

Lassen sich naturwissenschaftliche Lernstände im Grundschulalter mit schriftlichen Aufgaben valide erfassen?

Zusammenfassung: Im naturwissenschaftlichen Unterricht ist die Diagnostik von Schülerkonzepten im Sinne eines formativen Assessments eine notwendige Voraussetzung, um die konzeptuelle Entwicklung angemessen fördern zu können. Während qualitative Interviews zur Erfassung konzeptueller naturwissenschaftlicher Lernstände bei Grundschulern weit verbreitet sind, ist die Frage, inwiefern auch die ökonomischeren, geschlossenen schriftlichen Antwortformate eine valide Diagnostik individueller Konzepte ermöglichen, bisher noch wenig untersucht. Der vorliegende Beitrag zeigt an jeweils zwei Klassen der dritten Jahrgangsstufe der Grundschule (Alter: $M = 9.26$ Jahre, $SD = 0.42$) für die beiden naturwissenschaftlichen Themenbereiche *Schwimmen und Sinken* ($N = 41$) und *Verdunstung und Kondensation* ($N = 32$), dass ein substantieller Zusammenhang zwischen mündlichen Interviews und schriftlichen Testergebnissen besteht und somit – unter bestimmten Voraussetzungen – von einer validen Erfassung kindlicher Lernstände auch mit schriftlichen Aufgaben ausgegangen werden kann.

1. Lernstände im naturwissenschaftlichen Lernbereich individuell erfassen – eine Grundlage für die Förderung konzeptueller Entwicklung

Bereits vor der Auseinandersetzung mit naturwissenschaftlichen Inhalten im Unterricht entwickeln Schüler¹ vielfältige naturwissenschaftliche Konzepte (Driver, 1989), die häufig nicht mit wissenschaftlichen Erklärungen vereinbar sind und zudem fragmentiert und situationsabhängig sein können (Kleickmann, Pollmeier, Hardy & Möller, im Druck; Treagust & Duit, 2008). Auch in den in diesem Beitrag untersuchten Inhaltsbereichen wurden vielfältige naive Konzepte identifiziert (vgl. für *Schwimmen und Sinken* z.B. Engelen, Jonen & Möller, 2002; Smith, Maclin, Grosslight & Davis, 1997; für *Verdunstung und Kondensation* z.B. Bar & Travis, 1991; Tytler, Peterson & Prain, 2006). Es wird davon ausgegangen, dass das Lernen neuer, angemessenerer Konzepte auf vorhandenen konzeptuellen Strukturen aufbaut und deren Modifikation, Erweiterung oder Revision beinhalten kann (Scott, Asoko & Leach, 2007). Der Prozess der konzeptuellen Umstrukturierung kann insbesondere dadurch erschwert werden, dass vorunterrichtliche Konzepte sehr robust sind, da sie den Rahmen für die Wahrnehmung, Verarbeitung

1 Im Text werden bevorzugt geschlechtsneutrale Formulierungen verwendet. Um die Lesbarkeit zu verbessern, verwenden wir an einigen Stellen Begriffe wie Schüler oder Experten als generisches Maskulinum für beide Geschlechter.

und Interpretation von neuen Informationen bilden (Duit, 1999). Die Diagnose der Schülerkonzepte wird deswegen als wesentlicher Bestandteil eines auf individuelle Förderung konzeptueller Entwicklung abzielenden Unterrichts angesehen (Bell & Cowie, 2001).

Für die Diagnose von konzeptuellem Wissen im Grundschulalter werden überwiegend Interviews und freie Antwortformate verwendet (z.B. Carey, Evans, Honda, Jay & Unger, 1989). Als nicht oder weniger geeignet gelten Multiple-choice-Aufgaben (MC-Aufgaben) und Zuordnungsaufgaben, insbesondere in Form von Paper-pencil-Tests (Duit, Häußler & Prenzel, 2001). Einer der Hauptkritikpunkte an MC-Tests im Grundschulalter betrifft die Validität, die von dem noch nicht ausgereiften Leseverständnis beeinflusst sein könnte, da sich gerade in den ersten Grundschuljahren noch deutliche Entwicklungen in der Gedächtnisspanne und in Strategien zum effektiven und sinnerfassenden Lesen zeigen (Carr, Kurtz, Schneider, Turner & Borkowski, 1989). Befunde aus Leseleistungstests (z.B. ‚Ein Leseverständnistest für Erst- bis Sechstklässler‘ (ELFE 1-6; Lenhardt & Schneider, 2006) oder auch Studien wie IGLU (Internationale Grundschul-Lese-Untersuchung) oder VERA (VERgleichsArbeiten) deuten aber darauf hin, dass sinnerfassendes Lesen bereits im Grundschulalter möglich ist, allerdings von der Komplexität des Textes und unterstützenden Elementen (z.B. Abbildungen) beeinflusst wird (Helmke et al., 2007).

Schriftliche Aufgaben mit geschlossenem Antwortformat wären im Vergleich zu Interviews hinsichtlich ihrer Ökonomie als durchaus attraktiv einzustufen, wenn sie den Testgütekriterien genügen könnten. Bis dato gibt es einige wenige Studien zur Entwicklung (natur)wissenschaftlicher Kompetenzen, in denen sowohl mit offenen Fragen als auch mit geschlossenen schriftlichen Antwortformaten gearbeitet wurde (z.B. Bullock, Sodian & Koerber, 2009; Frède et al., 2011). In diesen beiden Studien finden sich signifikante Korrelationen zwischen beiden Antwortformaten, die für eine konvergente Validität, also die Messung des gleichen Konstrukts, sprechen. Allerdings weisen diese Studien bessere Ergebnisse bei den geschlossenen als bei den offenen Aufgaben auf, sodass möglicherweise eine Überschätzung der Schülerleistungen bei geschlossenen, schriftlichen Antwortformaten oder eine Unterschätzung der Schülerkompetenzen bei den offenen Interviewformaten stattfindet. Diese Befunde werden dahingehend diskutiert, dass schon früher, als bisher aufgrund von Interviewstudien angenommen, ein beginnendes Verständnis des untersuchten Konstrukts vorliegt, welches sich jedoch in spontanen Produktionsaufgaben erst in höherem Alter zeigt (Bullock et al., 2009). Ein systematischer Vergleich von geschlossenen Antwortformaten mit Interviewdaten im Grundschulalter für die Erfassung von konzeptuellem Wissen im Bereich der Naturwissenschaften steht jedoch bis dato aus. Die vorliegende Studie vergleicht daher die mittels schriftlicher, geschlossener Antwortformate und mittels mündlicher Verfahren in Form von Interviews erhobenen naturwissenschaftlichen konzeptuellen Lernstände im Grundschulalter.

