen Wissens*.

2 Kompetenzen im Grundschulalter im Bezug auf *Naturwissenschaftliches Wissen*

Die Conceptual-Change-Forschung beschäftigt sich seit langem mit entwicklungs- und instruktionsbedingten Verständnisprozessen im Bereich des naturwissenschaftlichen Lernens, wobei insbesondere der Einfluss von naiven Schülervorstellungen auf Lernprozesse untersucht wird. Aufgrund von Erfahrungen in der Alltagswelt entwickeln Kinder naive Konzepte, die zur Interpretation von Phänomenen in der Welt herangezogen werden. Diese sind in vielen Fällen nicht ver-

einbar mit wissenschaftlichen Erklärungsmodellen und überdauern häufig traditionelle Formen des naturwissenschaftlichen Unterrichts (Duit, 1999; Wandersee, Mintzes & Novak, 1994). Naturwissenschaftliches Lernen beinhaltet demnach eine fundamentale Umstrukturierung von Konzepten in neue, wissenschaftlich begründete Ideen (z.B. diSessa, 2006; Vosniadou, Ioannides, Dimitrakopoulou, & Papademetriou, 2001). Konzeptuelle Entwicklung wird mittlerweile allerdings nicht mehr als ein abrupter Wechsel zwischen naiven und wissenschaftlichen Vorstellungen angesehen, sondern vielmehr als gradueller Prozess, der Phasen mit sogenannten Zwischenvorstellungen beinhalten kann (siehe Vosniadou, Baltas & Vamvakoussi, 2007). So schlagen beispielsweise auch Treagust und Duit (2008a) vor, die Bezeichnung „konzeptuelle Umstrukturierung“ (*conceptual reconstruction*) zu verwenden anstatt des Begriffs Konzeptwechsel. Insbesondere die Forschung von Stella Vosniadou postuliert die Konstruktion von sogenannten synthetischen Modellen, in denen Elemente wissenschaftlicher Erklärungen und naiver Vorstellungen in einem mentalen Modell integriert werden. DiSessa hingegen spricht von einem Kontinuum der konzeptuellen Entwicklung, wobei sukzessive einzelne Wissenseinheiten in umfassendere Systeme integriert werden. Auch wir konnten in einigen Untersuchungen zeigen, dass Zwischenvorstellungen im Sinne der Kombination von Fehlvorstellungen mit wissenschaftlichen Konzepten bzw. dem Bezug zu Alltagskonzepten beim Verständnis von „Schwimmen und Sinken“ auftreten und sich aus diesen durch Latent-Class-Analysen ermittelten Profilen Vorhersagen für den weiteren Lernverlauf der Schülerinnen und Schüler ableiten lassen (Hardy, Schneider & Möller, 2006).

In Übereinstimmung mit Ergebnissen von Forschungen zur konzeptuellen Entwicklung in unterschiedlichen Inhaltsgebieten und Altersgruppen (z.B. Tytler, 1998; Vosniadou et al., 2001) können im Bereich *Naturwissenschaftlichen Wissens* drei Kompetenz-

niveaus angenommen werden: (1) Naive Vorstellungen sind nur sehr eingeschränkt tragfähig, d. h. sie halten einer empirischen Prüfung in unterschiedlichen Kontexten nicht stand. (2) Zwischenvorstellungen können bereits zahlreiche Phänomene erklären oder vorhersagen, sind jedoch in ihrer Gültigkeit immer noch eingeschränkt, da sie durch entsprechende Evidenzen widerlegt werden können. (3) Wissenschaftliche Vorstellungen schließlich beruhen auf in der Wissenschaft geteilten Konzepten bzw. sind mit diesen kompatibel. Ein zusätzliches, höheres Kompetenzniveau (3+) kann in der Integration von Vorstellungen im Sinne der simultanen Ablehnung von naiven Vorstellungen und der Annahme von wissenschaftlichen Vorstellungen gesehen werden (integriertes Verständnis; Hardy, Jonen, Möller, & Stern, 2006). Die drei postulierten Niveaus beschreiben also eine Progression von naiven Sichtweisen hin zu wissenschaftsnäheren Vorstellungen der Schülerinnen und Schüler.

Konzeptuelle Entwicklung im naturwissenschaftlichen Unterricht

Die produktive Weiterentwicklung naiver Vorstellungen kann insbesondere durch die Fokussierung wissenschaftlicher Methoden der Erkenntnisgewinnung, wie der Nutzung von empirischer Evidenz, und eine diskursive Orientierung im Sinne des wissenschaftlichen Argumentierens und Begründens erreicht werden (Vosniadou, 2007). Linn (2006) fordert, dass Schüler durch Unterricht Kriterien erfahren sollten, anhand derer sie zwischen multiplen Ideen unterscheiden können. Naive Konzepte können auf diese Weise mit weiterführenden Vorstellungen integriert oder aber als wenig anwendbar oder überzeugend aufgegeben werden (siehe auch Tytler, 2000). Auch in unseren Arbeiten zur konzeptuellen Entwicklung im Inhaltsgebiet „Schwimmen und Sinken“ im Grundschulalter zeigte sich, dass ein integriertes konzeptuelles Verständnis von

Dichte und Auftrieb erst durch intensiven Unterricht ermöglicht wurde, wohingegen bei Zwischenvorstellungen auch ohne Instruktion ein Zuwachs im Laufe eines Jahres zu verzeichnen war (Möller, Jonen, Hardy, & Stern, 2002; Hardy, Jonen et al., 2006). In zwei Unterrichtsstudien und einer Laborstudie wurde im Inhaltsgebiet „Schwimmen und Sinken“ untersucht, ob sich Strukturierungselemente in einem konstruktivistisch orientierten Unterricht förderlich auf den Erwerb physikalischer Basiskonzepte bei Grundschulkindern auswirken. Wir operationalisierten dabei Strukturierungselemente in Bezug auf die Sequenzierung der Unterrichtsinhalte und die Gesprächsführung (Möller et al., 2002) und andererseits in Bezug auf die Verwendung von externen Repräsentationen (Hardy, Jonen, Möller, & Stern, 2004). Es zeigte sich, dass das konzeptuelle Verständnis von Dichte und Auftrieb bei Grundschulkindern durch Unterricht nachhaltig gefördert werden kann und dass Strukturierungselemente entscheidend zum Aufbau eines langfristigen, integrierten konzeptuellen Verständnisses beitragen (Hardy, Jonen et al., 2006; Jonen, Hardy & Möller, 2003). Dass Grundschulkindern durch Instruktion ein höheres Kompetenzniveau erreichen können, als früher angenommen wurde, bestätigen Ergebnisse von Studien aus verschiedenen Inhaltsbereichen (Stern & Möller, 2004).

Entwicklung *Naturwissenschaftlichen Wissens* im Inhaltsbereich Schwimmen und Sinken

Als Vorläufer des Verständnisses von Schwimmen und Sinken kann ein Verständnis von Materie und der Gewichtsbezug angesehen werden (Carey, 1991). Allerdings stellen Singer, Kohn und Resnick (1997) schon im Vorschulalter die Fähigkeit zur qualitativen Unterscheidung von Dichte fest, die auf der Wahrnehmung visueller Unterschiede beruht. Eine Unterscheidung von Volumen und Masse im Dichtebegriff

mit einer einhergehenden „Theorie der Materie“ ist jedoch frühestens zu Beginn der Sekundarstufe zu erwarten (Smith, Carey & Wisner, 1985; siehe auch Smith (2007) für eine Untersuchung zum Dichtebegriff mit Schülerinnen und Schülern in der achten Klasse). Dennoch wird die mittlere Dichte nach entsprechender Instruktion von Schülerinnen und Schülern als Erklärung für das Schwimmen und Sinken von Gegenständen schon zu Ende der Grundschulzeit herangezogen (Hardy, Jonen et al., 2006). Naive Erklärungen zum Schwimmen und Sinken im Grundschulalter sind häufig durch eine eindimensionale Fokussierung auf Aspekte des Objekts bzw. der Luft gekennzeichnet und somit nicht mit der das Objekt umgebenden Flüssigkeit verbunden (z. B. Tytler & Peterson, 2004). Erst durch Instruktion werden fortgeschrittenere Erklärungsansätze mit Dichtevergleich und Auftriebskraft systematisch in verschiedenen Kontexten angewendet (Hardy, Jonen et al., 2006; Kawasaki, Herrenkohl, & Yearly, 2004; Smith, Maclin, Grosslight, & Davis, 1997), wobei insbesondere die Aufforderung zu Erklärungen einen entscheidenden Beitrag zur Integration von Variablen im Kontext von Schwimmen und Sinken zu leisten scheint (Siegler & Chen, 2008).

