

Kognitiv anregende Verhaltensweisen von Lehrkräften im naturwissenschaftlichen Sachunterricht – Konzeptualisierung und Erfassung

Inaugural-Dissertation
zur Erlangung des akademischen Grades
des Doktors in den Erziehungswissenschaften
an der Westfälischen Wilhelms-Universität
Münster

vorgelegt von:

Julia Katharina Vehmeyer

geboren in Bremen

2009

1. Gutachterin: Prof.in Dr. Kornelia Möller

2. Gutachter: Prof. Dr. Manfred Holodynski

Tag der mündlichen Prüfung: 19. August 2009

Danksagung

Meine Arbeit an dieser Dissertation wurde durch viele Personen unterstützt.

Zuforderst gilt mein herzlicher Dank Prof'in Dr. Kornelia Möller, die mir die Anregung zu diesem Thema gab und mich während der gesamten Zeit meiner Arbeit am Seminar für Didaktik des Sachunterrichts und während der Erstellung dieser Arbeit umfassend unterstützt und betreut hat.

Bedanken möchte ich mich auch bei Thilo Kleickmann, Christina Beinbrech, Eva Blumberg, Berenike Gais, Kim Lange und Anne Ewerhardy für zahlreiche konstruktiv- inhaltliche aber auch persönliche Gespräche.

Außerdem gilt mein Dank den Lehrkräften, die ihren Unterricht haben videografieren lassen. Die dabei entstandenen Unterrichtsvideos bilden die Grundlage des empirischen Teils der Arbeit. In diesem Zusammenhang gebührt auch den drei Raterinnen, die mit mir gemeinsam in vielen Stunden Arbeit die Unterrichtsvideos analysiert haben, mein Dank.

Ich danke ebenfalls den fünf Personen, die mir ihre Zeit gaben, indem sie sich dazu bereit erklärten, an der Expertenvaldierung teilzunehmen.

Brigitta Kober möchte ich ganz herzlich für stets aufbauende und ermutigende Worte danken, ebenso wie meinen Freundinnen und Freunden, die immer viel Verständnis für mich während der Entstehung dieser Arbeit hatten und mich mit zahlreichen Ablenkungen versorgt haben.

Meinen Dank möchte ich den Freundinnen und Freunden aussprechen, die sich Zeit genommen haben, jeweils Teile der Arbeit Korrektur zu lesen.

Mein ganz persönlicher Dank gilt meinem Freund André für seinen emotionalen Halt und seine immer wieder ermutigenden Wort vor allem in den letzten Wochen des Schreibens der Arbeit, aber auch für seine Unterstützung im Umgang mit den im Rahmen der Arbeit verwendeten mathematischen Formeln.

Vor allem gilt mein besonderer Dank meinen Eltern und meinem Bruder, die mich während meiner gesamten Promotionsphase unterstützt und dabei immer an mich geglaubt haben. Ohne ihre Unterstützung hätte ich diese Arbeit nicht schreiben können.

Inhaltsverzeichnis

Einleitung	05
1 Stand der Forschung	10
1.1 Kognitive Aktivierung in der Mathematikdidaktik	10
1.1.1 Ursprung, Definition und Ziele der kognitiven Aktivierung	10
1.1.2 Elemente kognitiver Aktivierung im Mathematikunterricht	12
1.1.3 Zu den Wirkungen eines kognitiv aktivierenden Mathematikunterrichts	15
1.1.4 Zusammenfassung und Überleitung zum anspruchsvollen naturwissenschaftlichen Lernen im Sachunterricht der Grundschule	22
1.2 Hinweise zur Gestaltung von kognitiv anregenden Lerngelegenheiten aus dem Bereich der Instruktionspsychologie	26
1.2.1 Lernen aus verschiedenen Perspektiven der Instruktionspsychologie	27
1.2.1.1 Lernen als Veränderung von Verhalten	27
1.2.1.2 Lernen als Erwerb von Wissen	28
1.2.1.3 Lernen als Konstruktion von Wissen	29
1.2.2 Die theoretische Basis zum Bereich Conceptual Change	29
1.2.2.1 Präkonzeptforschung	30
1.2.2.2 Lernen als Conceptual Change	33
1.2.3 Die theoretische Basis zum sozialen Konstruktivismus und zur situierten Kognition	35
1.2.3.1 Lernen aus sozial-konstruktivistischer Sicht	36
1.2.3.2 Lernen aus der Sicht der situierten Kognition	38
1.3 Hinweise zur Gestaltung von kognitiv anregenden Lernumgebungen aus didak- tischen Konzeptionen des Sachunterrichts	44
1.3.1 Genetisches Lernen	44
1.3.1.1 Genetisches Lehren und Lernen bei Martin Wagenschein	45
1.3.1.2 Die genetische Orientierung des Sachunterrichts bei Walter Köhnlein	47
1.3.1.3 Lernen in Phänomenkreisen	49

1.3.2	Problemorientiertes Lernen im Sachunterricht	50
1.3.2.1	Zum Begriff	50
1.3.2.2	Ansätze und Ziele problemorientierten Lernens im Sachunterricht	51
1.3.3	Entdeckendes Lernen	54
1.3.3.1	Zum Begriff	55
1.3.3.2	Formen und Ziele entdeckenden Lernens	56
1.3.3.3	'Pure discovery' versus 'guided discovery'	57
1.3.4	Anspruchsvolles, Konzeptwechsel-förderndes Lernen	58
1.3.5	Zusammenfassung der Inhalte zu Hinweisen zur Gestaltung kognitiv anregender Lernumgebungen aus der Instruktionspsychologie und aus Konzeptionen des Sachunterrichts	60
1.4	Folgerungen für das Lehren kognitiv anregenden Sachunterrichts	62
1.4.1	Verhaltensweisen von Lehrkräften zur Anregung aktiven und konstruktiven Lernens	63
1.4.2	Verhaltensweisen von Lehrkräften zur Umsetzung sozialen und kooperativen Lernens ..	68
1.4.3	Verhaltensweisen von Lehrkräften zur Umsetzung unterstützten Lernens	68
1.4.4	Verhaltensweisen von Lehrkräften zur Umsetzung situierten Lernens	71
1.4.5	Zusammenfassung	71
1.5	Forschungsstand zur Erfassung von Prozessmerkmalen des Unterrichts per Videoanalyse	72
1.5.1	Methoden der Videoanalyse	73
1.5.1.1	Niedrig- bis hoch-inferente Instrumente der Videoanalyse	73
1.5.1.2	Zum Entwicklungsprozess von Videoanalyse-Instrumenten	79
1.5.1.3	Zu den Gütekriterien bei Videoanalysen	83
1.5.2	Bisher bestehende Instrumente zur Erfassung kognitiv anregender Verhaltensweisen von Lehrkräften	91
1.5.3	Zusammenfassung	92

2 Zielsetzung und Fragestellung	94
3 Methoden	96
3.1 Herkunft der verwendeten Daten	96
3.2 Untersuchungsanlage	97
3.3 Stichprobenbeschreibung	99
3.4 Rating-Instrument zur Erfassung kognitiv anregender Handlungen von Lehrkräften ..	101
3.4.1 Konstruktion des Ratingverfahrens	101
3.4.1.1 Theoretische Basis und Entwicklung des Videoverfahrens	102
3.4.1.2 'Pretest' und Training der Rater	104
3.4.1.3 Durchführung des Ratings und anschließende Analyse	107
3.4.2 Darstellung der Ratingdimensionen	108
3.5 Methoden der Prüfung des Rating-Instruments	116
3.5.1 Zur Objektivität	116
3.5.2 Zur Reliabilität	116
3.5.2.1 Analyse der Interrater-Reliabilität mittels der Generalisierbarkeitstheorie	116
3.5.2.2 Analyse der internen Konsistenzen	121
3.5.3 Zur Validität	122
3.5.3.1 Kriteriumsvalidität	122
3.5.3.2 Konstruktvalidität	122
4 Ergebnisse	124
4.1 Ergebnisse der Generalisierbarkeitsanalyse	124
4.2 Ergebnisse der Item- und Skalenanalyse	133
4.3 'Concurrent validity': Ergebnisse der Analyse von Unterschieden im Unterrichts- handeln der Lehrkräfte in den Fortbildungsgruppen	139
4.4 Konstruktvalidität: Ergebnisse der Expertenvalidierung	143
4.5 Zusammenfassung der Ergebnisse	156

5 Diskussion und Ausblick	157
5.1 Zusammenfassende Diskussion der Ergebnisse der Testanalysen	158
5.2 Probleme der Studie	165
5.3 Zur Fachspezifität der konzeptualisierten und erfassten Lehrerhandlungen	168
5.4 Einsatz des Instruments zur Erfassung von Zusammenhängen zwischen Lehrervorstellungen und -handlungen	169
5.5 Ausblick	171
Literaturverzeichnis	180
Abbildungsverzeichnis	200
Tabellenverzeichnis	201
Anhang	202
Anhang A	
A.1 Ergebnisse der Generalisierbarkeitsanalyse	202
A.2 Ergebnisse der Item- und Skalenanalyse	207
A.3 Nachtrag: Ergebnisse der Item- und Skalenanalyse	213
A.4 Verteilung der Stichprobe in den Beobachtungsdimensionen – Boxplot	216
Anhang B	
B.1 Der Zusammenhang des absoluten Generalisierbarkeitskoeffizienten mit der Anzahl der Rater	223
B.2 Abschätzung der Varianzkomponenten	224
Anhang C	
C. 1 NRW-Population	225

Einleitung und Zielbestimmung

„Auf die Lehrer kommt es an!“
(STERN.DE, 2008)

Zu den Zielen (naturwissenschaftlichen) Sachunterrichts gehört 'Verstehen'. Verstehen meint das Lernen von Zusammenhangswissen, d.h. den Aufbau von verstandenem, anwendbarem und anschlussfähigem Wissen (kognitive Ebene) und das Lernen von Einstellungen und Haltungen (motivational-emotionale Ebene) (vgl. Einsiedler, 2000, S. 72f.). Ein auf Verstehen ausgerichtetes Lernen verlangt gemäß aktuellen Lerntheorien, dass sich die Lernenden aktiv mit dem Lerngegenstand auseinandersetzen und ihr Wissen zu einem gewissen Maße selbstständig und in sozialen Auseinandersetzungen mit den Lerninhalten konstruieren. Aus diesem Lernverständnis resultiert eine Sichtweise von Lehren, gemäß der die Lehrenden die Lernenden in ihren Konstruktionsprozessen durch die Gestaltung anspruchsvoller und anregender Lernumgebungen unterstützen und begleiten (vgl. z.B. Gerstenmaier & Mandl, 1995; Reinmann & Mandl, 1998; Duit, 1995). Auf welche Art und Weise sich für den Lernenden Gelegenheiten zu einer selbstständigen, kooperativen und aktiven Auseinandersetzung ergeben, hängt also entscheidend vom Lehrenden ab (vgl. Vermunt & Verloop, 1999).

Auch in der aktuellen bildungspolitischen Debatte stehen Lehrkräfte vor allem seit den regelmäßig wiederkehrenden Veröffentlichungen zu Schulleistungsstudien wie TIMSS, PISA oder IGLU im Zentrum des Interesses. „Auf die Lehrer kommt es an!“ Mit diesen Worten beginnt ein Kommentar zur PISA-Studie auf *STERN.DE*¹ vom 18.11.2008, genau wie ein Bericht auf *WELT ONLINE*² (03.12.2007). Diese Worte finden sich ebenfalls in Artikeln auf *ZEIT ONLINE*³ (19.05.2004) und *FAZ.NET*⁴ (11.09.2008). Auch Lipowsky betitelt einen seiner Aufsätze aus dem Jahr 2006 mit diesen Worten. Die Aussage steht stellvertretend für die in der Öffentlichkeit immer wieder hervorgebrachten Aussagen bezüglich der Relevanz, die der Qualität von Lehrkräften für Lernerfolg und Lernfreude von Schülern⁵ beigemessen wird.