2. Methode

2.1. Material

Die in der vorliegenden Untersuchung eingesetzten Aufgaben wurden im Rahmen des Projektes Science-P² entwickelt und pilotiert. Ziel des Projektes ist die theoriegeleitete Entwicklung und empirische Überprüfung eines Modells der Kompetenzentwicklung in der Primarstufe für die beiden Dimensionen *Naturwissenschaftliches Wissen* (d.h. bereichsspezifisches konzeptuelles Wissen) und *Wissen über Naturwissenschaften* (d.h. naturwissenschaftliche Denk- und Arbeitsweisen sowie Wissenschaftsverständnis) (Hardy et al., 2010). Die vorliegende Arbeit bezieht sich auf die Dimension *Naturwissenschaftliches Wissen*, innerhalb derer die beiden physikalischen Inhaltsbereiche *Schwimmen und Sinken*, mit den Komponenten Dichte, Verdrängung und Auftrieb, und *Verdunstung und Kondensation*, mit den Komponenten Verdunstung, Kondensation und Wasserkreislauf (Wolken, Regen), untersucht werden (Kleickmann et al., 2010). Die naturwissenschaftlichen Konzepte der Schüler in den untersuchten Themenbereichen wurden drei hierarchischen Kompetenzniveaus zugeordnet: (1) Naiven Vorstellungen (Fehlvorstellungen), die einer empirischen Prüfung in unterschiedlichen Kontexten nicht standhalten, (2) Zwischenvorstellungen, mit denen Phänomene begrenzt erklärt werden können, und (3) wissenschaftlichen Vorstellungen, die auf in der Wissenschaft geteilten Konzepten beruhen (vgl. Hardy et al., 2010). Auf der Basis dieses Struktur-Niveau-Modells wurden 120 schriftliche Aufgaben mit größtenteils geschlossenem Antwortformat entwickelt, wobei die vorgegebenen Antworten jeweils einem der postulierten Kompetenzniveaus zugeordnet werden können. Die Antwortalternativen wurden dabei aus der Präkonzeptforschung zu den beiden Themenbereichen abgeleitet und in einer so genannten ‚construct map‘, in der die Anforderungen für unterschiedliche Kompetenzausprägungen spezifiziert werden, den jeweiligen Niveaus zugeordnet (vgl. Wilson, 2005). Die Aufgaben erfordern die Erklärung von Phänomenen in Alltagskontexten bzw. leicht verständlichen Experimentalsituationen und thematisieren die Begründung von Annahmen und Entscheidungen.

Zu den eingesetzten Aufgabenformaten gehören Aufgaben, bei denen die beste aus zwei (forced-choice) oder mehreren (multiple-choice) vorgegebenen Antworten gewählt werden soll, sowie Aufgaben, bei denen mehrere Antwortalternativen separat als richtig oder falsch beurteilt werden sollen (multiple-select, vgl. Abb. 1). Zusätzlich wurden noch einige Aufgaben mit offenem bzw. graphischem Antwortformat entwickelt. Für die vorliegende Studie wurde für jeden der untersuchten Inhaltsbereiche ein Testheft

² 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). Das Projekt Science-P (I und II) wird von den Arbeitsgruppen Sodian/Mayer (München), Hardy (Frankfurt), Koerber (Freiburg), Schwippert (Hamburg) und Möller/Kleickmann/Pollmeier/Tröbst (Münster) gemeinsam durchgeführt.

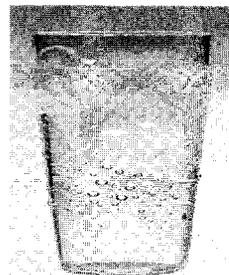
Kaltes Glas

Instruktion des Testleiters:

Du füllst ein Glas mit Leitungswasser und Eiswürfeln. Das Glas ist zunächst außen ganz trocken, aber nach einem kurzen Moment sieht es so aus wie auf dem Bild: Plötzlich sind Wassertropfen außen auf dem Glas.

Damit du dir das gut vorstellen kannst, werden wir den Versuch jetzt selbst ausprobieren (Demonstration mit den entsprechenden Materialien).

Plötzlich sind Wassertropfen außen auf dem Glas.
Wie sind die Wassertropfen außen an das Glas gekommen?



Kreuze nach jeder Antwort ‚Richtig‘ oder ‚Falsch‘ an!

	Richtig	Falsch
Die Wassertropfen sind durch feine Poren im Glas von innen nach außen gekommen. (NV)	<input type="checkbox"/>	<input type="checkbox"/>
Wasser aus der Luft ist durch die Kälte als Wassertropfen sichtbar geworden. (WV)	<input type="checkbox"/>	<input type="checkbox"/>
Die Wassertropfen sind wegen der Kälte aus der Luft entstanden. (ZV)	<input type="checkbox"/>	<input type="checkbox"/>

NV = Naive Vorstellung; ZV = Zwischenvorstellung; WV = Wissenschaftliche Vorstellung

Abb. 1: Multiple-select-Aufgabe für die Komponente Kondensation

mit 13 Aufgaben zusammengestellt (vgl. Tabelle 1), wobei die psychometrischen Kennwerte der Aufgaben in den Vorstudien sowie eine Repräsentation aller Inhaltsgebiete und Formate entsprechend ihrer relativen Gewichtung im Gesamtprojekt als Auswahlkriterien berücksichtigt wurden.

Für beide eingesetzten Testhefte gab es einen korrespondierenden Interviewleitfaden, in dem nach Vorgabe der gleichen Aufgabenstämme wie im Fragebogen nach einer offenen Antwort gefragt wurde.³ Diese halbstandardisierten, mündlichen Einzelinterviews wurden in Anlehnung an eigene Vorarbeiten entwickelt (vgl. Engelen et al., 2002).

³ Das Interview zum Inhaltsbereich Schwimmen und Sinken umfasste nur zwölf der dreizehn Aufgabenstämme, da die Aufgabe mit graphischem Antwortformat nur in Bezug auf das Niveau ihrer Bearbeitung mit einer Aufgabe gleichen Inhalts verglichen werden sollte, um so einen möglichen Effekt des graphischen Formats zu untersuchen.

Inhaltsbereich	Komponente	Aufgabenformat				Gesamt
		forced choice	multiple select	multiple choice	graphisch	
Schwimmen und Sinken	Dichte	–	–	2	–	2
	Verdrängung	2	–	–	1	3
	Auftrieb	5	3	–	–	8
Verdunstung und Kondensation	Verdunstung	2	3	–	–	5
	Kondensation	3	3	–	–	6
	Wasserkreislauf (Wolken, Regen)	1	1	–	–	2
Gesamt		13	10	2	1	26

Tab. 1: Zusammenstellung der Testhefte für die beiden Inhaltsbereiche

Als Kontrollvariablen wurden von allen Schülern das Leseverständnis (ELFE 1-6), die allgemeine Intelligenz (CFT 20-R: Untertests Topologische Schlussfolgerungen, Reihen fortsetzen; Weiß, 2006) sowie demographische Angaben erhoben. Zudem wurden der höchste Bildungsabschluss der Eltern, kodiert nach der International Standard Classification for Education (ISCED) (UNESCO Institute for Statistics, 1999), und der Status der aktuellen Beschäftigung der Eltern, kodiert nach dem International Socio-Economic Index of Occupational Status (ISEI) (Ganzeboom, de Graaf, Treiman & de Leeuw, 1992), über einen Elternfragebogen (vgl. Bos et al., 2005) erfasst.

2.2. Stichprobe

An der Studie nahmen drei Klassen der dritten Jahrgangsstufe aus Grundschulen in Nordrhein-Westfalen teil. Eine der Klassen bearbeitete die Instrumente zu beiden Inhaltsbereichen (*Schwimmen und Sinken*: $N = 20$; *Verdunstung und Kondensation*: $N = 16$).⁴ Es ergibt sich für *Schwimmen und Sinken* insgesamt eine Stichprobe von $N = 41$ Schülern (20 Jungen, 21 Mädchen; Alter: $M = 9.27$, $SD = 0.42$) und für *Verdunstung und Kondensation* von $N = 32$ Schülern (15 Jungen, 17 Mädchen; Alter: $M = 9.24$, $SD = 0.41$).