Die Erfassung *Naturwissenschaftlichen Wissens* und der Entwicklung dieses Wissens

In letzter Zeit wurden insbesondere die Konzeptualisierung und Erfassung von Konzeptwechseln in den Paradigmen der soziokulturellen bzw. situativen und kognitiven Forschung diskutiert (Special Issues der *Cultural Studies of Science Education*; z. B. Treagust & Duit, 2008a; 2008b und *Educational Psychologist*; z. B. Mason, 2007), welche die Unterschiede zwischen den Perspektiven bzgl. der Annahme von individuellen mentalen Modellen, der Rolle von diskursiven Praktiken und der Rolle von metakognitiven Prozessen und intentionalem Lernen deutlich machen. Entsprechend unterschiedlich

sind die Erhebungsmethoden, die zur Erfassung von konzeptueller Umstrukturierung herangezogen werden. Beispielsweise führt die Betonung der kontextuellen Einbindung von konzeptuellen Konstruktionen in diskursive Praktiken zu einer Erfassung „in situ“ im Sinne dialogischen Lernens bzw. distributiver Kognitionen (Mercer, 2008). Die Annahme eines Mechanismus der internen kognitiven Umstrukturierung und die Postulierung mentaler Modelle hingegen erlaubt die Erfassung von konzeptueller Entwicklung durch Einzelinterviews ohne die explizite Berücksichtigung des Dialogs bei der Auswertung (z. B. Tytler, 2000; Vosniadou, 2007). Dennoch beschränken sich vorliegende Erhebungsmethoden überwiegend auf Interview- und Mappingverfahren, teilweise unterstützt durch schriftliche offene Aufgabenformate bei Schülern der Sekundarstufe, welche insgesamt jedoch die längsschnittliche Modellierung von konzeptueller Entwicklung bei größeren Stichproben nahezu unmöglich machen. Dass die Erfassung von konzeptuellem Verständnis auch mit Multiple-Choice Aufgaben möglich ist, konnten wir mit unserem für das Inhaltsgebiet Schwimmen und Sinken entwickelten Test für Grundschul Kinder zeigen, dessen interne Aufgabenstruktur die Zuweisung von Einzelwerten für Fehlvorstellungen, Alltagsvorstellungen und wissenschaftliche Vorstellungen sowie eines integrierten Wertes erlaubte (Hardy, Jonen et al., 2006).

3 Das postulierte Kompetenzmodell für die Dimension *Naturwissenschaftliches Wissen*

Zur Modellierung naturwissenschaftlicher Kompetenz wird in Science-P ein Struktur-/Niveau-Modell zugrunde gelegt. Wie in Abschnitt 1 dargelegt werden zwei zentrale Dimensionen postuliert: *Naturwissenschaftliches Wissen* und *Wissen über Naturwissenschaften*. In beiden Dimensionen werden drei hierarchisch geordnete Kompetenzni-

		Level 1	Level 2	Level 3	Level 3+
		Naive Vorstellungen	Zwischen-Vorstellungen	Wissenschaftliche Vorstellungen	
				Wissenschaftl. Verständnis	Integriertes Verständnis
Schwimmen & Sinken	Auftrieb				
	Dichte				
	Verdrängung				
Verdunstung & Kondensation	Verdunstung				
	Kondensation				

Abb. 1: Kompetenzmodell für den Bereich Naturwissenschaftliches Wissen in den Inhaltsbereichen Schwimmen und Sinken und Verdunstung und Kondensation.

veaus postuliert: Naive Vorstellungen, Zwischenvorstellungen und wissenschaftliche Vorstellungen. Innerhalb der wissenschaftlichen Vorstellungen wird noch einmal unterschieden zwischen einem einfachen wissenschaftlichen Verständnis, bei dem Schüler wissenschaftlich akzeptable Vorstellungen äußern, und einem integrierten Verständnis, das zusätzlich durch die simultane Ablehnung naiver Vorstellungen gekennzeichnet ist. In diesem Beitrag wird wie eingangs beschrieben auf die erste Dimension, *Naturwissenschaftliches Wissen*, fokussiert. Innerhalb dieser postulierten Dimension wird wiederum *Naturwissenschaftliches Wissen* aus den physikbezogenen Inhaltsbereichen *Schwimmen und Sinken* und *Verdunstung und Kondensation* in den Blick genommen. Für jeden der beiden Inhaltsbereiche wurden relevante physikalische Konzepte festgelegt. Das für die Dimension *Naturwissenschaftliches Wissen* postulierte Kompetenzmodell (s. Abbildung 1) besteht also aus den beschriebenen Kompetenzniveaus und der Struktur der Inhaltsbereiche. Das Modell wird exemplarisch an den Inhaltsbereichen *Schwimmen und Sinken* sowie *Verdunstung und Kondensation* überprüft.

4 Vorgehen bei der Testkonstruktion und der empirischen Testung des Kompetenzmodells

Bei der Konstruktion von Testverfahren und der empirischen Testung des postulierten Kompetenzmodells greifen wir auf den von Wilson (2005) vorgeschlagenen Ansatz des „Construct Modeling“ zurück. Im Folgenden wird dieser Ansatz kurz vorgestellt und anschließend auf die Entwicklung und Evaluation des Kompetenzmodells im Rahmen von Science-P übertragen.

Der Ansatz des Construct Modeling nach Wilson

Ausgangspunkt in Wilsons Ansatz stellt eine sog. Construct Map dar, welche das zu erfassende Konstrukt definiert. Das Konstrukt wird als latente, kontinuierliche Variable verstanden, bei der aber ggf. verschiedene qualitative Niveaus unterschieden werden können. In der Construct Map werden die Ausprägungen des interessierenden Konstrukts auf diesen qualitativen Niveaus spezifiziert. Auf der Basis der Construct Map werden in einem zweiten Schritt Aufgaben-

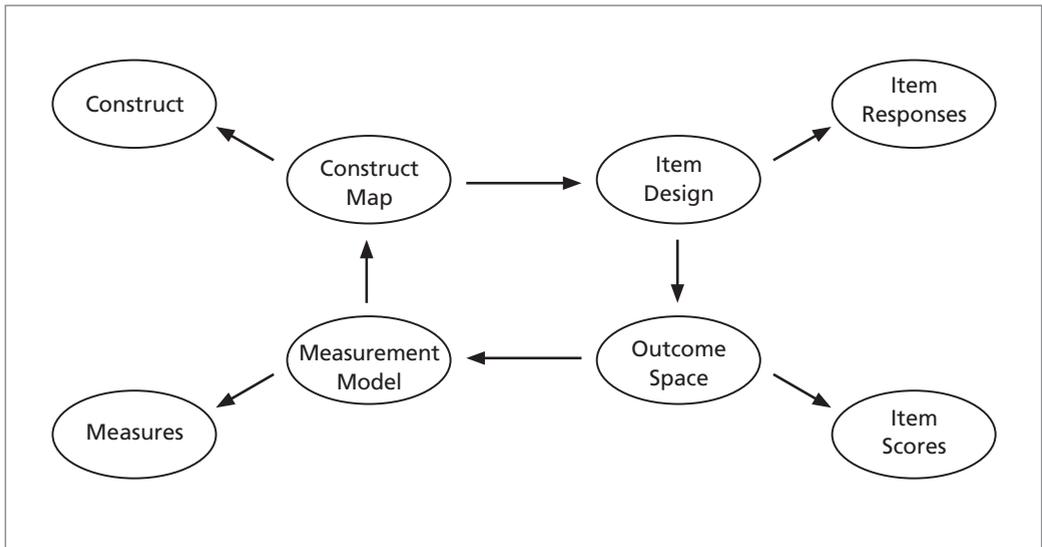


Abb. 2: Zyklus der konstruktbasierenden Instrumententwicklung nach Wilson (2005).