Allgemein wird davon ausgegangen, dass Lehrkräfte mit ihren unterrichtlichen Handlungen Einfluss auf kognitive und nicht kognitive Zieldimensionen von Unterricht haben. Doch finden sich auch Evidenzen für diese Annahmen? Ein Blick auf die aktuelle Unterrichtsforschung soll helfen, eine Antwort auf diese Frage zu geben.

¹URL: <http://www.stern.de/politik/deutschland/646048.html>; letzter Aufruf am 02.04.09.

²URL: http://www.welt.de/politik/article1424490/Schulstruktur_Auf_die_Lehrer_kommt_es_an.html; letzter Aufruf am 02.04.09.

³URL: http://www.zeit.de/2004/22/Glosse_22_a; letzter Aufruf am 02.04.09.

⁴URL: <http://www.faz.net/s/Rub117C535CDF414415BB243B181B8B60AE/Doc~E541D98744B9D40CE9B7CC596D9A9-BA5D~ATpl~Ecommon~Scontent.html>; letzter Aufruf am 02.04.09.

⁵In der vorliegenden Arbeit stehen männliche Personenbezeichnungen sowohl für männliche als auch für weibliche Formen.

Aktuelle Modelle der Unterrichtsforschung zur Wirkungsweise von Unterricht greifen die Sichtweise, Unterricht als Gelegenheitsstruktur für Lernprozesse zu betrachten, auf. Diese so genannten Angebots-Nutzungs-Modelle gehen von den folgenden zentralen Grundgedanken aus (vgl. Lipowsky, 2006, 2007a; Helmke, 2004; Helmke & Schrader, 2006a; Helmke & Weinert, 1997; Fend, 1981): Der von der Lehrkraft durchgeführte Unterricht wird als Lerngelegenheit (*Angebot*) betrachtet. Dieser führt nicht direkt, sondern erst durch die Wahrnehmung, aktive Nutzung und Verarbeitung der Lerngelegenheiten durch die Lernenden (*Nutzung*) zu Lernerfolgen (leistungsbezogene, motivationale, interessenbezogene und auf das Selbstkonzept bezogene Zielkriterien) (*Wirkung/Ertrag*). Die Qualität und Quantität der durch die Lehrkraft gestalteten Lernangebote hängt gemäß des Modells von ihren Kompetenzen (Professionswissen, Können, Überzeugungen, Motivation, selbstbezogene Kognitionen) ab. Ob und wie die Lernenden die Angebote wahrnehmen und nutzen, wird durch ihre individuellen Lernvoraussetzungen (kognitive Fähigkeiten, motivationale und emotionale Aspekte) (mit-)bestimmt. Neben den Kompetenzen der Lehrenden und den individuellen Voraussetzungen der Lernenden wirken weitere Kontextbedingungen auf Unterrichtsgestaltung, -nutzung und Lernerfolg ein. Dazu gehören kognitive und soziale Merkmale der Klasse, kulturelle, ökonomische und schulische Bedingungen sowie familiäre Voraussetzungen. Nicht zuletzt beeinflussen auch außerschulische Lerngelegenheiten die Lernaktivitäten der Lernenden.

Unterricht ist also in ein komplexes System von verschiedenen miteinander verbundenen Wirkungsgrößen eingebunden. Entsprechend gibt es eine Reihe von Faktoren, von denen man annimmt, dass sie den Lernerfolg beeinflussen, bzw. deren Lernwirksamkeit empirisch nachgewiesen ist (z.B. Helmke, 2004, 2006; Lipowsky, 2007a, 2007b). Lernerfolg gilt demnach als multipel determiniert. Empirisch gut belegt ist z.B. der Einfluss der individuellen Voraussetzungen der Lernenden (insbesondere ihr fachspezifisches Vorwissen) auf ihre fachlichen Leistungen (Helmke & Weinert, 1997). Aktuell wird auch der Lehrkraft mit ihren Kompetenzen und Handlungen im Unterricht eine wichtige Erklärungskraft für den Lernerfolg der Schüler beigemessen. Lipowsky (2006, 2007b) nennt einige value-added-Studien, die den Lernzuwachs der Lernenden über mehrere Jahre ermittelten und von äußeren Einflüssen und individuellen Lernvoraussetzungen bereinigten, so dass der Anteil am Lernerfolg übrig blieb, der durch den Lehrer oder den Unterricht erklärbar ist. Die Ergebnisse der Studien (Babu & Mendro, 2003; Rowan, Chiang & Miller, 1997; Rivkin, Hanushek & Kain, 2005; Goldhaber & Anthony, 2004) belegen, dass Merkmale der Lehrer- und Unterrichtsqualität bedeutsame Einflüsse auf den langfristigen Lernzuwachs der Schüler haben und vor allem leistungsschwächere Lernende und Grundschüler von einem guten Lehrer bzw. von einem guten Unterricht zu profitieren scheinen. Auch weitere Forschungen belegen, dass eine gute Lehrer- und Unterrichtsqualität einen weitaus größeren Einfluss auf das Lernen hat als in den 1970er Jahren noch angenommen wurde (z.B. Metaanalysen von Fraser et al., 1987; Wang, Haertel & Walberg, 1993 und Hattie, 2003; Studien aus der Unterrichtsforschung von Helmke & Weinert, 1997; Gruehn, 2000 und Einsiedler & Treinies, 1997). Es stellt sich daher die Frage, welche Merkmale den guten Lehrer und den guten Unterricht ausmachen.

Bezüglich der Frage, was guten Unterricht ausmacht, wird in der Unterrichtsqualitätsforschung zwischen der *Quantität* des Unterrichts (allgemeine und fachunspezifische Unterrichtsmerkmale) und der *Qualität* der Lernangebote (fachspezifische Unterrichtsmerkmale) unterschieden (vgl. z.B. Helmke & Schrader, 2006a). Über allgemeine und fachunspezifische Merkmale effektiven Unterrichts herrscht mittlerweile Konsens. Dazu gehört z.B. eine effiziente Klassenführung (z.B. Helmke, Hosenfeld & Schrader, 2002; Lipowsky, 2007a; Helmke, 2004; Holodynski, 1998) für eine intensive Nutzung der Lernzeit (time on task) zur wirksamen Vorbeugung von Disziplinstörungen (Kounin, 2006)⁶. Da Unterricht aber immer in ein spezifisches Fach eingebunden ist, interessieren auch fachspezifische Unterrichtsmerkmale (vgl. Helmke & Schrader, 2006a; Lipowsky, 2006).

Verbunden mit dem Interesse an fachspezifischen Unterrichtsmerkmalen ist die Frage, wie solche Elemente im Unterricht aussehen können. Wie kann die Lehrperson die Lernenden z.B. zu einer aktiven Auseinandersetzung mit dem Lerngegenstand anregen? Wie sollte sie sich verhalten, um die Schüler im Unterricht zu motivieren? Welche Möglichkeiten bieten sich der Lehrperson, um die Schüler sinnvoll zu unterstützen? Für den naturwissenschaftsbezogenen Sachunterricht gibt es hierzu bisher keine Studien. An dieser Stelle setzt die Dissertation an.

Das zentrale Ziel dieser Arbeit ist, kognitiv anregende Verhaltensweisen von Lehrkräften im naturwissenschaftsbezogenen Sachunterricht zu konzeptualisieren und mit einem Videoanalyse-Instrument (hoch-inferentes Rating) zu erfassen. Da es hierzu für den Sachunterricht der Grundschule kein Rating-Instrument gibt, wird im Rahmen dieser Arbeit ein solches entwickelt.

Entsprechend dem Ziel der Arbeit stehen die folgenden Fragen im Zentrum des Interesses: Gibt es empirische Hinweise auf die Bedeutung kognitiv anregenden Unterrichts? Welche Verhaltensweisen von Lehrkräften haben Potential, die Lernenden im naturwissenschaftsbezogenen Sachunterricht kognitiv anzuregen? Ist es möglich, ein Instrument zu entwickeln, welches kognitiv anregende Verhaltensweisen von Lehrkräften zuverlässig misst?

Die Arbeit ist eingebunden in die Studie „Veränderung des fachspezifisch-pädagogischen Wissens bei Grundschullehrkräften in Bezug auf Lehren und Lernen im naturwissenschaftlichen Unterricht durch Lehrerfortbildungen“. Das Projekt wurde im Rahmen des Schwerpunktprogramms BIQUA („Die Bildungsqualität von Schule: Fachliches und fächerübergreifendes Lernen im mathematisch-naturwissenschaftlichen Unterricht in Abhängigkeit von schulischen und außerschulischen Kontexten“) (vgl. Doll & Prenzel, 2004) von der Deutschen Forschungsgemeinschaft (DFG) im dritten Antragszeitraum (2004-2006) gefördert.

⁶Eine detaillierte Übersicht über fachübergreifende Unterrichtsmerkmale findet sich z.B. bei Helmke (2004).

Die Arbeit gliedert sich in fünf Kapitel.

Das **Kapitel 1** beinhaltet die theoretische Basis der Arbeit. Hier geht es um die Konzeptualisierung kognitiv anregenden Lehrerverhaltens im naturwissenschaftsbezogenen Sachunterricht. Im Zuge des Interesses an einem effektiven, auf Verständnis ausgerichteten Unterricht und der damit verbundenen Frage, wie im Unterricht Gelegenheitsstrukturen für verständnisvolles fachliches Lernen durch die Lehrkraft geschaffen werden können, wird in der deutschsprachigen mathematikdidaktischen Diskussion die *kognitive Aktivierung* der Lernenden als entscheidend betrachtet (vgl. Baumert et al., 1997; Kunter et al., 2005; Lipowksy, 2006, 2007a). Auf der Ebene der Unterrichtsgestaltung gilt die kognitive Aktivierung als ein *fachliches Unterrichtsqualitätsmerkmal* zur Förderung verständnisvoller Lernprozesse. Das erste Kapitel beginnt deshalb in **1.1** mit einer Auseinandersetzung des Konstrukts der kognitiven Aktivierung im Kontext der Mathematikdidaktik. Am Ende dieses Unterkapitels erfolgt eine Überleitung zum naturwissenschaftsbezogenen Sachunterricht. Hier werden Gründe für die aktuellen Forderungen nach einem anspruchsvollen, verständnisorientierten und denkfördernden naturwissenschaftlichen Sachunterricht und den damit verbundenen Zielen herausgearbeitet. Daran anknüpfend werden Hinweise zur Gestaltung kognitiv anregender Lernumgebungen in **1.2** aus dem Bereich der Instruktionspsychologie und anschließend in **1.3** aus Konzeptionen der Sachunterrichtsdidaktik herausgearbeitet. Das Unterkapitel **1.4** setzt sich mit den Folgerungen auseinander, die sich aus den herausgearbeiteten Implikationen zur Gestaltung von Unterricht ergeben. Ziel dieses Unterkapitels ist die Ableitung von Verhaltensweisen von Lehrkräften zur Anregung aktiven, konstruktiven, sozialen, kooperativen und situierten Lernens und zur Umsetzung unterstützenden Lernens. Videoanalysen spielen in der Unterrichtsforschung zur Beschreibung von Unterricht und zur Untersuchung der Wirksamkeit von Unterricht vor allem seit Ende der 1990er Jahre eine wichtige Rolle. Da die Erfassung der in 1.4 abgeleiteten Lehrerverhaltensweisen in dieser Arbeit mit Hilfe eines Videoanalyse-Instruments erfolgen soll, befasst sich das Unterkapitel **1.5** mit dem Forschungsstand der Videoanalyse.