⁴ Aufgrund krankheitsbedingter Ausfälle unterscheidet sich die Fallzahl der Klasse, die beide Themenbereiche bearbeitet hat, für die beiden Teilstudien.

2.3. Durchführung

Alle Schüler nahmen an Tests zum Leseverständnis und zur allgemeinen Intelligenz (Kontrollvariablen) sowie an beiden Tests zum *Naturwissenschaftlichen Wissen* (Interview/schriftlicher Fragebogen) teil. Die schriftlichen Tests wurden im Klassenverband im Abstand von circa zwei Wochen durchgeführt. Dabei wurden die Aufgaben zum *Naturwissenschaftlichen Wissen* zur Minimierung des Einflusses des Leseverständnisses von den Testleitenden vorgelesen und bei 14 der 26 Aufgaben die in den Aufgabenstäm- men dargestellten Situationen mit realen Gegenständen demonstriert. Die mündlichen Interviews wurden von drei geschulten Interviewenden während der Unterrichtszeit mit den einzelnen Kindern durchgeführt, die zu diesem Zweck den Unterricht für circa 25 bis 35 Minuten unterbrachen. Die Schüler wurden dabei randomisiert einer Gruppe zu- gewiesen: Eine Hälfte der Schüler bearbeitete das mündliche Interview vor dem schrift- lichen Fragebogen, während die andere Hälfte die Instrumente in umgekehrter Reihen- folge bearbeitete. Der Elternfragebogen wurde am Ende der Studie an die Eltern aus- gegeben.

2.4. Aufbereitung und Analyse des Datenmaterials

Die Antworten der Schüler auf die Fragen im Interview wurden einer qualitativen In- haltsanalyse in Anlehnung an Gropengießer (2007, nach Mayring, 2008) unterzogen. Die Auswertungen nahmen für jeden Themenbereich zwei Kodierende vor, die von Ex- perten hinsichtlich des methodischen und inhaltlichen Vorgehens geschult wurden.

Nach der Transkription der aufgezeichneten Interviews erfolgte eine Reduktion des Datenmaterials (vgl. Gropengießer, 2007) mit dem Ziel, redundante und irrelevante Textstellen zu eliminieren und wesentliche Informationen zu erhalten. Die redigierten Aussagen wurden anschließend nach ihrem Inhalt geordnet. Im letzten Schritt, der Struk- turierung, wurden die in den Schülerantworten identifizierten Konzepte anhand eines Kodierleitfadens Kategorien zugeordnet, die wiederum den drei postulierten Kompe- tenzniveaus zugewiesen waren. Ausgangspunkt des Kodierleitfadens war die ‚construct map‘, die auch der Entwicklung der schriftlichen Testaufgaben zugrunde lag (s.o.). Sie wurde in einer induktiven Vorgehensweise um Kategorien erweitert, die aus den Inter- viewdaten abgeleitet werden konnten.

Für jeden der Analyseschritte wurde ein schriftlicher Leitfaden in einem iterativen Prozess erarbeitet. Für die Strukturierung wurde zudem die Interrater-Reliabilität an- hand einer Sub-Stichprobe von acht Interviews je Themenbereich berechnet; Spearman's $\rho = 0.78$ (*Schwimmen und Sinken*: $\rho = 0.82$, *Verdunstung und Kondensation*: $\rho = 0.71$). Die Tabellen 2 und 3 geben eine Übersicht über die identifizierten Konzepte.

Konzeptname und Ankerbeispiel	Int	FB
<i>Naive Vorstellungen</i>		
Gewichts-Konzept: „Der Nagel geht nicht unter, weil der so leicht ist.“	73	38
Wassermenge-Konzept: „Mehr Wasser macht, dass der Nagel schwimmen kann.“	37	25
Gewicht-Verdrängung-Konzept: „Das Wasser wird bei diesem Klotz viel höher steigen, weil der ja viel schwerer ist.“	37	34
Antriebs-Konzept: „Die Enten gehen nicht unter, weil die mit den Füßen paddeln.“	34	18
Kleine Differenzierung von Gewicht und Dichte: „Der Klotz ist ja schwerer, deswegen ist der aus dem schwereren Material.“	27	11
Luft-aktiv-Konzept: „Die Luft zieht das Ding nach oben.“	26	8
Loch-Konzept: „Das Brett schwimmt, wegen den Löchern.“	18	–
Luft-im-Wasser-Konzept: „Im Wasser sind ja meistens so Blubberbläschen und die drücken den Baumstamm dann hoch.“	17	–
Wasser-im-Gegenstand-Konzept: „Wenn da Wasser drin ist, dann geht das Ding unter.“	12	–
Form-Konzept: „Die Enten sind ja unten geformt wie ein Boot, deshalb schwimmen die.“	11	12
Gewicht-verkehrt-Konzept: „Der Baumstamm schwimmt, weil schwere Sachen immer oben schwimmen.“	11	–
Wassergewicht-Konzept: „Das kann schwimmen, weil das Wasser so leicht ist.“	11	–
Wasserbewegung-Konzept: „Der Nagel schwimmt wegen der Wellen.“	9	–
Gleiches Material = Gleiches Gewicht: „Die Klötze sind aus dem gleichen Material, weil sie gleich schwer sind.“	9	15
Gleichgewichts-Konzept: „Das Holzbrett kann schwimmen, weil die Masse gleich verteilt ist.“	5	–
Wasserabweisend-Konzept: „Federn sind wasserdicht, deswegen können die Enten schwimmen.“	4	–
Schwerkraft-Konzept: „Im Wasser wirkt die Schwerkraft ein bisschen weniger.“	3	–
Natur-Konzept: „Der Baumstamm kommt ja aus der Natur, deswegen schwimmt er.“	3	–
Größen-Konzept: „Der Tischtennisball schwimmt, weil der so klein ist.“	2	16
Falsches Druck-Konzept: „Das Wasser versucht, den Apfel nach unten zu ziehen.“	2	21

Konzeptname und Ankerbeispiel	Int	FB
Schwerer-im-Wasser-Konzept: „Wenn man eine Schraube ins Wasser tunkt, dann kriegt die noch ein bisschen Gewicht dazu.“	2	–
Substanz-im-Wasser-Konzept: „Chlor im Wasser lässt schwimmen.“	2	–
Reine Deskription [Proband beschreibt Phänomen ohne Erklärung].	1	–
Schwerer-als-Wasser-Konzept: „Der Baum schwimmt, weil er schwerer als das Wasser ist.“	1	–
Oberflächen-Konzept: „Das Wasser muss ja nicht das ganze Gewicht des Baumstammes heben, sondern nur die Rinde, weil ja nur die beim Wasser drauf ist.“	1	–
Position-Konzept: „Wenn die Schraube hochkant reingestellt wird, dann schwimmt sie.“	1	–
Luftdruck-Konzept: „Das Wasser läuft über, weil der Luftdruck da rein geht, wenn man den Würfel reinlegt.“	1	–
Gesamt	360	198
<i>Zwischenvorstellungen</i>		
Luft-Konzept: „Die schwimmen, weil die auch ein bisschen Luft innen drin haben.“	91	49
Material-Konzept: „Der Baumstamm besteht aus Holz und Holz schwimmt.“	70	38
Leichter-als-Wasser-Konzept: „Der Ball schwimmt, weil der leichter als das Wasser ist.“	63	–
Naives Druck-Konzept: „Das Wasser hilft den Stamm hochzuheben.“	58	–
Leichter-im-Wasser-Konzept: „Man wird im Wasser leichter.“	42	11
Hohl-Konzept: „Die schwimmen, weil die Äpfel sind ja ein bisschen hohl von innen.“	33	37
Verdrängungs-Konzept: „Das Wasser wird vom Baumstamm verdrängt und dann kann der Baumstamm schwimmen.“	3	–
Vollkörperkonzept: „In dem Klotz ist ja kein Loch drin, wo das Wasser reinlaufen kann, deswegen läuft das Wasser über.“	1	–
Gesamt	364	135
<i>Wissenschaftliche Vorstellungen</i>		
Druck-Konzept: „Der schwimmt, weil das Wasser ihn nach oben drückt.“	60	67
Volumen-Verdrängung: „Das Wasser schwappt über, weil der Klotz Platz braucht.“	50	48