sets entwickelt, von denen angenommen werden kann, dass die Art und Weise ihrer Bearbeitung auf die jeweilige Ausprägung des Konstrukts bei der bearbeitenden Person zurückgeht. Das *Item Design* betrifft die Form, in der die Aufgabe den Untersuchungspersonen vorgelegt wird. Um aus den Antworten von Untersuchungspersonen auf die Items Schlussfolgerungen zur Ausprägung des Konstrukts bei den Personen ziehen zu können, müssen in einem dritten Schritt die Antworten kategorisiert und mit numerischen Werten auf dem Kontinuum der latenten Variable versehen werden (erster Inferenzschritt). Dazu muss festgelegt werden, welche Teile der Antworten für die Schlussfolgerung herangezogen werden und wie diese Teile dann kategorisiert und bepunktet werden. Dies ist der sogenannte *Outcome Space*, der im Falle von Aufgaben mit offenem Antwortformat oder im Falle von Interviews als Kodierschema repräsentiert sein kann. In einem zweiten Inferenzschritt werden die Punktwerte, die Personen bei den einzelnen Aufgaben erhalten, auf das Konstrukt zurückbezogen. Dies geschieht anhand eines psychometrischen Modells (Measurement Model) wie bspw. dem Rasch-Modell. Es spezifiziert,

wie die bepunkteten Antworten so zusammengefasst werden können, dass eine Aussage über die Ausprägung auf der latenten Variable gemacht werden kann. Die gewonnenen Maße können dann genutzt werden, um Rückschlüsse auf die Güte der Erfassung des interessierenden Konstrukts ziehen zu können. Dieser von Wilson beschriebene Ansatz zur Testentwicklung kann als Zyklus dargestellt werden (siehe Abb. 2).

Anwendung des Ansatzes im Rahmen von Science-P

Im Projekt Science-P wurde dieser Ansatz bei der Entwicklung und Evaluation eines Modells naturwissenschaftlicher Kompetenz im Bereich *Naturwissenschaftliches Wissen* wie folgt umgesetzt.

a) Construct Map: Definition des Konstrukts
Basis für die Construct Map stellt das o.g. theoretisch abgeleitete Kompetenzmodell dar (siehe Abb. 2). Für die Construct Map wurden die Zellen des Kompetenzmodells genauer definiert. Es wurde also jeweils spezifiziert, wodurch sich *Naturwissenschaftliches Wissen* in den betrachteten Inhaltsbereichen auf den

unterschiedlichen Kompetenzniveaus auszeichnet. Tabelle 1 zeigt dies exemplarisch für das Konzept Auftrieb im Inhaltsbereich Schwimmen und Sinken. Das Kompetenzniveau 3+ wird erreicht, wenn wissenschaftliche Vorstellungen angenommen und naive Vorstellungen abgelehnt werden.

b) Item Design: Generierung von Aufgaben

Die Entwicklung von Aufgaben geschieht in einem gestuften Prozess, der Vorpilotierungen sowie eine Pilotierungs- und eine Validierungsstudie einschließt. In einer ersten Phase wurde, basierend auf der Construct Map, ein Aufgabenpool von je 60 Aufgaben für die Inhaltsbereiche *Schwimmen und*

Sinken und Verdunstung und Kondensation entwickelt. Dabei wurden Aufgaben mit verschiedenen Formaten konstruiert: Bei Forced-Choice-Aufgaben (siehe Abb. 3) muss die bessere von zwei vorgegebenen Antwortalternativen gewählt werden. Dabei werden jeweils zwei nebeneinander liegende Kompetenzniveaus gegeneinander getestet. Diese Aufgaben liefern Informationen darüber, welches Niveau bei einem direkten Vergleich bevorzugt wird.

Bei integrierten Aufgaben müssen jeweils eine wissenschaftliche Erklärung und mehrere naive Erklärungen als richtig oder falsch beurteilt werden. Damit soll überprüft werden, ob das Niveau des integrierten Ver-

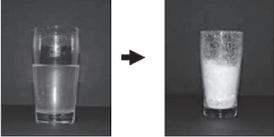
Tab. 1: Construct Map für das Konzept Auftrieb

Physikalisches Konzept	Niveau 1 Naive Vorstellungen	Niveau 2 Zwischenvorstellungen	Niveau 3 wissenschaftliche Vorstellungen
Auftrieb (in Flüssigkeiten, hier Wasser)	<p>Luft-aktiv-Konzept: Luft zieht Sachen nach oben, z.B. „Die Luft zieht das nach oben.“</p> <p>Formkonzept: Die spez. Form des Dings ist für das Schwimmen oder Sinken verantwortlich, z.B. „Weil die Enten sind ja unten geformt wie ein Boot.“</p> <p>Größenkonzept: alles was klein ist kann schwimmen, alles was groß ist geht unter</p> <p>Gewicht verkehrt: alles was leicht ist geht unter falsches Druckkonzept: Wasser saugt/drückt nach unten</p> <p>Gewichtskonzept: alles was schwer ist geht unter, z.B. „Der geht nicht unter, weil der so leicht ist.“ zu wenig Wasser: Sachen gehen unter, wenn nicht genug Wasser da ist</p>	<p>Hohlkörperkonzept: hohle Sachen schwimmen, z.B. „Weil die Äpfel sind ja ein bisschen hohl von innen.“</p> <p>Luft-Konzept: etwas schwimmt, weil Luft drin ist (nicht aktiv); auch Luft macht leichter, z.B. „Weil sie innen drin auch ein bisschen Luft haben.“</p> <p>leichter im Wasser: Dinge/Menschen werden im Wasser leichter</p> <p>Leichter-als-Wasser-Konzept: Dinge schwimmen, wenn sie leichter als Wasser sind (ohne Berücksichtigung des Volumens), z.B. „Ehm, weil der sch - schwerer als das Wasser ist.“</p> <p>Materialkonzept: etwas schwimmt, weil es aus Holz ist, z.B. „Und der besteht aus Plastik und Plastik schwimmt.“</p>	<p>Druckkonzept/Auftrieb: Wasser drückt alle Sachen nach oben, z.B. „Der schwimmt, weil der wird ja auch vom Wasser nach oben gedrückt.“</p> <p>Dichtevergleich: Der Gegenstand wiegt weniger als die gleiche Menge Wasser</p> <p>Archimedisches Prinzip: Der Gegenstand wiegt weniger als das von ihm verdrängte Wasser</p> <p>Vergleich Gewichtskraft – Auftriebskraft: Das Wasser drückt stärker nach oben, als der Gegenstand von seinem Gewicht nach unten gedrückt/gezogen wird</p>

Salzwasser

Instruktion des Testleiters:
 Paul hat Salz und Wasser in einem Glas vermischt. Er stellt das Glas in die Mikrowelle. Nach einer Dreiviertelstunde sieht das Glas so aus, wie auf dem Foto: das Wasser ist weg und im Glas ist nur noch das harte Salz.

Was ist mit dem Wasser passiert?
 Kreuze die bessere von beiden Antworten an!
 Das Wasser...



... ist in das Salz eingezogen und ist jetzt dort gespeichert.	<input type="checkbox"/>	<input type="checkbox"/>	... ist in die Luft gestiegen und ist jetzt nicht mehr sichtbar.