In **Kapitel 2** werden Zielsetzung und Fragestellungen sowie die erwarteten Ergebnisse der Arbeit konkretisiert.

In **Kapitel 3** werden die methodischen Grundlagen der Arbeit dargestellt. In einem ersten Schritt wird die Herkunft der im Rahmen dieser Arbeit verwendeten Daten und damit die Anbindung der Arbeit an das oben genannte DFG-Projekt aus BIQUA kurz erläutert (**3.1**). Es folgen die Erläuterung der Untersuchungsanlage (**3.2**) und die Stichprobenbeschreibung (**3.3**). Insbesondere wird in diesem Kapitel die inhaltliche und methodische Entwicklung des Videoanalyse-Instruments zur Erfassung kognitiv anregenden Lehrerverhaltens im naturwissenschaftsbezogenen Sachunterricht berichtet (**3.4**). Des Weiteren erfolgt in **3.5** die Beschreibung der in der Arbeit genutzten Methoden zur Prüfung der Güte des entwickelten Instruments.

In **Kapitel 4** werden die zentralen Ergebnisse zur Güteprüfung des entwickelten Videoanalyse-Instruments vorgestellt.

In **Kapitel 5** werden in einem ersten Schritt die Ergebnisse der Testanalysen diskutiert (**5.1**). Daran anschließend werden Probleme der Studie besprochen, die sich vor allem aus der methodischen Anlage der Arbeit ergeben (**5.2**). Es folgen Ausführungen zur Fachspezifität der konzeptualisierten und erfassten unterrichtlichen Handlungsweisen von Lehrkräften (**5.3**). Das in dieser Arbeit vorgestellte Videoanalyse-Instrument konnte im Rahmen der Münster'schen DFG-Studie erfolgreich zur Analyse von Zusammenhängen zwischen Vorstellungen zum Lehren und Lernen von Lehrkräften mit ihren unterrichtlichen Verhaltensweisen eingesetzt werden. In Abschnitt **5.4** wird auf vorliegende und bereits veröffentlichte Ergebnisse zu dieser Fragestellung eingegangen. Ein Ausblick (**5.5**) rundet die Arbeit ab. Hier werden Folgerungen, die sich aus der methodischen Anlage der Arbeit ergeben und Folgerungen für weiterführende inhaltliche Überarbeitungen des Instruments, weiterführende Fragestellungen und Einsatzmöglichkeiten von Videoanalyse-Instrumenten über Fragestellungen der Unterrichtsforschung hinaus diskutiert.

1 Stand der Forschung

1.1 Kognitive Aktivierung in der Mathematikdidaktik

In der Einleitung wurde bereits darauf hingewiesen, dass die 'kognitive Aktivierung' in der deutschsprachigen Mathematikdidaktik als ein bedeutsames Merkmal im Zuge der Diskussion um einen fachlich anspruchsvollen Mathematikunterricht gilt. Ziel dieses Kapitels ist, näher zu beleuchten, weshalb dies so ist. Dazu wird zuerst herausgearbeitet, in welchem Kontext die kognitive Aktivierung entstanden ist, wie das Konstrukt definiert wird und welche Ziele mit einem kognitiv aktivierenden Unterricht verfolgt werden (1.1.1). Daran anschließend werden Unterrichtselemente vorgestellt, von denen man in der Mathematikdidaktik annimmt, dass sie die Lernenden kognitiv aktivieren (1.1.2). Es folgt die Befundlage zu den Wirkungen eines kognitiv aktivierenden Unterrichts im Fach Mathematik (1.1.3). Schließlich erfolgt nach einer Zusammenfassung der dargestellten Inhalte eine Überleitung zum naturwissenschaftsbezogenen Sachunterricht (1.1.4). In der Überleitung wird herausgestellt, dass auch im Sachunterricht die Forderung nach einem anspruchsvollen, verständnisorientierten und denkfördernden Lernen besteht, die sich aus verschiedenen Perspektiven begründen lässt.

1.1.1 Ursprung, Definition und Ziele der kognitiven Aktivierung

Das Konstrukt der kognitiven Aktivierung ist im Kontext von Untersuchungen zum Mathematiklernen (Sekundarstufe I) entstanden. Insbesondere durch die als Teilelement von TIMSS (Third International Mathematics and Science Study) 1995 durchgeführte TIMS-Video-Studie (Stigler et al., 1999; Stigler & Hiebert, 1998, 1999) hat der Begriff Einzug in die mathematikdidaktische Diskussion und in die Unterrichtsqualitätsforschung erhalten.

TIMSS-Video zielte darauf ab, den Mathematikunterricht in Japan als Land, dessen Schüler im internationalen Vergleich hohe Testleistungen erreichten, und den Mathematikunterricht in Deutschland und den USA als leistungsschwächere Länder zu vergleichen (vgl. Clausen, Reusser & Klieme, 2003; Hugener, Pauli & Reusser, 2007). Mit dem Ziel, objektive Indikatoren zur Beschreibung von Unterrichtsprozessen für den Mathematikunterricht in den drei Ländern zu finden (vgl. Baumert et al., 1997) und Merkmale guten Unterrichts bestimmen zu können, wurden 231 zufällig aus der TIMSS-Stichprobe gezogene achte Jahrgangsklassen (N=100 in Deutschland, N=50 in Japan, N=81 in den USA) hinsichtlich ihrer Stundenstrukturen und -abläufe, methodisch-didaktischen Aspekte, Interaktionsprozesse, Unterrichtsgespräche und Inhalte beschrieben und miteinander verglichen (vgl. Stigler et al., 1999; Stigler & Hiebert, 1999; Klieme & Bos, 2000; Clausen, Reusser & Klieme, 2003).

Die Befunde bestätigen die Annahme, dass sich der Mathematikunterricht der Kulturen durch landesspezifische, inhaltliche und kognitive Schwerpunkte auszeichnet (vgl. Klieme & Bos, 2000). Entspre-

chend den Ländern zeigten sich drei unterschiedliche Unterrichtsmuster⁷: Ein problemlösend-entdeckendes Unterrichtsmuster in Japan, ein fragend-entwickelndes deutsches und ein darstellendes amerikanisches Unterrichtsmuster (vgl. z.B. Hugener, Pauli & Reusser, 2007). Bezüglich des Vergleichs des deutschen und japanischen Mathematikunterrichts lassen sich die Kernergebnisse wie folgt zusammenfassen: Japanischer Unterricht zielt auf die Schulung mathematischen Verständnisses ab, wohingegen der Mathematikunterricht in Deutschland hauptsächlich auf den Erwerb von mathematischen Fertigkeiten und Routinen ausgerichtet ist. In japanischen Unterrichtsstunden bildet das Problemlösen ein zentrales Element. Anspruchsvolle Problem- und Denkaufgaben, die den Lernenden eigene Lösungen erlauben, bilden einen hohen Anteil in Schülerarbeitsphasen, wohingegen in Deutschland größtenteils Routineaufgaben mit festgelegten Lösungsmustern vorherrschen. Zudem nutzen japanische Lehrkräfte im Vergleich zu ihren deutschen Kollegen deutlich häufiger durch Schüler entwickelte und dargestellte alternative Lösungswege, um ein mathematisches Problem zu lösen (vgl. u.a. Baumert et al., 1997; Klieme & Bos, 2000; Klieme, Schümer & Knoll, 2001; Clausen, Reusser & Klieme, 2003). Somit wird dem japanischen Mathematikunterricht insgesamt ein höheres Niveau an mathematischen Anforderungen und kognitiven Herausforderungen zugesprochen als dem deutschen Mathematikunterricht (vgl. Klieme & Bos, 2000).

Die Ergebnisse der TIMS-Video-Studie führten zu der Hypothese, dass die im japanischen Unterricht enthaltenen Merkmale im Zusammenhang mit den hohen Testleistungen der japanischen Schüler stehen könnten (Klieme & Bos, 2000)⁸. Die Konsequenz daraus war, dass die japanischen Unterrichtsmerkmale als wünschenswerte Zielvorstellungen für einen kognitiv anspruchsvollen und leistungsfördernden Mathematikunterricht in Deutschland betrachtet wurden (vgl. Blum, 1998; Baumert & Köller, 2000; Klieme, Schümer & Knoll, 2001).

In der mathematikdidaktischen Diskussion wurde ein Unterricht zum Leitbild, der den Lernenden Raum für eigenständige und kooperative Problemlöseprozesse bei der Bearbeitung komplexer Aufgaben in möglichst authentischen Lernsituationen lässt, zu einer kognitiv aktiven Auseinandersetzung mit dem Lerngegenstand anregt und verständnisvolles fachliches Lernen stimuliert, so dass die Lernenden zu einem Verständnis von Prinzipien anstelle von nicht anwendbaren schematisierten und frag-

⁷Unterrichtsmuster (typische Muster unterrichtlicher Aktivitäten; Seidel et al., 2006a) oder Inszenierungsmuster (Verlaufsmuster von Unterricht; Hugener, 2008), im Englischen 'patterns of teaching' (Stigler & Hiebert, 1998) beschreiben die Oberflächenstrukturen bzw. die Sichtstrukturen des Unterrichts, die im Verhalten der am Unterricht involvierten Personen sichtbar werden, also die Ebene der konkreten Unterrichtsgestaltung. Dazu gehören Sozialformen und unterschiedliche Lehrer- und Schüleraktivitäten und deren Anordnung im Verlauf der Unterrichtsstunden (vgl. Seidel et al., 2006a; Hugener, 2008 in Anlehnung an Reusser, 2001 und Oser & Patry, 1990). Im Zusammenhang des Drei-Länder-Vergleichs in TIMSS 1995 Video (und in der sich 1999 anschließenden Sieben-Länder TIMS Videostudie) wurde auch von 'kulturellen Skripts' ('country specific scripts' oder 'cultural patterns') gesprochen. Darunter wurde verstanden, dass das Gesamtbild des Unterrichts in einem Land den in der Kultur verankerten Vorstellungen und Umsetzungen von 'gutem' Unterricht ähnelt (vgl. Baumert et al., 1997; Stigler & Hiebert, 1998).

⁸Eine empirische Überprüfung eines direkten Zusammenhangs der Unterrichtsmuster mit den Leistungen der Lernenden war auf der Basis der zur Verfügung stehenden Datenbasis nicht möglich (vgl. Baumert & Köller, 2000).

mentierten Wissenseinheiten kommen (vgl. Klieme et al., 2006; Klieme & Reusser, 2003; Cobb, Wood & Yackel, 1993).

Zur Beschreibung derartiger Lernumgebungen hat sich in der deutschsprachigen Mathematikdidaktik der Begriff der kognitiven Aktivierung durchgesetzt (vgl. Kunter et al., 2005; Baumert & Köller, 2000). Demnach beschreibt dieses Konstrukt einen fachlich anspruchsvollen Unterricht, der die Lernenden – so Lipowsky (2006, S. 60) – „zum vertieften Nachdenken und zu einer elaborierten Auseinandersetzung mit dem Unterrichtsgegenstand anregt“.

Theoretische Basis für das Konzept der kognitiven Aktivierung bzw. für verständnisvolle Lernprozesse ist die Auffassung, dass Lernprozesse aktive und selbstständige Konstruktionsleistungen der Lernenden sind, die situiert und kontextuiert sein müssen und motivationale Merkmale berücksichtigen (vgl. Kunter et al., 2005; Baumert & Köller, 2000; Greeno, Collins & Resnick, 1996). Die Nähe zu konstruktivistisch orientierten Sichtweisen des Lernens wird deutlich (vgl. dazu die Ausführungen in Kapitel 1.2).