Konzeptname und Ankerbeispiel	Int	FB
Dichte als Materialeigenschaft: „Die Klötze sind gleich groß, aber unterschiedlich schwer, weil der kleine Klotz ja aus einem anderen Material ist.“	18	31
Quantitatives Dichteverständnis: „Der Klotz ist ja dreimal so groß wie der andere, wenn der aus dem gleichen Material wäre, würde der sechs Kilo wiegen und keine fünf.“	18	25
Differenzierung Gewicht-Dichte: „Der Klotz ist ja für seine Größe schwerer als der andere.“	14	–
Dichtevergleich: „Der Baumstamm schwimmt, weil er weniger wiegt als die gleiche Menge Wasser.“	–	19
Archimedisches Prinzip: „Ein Apfel schwimmt, weil er weniger wiegt als das Wasser, das er wegdrängt.“	–	18
Vergleich Gewicht – Auftrieb: „Das Holzbrett schwimmt, weil es schwächer vom eigenen Gewicht nach unten gezogen wird als das Wasser es nach oben drückt.“	–	24
Gesamt	160	232

Wurde ein Konzept im Interview nicht geäußert, wurde eine Antwortalternative des Fragebogens als Ankerbeispiel übernommen. Int = Interview; FB = Fragebogen

Tab. 2: Anzahl der Nennungen aller Konzepte in Fragebogen und Interview im Themenbereich Schwimmen und Sinken

Konzeptname und Ankerbeispiel	Int	FB
<i>Naive Vorstellungen</i>		
Inadäquate Fehldeutungen (V): „Die Pflanzen und Fische haben das Wasser getrunken.“	77	56
Falsche Ursachen (K): „Es bilden sich Tröpfchen, weil es so warm ist.“	43	22
Falscher Ortswechsel (K): „Die Wassertropfchen kommen oben aus dem Glas raus und gehen direkt an das Glas.“	41	44
Phänotypische Analogiebildung (V): „Das ganze Wasser ist heruntergetropft.“	23	–
Absorption (V): „Das Wasser ist in das Geschirr eingezogen.“	22	–
Falsche Deutung der Zustandsänderung (K): „Die Kälte ist zu Wasser geworden“	18	8
Falsche Substanz (Wk): „Die Wolken bestehen aus Abgasen.“	18	35
Anthropomorphe Deutung (V): „Die Sonne zieht das Wasser hoch.“	16	15

Konzeptname und Ankerbeispiel	Int	FB
Falsche Ursachen (Wk): „Die Wolken entstehen oben am Himmel, weil da nichts anderes ist.“	15	22
Anthropomorphe Deutung (K): „Der Spiegel zieht das Wasser an.“	14	–
Keine Massenerhaltung (K): „In der Flasche ist jetzt mehr Wasser als am Montag.“	11	6
Durch Materie hindurch (K): „Die Tröpfchen kommen durch das Glas nach außen.“	8	16
Keine Massenerhaltung (V): „Das Wasser ist einfach verschwunden.“	6	13
Anthropomorphe Deutung (Wk): „Die Wolken entstehen, weil die Sonne das Wasser hochzieht.“	5	–
Reine Deskription [Proband beschreibt Phänomen ohne Erklärung].	3	–
Leerer Begriff [Proband verwendet die Begriffe „Verdunsten“/„Kondensieren“, ohne sie erklären zu können].	3	–
Falsche Deutung der Zustandsänderung (V): „Das Wasser ist zu Wärme geworden.“	2	18
Gesamt	325	255
<i>Zwischenvorstellungen</i>		
Richtiger Ortswechsel (V): „Das Wasser ist in die Luft aufgestiegen.“	52	32
Richtige Substanz (Wk): „Die Wolken bestehen aus Wasser“	47	29
Richtige Ursachen (K): „Die Tröpfchen kommen an das Glas, weil das kälter ist als die Luft.“	33	25
Richtige Ursachen (V): „Das Wasser verschwindet schneller wegen der Wärme.“	23	–
Richtiger Ortswechsel (K): „Das Wasser kommt aus der Luft.“	17	–
Richtige Zustandsänderung (V): „Das Wasser kann man jetzt nicht mehr sehen, es wird zu einem dünnen Dampf.“	12	26
Richtige Zustandsänderung (K): „Die Tröpfchen entstehen aus Wasserdampf.“	8	–
Richtige Ursache (Wk): „Die Wolken entstehen wegen der Kälte.“	5	5
Richtige Zustandsänderung (Wk): „Wolken entstehen aus Wasserdampf.“	2	–
Gesamt	199	117
<i>Wissenschaftliche Vorstellungen</i>		
Verdunstung: „Das Wasser steigt in die Luft und wird zu Dampf.“	27	88

Konzeptname und Ankerbeispiel	Int	FB
Wolkenstruktur: „Wolken sind aus kleinen verdunsteten Wassertröpfchen.“	16	30
Wolkenentstehung: „Wasser aus der Luft sammelt sich wegen der Kälte als Wassertröpfchen.“	15	27
„Gasförmiges“ Wasser in der Luft: „In der Luft ist immer Wasser.“	11	22
Kondensation: „Das Wasser kommt aus der Luft, vorher war das so etwas wie Nebel. Das Wasser setzt sich auf das Glas, weil das kälter ist.“	7	69
Gesamt	76	262

Int = Interview; FB = Fragebogen; V = Verdunstung; K = Kondensation; Wk = Wasserkreislauf

Tab. 3: Anzahl der Nennungen aller Konzepte in Fragebogen und Interview im Themenbereich Verdunstung und Kondensation

3. Ergebnisse

3.1. Themenbereich Schwimmen und Sinken

Bei diesem Themenbereich konnte auf dem unteren und mittleren Niveau im Interview eine größere Anzahl inhaltlicher Konzeptkategorien identifiziert werden als im Fragebogen (siehe Tab. 2). Dieser Unterschied ist auf dem Niveau der naiven Vorstellungen besonders deutlich: Hier wurden im Interview von den Schülern annähernd dreimal so viele unterschiedliche Konzeptkategorien geäußert wie im Fragebogen. Einige der im Interview erfassten Zwischenvorstellungen („Leichter-als-Wasser-Konzept“ und „naives Druck-Konzept“ mit 63 bzw. 58 Nennungen) waren im Fragebogen nicht als Alternativen angeboten worden (vgl. Tab. 2). Auch gab es Konzepte auf dem wissenschaftlichen Niveau („Dichtevergleich“, „Archimedisches Prinzip“ und „Vergleich Gewicht – Auftrieb“), die zwar im Fragebogen angenommen wurden (19, 18 bzw. 24 Nennungen), nicht aber im Interview geäußert wurden.