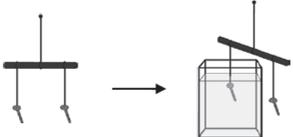
Abb. 3: Forced-Choice-Aufgabe aus dem Inhaltsbereich Verdunstung und Kondensation mit einer Antwort auf Niveau 1 und einer Antwort auf Niveau 3.

Schraubenwaage

Instruktion des Testleiters:
 Ich habe mir hier mit einem Stab, zwei Schrauben und drei Stücken Schnur eine einfache Waage gebastelt. Wie ihr seht, befindet sich die Waage im Gleichgewicht, d.h. der Stab ist gerade. Die beiden Schrauben sind also gleich schwer. Nun tauche ich eine der beiden Schrauben vorsichtig in das Wasserbecken. Jetzt gerät die Waage aus dem Gleichgewicht.

Die Waage ist nicht mehr im Gleichgewicht, wenn eine Schraube im Wasser ist. Wieso ist das so?
 Kreuze nach jeder Antwort ‚Richtig‘ oder ‚Falsch‘ an!

Die Waage ist nicht mehr im Gleichgewicht, weil...



	Richtig	Falsch
... das Wasser die eingetauchte Schraube nach oben drückt.	<input type="checkbox"/>	<input type="checkbox"/>
... kleine Luftbläschen die eingetauchte Schraube nach oben ziehen.	<input type="checkbox"/>	<input type="checkbox"/>
... die eingetauchte Schraube eine ganz kleine Oberfläche hat.	<input type="checkbox"/>	<input type="checkbox"/>
... die eingetauchte Schraube im Wasser leichter wird.	<input type="checkbox"/>	<input type="checkbox"/>

Abb. 4: Partial-Credit-Aufgabe aus dem Inhaltsbereich Schwimmen und Sinken.

ständnisses erreicht wird, also der wissenschaftlichen Erklärung zugestimmt wird und gleichzeitig die naiven Erklärungen ablehnt werden. In Partial-Credit-Aufgaben (siehe Abb. 4) müssen mehrere Antwortalternativen

als richtig oder falsch beurteilt werden, die jeweils eines der postulierten drei Kompetenzniveaus repräsentieren. Aufgaben mit offenem Antwortformat werden ebenfalls aufgenommen, um die freie

Konstruktion von Erklärungen bei den Kindern zu berücksichtigen. Einige wenige Aufgaben haben ein grafisches Antwortformat. Die Aufgaben wurden zunächst in kleineren Vortests auf die Eignung für Gruppentestung und auf die Verständlichkeit der Aufgaben hin überprüft. Eine Pilotierung der Aufgaben in 42 Klassen der zweiten und vierten Jahrgangsstufe wurde genutzt, um Gütekriterien auf der Basis der klassischen Testtheorie zu ermitteln und inadäquate Aufgaben aussondern bzw. überarbeiten zu können. Für diese Pilotierung wurden die Aufgaben auf 10 Testhefte verteilt. Dabei wurde darauf geachtet, dass die verschiedenen Aufgabenformate in den Heften gleichmäßig repräsentiert sind. Jedes Testheft wurde nur mit Aufgaben aus einem Inhaltsbereich (Schwimmen und Sinken oder Verdunstung und Kondensation) versehen. Testanalysen zu einem dieser Testhefte werden exemplarisch im folgenden Kapitel berichtet.

In einer Validierungsstudie soll die konvergente und diskriminante Validität der Aufgaben ermittelt werden. Dazu werden einer Stichprobe von ca. 40 Zweit- und 40 Viertklässlern sowohl Testaufgaben, die sich in der Pilotierung bewährt haben, als auch Tests zu allgemeinen kognitiven Fähigkeiten und zum Leseverständnis vorgelegt. Für die Analysen der Daten der Pilotierungs- und der Validierungsstudie ist bereits eine Definition des Outcome Space erforderlich. Darauf wird im folgenden Abschnitt eingegangen.

c) Outcome Space: Kategorisierung von Antworten und Bepunktung

Die Bepunktung der Antworten der Schülerinnen und Schüler geschieht in zwei Schritten. In einem ersten Schritt werden die Antworten den postulierten Kompetenzniveaus zugeordnet. Im Falle der Aufgaben mit vorgegebenem Antwortformat wurden die Antwortalternativen bereits entsprechend eines der drei Kompetenzniveaus konstruiert.

Im Falle der Aufgaben mit offenem Antwortformat dient ein Kodierschema der Zuord-

nung der Schülerantworten zu den Kompetenzniveaus, welches wiederum an die Construct Map angelehnt ist.

In einem zweiten Schritt wird dann die eigentliche Bepunktung der Schülerantworten vorgenommen. Im Falle der Forced-Choice-Aufgaben wird die Wahl der Antwortalternative bepunktet, die dem höheren Kompetenzniveau entspricht (dichotome Bepunktung, 0-1). Bei den Aufgaben zum integrierten Verständnis wird bepunktet, wenn gleichzeitig eine wissenschaftliche Vorstellung angenommen wird und alle naiven Vorstellungen abgelehnt werden (dichotome Bepunktung, 0-1). Bei den Partial-Credit-Aufgaben wird die Wahl der besten Antwortalternative für die Bepunktung herangezogen. Hier werden den postulierten Kompetenzniveaus entsprechend 0 Punkte für naive, 1 Punkt für Zwischen- und 2 Punkte für wissenschaftliche Vorstellungen zugeordnet. Bei den Aufgaben mit offenem Antwortformat ist eine analoge Bepunktung vorgesehen.

d) Measurement Model:

Statistische Modellierung

Die statistische Modellierung soll anhand einer Variante des Rasch-Modells (Partial-Credit-Modell; Rost, 2004) erfolgen. Es wird erwartet, dass die empirischen Itemschwierigkeiten bzw. die Schwellenparameter im Falle der Partial-Credit-Aufgaben die im Kompetenzmodell postulierten Kompetenzniveaus widerspiegeln. Aufgaben zum integrierten Verständnis sollten demnach die schwierigsten Aufgaben sein, wohingegen Forced-Choice-Aufgaben, die zwischen Niveau 1 und 2 differenzieren, und die Schwellen zwischen Niveau 1 und 2 bei den Partial-Credit-Aufgaben die geringste Schwierigkeit aufweisen sollten.

Wegen der anzunehmenden Bereichsspezifität konzeptuellen *Naturwissenschaftlichen Wissens* sind in weiterführenden Analysen, die die beiden Inhaltsbereiche *Schwimmen und Sinken* und *Verdunstung und Kondensation* simultan berücksichtigen, ein- und mehrdimensionale Modellie-

Tab. 2: Stichprobe

	N	Geschlecht*	Alter**
2. Klasse	45	17/28	7.5 (0.5)
4. Klasse	55	25/30	9.9 (0.6)
Gesamt	100	42/58	8.8 (1.3)

* Anzahl Jungen/Anzahl Mädchen

** Mittelwert (Standardabweichung)

rungen vorgesehen. Auf diese Weise soll geprüft werden, ob die Inhaltsbereiche *Schwimmen und Sinken* und *Verdunstung und Kondensation* zwei Dimensionen des Kompetenzbereichs *Naturwissenschaftliches Wissen* bilden.

5 Erste Ergebnisse der Pilotierungsstudie

In diesem Beitrag werden exemplarisch für eines der 10 in der Pilotierung eingesetzten Testhefte Ergebnisse aus Analysen berichtet, die auf der klassischen Testtheorie basieren. Das Testheft beinhaltet elf Aufgaben aus dem Inhaltsbereich Schwimmen und Sinken, davon sechs Forced-Choice-, zwei Partial-Credit- und zwei Integrierte Aufgaben sowie eine Aufgabe mit offenem und eine mit grafischem Antwortformat. Da die Kodierung der Antworten auf die offenen Fragen bisher noch nicht abgeschlossen ist, werden nur Ergebnisse zu den geschlossenen Antwortformaten dargestellt. Eine weitere Aufgabe (Integrierte Aufgabe) wurde in die folgenden Analysen nicht mit einbezogen, da sie nicht von der ganzen Stichprobe (s. u.) bearbeitet wurde. Eine Forced-Choice-Aufgabe wurde ausgeschlossen, da sie aus plausiblen Gründen eine negative Trennschärfe (korrigierte Skala-Item-Korre-

lation) aufwies.¹ Das hier analysierte Testheft umfasst also acht Aufgaben.