Das Ziel kognitiv aktivierender Lernanlässe ist der Aufbau eines konzeptuellen Verständnisses bzw. der Erwerb konzeptuellen Wissens (vgl. Kunter et al., 2005; Brunner et al., 2006), welches sich durch die Vernetzung von Wissen, Fertigkeiten und Fähigkeiten auszeichnet, auf neue Situationen anwendbar ist und welches das Werkzeug für Problemlöseprozesse darstellt (vgl. Kunter et al., 2005, S. 503). Kognitive Aktivierung beschränkt sich aber nicht nur auf rein kognitive Ziele. Neben dem direkten Effekt auf den Lernzuwachs und einer qualitativen Verbesserung der Wissensstrukturen durch eine stärkere Vernetzung gelernter Wissensinhalte soll durch eine kognitive Aktivierung der Lernenden auch ihr Interesse und ihre Motivation an den Unterrichtsinhalten und am Unterricht selbst gefördert werden (vgl. Clausen 2002, S.17).

1.1.2 Elemente kognitiver Aktivierung im Mathematikunterricht

Die im Anschluss an TIMSS-Video gestellte Forderung, der deutsche Mathematikunterricht müsse mehr Unterrichtsmerkmale nach japanischem Vorbild enthalten, führte dazu, dass in den letzten Jahren Elemente für den Mathematikunterricht herausgearbeitet worden sind, von denen angenommen wird, dass sie die Lernenden kognitiv aktivieren. Dazu zählen:

- Die Auswahl und Implementierung von Aufgaben, da diese im Mathematikunterricht ein besonderes Potential zur Steuerung verständnisvollen Lernens bieten (vgl. Brunner et al., 2006; Kunter, 2005; Kunter et al., 2005). Es sollten herausfordernde und kognitiv *anspruchsvolle Aufgaben* mit unterschiedlichen Lösungsmöglichkeiten anstelle von Routineaufgaben verwendet werden. Die Aufgaben sollten an das Vorwissen der Lernenden anschließen und bestehen-

de Konzepte und Kompetenzen weiterentwickeln (vgl. Baumert et al., 1997; Kunter, 2005; Kunter et al., 2005, Klieme et al., 2006; Klieme, Schümer & Knoll, 2001)⁹.

- *Anspruchsvolle Unterrichtsgespräche* (nach genetisch-sokratischem Vorbild¹⁰), welche eine wesentliche aktive Beteiligung der Lernenden zulassen (vgl. Brunner et al., 2006; Kunter et al., 2005; Pauli et al., 2008). Darunter fällt auch ein *positiver Umgang mit Fehlern*, d.h. ein Verständnis fördernder Umgang mit Fehlern durch Nutzung von Fehlern als Lerngelegenheiten (vgl. Klieme et al., 2006; Kunter et al., 2005; Gruehn, 2000; Kunter, 2005).
- Das *anspruchsvolle Üben* mit variierenden mathematischen Kontexten anstelle von repetitiven Übungsphasen (vgl. z.B. Kunter et al., 2005; Gruehn, 2000; Kunter, 2005; Neubrand, 2002; Klieme, Schümer Knoll, 2001; Clausen, Reusser & Klieme, 2003).

Dazu zählt ebenfalls:

- Die *Motivierungsfähigkeit der Lehrperson*, d.h. eine generell interessante und anregende Unterrichtsgestaltung durch die Lehrperson (vgl. Klieme, Schümer & Knoll, 2001; Clausen, Reusser & Klieme, 2003; Kunter, 2005).
- Das Verhalten der *Lehrperson als 'Mediator'*, d.h. die Lehrperson gibt den Lernenden die Gelegenheit, eigene Gedanken zu entfalten und auszutauschen, indem sie den Lernenden z.B. Zeit gibt, Ideen zu entwickeln, indem sie Beiträge von Lernenden aufeinander bezieht und die Lernenden bei der Ausformulierung von Ideen unterstützt (vgl. Klieme & Clausen, 1999; Klieme, Schümer & Knoll, 2001).
- Ein angemessenes Tempo bei der Vermittlung des Unterrichtsinhaltes (*Pacing*), d.h. das Vorranschreiten im Unterricht ist am Leistungsstand der Lernenden orientiert (vgl. Kunter, 2005; Klieme & Clausen, 1999; Clausen, Reusser & Klieme, 2003).
- Strukturelle Klarheit bezüglich fachlicher und/oder organisatorischer Elemente im Unterricht (vgl. Clausen, Reusser & Klieme, 2003; Drollinger-Vetter & Lipowsky, 2006).

Folgte man im Zuge der Diskussion um die Ergebnisse von TIMSS 1995 Video der Überlegung, dass kognitiv aktivierende Unterrichtselemente vermehrt in einem problemlösend-entdeckenden Unterricht zu beobachten seien (vgl. 1.1.1), tendiert man in der neueren Literatur dazu, kognitive Aktivierung unabhängig von bestimmten Inszenierungsmustern zu betrachten, d.h. kognitiv aktivierende Unterrichtselemente sind nicht an bestimmte Methoden oder Sozialformen gebunden, sondern werden innerhalb dieser realisiert (vgl. Brunner et al., 2006, S. 58). Demnach sollte sich ein kognitiv aktivierender Unterricht also gleichermaßen in einem durch den Lehrenden enger geführten oder offener gestalteten Unterricht verwirklichen lassen, wie auch in Schülerarbeitsphasen oder gemeinsamen Unterrichtsgesprächen. „Für die normative Privilegierung bestimmter Inszenierungsmuster gibt es“ – so Baumert et

⁹Weiterführende Hinweise zur Einführung und zum Einsatz von Aufgaben im Mathematikunterricht sind nachzulesen in Knoll (2003) und Blum et al. (2006).

¹⁰Zum genetischen Lehren und Lernen siehe Kapitel 1.3.1.

al., 2004, S. 316 in Bezug auf Stebler & Reusser, 2000 – „keine belastbaren empirischen Belege“. Ob nun lehrerzentrierter oder schülerorientierter Unterricht, man kann wie Helmke formuliert, „das eine wie das andere dilettantisch oder vorzüglich tun“ (2006, S. 43). Wie viel Potential ein Unterricht hat, die Lernenden kognitiv zu aktivieren, hängt davon ab, in welcher Art und Weise Aufgaben eingeführt und behandelt werden (vgl. Kunter et al., 2005) und wie mit Vorstellungen, Ideen oder Erklärungen der Lernenden im Unterricht umgegangen wird.

In der schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“¹¹ wurde an die Diskussion um die Über- bzw. Unterlegenheit bestimmter methodischer Verfahren angeknüpft, indem verschiedene Inszenierungsmuster aus Unterrichtseinheiten zur Einführung des Satz des Pythagoras hinsichtlich ihres kognitiv aktivierenden Potentials durch Unterrichtsbeobachtungen eingeschätzt wurden (vgl. Pauli et al., 2008; Hugener, 2008). In der Studie wurden drei Inszenierungsmuster für Einführungsphasen identifiziert (Hugener, 2008): Ein *darstellendes Verfahren* (die Lehrperson erklärt den Satz des Pythagoras), ein *problemlösend-entwickelndes Vorgehen* (der Satz des Pythagoras wird anhand einer Problemstellung eingeführt und unter Führung der Lehrperson erarbeitet) und ein *problemlösend-entdeckendes Verfahren* (zu einer Problemstellung suchen die Lernenden zuerst selbstständig nach Lösungen, die im Klassenverband vorgestellt und diskutiert werden). Pauli et al. (2008) fassten das darstellende und das problemlösend-entwickelnde Verfahren als *geführtes Vorgehen* (N=19) zusammen, stellten es dem *problemlösend-entdeckenden Vorgehen* (N=19) gegenüber und ließen diese 38 Unterrichtseinheiten¹² mittels hoch-inferenter Expertenurteile hinsichtlich des Ausmaßes an (1) Gelegenheiten zur Auseinandersetzung mit anspruchsvollen Aufgaben in Schülerarbeitsphasen, (2) Klassengesprächen mit substanzieller Schülerbeteiligung und (3) inhaltlich-struktureller Klarheit in Theoriephasen¹³ beurteilen. Die Ergebnisse zeigen keine systematischen Zusammenhänge zwischen dem methodischen Vorgehen beim Begriffsaufbau in der Einführungsphase und den untersuchten Dimensionen eines kognitiv aktivierenden Unterrichts. In welchem Ausmaß die drei erfassten Qualitätsmerkmale in Einführungsphasen realisiert werden, hängt also nicht vom Instrukti-
onsansatz ab. Die Untersuchungsergebnisse werden von den Autoren als Hinweis für die Berücksichtigung methodenunabhängiger Qualitätsmerkmale gesehen. Auch Hugener (2008) ließ im selben Projekt im Rahmen ihrer Dissertation neben den oben genannten Einführungsmustern auch Übungsmuster (Ü1 kaum bzw. keine Übungsphase vs. Ü2 kurze, repetitive Übungsphase vs. Ü3 lange, selbstgesteuerte, repetitive und anspruchsvolle Übungsphase) hinsichtlich kognitiv aktivierender Unterrichtselemente durch externe Beobachter einschätzen. Als Indikatoren für kognitive Aktivierung galten die Dimensionen „Herausfordernde Aufgaben und Probleme“, „Aktivierung des Vorwissens“, „Exploration der

¹¹Das Kooperationsprojekt wurde auf deutscher Seite von E. Klieme (DIPF) und auf schweizerischer Seite von K. Reusser und C. Pauli (Universität Zürich) geleitet.

¹²Die 38 Unterrichtseinheiten zur Einführung des Pythagoras setzten sich aus jeweils drei Mathematikstunden von Lehrpersonen aus der Schweiz (N=19) und aus Deutschland (N=19) zusammen.

¹³Vgl. dazu auch Kapitel 1.1.3 zu den Wirkungen eines kognitiv aktivierenden Unterrichts. Detaillierte Angaben zu den genannten Indikatoren für kognitive Aktivierung sind nachzulesen in Hugener, Pauli & Reusser (2007); Lipowsky et al. (2007); Drollinger-Vetter & Lipowsky (2006).

Denkprozesse“, „Evolutionärer Umgang mit vorhandenen Ideen“ und „rezeptives Lernverständnis“. Ihre Befunde zeigen für die Übungsmuster keine systematischen Zusammenhänge. Problemorientiert-entdeckende Einführungsmuster gehen jedoch entgegen den Ergebnissen von Pauli et al. (2008) mit einer günstigeren Beurteilung der kognitiven Aktivierung einher. Zwar verlaufen die Ergebnisse nicht völlig konform, sie können aber – vor allem auch in Anbetracht der geringen Effekte von Inszenierungsmustern auf Schülerleistungen (vgl. Folgekapitel 1.1.3) als Hinweis dafür gesehen werden, dass man das Augenmerk mehr auf einzelne Unterrichtsmerkmale und nicht so sehr auf das Gesamtarrangement von Unterricht lenken sollte.

1.1.3 Zu den Wirkungen eines kognitiv aktivierenden Mathematikunterrichts

Mit dem Ziel, schulische Leistungsunterschiede von Lernenden in Mathematik durch das Ausmaß an kognitiver Aktivierung im Unterricht erklären zu können, wurden in mehreren Studien die Wirkungen von Unterrichtselementen, von denen man annimmt, dass sie die Lernenden kognitiv aktivieren (vgl. 1.1.2), in Bezug auf die fachlichen Leistungen der Lernenden analysiert. In einigen Studien wurden auch nicht kognitive Zieldimensionen berücksichtigt; es wurde die Wirkung eines kognitiv aktivierenden und fachlich anspruchsvollen Unterrichts auf motivationale Zielkriterien untersucht. Die bisher vorliegenden Forschungsbefunde sollen im Folgenden dargelegt werden. Dabei werden in Anlehnung an Hugener (2008) zuerst Untersuchungen aufgeführt, in denen die Analyse des Unterrichts hinsichtlich seines kognitiv aktivierenden Potentials auf der Basis von Befragungen (Lehrende und Lernende) erfolgte. Daran anschließend werden Ergebnisse aus Studien auf der Basis von Unterrichtsbeobachtungen (Videostudien) dargestellt. Alle hier dargestellten Studien stammen aus dem deutschsprachigen Raum.