Insgesamt waren im Fragebogen 565 Antwortreaktionen, im Interview 884 Konzeptnennungen zu verzeichnen. Von diesen entfielen im Interview 360 Nennungen (40,72%), im Fragebogen hingegen nur 198 Nennungen (35,04%)⁵ auf das Niveau der naiven Vorstellungen. Am seltensten wurde im Fragebogen das Niveau der Zwischenvorstellungen mit 135 Nennungen (23,89%) kodiert, welches im Interview dagegen mit 364 Nennungen (41,18%) am häufigsten vorkam. Das Niveau der wissenschaftlichen Vorstellungen wurde im Fragebogen mit 232 Nennungen (41,02%) am häufigsten ko-

⁵ Bei der Betrachtung der Antworthäufigkeiten ist zu bedenken, dass nur sechs Antwortmöglichkeiten des Fragebogens das mittlere Niveau anboten, während das naive und das wissenschaftliche Niveau 19 bzw. 20 Mal angeboten wurden.

diert, im Interview kam es mit 160 Nennungen (18,10%) dagegen seltener vor. Ein Chi-Quadrat-Test ergab, dass sich die Verteilungen der Antworten auf die Niveaus in den beiden Instrumenten signifikant voneinander unterscheiden ($\chi^2(2) = 99.97, p < 0.01$).

Für die Analyse der Zusammenhänge zwischen den Instrumenten wurden die Personenschätzer aus separat an die dichotom kodierten Items angepassten eindimensionalen Rasch-Modellen⁶ untereinander sowie mit den Testwerten der Kontrollvariablen korreliert (siehe Tab. 4). Für den Fragebogen *Schwimmen und Sinken* lag die Reliabilität der Schätzung der Personenparameter (Warms weighted likelihood estimates – WLE) mit 21 Items⁷ bei 0.70. Die WLE-Schätzer des Interviews erreichten mit 16 Items eine Reliabilität von 0.62. Die Korrelation zwischen den Personenschätzungen von Fragebogen und Interview wiesen einen mittleren Zusammenhang auf ($r = 0.44, p < 0.01$).

Für keines der beiden Instrumente war ein bedeutsamer Zusammenhang mit dem Leseverständnis zu finden. Für den Fragebogen zeigte sich allerdings ein signifikanter Zusammenhang mit der Intelligenz ($r = 0.39, p < 0.05$), der beim Interview zu diesem Themenbereich nicht nachzuweisen war. Beide Modi der Aufgabebearbeitung wiesen zudem eine signifikante Korrelation zum ISCED (Fragebogen: $r = 0.43, p < 0.01$; Interview: $r = 0.42, p < 0.01$) und zum ISEI (Fragebogen: $r = 0.39, p < 0.05$; Interview: $r = 0.32, p < 0.05$) auf.

		N _{Item}	WLE-Reliabilität	Personenfähigkeit***	Korrelationskoeffizienten				
					Interview	ELFE	CFT	ISCED	ISEI
S&S	Fragebogen	21	0.70	0.29 (0.96)	0.44**	0.22	0.39*	0.43**	0.39*
	Interview	16	0.62	0.17 (1.02)	–	0.30	0.26	0.42**	0.32*
V&K	Fragebogen	18	0.70	0.68 (1.30)	0.49**	0.12	0.31	0.11	-0.11
	Interview	25	0.61	-0.86 (0.84)	–	0.02	0.11	0.04	0.22

S&S = Schwimmen und Sinken; V&K = Verdunstung und Kondensation; * $p < 0.05$, ** $p < 0.01$, *** M(SD)

Tab. 4: Deskriptive Statistiken und Ergebnisse der Zusammenhangsanalysen für den Themenbereich Schwimmen und Sinken

- 6 Im Projekt Science-P werden zur Überprüfung des postulierten Struktur-Niveau-Modells IRT-Modelle verwendet. Vor diesem Hintergrund wurde dieser Auswertungsansatz mit dem Rasch-Modell auch für die vorliegende Studie beibehalten.
- 7 Da pro Aufgabenstamm teilweise mehr als eine Antwortreaktion von den Probanden verlangt war (z.B. separate Beurteilung mehrerer Antwortalternativen), übersteigt die Anzahl an Items die Anzahl der Aufgaben. Auch die Anzahl von Items in Fragebogen und Interview unterschied sich.

3.2. Themenbereich Verdunstung und Kondensation

Auch in diesem Themenbereich wurde im Interview eine größere Anzahl unterschiedlicher Konzepte auf dem unteren und mittleren Kompetenzniveau als im Fragebogen gefunden (vgl. Tab. 3). Im Fragebogen wurden einige Konzepte angeboten, die dort häufig gewählt (auf dem naiven Niveau: „falsche Deutung der Zustandsänderung (V)“ mit 18 Nennungen; auf dem wissenschaftlichen Niveau: „Kondensation“ mit 69 Nennungen), im Interview allerdings nur selten geäußert wurden (2 bzw. 7 Nennungen; siehe auch Tab. 3). Hingegen wurden andere Vorstellungen, die im Interview geäußert wurden, nicht durch den Fragebogen erfasst (z.B. auf dem mittleren Niveau: „richtige Ursachen (V)“ mit 23 Nennungen im Interview).

Insgesamt gab es im Interview 600 Konzeptnennungen, während im Fragebogen 634 Antwortreaktionen zu verzeichnen waren. 255 Nennungen (40,22%)⁸ des Fragebogens entfielen davon auf das unterste Niveau; im Interview wurde dieses Niveau mit 325 Nennungen (54,17%) am häufigsten kodiert. Mit 117 Nennungen (18,45%) im Fragebogen war das Niveau der Zwischenvorstellungen das am seltensten kodierte Niveau. Im Interview wurde dieses Antwortniveau mit 199 Nennungen (33,17%) im Unterschied zum Fragebogen am zweithäufigsten zugewiesen. Das wissenschaftliche Niveau war mit 76 Nennungen (12,58%) im Interview das seltenste, im Fragebogen mit 262 Nennungen (41,32%) das häufigste Niveau. Auch in diesem Themenbereich war der Unterschied zwischen den Verteilungen der Niveaus in den beiden Instrumenten statistisch bedeutsam ($\chi^2(2) = 131.24$, $p < 0.01$).

Die Reliabilität der WLE-Schätzer des Fragebogens *Verdunstung und Kondensation* lag mit 23 Items bei 0.70. Die Reliabilität der Schätzung für 25 Items des Interviews erreichte einen Wert von 0.61. Fragebogen und Interview wiesen einen mittleren Zusammenhang untereinander auf ($r = 0.49$, $p < 0.01$). Darüber hinaus fanden sich keine Zusammenhänge mit dem Leseverständnis, der Intelligenz und dem sozio-ökonomischen Status (vgl. Tab. 4).