Die Stichprobe, an der das hier beschriebene Testheft eingesetzt wurde, umfasst 45 Kinder der zweiten und 55 Kinder der vierten Klasse (siehe Tab. 2). Die Bearbeitungsdauer lag in Klasse 2 bei 36 Minuten, in Klasse 4 bei 24 Minuten für die analysierten acht Aufgaben einschließlich aller Instruktionen.

In Tabelle 3 sind zunächst deskriptive Werte zu den acht Items angegeben. Die Mittelwerte deuten an, dass es bei den acht Items weder Decken- noch Bodeneffekte zu geben scheint und dass ausreichend Variabilität in den Werten liegt. Nimmt man die Mittelwerte als Schätzer für die Itemschwierigkeit, so deutet sich an, dass diese nicht in jedem Falle den durch das Kompetenzmodell postulierten Schwierigkeiten entspricht. So ist Aufgabe 4 bspw. relativ leicht, obwohl sie das dritte Kompetenzniveau (wissenschaftliche Vorstellungen) erfasst.

Die interne Konsistenz des Tests aus den acht Items ist für die Gesamtstichprobe von 100 Kindern noch zufrieden stellend (Cronbach's $\alpha = .607$). In der Gruppe der Zweitklässler beträgt die interne Konsistenz jedoch nur $.402$, bei den Viertklässlern liegt der Wert bei $.677$. Die in Tabelle 3 angegebenen Trennschärfen sind bei einigen Items sehr zufriedenstellend (Aufgaben 3 und 5), bei anderen (Aufgaben 1, 2 und 4) niedrig.

1 Im Unterricht war den Kindern die vorgegebene Antwort des niedrigeren Kompetenzniveaus (Level 2) als die korrekte vermittelt worden („Das Wasser trägt.“ Laut Construct-Map läge diese Antwort nur auf Level 2, nicht auf 3.)

Tab. 3: Deskriptive Itemstatistiken

Aufgabe	Format*	Bepunktung	Mittelwert (Standardabweichung)	korrigierte Skala- Item-Korrelation**
1	forced choice, L1/L2	0-1	0.64 (0.482)	0.154
2	forced choice, L1/L2	0-1	0.75 (0.429)	0.153
3	grafisches Antwortformat L1/L3	0-1	0.21 (0.411)	0.597
4	forced choice, L1/L3	0-1	0.76 (0.429)	0.153
5	multiple choice, alle Level	0-1-2	0.62 (0.850)	0.540
6	forced choice, L1/L2	0-1	0.39 (0.490)	0.275
7	integriertes Verständnis, L1 und L3	0-1-2***	0.50 (0.532)	0.335
8	forced choice, L1/L3	0-1	0.40 (0.492)	0.302

* Angabe des Antwortformates und des Kompetenzniveaus (L1 = naive Vorstellungen, L2 = Zwischenvorstellungen, L3 = wissenschaftliche Vorstellungen)

** Angabe der Trennschärfen bezogen auf die Analyse der internen Konsistenz für die gesamte Stichprobe, s. folgender Abschnitt

*** Bewertung der Aufgaben für das integrierte Verständnis: 0 Punkte = kein Annehmen der wissenschaftlichen Vorstellung, 1 Punkt = Annehmen der wissenschaftlichen Vorstellung bei simultaner Annahme von einer oder mehreren naiven Vorstellungen, 2 Punkte = Annehmen der wissenschaftlichen Vorstellung bei gleichzeitiger Ablehnung der naiven Vorstellungen

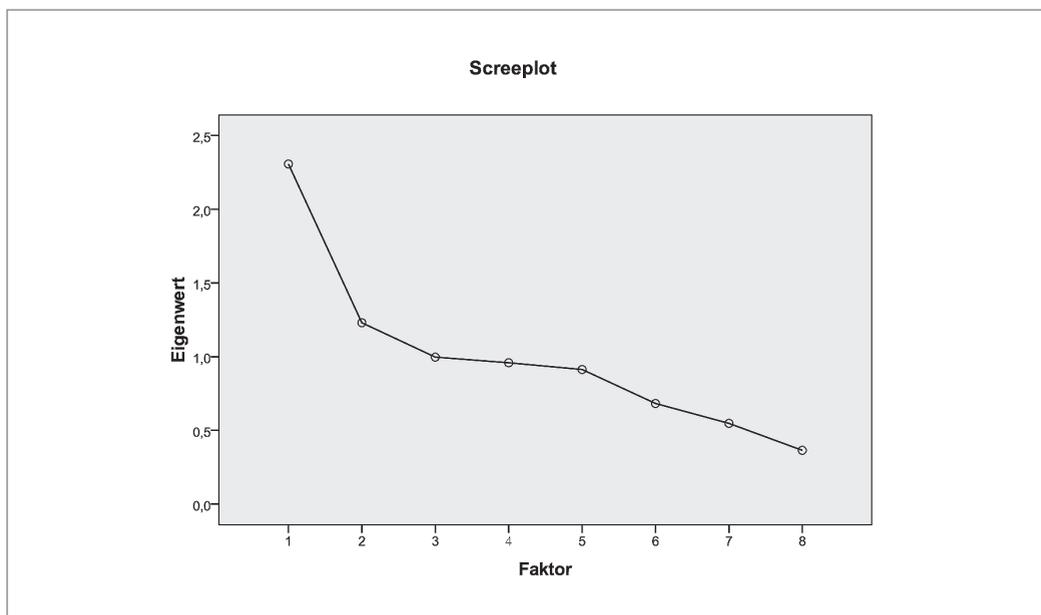


Abb. 5: Lösung der Faktorenanalyse mit Hauptkomponentenanalyse (8 Items) für die gesamte Stichprobe.

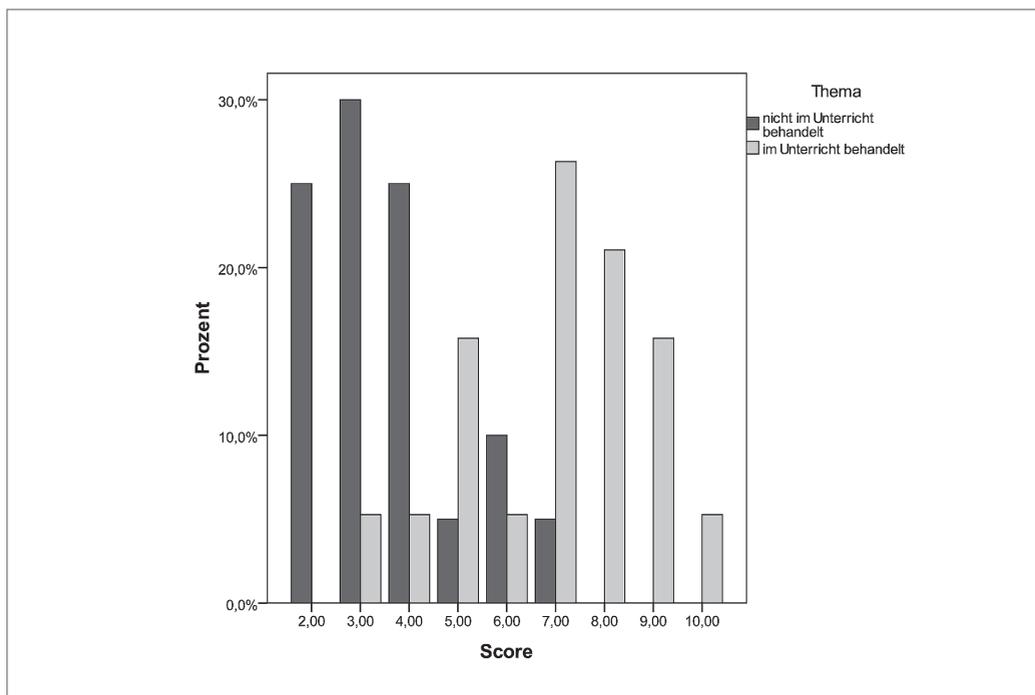


Abb. 6: Verteilung der Testwerte in einer unterrichteten und einer nicht unterrichteten Klasse.