Studien auf Basis von Befragungen (Lernende, Lehrende)

Gruehn (2000) konnte im Rahmen der BIJU-Studie (Bildungsverläufe und psychosoziale Entwicklung im Jugendalter) für den Mathematikunterricht der siebten Jahrgangsstufe (und für den Physik- und Biologieunterricht) auf der Basis von Schülerbefragungen zeigen, dass kognitiv aktivierende Unterrichtselemente einen bedeutsamen Effekt auf den Leistungszuwachs der Lernenden im Mathematikunterricht an Gymnasien haben. Gemessen wurde die kognitive Aktivierung anhand von Items zum genetisch-sokratischen Umgang mit Vorstellungen und Fehlern von Lernenden und zum anspruchsvollen Üben (Einsatz von Aufgaben, die der Prüfung des Verständnisses von Konzepten und Prinzipien dienen; vgl. Gruehn, 2000).

Baumert und Köller (2000) untersuchten im Rahmen von TIMSS 1995 ebenfalls die Wirkung verschiedener, per Schülerfragebogen erfasster Unterrichtsmerkmale auf die Leistungen der Lernenden im Mathematikunterricht der gymnasialen Oberstufe. Befunde auf der Basis von Mehrebenenanalysen belegen, dass ein Unterricht, in dem die Lernenden didaktische Maßnahmen wahrnehmen, die auf Verständnisorientierung durch kognitive Herausforderungen und Eigentätigkeit weisen (z.B. „Zusammen-

hänge darstellen und analysieren“) mit höheren Testleistungen der Lernenden einhergehen. Demgegenüber ist ein Unterricht, der nach Aussagen der Lernenden auf Rezeptivität und Reproduktivität angelegt ist (z.B. *„Der Lehrkraft zusehen, wie sie zeigt, wie mathematische Aufgaben und Prozeduren behandelt werden“*) dem Lernerfolg eher abträglich (Baumert & Köller, 2000, S. 276f. und 282). Weitere Analysen zeigen, dass verständnisorientierte Unterrichtsstrategien in einem systematischen positiven Zusammenhang mit motivationalen Zielkriterien des Unterrichts stehen, während rezeptiver Unterricht mit negativen Effekten im motivationalen Bereich einhergeht (S. 305).

Im Rahmen einer schweizerischen Vertiefung zu TIMSS 1995 untersuchten Stebler und Reusser (2000) die didaktische Gestaltung von Mathematikunterricht der Sekundarstufe I. Mittels Clusteranalysen auf der Basis von Lehrerangaben (N=102) ließen sich drei Typen von Unterrichtsmustern bilden, die sich bezüglich der Gelegenheiten der Lernenden zum Artikulieren, Reflektieren und Explorieren mathematischer Sachverhalte sowie in Bezug auf die Leistungsbeurteilung und die Lehrerrolle unterscheiden. Die Formen der Unterrichtsgestaltung haben gemäß den Studienergebnissen keinen Einfluss auf die Mathematikleistungen der Lernenden.

Moser et al. konnten 1997 in der Schweiz im Rahmen von TIMSS+ (Vertiefung zu TIMSS) auf der Basis von Schülerbefragungen und Leistungstests keinen direkten Effekt von Unterrichtsmerkmalen auf Mathematikleistungen feststellen, wohl aber einen indirekten Einfluss: Die Unterrichtsgestaltung (Merkmal „Selbstständigkeit“, z.B. *„Unser Lehrer gibt uns oft die Gelegenheit selbstständig zu denken“*) hatte einen Effekt auf das Interesse der Schüler; für das Interesse konnte wiederum ein Einfluss auf die Leistung nachgewiesen werden (zit. n. Hugener, 2008, S. 76f.).

An dieser Stelle sei angemerkt, dass bei der Betrachtung der Ergebnisse zu beachten ist, dass es sich in allen genannten Studien bei den Einschätzungen des Unterrichts hinsichtlich seines kognitiv aktivierenden Potentials entweder um Schüler- oder um Lehrerwahrnehmungen handelt, also um Berichte von unmittelbar am Unterrichtsgeschehen beteiligter Personen, die wie alle Selbstberichte Verzerrungen unterworfen sind (vgl. z.B. Baumert et al., 2004). Die Unterrichtsforschung hat gezeigt, dass die Frage, mittels welcher Informationsquelle Unterricht erfasst werden sollte – ob durch Lehrer, Schüler oder externe Beobachter – von den zu untersuchenden Merkmalen und Fragestellungen abhängt (vgl. z.B. Clausen, 2002)¹⁴. Geht es um eine neutrale Erfassung und Bewertung von Unterrichtsqualitätsmerkmalen, die sich auf Aspekte des unterrichtlichen Lehrerhandelns bezieht, wie es bei dem Konstrukt der kognitiven Aktivierung der Fall ist, empfiehlt sich gemäß Untersuchungsbefunden von Clausen (2002) der Einsatz externer Beobachter. Lehrerurteile weisen hingegen bei Aspekten, die eine direkte Selbstbewertung einschließen, eingeschränkte Gültigkeit auf, was sich nach Clausen (2002, S. 188) durch die „selbstdienlichen Verzerrungstendenzen“ erklärt. Auch Schülerurteile sind für eine neutrale Beobachtung des Unterrichtsgeschehens weniger valide, da es scheint, dass Schüler eher eine

¹⁴Clausen (2002) untersuchte im Rahmen seiner Dissertation auf der Basis von Daten aus TIMSS 1995 Video den Einsatz verschiedener Datenquellen für die Erfassung von Unterrichtsmerkmalen.

globale Beurteilung ihrer Lehrkräfte vornehmen, die durch die wahrgenommene affektive Unterrichtsqualität beeinflusst wird (vgl. Clausen, 2002). Zudem zeigt ein Untersuchungsbefund aus dem Projekt COAKTIV¹⁵, dass Lernende scheinbar den im Unterricht erlebten Schwierigkeitsgrad berichten und nicht das tatsächliche Angebotsniveau (vgl. Baumert et al., 2004; Kunter, 2005)¹⁶. Die Erhebung von Schülerwahrnehmungen empfiehlt sich nach Baumert et al. (2004), wenn es um das subjektive Erleben des Unterrichts hinsichtlich affektiver und sozialer Merkmale geht.

Studien auf der Basis von externen Beobachtungen

Renkl und Stern (1994) untersuchten auf der Basis von Daten aus SCHOLASTIK (Schulorganisierte Lernangebote und Sozialisation von Talenten, Interessen und Kompetenzen; z.B. Helmke & Weinert, 1997) u.a. die Bedeutung der Art der im Mathematikunterricht eingesetzten Lernaufgaben für die Erklärung von Leistungsunterschieden bei Grundschulkindern. Es wurde durch externe Unterrichtsbeobachter mittels eines niedrig-inferenten Beobachtungsinstruments im event-sampling¹⁷ die Häufigkeit von einfachen und komplexen Textaufgaben pro Unterrichtsstunde (N=33) bestimmt (Renkl & Stern, 1994). Die Ergebnisse zeigen, dass sich konzeptuell anspruchsvolle Aufgaben im Mathematikunterricht der Grundschule als lernförderlich erweisen. Staub und Stern (2002) konnten diesen Befund auch für eine Subgruppe aus der von Renkl und Stern (1994) genutzten Stichprobe (N=22) bestätigen. Auch ihre Befunde zeigen, dass der Gebrauch von anspruchsvollen Aufgaben im Unterricht durch die Lehrkraft ein wichtiger Aspekt ist, um die Lernenden im Lösen von komplexen mathematischen Textaufgaben zu unterstützen.

Klieme und Clausen (1999) untersuchten 147 Unterrichtslektionen aus 82 an der TIMS-Studie beteiligten Klassen mittels hoch-inferenten Ratings auf Merkmale kognitiv aktivierenden Mathematikunterrichts¹⁸. Analysen ergaben keinen direkten Einfluss der Indikatoren auf die Lernleistungen der Lernenden. Jedoch konnte ähnlich wie bei Moser (1997) ein direkter positiver Zusammenhang der Items „Schülerdiskussion“, „Lehrperson als Mediator“, „Offenheit der Problemstellungen und Lösungen“ und „Schülerkooperation“ mit der Interessenentwicklung der Lernenden im Fach Mathematik nachgewiesen werden.

¹⁵Das Projekt COAKTIV („Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz“) war eingebunden in PISA 2003/2004 der OECD. Ziel des Projekts war die mehrperspektivische Beschreibung der Unterrichtsqualität (vgl. Baumert et al., 2004; Krauss et al., 2004; Kunter et al., 2005).

¹⁶Die Untersuchung zeigt, dass eine kognitiv aktivierende Unterrichtsgestaltung allgemein selten und in Gymnasien häufiger als in Realschulen und dort häufiger als in Hauptschulen beobachtbar ist. Demgegenüber erleben sich Gymnasial- und Realschüler als durchschnittlich wenig kognitiv aktiviert, während sich Hauptschüler als kognitiv sehr aktiviert erleben. Dies wird als Hinweis gesehen, dass die Lernenden den erlebten Schwierigkeitsgrad berichten und nicht das tatsächliche Unterrichtsangebot (vgl. Baumert et al., 2004; Kunter et al., 2005).

¹⁷Zum Begriff der Inferenz und zu Verfahren der Videoanalyse siehe Kapitel 1.5.1.

¹⁸Erfasst wurden neun Items, die sich auf drei Skalen verteilen: student discussion, teacher as mediator, openness, student cooperation (Skala I: argumentative and reflective interaction); focusedness, problem solving process, knowledge anchoring (Skala II: goal-oriented and focussed solution approaches) und authentic embedding, multiple contexts (Skala III: problems in multiple natural contexts) (Klieme & Clausen, 1999, S. 10).

Klieme, Schümer und Knoll (2001) konnten nach einer Analyse der Unterrichtsaufzeichnungen aus der TIMS 1995 Videostudie nachweisen, dass der Lernzuwachs in den Klassen (gemessen mit dem TIMSS-Leistungstest) in schwach positivem Zusammenhang mit kognitiv aktivierendem Unterricht steht (Korrelation von .22). Das Konstrukt der kognitiven Aktivierung spiegelt nach Klieme, Schümer und Knoll (2001, S. 51) die Komplexität von Aufgabenstellungen und Argumentationen sowie die Intensität des fachlichen Lernens wider und wurde in dieser Studie über hoch-inferente Ratings durch geschulte externe Beobachter anhand der Skalen „Genetisch-sokratisches Vorgehen“, „Anspruchsvollen Üben“, „Motivierungsfähigkeit des Lehrers“ und in umgekehrter Polung „Repetitives Üben“ erfasst.

Kunter (2005) konnte im Rahmen einer Re-Analyse von 80 deutschen Unterrichtsvideos aus der TIMS 1995 Videostudie (8. Jahrgangsstufe) in Mehrebenenanalysen auf Individualebene und unter Kontrolle der Eingangsvoraussetzungen (Leistung im vorausgehenden Schuljahr) und der Klassenführung (auf der Basis von Schülerbefragungen) keine signifikanten Effekte zwischen den hoch-inferent eingeschätzten Unterrichtsmerkmalen „Konstruktion des Wissens“, „Relevanz und Bedeutung der Lernerfahrungen“ und „Unterstützung der Lernenden beim eigenständigen Lernen“¹⁹ auf die Leistungsentwicklung nachweisen. Auch für die Interessenentwicklung konnten keine signifikanten Effekte nachgewiesen werden. Allerdings geht eine aktive Konstruktion des Wissens (der Bereich beschreibt die Exploration von Vorwissen und Denkweisen der Lernenden, die Art der Aufgabenstellungen und die Gesprächsführung bei der Vermittlung von Inhalten) tendenziell mit etwas höheren Leistungszuwächsen einher.