4. Diskussion

4.1. Wie unterscheidet sich die Erfassung des konzeptuellen Verständnisses mittels Fragebogen bzw. Interview?

Als Ursache für die größere Vielfalt der geäußerten Konzeptkategorien auf dem unteren und mittleren Niveau in beiden Interviews kann angenommen werden, dass seltener vorkommende Präkonzepte, die in der zur Aufgabenentwicklung herangezogenen Literatur nicht regelmäßig berichtet wurden, im Fragebogen nicht als Antwortalternativen angeboten worden waren. Beim Bearbeiten der schriftlichen Aufgaben müssen die

⁸ Auch im Fragebogen Verdunstung und Kondensation unterschied sich die Anzahl an Gelegenheiten, die unterschiedlichen Niveaus zu wählen: Das naive Niveau wurde 23 Mal, das mittlere Niveau sieben Mal und das wissenschaftliche Niveau elf Mal angeboten.

Schüler, die solche Vorstellungen halten, andere Konzepte des gleichen oder sogar eines anderen Niveaus auswählen. Dass durch vorgegebene Antwortalternativen im Fragebogen die Annahme von Konzepten auch suggeriert werden kann, zeigt sich deutlich bei solchen Vorstellungen, die im Interview selten oder gar nicht, im Fragebogen dagegen aber häufig geäußert wurden. Dies kann zu einer fehlerhaften Einschätzung des konzeptuellen Verständnisses der Schüler führen.

Neben der größeren inhaltlichen Vielfalt sind im Interview zum Bereich *Schwimmen und Sinken* insgesamt auch wesentlich mehr Antwortreaktionen zu verzeichnen als im Fragebogen, da die Schüler im Interview bei einer Aufgabe parallel mehrere unterschiedliche Konzepte äußern, während diese Möglichkeit im Fragebogen durch das gewählte Antwortformat begrenzt ist. Im Themenbereich *Verdunstung und Kondensation* hingegen gibt es sogar eine leichte Mehrheit von Nennungen im Fragebogen. Dies könnte ein Hinweis darauf sein, dass im Themenbereich *Schwimmen und Sinken* das Wissen stärker fragmentiert und kontextualisiert ist und somit mehrere nur lose miteinander verknüpfte Wissens Elemente in Abhängigkeit vom Aufgabenstamm im Interview geäußert werden. Die deutlich geringere Anzahl unterschiedlicher Konzeptäußerungen bei *Verdunstung und Kondensation* könnte dementsprechend auf eine stärkere inhaltliche Kohärenz der Vorstellungen in diesem Bereich hindeuten, sodass hier weniger und theorieähnlich miteinander verknüpfte Konzepte geäußert werden. Ob solche Unterschiede in der Struktur des Wissens tatsächlich vorliegen, kann auf der Basis der vorliegenden Studie nicht überprüft werden; hierzu sind weitergehende Auswertungsansätze mit einem größeren Stichprobenumfang erforderlich (vgl. Kleickmann et al., im Druck).

Die Verteilungen der Antworten auf die drei untersuchten Niveaus unterscheiden sich in den jeweiligen Instrumenten signifikant: Im Fragebogen geben die Schüler wesentlich mehr Antworten auf dem wissenschaftlichen Niveau, während sie im Interview am häufigsten auf dem Niveau der naiven Vorstellungen (*Verdunstung und Kondensation*) oder Zwischenvorstellungen (*Schwimmen und Sinken*) antworten. Ähnliche Befunde sind bei Vergleichen offener und geschlossener schriftlicher Antwortformate mit einem ähnlichen Fragebogen zum *Schwimmen und Sinken* (Hardy, Jonen, Möller & Stern, 2006) sowie beim Vergleich offener und geschlossener Aufgaben zur Methodenkompetenz (Bullock et al., 2009) aufgetreten. Die freie Produktion wissenschaftlich angemessener Erklärungen stellt also an die Schüler deutlich höhere Anforderungen als das Erkennen bzw. die Akzeptanz vorgegebener wissenschaftlicher Erläuterungen. Um eine mögliche Überschätzung des Schülerverständnisses bei den geschlossenen Fragebogenformaten zu vermeiden, sollten auch bei schriftlichen Formaten Kurzbegründungen punktuell eingefordert werden. Zusammenfassend sind hinsichtlich der Erfassung des konzeptuellen Verständnisses mit dem Fragebogen im Vergleich mit dem Interview folgende Unterschiede festzuhalten: Im Fragebogen kann die Vielfalt unterschiedlicher naiver Vorstellungen und Zwischenvorstellungen nur begrenzt erfasst werden, eine mögliche Andersartigkeit der Struktur des Wissens in verschiedenen Themenbereichen wird nicht abgebildet und die Fähigkeit der Schüler könnte überschätzt werden.

4.2. Wie valide erfassen die Fragebögen das konzeptuelle Wissen?

Obwohl sich die Erfassung des Wissens, wie oben beschrieben wurde, in einigen Merkmalen unterscheidet, konnte bei beiden Themenbereichen ein mittlerer Zusammenhang zwischen den Instrumenten gefunden werden, was die konvergente Validität der Fragebögen belegt.

Die Bearbeitung der Aufgaben wurde darüber hinaus nicht durch das Leseverständnis beeinflusst, obwohl zu erwarten gewesen wäre, dass bei der Bearbeitung der Fragebögen ein gutes Leseverständnis von Vorteil sein könnte. Demnach scheinen Maßnahmen, wie das Vorlesen aller Aufgaben und Antwortalternativen durch die Testleitenden und das Demonstrieren wichtiger Vorgänge, erfolgreich zur Minimierung des Einflusses des Leseverständnisses beigetragen zu haben.

Bei *Verdunstung und Kondensation* war darüber hinaus kein Zusammenhang mit der allgemeinen Intelligenz nachzuweisen. Für diesen Themenbereich scheint es also gelungen zu sein, das inhaltliche Wissen der Schüler weitgehend frei von den Einflüssen allgemeinerer Kompetenzen zu erfassen. Der Fragebogen *Schwimmen und Sinken* wies im Gegensatz dazu jedoch einen signifikanten Zusammenhang mit der Intelligenz auf, der durch die Aufgaben, die speziell das Verständnis von Dichte und Verdrängung testen sollten, zustande gekommen sein könnte, da diese eine größere Ähnlichkeit mit den Aufgaben des CFT 20-R aufweisen als die Aufgaben zu *Verdunstung und Kondensation* (vgl. Abb. 2).

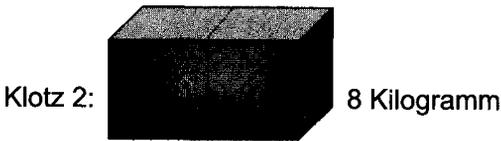
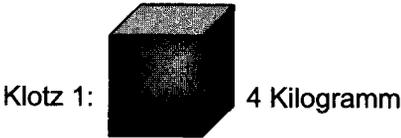
Entsprechend den Befunden großer Schulleistungsstudien, in denen regelmäßig ein starker Zusammenhang zwischen sozialer Herkunft und Leistung nachgewiesen werden kann (vgl. z.B. Bensen, Frey & Bos, 2008), wiesen die Instrumente zum *Schwimmen und Sinken* darüber hinaus signifikante Zusammenhänge zum ISCED und zum ISEI auf, während bei *Verdunstung und Kondensation* keine entsprechenden Korrelationen gefunden wurden. Beide untersuchten Themenbereiche umfassen zwar Phänomene, die im Alltagsleben von Grundschulern wiederzufinden sind – die Phänomene zum *Schwimmen und Sinken* schienen bei der Durchführung der Interviews für die Kinder aber interessanter zu sein. Daher ist anzunehmen, dass sie für diese Phänomene in ihren Familien häufiger Erklärungen einfordern, was in bildungsnäheren Schichten zu einem entsprechend stärkeren Aufbau von Wissen in diesem Bereich geführt haben kann. Erklärungen für den Wasserkreislauf sind dagegen häufig in den Medien anzutreffen, die unabhängig vom Elternhaus von Kindern aller Schichten rezipiert werden.