Tab. 4: Stichprobe für den Gruppenvergleich zwischen unterrichteter und nicht unterrichteter Klasse

	N	Geschlecht*	Alter**	Score**
Thema nicht behandelt	20	10/10	9.9 (0.7)	3.60 (1.47)
Thema behandelt	19	9/10	9.9 (0.4)	6.95 (1.87)
Gesamt	39	19/20	9.9 (0.6)	5.23 (2.37)

* Anzahl Jungen/Anzahl Mädchen

** Mittelwert (Standardabweichung)

Anhand einer explorativen Faktorenanalyse wurde überprüft, inwiefern der konstruierte Test wie angenommen nur eine Kompetenz erfasst. Aus dem in Abbildung 5 dargestellten Screeplot kann entnommen werden, dass dem Test tatsächlich eine starke Hauptkomponente zugrunde liegt. Dieser Faktor klärt 28,8 Prozent der Varianz in den Antworten auf. Nach Kaiser-Kriterium und nach Analyse des Screeplots wäre allerdings ggf. noch eine zweite Komponente

zu berücksichtigen, die auch noch 15,4% an Varianz aufklärt.

Um Hinweise auf die Sensitivität des Tests für Unterrichtseffekte zu erhalten, wurde eine Klasse, die das Thema Schwimmen und Sinken bereits im Unterricht durchgenommen hatte, mit einer nicht unterrichteten Klasse verglichen. Erwartet wurde ein substantiell besseres Testergebnis der unterrichteten Klasse. Zur Konstanzhaltung anderer Einflussgrößen wurden zwei vierte Klassen

Tab. 5: Mittelwerte bei Forced-Choice-Aufgaben (L1 vs. L2) getrennt nach Klassenstufen

Inhaltsbereich	Aufgabenlabel	Mittelwerte	
		Klasse 2	Klasse 4
Schwimmen und Sinken	Gummiente	0,82	0,87
	Überlaufendes Wasser	0,29	0,77
	Springender Ball	0,44	0,31
	Zwei Formen in Wasser getaucht	0,65	0,73
	Papierschiff	0,51	0,75
	Klotz am Gummiband	0,67	0,84
	Knetgummi-Schale	0,42	0,36
	Styroporwürfel	0,33	0,20
	Enten auf einem See	0,13	0,74
	Draht	0,50	0,56
	Krümel aus Knete	0,33	0,08
	Nilpferd	0,67	0,79
	Tischtennisball	0,48	0,81

ausgewählt, die – an derselben Schule – von derselben Sachunterrichtslehrerin unterrichtet werden (Kontrollvariablen wie allgemeine kognitive Fähigkeiten standen in dieser Pilotierung leider nicht zur Verfügung). Abbildung 6 zeigt zunächst die Verteilungen der Testwerte in den beiden Klassen und in Tabelle 4 sind zusätzlich die Mittelwerte der Testwerte in den beiden Klassen angegeben. Schon hier deutet sich eine Überlegenheit der unterrichteten Klasse an.

Ein T-Test für unabhängige Stichproben zeigt, dass der Unterschied zwischen den beiden Klassen signifikant ist ($t(37) = -6,240$, $p = 0,000$; Freiheitsgrade und p für Stichproben mit gleichen Varianzen, da Levene-Test n.s.). Die Stärke des Effekts beträgt $d = 1,18$ und ist damit als groß einzuschätzen.

Da bereits gezeigt werden konnte (Hardy, Jonen et al., 2006), dass Grundschul Kinder Zwischenvorstellungen (zweites Kompetenzniveau) über die Zeit auch ohne gezielte Instruktion aufbauen, wurde des Weiteren geprüft, ob Viertklässler im Vergleich

zu Zweitklässlern eher Kompetenzniveau 2 erreichen. Dazu wurden für diesen Beitrag sämtliche Forced-Choice-Aufgaben aus allen 10 Testheften zur Differenzierung zwischen Niveau 1 und 2 herangezogen und die Mittelwerte in Klasse 2 und 4 verglichen (siehe Tab. 5).

Es zeichnet sich tendenziell ab, dass die Mittelwerte tatsächlich in Klasse 4 höher liegen als in Klasse 2. Bei neun der 13 Aufgaben ist dies der Fall, wobei bei zwei der neun Aufgaben der Unterschied marginal ist. Insgesamt scheinen aber Viertklässler eher das zweite postulierte Kompetenzniveau zu erreichen, während Zweitklässler wie erwartet eher auf dem ersten Kompetenzniveau antworten.

6 Diskussion

Ziel der beschriebenen Projektphase von Science-P ist die Entwicklung von gruppentestfähigen Aufgaben zur Erfassung

naturwissenschaftlicher Kompetenzen im Grundschulalter, die durch ein theoretisch abgeleitetes Kompetenzmodell beschrieben werden. In diesem Beitrag wurde dargestellt, wie der Ansatz des Construct Modeling (Wilson, 2005) bei der Testkonstruktion genutzt wird. Die entwickelten 120 Aufgaben wurden im Rahmen einer Pilotierungsstudie auf 10 Testhefte verteilt. Berichtet wurden hier Analysen eines ausgewählten Testhefts aus dem Inhaltsbereich Schwimmen und Sinken. Bei der Interpretation der im vorigen Kapitel berichteten Ergebnisse ist zu berücksichtigen, dass das Testheft, das hier wie ein eigenständiger Test behandelt wurde, eine erste Zusammenstellung von Aufgaben darstellt, die abgesehen von der Wahl der Aufgabenformate, die in den Heften gleich verteilt sein sollte, zufällig getroffen wurde. Die Analyse der einzelnen Testhefte auf der Basis der klassischen Testtheorie dient dazu, Hinweise auf die Überarbeitung oder Entfernung von Aufgaben für die weitere Instrumententwicklung zu liefern.

Die interne Konsistenz und die Trennschärfen der analysierten acht Aufgaben geben erste Hinweise darauf, dass mit den Aufgaben eine konsistente Fähigkeit gemessen wird, auch wenn die Trennschärfen der Aufgaben 1, 2 und 4 gering sind. In Klasse zwei ist das Antwortverhalten weniger konsistent als in Klasse vier. Möglicherweise ist dies auf stärkeres Raten der jüngeren Kinder zurückzuführen. Die Ergebnisse der explorativen Faktorenanalyse unterstützen die Feststellung, dass dem Antwortverhalten der Kinder bei den gewählten acht Aufgaben wie postuliert eine latente Fähigkeit zugrunde liegt. Es zeichnet sich eine starke Hauptkomponente ab, die 28,8 % Varianz aufklärt. Ein Vergleich einer im Thema Schwimmen und Sinken unterrichteten mit einer nicht unterrichteten Klasse ergab, dass die unterrichteten Kinder wie angenommen substanziell bessere Testleistungen zeigten. Dies kann als Hinweis auf Validität des Instruments interpretiert werden. Zudem scheint das analysierte Instrument sensitiv für Unterrichtseffekte zu sein, da die unter-

suchten Viertklässler mit bereits erfolgtem Unterricht zum Schwimmen und Sinken in ihrem Antwortverhalten signifikant über den Schülerinnen und Schülern einer Vergleichsklasse lagen.