Im Rahmen der bereits in Kapitel 1.1.2 genannten schweizerisch-deutschen videobasierten Unterrichtsstudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“ wurden verschiedene Analysen bezüglich der Bedeutung kognitiv aktivierender Unterrichtselemente für die Leistungs- und Motivationsentwicklung der Lernenden untersucht.

So schätzten externe Beobachter im Projekt als Indikatoren für kognitiv aktivierenden Unterricht u.a. ein, inwieweit die Lehrkraft das Vorwissen der Schüler exploriert, inwieweit sie herausfordernde Aufgaben stellt und inwieweit sie sich für die Gedankengänge der Schüler interessiert. Es wurde ein schwach positiver signifikanter Zusammenhang zwischen der hoch-inferent eingeschätzten kognitiven Aktivierung und dem Lernerfolg der Schüler während einer Unterrichtseinheit (3 Stunden) zum Themengebiet 'Satz des Pythagoras' festgestellt (vgl. Lipowsky, 2006 und Lipowsky et al., 2009). Zudem zeigt sich hier ein Interaktionseffekt zwischen kognitiven und motivationalen Faktoren: Vor allem die stärker an Mathematik interessierten Lernenden profitieren von kognitiv aktivierendem Unterricht (vgl. Klieme et al., 2006; Lipowsky et al., 2009).

¹⁹Eine ausführliche Erläuterung der Skalen findet sich bei Kunter (2005, S. 211ff).

In einer weiteren Publikation des Projekts (Pauli et al., 2008) wurde untersucht, ob die folgenden drei Merkmale eines kognitiv aktivierenden Unterrichts einen Effekt auf die Leistungsentwicklung der Lernenden haben: „Auseinandersetzung mit anspruchsvollen Aufgaben in Schülerarbeitsphasen“ (anspruchsvolles vs. repetitives Üben; Analyseeinheit: Schülerarbeitsphasen), „Klassengespräche mit substanzieller Schülerbeteiligung“ (durchschnittliche Länge eines Schülerbeitrags in Anzahl von Worten) und „inhaltliche Kohärenz und Transparenz des Unterrichts“ (Beispielitem „*Kohärenz von Satz und Beweis ist nachvollziehbar*“; 1 = niedrige bis 4 = hohe Ausprägung, vgl. Drollinger-Vetter & Lipowsky, 2006). Die Ergebnisse von Mehrebenenanalysen zeigen, dass sich eine höher eingeschätzte inhaltlich-strukturelle Klarheit in Theoriephasen und längere Schülerbeiträge im Klassengespräch positiv auf die Leistungen auswirken. Ein höherer zeitlicher Anteil anspruchsvoller Schülerarbeitsphasen leistet keinen Beitrag zur Verständnisentwicklung.

Merkmale der Schülerbeteiligung im Unterricht wurde auch in weiteren Auswertungen nachgegangen. Frühere Analysen zeigen, dass eine gleichmäßige Beteiligung aller Schüler an Klassengesprächen einen tendenziell positiven Effekt auf den kurzfristigen Lernerfolg hat (Lipowsky et al., 2007). Für die Beteiligung aller Schüler an Lehrer-Schülerinteraktionen in Schülerarbeitsphasen trifft dies nicht zu (Lipowsky et al., 2008).

Eine weitere Veröffentlichung (Klieme et al., 2006) zeigt, dass auch ein prozessorientierter, diskursiver, auf konzeptuelles Verständnis ausgerichteter Umgang mit Hausaufgaben im Mathematikunterricht Einfluss auf die Leistungsentwicklung im Verlauf eines Schuljahres hat²⁰.

Hugener ging im Rahmen ihres Dissertationsvorhabens (2008) der Frage nach Zusammenhänge zwischen Inszenierungsmustern und Lernleistungen der Lernenden nach. Wie bereits in Kapitel 1.1.2 erläutert, untersuchte sie die Einführungs- und Übungsmuster getrennt voneinander. Die Ergebnisse zeigen, dass weder die Einführungs- noch die Übungsmuster unter Kontrolle des Vorwissens auf Individual- und Klassenebene einen signifikanten Effekt auf die Nachtestleistungen der Lernenden (erhoben unmittelbar nach der Videoaufzeichnung) haben. Die Ergebnisse verlaufen konform zu den Ergebnissen aus der erwähnten Studie von Stebler und Reusser (2000). Hugener selbst diskutiert ihre Ergebnisse mit Bezug auf die Studien von Kunter (2005) und Seidel et al. (2006b), die ebenfalls den fehlenden Einfluss von Oberflächenmerkmalen auf die Leistungsentwicklung der Lernenden bestätigen, dahingehend, dass prinzipiell verschiedene Inszenierungsmuster als leistungsförderlich gelten können und dass es möglicherweise weitere lernförderliche Unterrichtsmerkmale gibt, die weniger auf der Oberflächenebene, sondern eher auf der Ebene der Tiefenstruktur von Unterricht (Lehrer- und Schülerverhalten) anzusiedeln sind. Diese Argumentation unterstützt also die in Kapitel 1.1.2 angesprochene Befürwortung methodenunabhängiger Qualitätsmerkmale.

²⁰Die Skala ist näher beschrieben in Klieme et al. (2006).

Betrachtet man die Ergebnisse in Zusammenhang mit Ergebnissen zu Schülerwahrnehmungen von Lernprozessmerkmalen, verwundert zumindest der ausgebliebene Effekt der problemlösend-entdeckenden Einführungsmuster auf die Lernleistungen der Schüler nicht: Die Ergebnisse zeigen, dass problemlösend-entdeckende Einführungsmuster bei den Schülern zu einer negativen emotionalen Empfindsamkeit führen. Zudem berichten die Schüler in einem derartigen Unterricht im Vergleich zu einem darstellenden Unterricht von einem signifikant geringeren Verstehen (vgl. Hugener, 2008, S. 221). Folgerichtig schließt Hugener (2008, S. 229) daraus, dass die Ergebnisse möglicherweise auf eine Überforderung der Lernenden durch zu viel Selbststeuerung hinweisen. Damit knüpft Hugener an die schon seit Jahren geführte Diskussion um das Für und Wider einer instruktionalen Unterstützung durch die Lehrkraft im Unterricht an (vgl. Abschnitt 1.3.3.3).

Fazit

Insgesamt zeigen die berichteten Forschungsbefunde zu den Wirkungen eines kognitiv aktivierenden Mathematikunterrichts kein völlig einheitliches Bild. Es gibt Studien, in denen eine positive Wirkung eines derartigen Unterrichts in Mathematik bestätigt werden konnte. Es gibt aber auch Studien, in denen keine Zusammenhänge zwischen kognitiv aktivierenden Unterrichtsmerkmalen und Zielkriterien des Unterrichts nachgewiesen werden konnten. Welche Schlüsse lassen sich nun daraus ziehen?

Die Ausführungen zur Entstehung des Konstrukts der kognitiven Aktivierung haben gezeigt, dass der Begriff ein empirisch gewonnener ist und theoretisch – wenn er auch Bezüge zu konstruktivistisch orientierten Lerntheorien aufweist – kaum fundiert ist. Dies hat möglicherweise dazu geführt, dass die aus der Praxis gewonnenen Eindrücke zur Erkennung kognitiver Aktivierung in den einzelnen Studien unterschiedlich operationalisiert wurden, was zu unterschiedlichen Befunden bei der Untersuchung von Effekten kognitiv aktivierender Unterrichtselemente führen kann. Nicht alle operationalisierten Elemente eignen sich zur Erhebung kognitiv aktivierenden Unterrichts. Somit müssen negative Ergebnisse nicht bedeuten, dass kognitive Aktivierung keinen Einfluss auf die Schülerleistungen hat. Vielmehr könnten sie auf unzureichende Operationalisierung der kognitiven Aktivierung hinweisen. Auf das Problem der mangelnden Operationalisierung verweist auch Hugener (2008, S. 82). Es ist ihr zuzustimmen, dass die Operationalisierung des Konstrukts und damit verbunden die Theoriearbeit als noch nicht abgeschlossen gelten kann.

Außerdem können methodische Probleme auftreten. Auf die Problematik der 'Verzerrungen' bei der Erfassung kognitiver Aktivierung per Lehrer- oder Schülerfragebogen wurde bereits hingewiesen. Aber auch die Videoanalyse birgt Schwierigkeiten. Die Urteile aus externer Beobachterperspektive sind natürlich ebenfalls nicht fehlerfrei. Des Weiteren verweist Hugener (2008, 2009) auf weitere Aspekte, die bezüglich der videogestützten Forschung berücksichtigt werden müssen: In unterschiedlichen Studien werden verschiedene Erhebungsmethoden und Analyseeinheiten (z.B. ganze Stunde oder Klassengespräch) genutzt (siehe dazu die Ausführungen in Kapitel 1.5.1 zu Methoden der Videoanalyse). Zudem wurde in videogestützten Studien meist nur eine Lektion pro Lehrperson gefilmt. Oft wur-

de das Thema der gefilmten Stunden nicht vereinheitlicht. Hinzu kommt – und das ist ein sehr bedeutender Punkt, den Hugener nennt – dass zur Erfassung der Lernleistungen (außer bei Lipowsky et al., 2009) meist internationale Leistungstests aus PISA oder TIMSS oder sonstige eher allgemein gefasste Tests benutzt wurden, welche allgemein mathematische (Problemlöse-)Kompetenzen und Fähigkeiten messen, aber in keinem direkten Zusammenhang mit dem beobachteten Unterricht stehen. Aussagen zur mikrogenetischen Leistungsentwicklung mit direktem Bezug zum analysierten Unterricht sind daher nicht möglich.

Dennoch, trotz uneinheitlicher Forschungslage und trotz der kurz angerissenen Probleme des Forschungsfeldes weisen die berichteten positiven Befunde darauf hin, dass es Elemente im Unterricht losgelöst vom unterrichtlichen Gesamtarrangement gibt, die durchaus Einfluss auf den Lernerfolg der Schüler haben können. Hinweise für die Bedeutung der fachdidaktischen Qualität von Unterricht finden sich auch in der internationalen Forschung, darauf macht Lipowsky aufmerksam, indem er z.B. auf das britische Projekt CAME (Cognitive Acceleration in Mathematics Education) verweist: Die Befunde des Projekts „verweisen ebenfalls auf die positive Wirkung eines fachlich anspruchsvollen, kognitiv aktivierenden Mathematikunterrichts. Die Studie berichtet über eine mittlere Effektgröße von $d=.40$ für Schüler, deren Lehrer an einem Programm zur Steigerung des kognitiven Niveaus im Unterricht teilgenommen hatten, gegenüber Kontrollgruppenschülern, deren Lehrer nicht an diesem Programm teilgenommen hatten (Adhami, 2002; Shayer & Adhami, 2007)“ (Lipowsky, 2007b, S. 43).

Somit kann Lipowsky (2006) zugestimmt werden, dass es sinnvoll und wichtig ist, neben allgemein fachunabhängigen Unterrichtsmerkmalen (wie z.B. einer effiziente Klassenführung) auch Merkmale von Unterricht zu erfassen, die auf eine vertiefende inhaltliche Auseinandersetzung der Lernenden mit dem Unterrichtsgegenstand deuten.