Die Ergebnisse der vorliegenden Studie sind zwar aufgrund der geringen Stichprobengröße nur bedingt generalisierbar. Die Zusammenhänge, die wir zwischen den jeweiligen Interviews und Fragebögen finden konnten, weisen jedoch darauf hin, dass bei sorgfältiger Entwicklung der Aufgaben und theoretischer Ableitung der Antwortalternativen eine Erfassung naturwissenschaftlicher Konzepte von Grundschulern auch mit geschlossenen, schriftlichen Antwortformaten valide umsetzbar ist. Darüber hinaus sprechen die Befunde zu *Verdunstung und Kondensation* auch für eine diskriminante Validität des schriftlichen Fragebogens, da keine Zusammenhänge mit den weiteren untersuchten Variablen festzustellen waren. Die Aufgaben zum Themenbereich *Schwim-*

Klötze-Vergleich

Instruktion des Testleiters:

Ich habe hier zwei Klötze. Das ist Klotz 1. Klotz 1 wiegt 4 Kilogramm. Und das ist Klotz 2. Klotz 2 wiegt 8 Kilogramm (Präsentation der entsprechenden Materialien).



Welcher Klotz ist aus dem schwereren Material gemacht?

Kreuze die beste Antwort an!

Klotz 1 (NV)	<input type="checkbox"/>
Klotz 2 (NV)	<input type="checkbox"/>
Beide sind aus dem gleichen Material (WV)	<input type="checkbox"/>

NV = Naive Vorstellung; WV = Wissenschaftliche Vorstellung

Abb. 2: Multiple-choice-Aufgabe für die Komponente Dichte

men und Sinken wiesen zwar Korrelationen mit den Kontrollvariablen auf, diese sind aber im Hinblick auf Eigenschaften der Aufgaben und des Themengebietes plausibel zu erklären.

Auch eine parallele Studie des Projektes Science-P, in der die konvergente Validität von schriftlichen Gruppentests mit geschlossenen Antwortformaten und Einzelinterviews für die Dimension *Wissen über Naturwissenschaften* untersucht wurde, belegt, dass eine valide Erfassung naturwissenschaftlicher Kompetenz mit geschlossenen, schriftlichen Antwortformaten möglich ist (Koerber et al., 2010).

4.3. *Wie können schriftliche Methoden der Lernstandserfassung in den Unterricht integriert werden?*

Schriftliche Messverfahren werden häufig in Verbindung mit summativer Leistungsmessung, d.h. zur abschließenden Beurteilung der Lernzielerreichung, verwendet. Sie eignen sich aber auch als ein auf individuelle Lernstände abgestimmtes, in den Unterrichtsverlauf integriertes Lernangebot, das eine Förderung konzeptueller Entwicklung zum Ziel hat. In der anglo-amerikanischen Literatur wird eine solche Diagnostik als formatives Assessment (Black & William, 2009; Pellegrino, Chudowsky & Glaser, 2001; Shavelson et al., 2008) bezeichnet, wobei eine Adaption des Lehr-Lernprozesses an das aktuelle Verständnisniveau der Schüler angestrebt wird (Maier, 2010).

So verweisen Duschl und Gitomer (1997) in diesem Zusammenhang auf die Bedeutung von ‚assessment conversations‘. Dies sind Unterrichtsaktivitäten, welche Schüler zur Diskussion unterschiedlicher (naturwissenschaftlicher) Ideen und Repräsentationen anregen und durch die Koordination von Theorie und Evidenz zum Erreichen von Lernzielen in den Naturwissenschaften führen. Dabei wird der Begriff ‚conversation‘ sehr breit als jegliche Verhandlung wissenschaftlicher Ideen, wie Argumentationen, Modellierungen, Zeichnungen und schriftliche Aufzeichnungen, verstanden. Eine Verwendung von schriftlichen Aufgaben als Grundlage von ‚assessment conversations‘ erscheint als ein fruchtbarer Weg, um Informationen über den Lernstand von Schülern einzuholen und mit den Schülern gemeinsam adaptive Möglichkeiten des weiteren Unterrichts- und Lernverlaufs zu bestimmen. Schriftliche Aufgaben können auch eingesetzt werden, um den Grad der individuell erfolgten Umstrukturierung durch einen Vergleich von Vorher-Nachher-Testungen zu erfassen und die hieraus gewonnenen Informationen zur Identifizierung von Lernschwierigkeiten zu nutzen. Schüler können anhand der gewonnenen Informationen eigene Umstrukturierungsprozesse reflektieren, was eine Integration von naiven Vorstellungen und im Unterricht Erlerntem fördern sollte. Schließlich können Aufgaben zur Erfassung konzeptueller Lernstände zur Förderung von Transferleistungen, beispielsweise auch im Rahmen von Differenzierungsmaßnahmen, eingesetzt werden. Als Form des formativen Assessments sollten komplexe, diagnostische Aufgaben deshalb ein integraler Bestandteil eines schülerorientierten Unterrichts sein (Bell & Cowie, 2001).

Insgesamt sollten schriftliche Testverfahren neben Multiple-choice-Aufgaben auch aus Aufgabenformaten mit freien schriftlichen Antworten – insbesondere in den höheren Klassen der Grundschule – bestehen, da die hierbei produzierten Antworten den in Interviews gegebenen Antworten näher kommen. Allerdings ist dabei zu bedenken, dass die Produktion offener schriftlicher Antworten durch die schriftsprachlichen Fähigkeiten der Schüler beeinträchtigt sein könnte und dass die Auswertung einen höheren Aufwand mit sich bringt. Auch erfordert die Bewertung offener schriftlicher Antworten eine hohe diagnostische Kompetenz der Lehrkräfte, um die gegebenen Antworten den Niveaus zuordnen zu können. Über Beispielantworten und eine Einführung in die zugrunde liegenden Theorien der konzeptuellen Umstrukturierung kann diese Kompetenz allerdings geschult werden.