Da bei den beiden hier betrachteten Klassen kein Vorwissen erfasst wurde, sind die berichteten Ergebnisse sicher nur als erste empirische Hinweise zu verstehen, die weiterer Untersuchung bedürfen. Im Rahmen der in Kapitel 2 genannten BiQua-Studie (Möller et al., 2002; Hardy, Jonek et al., 2006) konnte allerdings bereits nachgewiesen werden, dass mit Testaufgaben, die weitgehend wie die hier dargestellten Aufgaben konstruiert waren, Unterschiede zwischen zwei Gruppen nachgewiesen werden konnten, die beide konstruktivistisch orientierten Unterricht erhalten hatten, der sich aber im Grad der Strukturierung unterschied (Hardy, Jonek et al., 2006). Einzelitem-Analysen im Rahmen der hier berichteten Pilotstudie deuten an, dass die unterrichteten Kinder insbesondere im Geben wissenschaftsnaher Erklärungen (drittes Kompetenzniveau) und in der Integration des Wissens (Niveau 3+) besser abschneiden als nicht unterrichtete Kinder. Auch dieser Befund ist kongruent mit den in BiQua gefundenen Ergebnissen.

Analysen aller im Bereich Schwimmen und Sinken eingesetzten Forced-Choice-Aufgaben, die zwischen Kompetenzniveau 1 und 2 differenzieren, weisen darauf hin, dass Zwischenvorstellungen (zweites Niveau) von Kindern der 4. Klasse eher geäußert werden als von Zweitklässlern, auch wenn beide Gruppen noch keinen Unterricht zum Thema Schwimmen und Sinken erhalten haben. Dieser Befund steht ebenfalls in Übereinstimmung mit den in Kapitel 2 referierten Befunden, die zeigen, dass das Niveau von Zwischenvorstellungen bei Grundschulkindern über die Zeit auch ohne spezifische Instruktion erreicht werden kann (Hardy, Jonek et al., 2006).

Anhand der Mittelwerte der im Rahmen dieses Beitrags untersuchten acht Items können bereits erste Hinweise auf die Gültigkeit der im postulierten Kompetenzmo-

dell beschriebenen drei Kompetenzniveaus gewonnen werden. Es zeigt sich, dass Aufgaben, deren Lösung die theoretisch postulierten höheren Kompetenzniveaus (wissenschaftliche Vorstellungen und integriertes Verständnis) erfordern, tendenziell auch schwerer zu sein scheinen. Für Item 4 trifft dieser Trend allerdings nicht zu. Vor der Modelltestung wird, wie in Kapitel 4 beschrieben, in einer separaten Validierungsstudie der konvergenten und diskriminanten Validität der Aufgaben mit Hilfe eines Multi-Trait-Multi-Method-Verfahrens nachgegangen. In einem zyklischen Vorgehen werden die entwickelten Aufgaben weiter revidiert. Insbesondere Items mit geringer Trennschärfe werden nochmals auf Verständlichkeit hin überprüft und ggf. entfernt.

Erst mit Hilfe von Rasch-Analysen, in denen Person- und Itemparameter getrennt geschätzt werden, wird eine Überprüfung der angenommenen Kompetenzniveaus möglich sein und es werden Einschätzungen der Funktionsweise einzelner Items adäquater vorgenommen werden können. An einer größeren Stichprobe werden dazu mit einem rotierten Testheft-Design und auf der Basis von Rasch-Analysen Prüfungen der im Kompetenzmodell getroffenen Annahmen (Kompetenzniveaus und Dimensionalität) vorgenommen, welche dem von Wilson (2005) als Measurement Model beschriebenen Schritt entsprechen.

Danksagung

Diese Veröffentlichung wurde ermöglicht durch eine Sachbeihilfe der Deutschen Forschungsgemeinschaft (Kennz.: MO 942/4-1) im Schwerpunktprogramm „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“ (SPP 1293).

Literatur

- American Association for the Advancement of Science (AAAS) (1993). *Benchmarks for Science Literacy*. Washington, DC: AAAS Project 2061.
- Bos, W., Lankes, E., Prenzel, M., Schwippert, K., Walther G., & Valtin, R. (2003). *Erste Ergebnisse aus IGLU: Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change? In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 257–291). Hillsdale, NJ: Lawrence Erlbaum Associates.
- diSessa, A. (2006). A history of conceptual change research. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 265–281). Cambridge: Cambridge University Press.
- Duit, R. (1999). Conceptual change approaches in science education. In W. Schnotz, S. Vosniadou, & M. Carretero (Eds.), *New perspectives on conceptual change* (pp. 263–282). New York: Pergamon.
- Duit, R., Häußler, P., & Prenzel, M. (2001). Schulleistungen im Bereich der naturwissenschaftlichen Bildung. In F. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 169–186). Weinheim: Beltz.
- Gesellschaft für Didaktik des Sachunterrichts (GDSU) (2002). *Perspektivrahmen Sachunterricht*. Bad Heilbrunn: Klinkhardt.
- Grygier, P. (2008). *Wissenschaftsverständnis von Grundschulern im Sachunterricht*. Bad Heilbrunn: Klinkhardt.
- Hammann, M. (2004). Kompetenzentwicklungsmodelle. Merkmale und ihre Bedeutung – dargestellt anhand von Kompetenzen beim Experimentieren. *MNU*, 57(4), 196–203.
- Hardy, I., Jonen, A., Möller, K., & Stern, E. (2004). Die Integration von Repräsentationsformen in den Sachunterricht der Grundschule. In J. Doll & M. Prenzel (Eds.), *Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (pp. 267–283). Münster: Waxmann.
- Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "Floating and Sinking". *Journal of Educational Psychology*, 98(2), 307–326.

- Hardy, I., Schneider, M., & Möller, K. (2006). *Patterns of conceptual change: The development of elementary school students' explanations of "Floating and Sinking"*. Vortrag auf der Tagung der EARLI Special Interest Group für Conceptual Change, Stockholm, Schweden.
- Helmke, A., & Hosenfeld, I. (2004). Vergleichsarbeiten – Standards – Kompetenzstufen: Begriffliche Klärung und Perspektiven. In M. Wosnitza (Ed.), *Lernprozess, Lernumgebung und Lerndiagnostik: Wissenschaftliche Beiträge zum Lernen im 21. Jahrhundert* (pp. 56–75). Landau: Verlag Empirische Pädagogik.
- Jonen, A., Hardy, I., & Möller, K. (2003). Schwimmt ein Holzbrett mit Löchern? – Erklärungen von Kindern zum Schwimmen und Sinken verschiedener Gegenstände vor und nach dem Unterricht. In A. Speck-Hamdan, H. Brügelmann, M. Fölling-Albers, & S. Richter (Eds.), *Kulturelle Vielfalt. Religiöses Lernen* (pp. 159–164). Seelze: Kallmeyersche Verlagsbuchhandlung.
- Kawasaki, K., Herrenkohl, L. R., & Yearly, S. A. (2004). Theory building and modeling in a sinking and floating unit: a case study of third and fourth grade students' developing epistemologies of science. *International Journal of Science Education*, 26(11), 1299–1324.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M. et al. (2003). *Expertise zur Entwicklung nationaler Bildungsstandards*. Berlin: BMBF.
- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Überarbeitete Fassung des Antrags an die DFG auf Einrichtung eines Schwerpunktprogramms.
- Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme & D. Leutner (Eds.), *Assessment of Competencies in Educational Contexts* (pp. 3–22). Göttingen: Hogrefe & Huber Publishers.
- Labudde, P. (2008). Developing and Implementing New National Standards in Science Education: The Role of Science Educators. In B. Ralle, & I. Eilks (Hrsg.), *Didaktik der Naturwissenschaften: Promoting Successful Science Education – The Worth of Science Education Research* (S. 63–74), Aachen: Shaker.
- Linn, M. (2006). The knowledge integration perspective on learning and instruction. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 243–264). Cambridge: Cambridge University Press.
- Mason, L. (2007). Introduction: Bridging the Cognitive and Sociocultural Approaches in Research on Conceptual Change: Is it Feasible? *Educational Psychologist*, 42(1), 1–7.
- Mercer, N. (2008). Changing our minds: a commentary on 'Conceptual change: a discussion of theoretical, methodological and practical challenges for science education'. *Cultural Studies of Science Education*, 3(2), 351–362.
- Möller, K., Jonen, A., Hardy, I., & Stern, E. (2002). Die Förderung von naturwissenschaftlichem Verständnis bei Grundschulkindern durch Strukturierung der Lernumgebung. *Zeitschrift für Pädagogik*, 45. Beiheft, 176–191.
- National Science Education Standards (1996). *National Research Council, National Committee on Science Education Standards and Assessment*. Washington, DC.
- Norris, S., & Philipps, L. (2003). How Literacy in its Fundamental Sense is Central to Scientific Literacy. *Science Education*, 87(2), 224–240.
- Pollmeier, J., Kleickmann, T., Hardy, I., Tröbst, S., Möller, K., & Schwippert, K. (2009). Entwicklung naturwissenschaftlicher Kompetenz in der Grundschule (Science-P): Naturwissenschaftliches Wissen. In C. Röhner, C. Henrichwark, & M. Hopf (Hrsg.), *Europäisierung der Bildung. Konsequenzen und Herausforderungen für die Grundschulpädagogik* (S. 199–203), Wiesbaden: VS Verlag für Sozialwissenschaften.
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hamann, M., Klieme, E. et al. (2007). *PISA 2006 in Deutschland – Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.
- Qualifications and Curriculum Authority. (2000). Department for Education and Employment (DfEE). *Science: The National Curriculum for England: Key Stages 1–4*.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*, 2. Aufl. Bern: Huber.
- Siegler, R. S., & Chen, Z. (2008). Differentiation and integration: guiding principles for analyzing cognitive change. *Developmental Science*, 11(4), 433–453.
- Singer, J. A., Kohn, A. S., & Resnick, L. B. (1997). Knowing about proportions in different contexts. In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics: An international perspective* (pp. 115–132). Hove, England: Psychology Press.
- Smith, C. (2007). Bootstrapping Processes in the Development of Students' Commonsense Matter Theories: Using Analogical Mappings, Thought Experiments, and Learning to Measure to Promote Conceptual Restructuring. *Cognition and Instruction*, 25(4), 337–398.

- Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition*, 21, 177–237.
- Smith, C., Maclin, D., Grosslight, L., & Davis, H. (1997). Teaching for understanding. A study of students' preinstruction theories of matter and a comparison of the effectiveness of two approaches to teaching about matter and density. *Cognition and Instruction*, 15, 317–393.
- Sodian, B., Jonek, A., Thoermer, C., & Kircher, E. (2006). Die Natur der Naturwissenschaften verstehen – Implementierung wissenschaftstheoretischen Unterrichts in der Grundschule. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms*. Münster: Waxmann.
- Stathopoulou, C., & Vosniadou, S. (2007). Exploring the relationship between physics-related epistemological beliefs and physics understanding. *Contemporary Educational Psychology*, 32, 255–281.
- Stern, E., & Möller, K. (2004). Der Erwerb anschlussfähigen Wissens als Ziel des Grundschulunterrichtes. In D. Lenzen, J. Baumert, R. Watermann & U. Trautwein (Hrsg.), *PISA und die Konsequenzen für die erziehungswissenschaftliche Forschung* (S. 25–36), Wiesbaden: VS (= Zeitschrift für Erziehungswissenschaft. 3. Beiheft).
- Treagust, D. F. & Duit, R. (2008a). Compatibility between cultural studies and conceptual change in science education: There is more to acknowledge than to fight straw men! *Cultural Studies of Science Education*, 3(2), 387–395.
- Treagust, D.F., & Duit, R. (2008b). Conceptual change: A discussion of theoretical, methodological and practical challenges for science education. *Cultural Studies of Science Education*, 3, 297–328.
- Tyler, R. (2000). A comparison of year 1 and year 6 students' conceptions of evaporation and condensation: dimensions of conceptual progression. *International Journal of Science Education*, 22(5), 447–467.
- Tyler, R. (1998). The nature of students' informal science conceptions. *International Journal of Science Education*, 20(8), 901–927.
- Tyler, R., & Peterson, S. (2004). From „try it and see“ to strategic exploration: Characterizing young children's scientific reasoning. *Journal of Research in Science Teaching*, 41(1), 94–118.
- Vosniadou, S. (2007). Conceptual Change and Education. *Human Development*, 50(1), 47–54.
- Vosniadou, S, Baltas, A., & Vamvakoussi, X. (2007). *Re-Framing the Conceptual Change Approach in Learning and Instruction. Advances in Learning and Instruction Series*, Amsterdam: Elsevier.
- Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction*, 15, 317–419.
- Wandersee, J., Mintzes, J., & Novak, J. (1994). Research on alternative conceptions in science. In D. Gabel (Ed.), *Handbook of Research on Science Teaching and Learning* (pp. 177–210). New York: Macmillan Publishing Company.
- Weinert, F. E. (2001). A concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Seattle: Hogrefe & Huber.
- Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. Mahwah: Lawrence Erlbaum Associates.

Kontakt

Dr. Thilo Kleickmann
Max-Planck-Institut für Bildungsforschung
Forschungsbereich Erziehungswissenschaft
und Bildungssysteme
Lentzeallee 94
14195 Berlin
kleickmann@mpib-berlin.mpg.de

Autoreninformation

Dr. Thilo Kleickmann ist wissenschaftlicher Mitarbeiter im Forschungsbereich Erziehungswissenschaft und Bildungssysteme am Max-Planck-Institut für Bildungsforschung. Seine Forschungsschwerpunkte liegen in der empirischen Lehr-Lern-Forschung, insbesondere im Bereich Naturwissenschaften in der Grundschule, sowie in Forschungen zur professionellen Kompetenz von Lehrkräften.

Prof. Dr. Ilonca Hardy ist Professorin für Erziehungswissenschaft mit dem Schwerpunkt Grundschulpädagogik an der Johann Wolfgang Goethe-Universität in Frankfurt. Ihre Forschungsschwerpunkte sind kognitive Strukturierung im mathematisch-naturwissenschaftlichen Unterricht, Entwicklung und Förderung naturwissenschaftlicher Kompetenzen im Grundschulalter, Lern- und Unterrichtsprozesse im Elementarbereich sowie Zweisprachigkeit im schulischen Kontext.

Prof. Dr. Kornelia Möller ist Professorin für Didaktik des Sachunterrichts an der Westfälischen Wilhelms-Universität Münster und Mitglied der DFG-Forschergruppe und des Graduiertenkollegs Naturwissenschaftlicher Unterricht an der Universität Duisburg-Essen. Ihre Forschungsschwerpunkte sind Lehr-Lernprozesse und Unterrichtsqualität im naturwissenschaftlichen Unterricht der Primarstufe, Aufbau und Entwicklung naturwissenschaftlicher Konzepte im Grundschulalter, Professionswissen und dessen Förderung durch Aus- und Fortbildung bei Grundschullehrkräften im Bereich des naturwissenschaftlichen und technischen Lernens.

Judith Pollmeier ist Diplom-Psychologin und wissenschaftliche Mitarbeiterin am Seminar für Didaktik des Sachunterrichts in Münster. Sie ist

Doktorandin im Projekt Science-P und promoviert zum Thema Modellierung naturwissenschaftlicher Kompetenz in der Grundschule.

Steffen Tröbst ist Diplom-Psychologe und wissenschaftlicher Mitarbeiter am Seminar für Didaktik des Sachunterrichts in Münster.

Dr. Christina Beinbrech ist Lehrerin für die Primarstufe, Fach Sachunterricht. Promotion in der Didaktik des Sachunterrichts. Bis 2009 als Hochschulassistentin am Seminar für Didaktik des Sachunterrichts, Westfälische Wilhelms-Universität Münster tätig. Ihre Forschungsschwerpunkte liegen in der fachdidaktischen Lehr-Lern-Forschung zum naturwissenschaftlich-technischen Sachunterricht, insbesondere in der videobasierten Analyse von Unterrichtsgesprächen.