Auf der Basis des Forschungsstands zur kognitiven Aktivierung herrscht in den Arbeitsgruppen bedeutender deutscher Unterrichtsforscher wie Baumert, Klieme und Reusser allgemein Konsens darüber, die kognitive Aktivierung als ein entscheidendes Unterrichtsqualitätsmerkmal zu betrachten. In Abgrenzung zu anderen Basisdimensionen von Unterrichtsqualität aus allgemein-fachdidaktischer Sicht, welche die Grundlage für Lernprozesse in allen Unterrichtsfächern bilden, handelt es sich bei der kognitiven Aktivierung um eine Dimension guten Unterrichts aus fachdidaktischer Sicht (vgl. z.B. Klieme, Schümer & Knoll, 2001; Clausen, 2002; Kunter et al., 2005; Klieme et al., 2006; Brunner et al., 2006; Lipowsky, 2007a; Klieme & Thußbas, 2001). Das Konstrukt 'kognitive Aktivierung' steht dabei in Zusammenhang mit anderen Unterrichtsqualitätsmerkmalen. So ist gemäß Klieme, Schümer und Knoll (2001, S. 53) zum Beispiel eine Schülerorientierung ein wichtiger Faktor, um die Lernmotivation der Schüler zu fördern. Der Lernzuwachs wird durch das Ausmaß an kognitiver Aktivierung bestimmt. Die Effizienz der Unterrichtsführung, d.h. eine strukturierte und störungsfreie Lernumgebung, sowie ein Mindestmaß an Aufmerksamkeit bilden eine notwendige, wenn auch nicht hinreichende Voraussetzung für die kognitive Aktivierung.

1.1.4 Zusammenfassung und Überleitung zum anspruchsvollen naturwissenschaftlichen Lernen im Sachunterricht der Grundschule

Das Konstrukt der kognitiven Aktivierung ist im Kontext von Untersuchungen zum Mathematiklernen (Sekundarstufe I) entstanden bzw. empirisch entwickelt und in die deutschsprachige fachdidaktische Diskussion um Mathematikunterricht eingebracht worden. Kognitiv aktivierende Lernanlässe gelten als Voraussetzung für verständnisvolles fachliches Lernen und beschreiben ein anspruchsvolles, verständnisbezogenes und aktives Lernen. Ob die Lernenden tatsächlich kognitiv aktiviert sind, kann nicht direkt, sondern nur indirekt über Merkmale erfasst werden, von denen man annimmt, dass sie die Lernenden kognitiv aktivieren. Solche Unterrichtselemente sind in den letzten Jahren für den Mathematikunterricht entwickelt und auf ihre Lernwirksamkeit hin untersucht worden. Studienbefunde zeigen, dass Unterschiede in den Leistungen der Lernenden durch ein unterschiedlich hohes Maß an kognitiv aktivierenden Unterrichtselementen erklärt werden können, wobei die kognitive Aktivierung mittels verschiedener Indikatoren (z.B. Aufgaben mit adäquatem kognitiven Potential, anspruchsvolles vs. repetitives Üben oder Klassengespräche mit einem hohen Maß an Schülerbeteiligung) und durch verschiedene Datenquellen (Lehrer-Schüler-Befragungen und Unterrichtsbeobachtungen durch externe Beobachter) erfasst wurde. Darüber hinaus konnte nachgewiesen werden, dass ein kognitiv aktivierender, auf Verständnis ausgerichteter Mathematikunterricht nicht nur für den kognitiven Lernerfolg der Lernenden wichtig ist, sondern auch auf die Motivation und das Interesse der Lernenden einen positiven Effekt hat. Basierend auf den Ergebnissen gilt die kognitive Aktivierung heute als bedeutendes fachliches Unterrichtsqualitätsmerkmal für den Mathematikunterricht.

Anspruchsvolles naturwissenschaftsbezogenes Lernen im Sachunterricht

Nicht nur für den Mathematikunterricht (und daran anschließend für die naturwissenschaftlichen Unterrichtsfächer) der Sekundarstufe werden verständnisorientierte Lernanlässe gefordert. Diese Forderung besteht auch für das naturwissenschaftliche Lernen in der Grundschule. Dabei sind vor allem Forderungen nach einem anspruchsvollen, verständnisorientierten und denkfördernden naturwissenschaftsbezogenen Lernen im Vorfeld der so genannten 'harten' Naturwissenschaften (Physik, Chemie) laut geworden.

Ein Rückblick auf die Entwicklung des Faches Sachunterricht zeigt, dass diese Forderungen nicht neu sind. In Deutschland existiert die Forderung nach einer fundierten frühen naturwissenschaftlichen Bildung seit den späten 1960er und frühen 1970er Jahren (vgl. Möller, 2002, 2004a). Auf der Basis von Empfehlungen des Deutschen Bildungsrates wurden in den 1970er Jahren so genannte wissenschaftsorientierte Curricula nach amerikanischem Vorbild eingeführt. Bereits Ende der 1970er Jahre scheiterten die Bemühungen infolge der Vernachlässigung kindlicher Denkweisen und Interessen. Die Kinder waren kognitiv überfordert, so dass ein Verständnis naturwissenschaftlicher Konzepte und Begriffe nicht erreicht wurde (vgl. Soostmeyer, 1988; Möller, 2004a; Lauterbach, 1992). In der Folgezeit setzte

eine Hinwendung zu lebensweltorientierten Ansätzen ein. Physikalische und chemische Inhalte wurden fast komplett aus den Lehrplänen eliminiert (vgl. Möller, 2004a; Köhnlein, 1984). Heute besteht wieder Konsens darin, dass bereits Grundschulkindern ein verstehen- und interessenförderndes naturwissenschaftliches Lernen ermöglicht werden muss. International wird diese Forderung unter dem Begriff 'scientific literacy' geführt. Das Konzept meint allgemein „die Grundlegung eines *für alle* gemeinsamen naturwissenschaftlich auslegbaren Wissens, Verstehens und Handelns“ (Lauterbach, 2005, S. 572). Die Grundschule als erste Schulstufe des Bildungswesens wird als bedeutender Ort für eine naturwissenschaftliche Grundbildung gesehen (vgl. Prezel et al., 2003).

In Deutschland hat die Forderung dazu geführt, dass naturwissenschaftliche Themen in den aktuellen Lehrplänen wieder vermehrt zu den verpflichtenden Inhalten gehören. Im Perspektivrahmen Sachunterricht der Gesellschaft für Didaktik des Sachunterrichts (GDSU) (2002), der inzwischen in vielen Bundesländern als Orientierung bei Lehrplanrevisionen diente, werden naturwissenschaftliche Inhalte in einem eigenen Inhaltsschwerpunkt (Perspektive) berücksichtigt. Zudem sind in den letzten Jahren Programme zur Förderung der naturwissenschaftlichen Bildung entstanden, wie z.B. „SINUS-Transfer Grundschule“ als Weiterentwicklung des BKL-Modellversuchs „SINUS“ („Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“).

Die Gründe für die Forderung nach und die Bemühungen für die Umsetzung von anspruchsvollem naturwissenschaftsbezogenem Lernen im Sachunterricht sind vielfältig und sollen im Folgenden näher beleuchtet werden.

Reaktionen auf die Ergebnisse internationaler Schulleistungsstudien. Nach der Abwendung des Sachunterrichts von naturwissenschaftlichen Inhalten in den 1980er Jahren, sind Forderungen nach einem anspruchsvollen (naturwissenschaftlichen) Sachunterricht vereinzelt immer wieder laut worden. Stellvertretend dafür steht z.B. die Aussage von Schreier:

„Die in den frühen siebziger Jahren vom Deutschen Bildungsrat angesagte Wissenschaftsorientierung, so könnte man meinen, hat die Grundschulpädagogen derart verschreckt, dass es zu einer Gegenbewegung gekommen ist, die immer noch anhält und rigoros alles aus dem Unterricht entfernt, was Ratio, Intellekt und Logik präsentiert“ (Schreier, 1990, S. 9).

Einen großen Anstoß erhielt die Debatte um einen anspruchsvollen und verständnisorientierten naturwissenschaftlichen Sachunterricht allerdings erst gut 10 Jahre später im Zuge der Diskussion um die unbefriedigenden Ergebnisse der Schulleistungsstudien wie TIMSS und PISA bezüglich der deutschen Sekundarstufenschüler im mathematisch-naturwissenschaftlichen Bereich. Diese führten Ende der 1990er Jahre nicht nur zu einer Diskussion um die unzureichende Bildungsqualität in Mathematik und in den Naturwissenschaften deutscher weiterführenden Schulen, sie hatten auch Auswirkungen auf die Grundschule und dort – wie schon knapp 40 Jahre zuvor – insbesondere auf Diskussion um die naturwissenschaftliche Bildung, die in Deutschland ein Teilelement des Sachunterrichts darstellt. Gründe für das schlechte Abschneiden der Sekundarstufenschüler wurden – ob nun gerechtfertigt oder nicht, dies soll an dieser Stelle nicht weiter diskutiert werden – unter anderem in der unzureichenden Ausbil-

derung in den Grundschulen gesehen (vgl. z.B. Stern & Möller, 2004). Von einer Verbesserung der naturwissenschaftlichen Grundschulbildung versprach man sich daher auch eine Verbesserung der naturwissenschaftlichen Lernleistungen in den weiterführenden Schulen.

Bildungsanspruch des Sachunterrichts. Die Legitimation einer frühen naturwissenschaftlichen Bildung im Zusammenhang mit TIMSS, PISA und IGLU nimmt demnach auf die Grundlegung für weiterführendes Lernen Bezug (vgl. Möller, 2002). Selbstverständlich hat ein anspruchsvolles naturwissenschaftsbezogenes Lernen in der Grundschule nicht nur fachpropädeutische Funktion. Dies verbietet der Bildungsanspruch des Sachunterrichts. In erster Linie ist es die Aufgabe des Sachunterrichts, die Lernenden bei der Bewältigung ihrer Lebenswelt bildungswirksam zu unterstützen. Dabei werden auch die Grundlagen für weiterführendes Lernen gelegt. Zudem muss der Sachunterricht die Lernbedürfnisse, Fragen und Interessen der Kinder berücksichtigen (GDSU, 2002, S. 2). Die Aufgabe des Sachunterrichts ist also, Anschlussfähigkeit sowohl an die Lebenswelterfahrungen der Kinder und ihre Interessen als auch an die Fächer der weiterführenden Schulen zu sichern (S. 3). Damit bewegt sich das naturwissenschaftliche Lernen zwischen den Polen einer Wissenschafts- und Kindorientierung (siehe dazu die folgenden Ausführungen zu den Zielen naturwissenschaftsbezogenen Sachunterrichts).

Neue entwicklungspsychologische Erkenntnisse. Heute werden die Forderungen nach einem anspruchsvollen frühen naturwissenschaftlichen Lernen durch neue entwicklungspsychologische Erkenntnisse gestützt (vgl. Möller, Jonen, Hardy & Stern, 2002), die zeigen, dass Grundschulkindern zu anspruchsvollen Lernprozessen fähig sind (Sodian, 1995; Stern, 2002). Die weit verbreiteten Auffassungen aus Piagets Theorien, denen zufolge das Denken des Grundschulkindes auf das so genannte 'Konkrete' beschränkt bleibt, gelten als nicht mehr haltbar (Stern & Möller, 2004). Zudem belegen jüngere empirische Studien, dass naturwissenschaftliche Themen im Sachunterricht bei entsprechender Unterrichtsgestaltung mit gutem Lernerfolg behandelt werden können (z.B. Möller, Jonen, Hardy & Stern, 2002; Sodian, Thoermer, Kircher Grygier & Günther, 2002). Nach dem Stand der Forschung sind Grundschulkindern also durchaus in der Lage, naturwissenschaftliche Fragestellungen zu klären und zu durchdenken.