Literatur

- Bar, V., & Travis, A. S. (1991). Children's views concerning phase changes. *Journal of Research in Science Teaching*, 28, 363-382.
- Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, 85, 536-553.
- Black, P., & William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5-31.
- Bonsen, M., Frey, K. A., & Bos, W. (2008). Soziale Herkunft. In W. Bos, M. Bonsen, J. Baumert, M. Prenzel, C. Selter & G. Walther (Hrsg.), *TIMSS 2007 Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 141-156). Münster: Waxmann Verlag.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R., Voss, A., & Walter, G. (Hrsg.) (2005). *IGLU – Skalenhandbuch zur Dokumentation der Erhebungsinstrumente*. Münster: Waxmann Verlag.
- Bullock, M., Sodian, B., & Koerber, S. (2009). Doing experiments and understanding science: Development of scientific reasoning from childhood to adulthood. In W. Schneider & M. Bullock (Hrsg.), *Human development from early childhood to early adulthood. Findings from the Munich Longitudinal Study* (S. 173-197). Mahwah, NJ: Erlbaum.
- Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). An experiment is when you try it and see if it works. A study of junior high school students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, 11, 514-529.
- Carr, M., Kurtz, B. E., Schneider, W., Turner, L. A., & Borkowski, J. G. (1989). Strategy acquisition and transfer among American and German children: Environmental influences on metacognitive development. *Developmental Psychology*, 25, 765-771.
- Driver, R. (1989). Students' conceptions and the learning of science. *International Journal of Science Education*, 11, 481-490.
- Duit, R. (1999). Conceptual change approaches in science education. In W. Schnotz, S. Vosniadou & M. Carretero (Hrsg.), *New perspectives on conceptual change* (S. 263-282). New York: Pergamon Press.
- Duit, R., Häußler, P., & Prenzel, M. (2001). Schulleistungen im Bereich der naturwissenschaftlichen Bildung. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 169-185). Weinheim/Basel: Beltz Verlag.
- Duschl, R., & Gitomer, D. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment*, 4, 37-73.
- Engelen, A., Jonen, A., & Möller, K. (2002). Lernfortschrittsdiagnosen durch Interviews – Ergebnisse einer Pilotstudie zum „Schwimmen und Sinken“ im Sachunterricht der Grundschule. In K. Spreckelsen, K. Möller & A. Hartinger (Hrsg.), *Ansätze und Methoden empirischer Forschung zum Sachunterricht*. Bd. 5: *Forschungen zur Didaktik des Sachunterrichts* (S. 155-173). Bad Heilbrunn: Klinkhardt Verlag.
- Frède, V., Nobs, G., Frappart, S., Panagiotaki, G., Troadec, B., & Martin, A. (2011). The acquisition of scientific knowledge: the influence of methods of questioning and analysis on the interpretation of children's conceptions of the earth. *Infant and Child Development*. Early View (advanced online publication).
<http://onlinelibrary.wiley.com/doi/10.1002/icd.730/full> [02.05.2011].
- Ganzeboom, H. B. G., de Graaf, P. M., Treiman, D. J., & de Leeuw, J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21, 1-56.
- Gropengießer, H. (2007). *Didaktische Rekonstruktion des Sehens. Wissenschaftliche Theorien und die Sicht der Schüler in der Perspektive der Vermittlung* (2. überarb. Aufl. Nachdr.). Beiträge zur Didaktischen Rekonstruktion, Bd. 1. Oldenburg: Didaktisches Zentrum.

- 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, 307-326.
- Hardy, I., Kleickmann, T., Koerber, S., Mayer, D., Möller, K., Pollmeier, J., Schwippert, K., & Sodian, B. (2010). Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (S. 115-125). 56. Beiheft der Zeitschrift für Pädagogik. Weinheim/Basel: Beltz Verlag.
- Helmke, A., Hosenfeld, I., Ophoff, J., Halt, A. C., Henk, F., Isaac, K., Koch, U., & Scherthan, F. (2007). *Ergebnisbericht VERA 2006*. Bremen Universität Koblenz Landau.
- Kleickmann, T., Hardy, I., Möller, K., Pollmeier, J., Tröbst, S., & Beinbrech, C. (2010). Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter: Theoretische Konzeption und Testkonstruktion. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 263-281.
- Kleickmann, T., Pollmeier, J., Hardy, I., & Möller, K. (im Druck). Die Struktur naturwissenschaftlichen Wissens von Grundschulkindern – eine person- und variablenzentrierte Analyse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*.
- Koerber, S., Sodian, B., Mayer, D., Kropf, N., Schwippert, K., & Möller, K. (2010, Juli). *Development of scientific reasoning and science understanding in elementary school*. Paper presented at the 21st Biennial International Congress of the International Society for the Study of Behavioural Development, Lusaka, Zambia.
- Lenhard, W., & Schneider, W. (2006). ELFE 1-6. *Ein Leseverständnistest für Erst- bis Sechstklässler*. Göttingen: Hogrefe Verlag.
- Maier, U. (2010). Formative Assessment – Ein erfolgversprechendes Konzept zur Reform von Unterricht und Leistungsmessung? *Zeitschrift für Erziehungswissenschaft*, 13, 293-308.
- Mayring, P. (2008). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (10. Aufl.). Weinheim/Basel: Beltz Verlag.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know. The science and design of educational assessment*. Washington, DC: National Academic Press.
- Scott, P., Asoko, H., & Leach, J. (2007). Student conceptions and conceptual learning in science. In S. K. Abell & N. G. Ledermann (Hrsg.), *Handbook of research on science education* (S. 31-56). New Jersey: Lawrence Erlbaum.
- Shavelson, R., Young, D., Ayala, C., Brandon, P., Furtak, E., Ruiz-Primo, M., Tomita, M., & Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education*, 21, 295-314.
- 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.
- Treagust, D. F., & Duit, R. (2008). Conceptual change: A discussion of theoretical, methodological and practical challenges for science education. *Cultural Studies of Science Education*, 3, 297-328.
- Tytler, R., Peterson, S., & Prain, V. (2006). Picturing evaporation: Learning science literacy through a particle representation. *Teaching Science*, 52(1), 12-17.
- UNESCO Institute for Statistics (1999). *Operational Manual for ISCED-1997 (International Standard Classification of Education)*. <http://dannagrace.com/UNESCO/UNESCO.pdf> [06.04.2011].
- Weiß, R. H. (2006). *Grundintelligenztest Skala 2 – Revision – (CFT 20-R)*. Göttingen: Hogrefe Verlag.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah: Lawrence Erlbaum.

Abstract: In science instruction, diagnosing students' concepts in the sense of a formative assessment constitutes a necessary prerequisite for an adequate promotion of conceptual development. While qualitative interviews aiming at surveying conceptual learning levels in natural science among elementary school students are well established, hardly any studies exist that examine in how far the more economic, closed written answer formats allow for a valid diagnosis of individual concepts. Based on a survey carried out in two classes of third-grade elementary school students (Age: $M = 9.26$ years, $SD = 0.42$) with regard to the two science units "Floating and Sinking" ($N = 41$) and "Evaporation and Condensation" ($N = 32$), the present contribution shows that there is a substantial correlation between oral interviews and written test results and that, therefore, – given certain preconditions – it can be assumed that written tests, too, allow for a valid recording of children's learning levels.

Anschrift der Autorinnen

Dipl. Psych. Judith Pollmeier, Westfälische Wilhelms-Universität, Seminar für Didaktik des Sachunterrichts, Leonardo Campus 11, 48149 Münster, Deutschland
E-Mail: j.pollmeier@uni-muenster.de

Prof. Dr. Ilonca Hardy, Goethe-Universität Frankfurt, Fachbereich Erziehungswissenschaften, Institut für Pädagogik der Elementar- und Primarstufe, Senckenberganlage 15, 60054 Frankfurt a.M., Deutschland
E-Mail: hardy@em.uni-frankfurt.de

Prof. Dr. Susanne Koerber, Institut für Psychologie, Abteilung Frühe Bildung, Pädagogische Hochschule Freiburg, Kunzenweg 21, 79117 Freiburg, Deutschland
E-Mail: Susanne.koerber@ph-freiburg.de

Prof. Dr. Kornelia Möller, Seminar für Didaktik des Sachunterrichts, Westfälische Wilhelms-Universität Münster, Leonardo Campus 11, 48149 Münster, Deutschland
E-Mail: sachunterricht@uni-muenster.de