Ziele naturwissenschaftlichen Lernens in der Grundschule. Zur Präzisierung des Bildungsanspruchs von Sachunterricht wurden im Perspektivrahmen für jeden Inhaltsbereich Kompetenzen als Zielvorgaben formuliert. Zur naturwissenschaftlichen Perspektive gehören die folgenden Kompetenzen: „Naturphänomene sachorientiert wahrnehmen, beobachten, benennen und beschreiben“; „Ausgewählte Naturphänomene auf physikalische, chemische und biologische Gesetzmäßigkeiten zurückführen und zwischen Erscheinungen der belebten und unbelebten Natur unterscheiden können“; „Fragehaltungen aufbauen, Probleme identifizieren und Verfahren der Problemlösung anwenden“; „die Regelmäßigkeit der unbelebten Natur auch als Bedingungen für die Existenz der belebten Natur verstehen“ und „Gründe für einen verantwortlichen Umgang mit der Natur erfassen“ (GDSU, 2002, S. 15f.). Die formulierten Kompetenzen zielen über den Erwerb bloßer Erkenntnisse und Fertigkeiten hinaus auf die Förde-

rung von Verstehen (S. 4). Es steht also die Auseinandersetzung mit der „Qualität des Wissens“ im Vordergrund (S. 2). Ein verstehendes Lernen zur Sicherung der Qualität des Wissens und des Lernens wird in der sachunterrichtsdidaktischen Diskussion seit Bestehen des Faches gefordert. In Anlehnung an Wagenschein formuliert Köhnlein 'Verstehen lehren' als übergreifende Aufgabe des Sachunterrichts (Köhnlein, 1991). Der Aufbau nachhaltigen Wissens als Wissen, welches aus eigenen geistigen Aktivitäten entsteht, verlangt 'Akte des Verstehens' bzw. 'Akte der Verinnerlichung'. Verstehen meint das (Nach-)Konstruieren in Gedanken und das Herstellen von Zusammenhängen, das Erfassen von Strukturen und das Zurückführen von Sachverhalten auf etwas Bekanntes oder Einfacheres (Köhnlein, 1991). „Die Schule ist [...] der Ort des Verstehens, des kritischen Denkens“, so formuliert Köhnlein, „ihr sollte das Entstehen der Erkenntnis nicht weniger wichtig sein als das fertige Ergebnis“. Verstehen ist über alle Erfahrungen hinaus wichtig, „sonst bleiben wir stehen bei der alten Kunde“ (Köhnlein, 1991, S. 12f.). Im Vordergrund steht 'formatio', d.h. der geistige Formungsprozess und nicht der Zustand des Gebildetsein (S. 10).

Neben der Förderung eines naturwissenschaftlichen Verständnisses und der Einführung in naturwissenschaftliche Arbeitsweisen gehört es ebenfalls zu den Zielen einer frühen naturwissenschaftlichen Bildung, die Interessen der Lernenden zu berücksichtigen und Interesse und Freude am Nachdenken über naturwissenschaftliche Phänomene zu entwickeln und Selbstvertrauen aufzubauen (vgl. Möller, 2006). Im Perspektivrahmen Sachunterricht wird dieser nicht kognitive Zielbereich etwas versteckt berücksichtigt, indem betont wird, dass der Sachunterricht „Fragen, Interessen und Lernbedürfnisse von Kindern berücksichtigen“ muss, um seiner Aufgabe gerecht zu werden (GDSU, 2002, S. 2). In diesem Zusammenhang wird auch von einer multikriterialen Zielerreichung gesprochen (vgl. Möller, 2006; für eine intensive Auseinandersetzung mit einer multikriterialen Zielerreichung im Sachunterricht siehe Blumberg, 2008).

Ähnliche Zielbereiche werden auch im Rahmen des Konzepts 'scientific literacy' formuliert. Scientific literacy beruht nach Prenzel et al. (2003, S. 146f.) auf:

- Naturwissenschaftlichen Begriffen und Prinzipien (Wissen bzw. Verständnis zentraler naturwissenschaftlicher Begriffe)
- Naturwissenschaftlichen Untersuchungsmethoden und Denkweisen (Verständnis naturwissenschaftlicher Prozesse, grundlegende Fertigkeiten, Denkhaltungen)
- Vorstellungen über die Besonderheit der Naturwissenschaft (Verständnis der 'Nature of science', epistemologische Vorstellungen, Wissen über die Grenzen der Naturwissenschaft)
- Vorstellungen über die Beziehungen zwischen Naturwissenschaft, Technik und Gesellschaft (Verständnis des 'Unternehmens Naturwissenschaft' im sozialen, ökonomischen, ökologischen Kontext)

Neben der Nennung kognitiver Aspekte wird auch im Konzept scientific literacy die Förderung motivationaler Komponenten (wie z.B. Interesse, Aufgeschlossenheit, Einstellungen und Überzeugungen) mit berücksichtigt (vgl. Blumberg, 2008 in Anlehnung an Gräber et al., 2002; Prenzel et al., 2003).

Zusammenfassend lässt sich also festhalten: Ein kognitiv anspruchsvoller Sachunterricht wird gefordert. Aktuelle entwicklungspsychologische Erkenntnisse belegen, dass Grundschüler durch diese Art von Unterricht – eine angemessene Unterrichtsgestaltung vorausgesetzt – nicht überfordert werden. Die im Perspektivrahmen zur Präzisierung des Bildungsanspruchs von Sachunterricht formulierten Kompetenzen als Zielvorgaben zielen über den Erwerb bloßer Kenntnisse und Fertigkeiten hinaus auf die Förderung von Verstehen ab. Will der Sachunterricht seinem Bildungsanspruch gerecht werden und die formulierten Kompetenzen erfolgreich fördern, muss er kognitiv anregende Lernanlässe schaffen oder anders formuliert: Die Lehrperson muss den Lernenden Lerngelegenheiten bieten, welche die Lernenden in ihrem Denken fordern, so dass sie ein qualitatives Verständnis von Inhalten aufbauen, sich im kritischen Gespräch untereinander üben, anwendbares und anschlussfähiges Wissen aufbauen, Stoffe kognitiv durchdringen und zudem Interesse für den Unterrichtsgegenstand entwickeln können. Doch welche Möglichkeiten bieten sich der Lehrperson für eine derartige Unterrichtsgestaltung?

Die Mathematikdidaktik hat auf die Bedeutung fachlicher Unterrichtsqualitätsmerkmale verwiesen. Im Rahmen dieser Arbeit soll daher der Versuch unternommen werden, kognitiv anregende Lehrerverhaltensweisen für den naturwissenschaftsbezogenen Sachunterricht zu konzeptualisieren und anschließend zu erfassen. Die Frage, welche Unterrichtselemente als kognitiv anregend gelten können, ist gemäß den Ausführungen zur kognitiven Aktivierung von fachlichen Kenntnissen abhängig (vgl. auch Kunter et al., 2005). Kognitiv aktivierende bzw. kognitiv anregende Unterrichtselemente sehen demnach in jedem Fach anders aus. Um Elemente kognitiv anregenden Verhaltens für den Sachunterricht bestimmen zu können, ist also ein fachspezifischer Zugang notwendig. In den Folgekapiteln wird deshalb Hinweisen zur Gestaltung von Unterricht aus dem Bereich der Instruktionspsychologie, die für naturwissenschaftliches Lernen von Bedeutung sind, und aus Konzeptionen der Sachunterrichtsdidaktik nachgegangen.

1.2 Hinweise zur Gestaltung von kognitiv anregenden Lerngelegenheiten aus dem Bereich der Instruktionspsychologie

Das übergeordnete Ziel dieses Kapitels ist, Hinweise für kognitiv anregendes Verhalten von Lehrkräften im naturwissenschaftsbezogenen Sachunterricht aus der Instruktionspsychologie herauszuarbeiten. Anzumerken ist an dieser Stelle, dass der Terminus 'Instruktion' nicht als Gegensatz zum Begriff 'Konstruktion' verstanden wird. Der aus dem Englischen stammende Begriff 'Instruktion' bezieht sich gemäß Leutner (2006, S. 261) auf institutionalisierte Situationen (z.B. Unterricht), die zum Zwecke des Lernens und zur Beeinflussung von Lernvorgängen hergestellt werden. Instruktion ist demnach eine

spezifische Form von Intervention. Entsprechend umfasst die 'Instruktionspsychologie' (engl. *instructional psychology*) das Gebiet der Psychologie, welches den Bereich des Lernens unter Einfluss des Lehrens thematisiert. Sie setzt sich somit mit psychologisch begründeten instruktionalen Interventionen auseinander (vgl. Leutner, 2006).

Die Frage nach instruktionaler Förderung ist abhängig von den Vorstellungen darüber, wie Lernende optimal lernen. Ein Blick auf die Entwicklung der Instruktionspsychologie zeigt, dass sich die Vorstellungen von Lernen und die damit verbundenen Bilder der Lernenden und Lehrenden im Verlauf des 20. Jahrhunderts enorm verändert haben. Wurde Lernen zu Beginn des 20. Jahrhunderts noch als rezeptiver Prozess verstanden und wurden Lernende als passive Stimuli-Empfänger angesehen, gelten diese Sichtweisen heute als nicht mehr angemessen. Seit den 1980er Jahren haben sich konstruktivistische Instruktionstheorien durchgesetzt, die zum Ziel haben, den Lernenden zu weitgehend eigenständigem Lernen zu befähigen und die ihren Fokus in der Forschung auf den Lernenden lenken (vgl. z.B. Klauer & Leutner, 2007; Landwehr, 1995; Reinmann & Mandl, 2006; Gerstenmaier & Mandl, 1995).

Für die Gestaltung konstruktivistischer Lernumgebungen im naturwissenschaftlichen Bereich können Hinweise zum einen aus der Conceptual Change-Forschung und zum anderen aus sozial-konstruktivistischen Theorien und Ansätzen zur situierten Kognition abgeleitet werden (vgl. Hardy, 2007). Die Ausführungen dieses Kapitels gliedern sich wie folgt: In einem ersten Abschnitt wird in Form eines kurzen historischen Rückblicks die Entwicklung von behavioristischen Lerntheorien hin zu konstruktivistischen Sichtweisen des Lernens dargelegt (1.2.1). Dieses Unterkapitel dient der Orientierung und kann vom kundigen Leser auch übergangen werden. Der Lesbarkeit wegen beziehen sich die Ausführungen ausschließlich auf Sekundärliteratur. Es folgen die Darlegung der theoretischen Basis zum Bereich Conceptual Change (1.2.2) und daran anschließend die Beschreibung der theoretischen Basis zum sozialen Konstruktivismus und zur situierten Kognition (1.2.3). Eine Zusammenfassung der Inhalte dieses Unterkapitels erfolgt am Ende des Unterkapitels 1.3, um unnötige Doppelungen zu vermeiden.

1.2.1 Lernen aus verschiedenen Perspektiven der Instruktionspsychologie

1.2.1.1 Lernen als Veränderung von Verhalten

Die Tatsache, dass Instruktion oft als Gegenteil von Konstruktion verstanden wird, ist möglicherweise darauf zurückzuführen, dass der Begriff Instruktionspsychologie behavioristischen Positionen entstammt, die zu Beginn des 20. Jahrhunderts dominierten und die Lernen als Veränderung von Verhaltenswahrscheinlichkeiten bzw. als Erwerb neuer Verhaltensweisen und Reaktionen in bestimmten Situationen als Ergebnis von Erfahrungen verstehen. Der Auslöser für ein Verhalten ist dabei immer ein Reiz, auf den der Lernende reagiert. Ein Reiz (Stimulus) aus der Umwelt löst beim Lernenden eine Reaktion in Form eines bestimmten Verhaltens aus (Reiz-Reaktions-Schema). Dementsprechend sind Wissensinhalte aus behavioristischer Sicht Reiz-Reaktions-Verknüpfungen durch äußere Kontrolle

(vgl. Leutner, 2006; Klauer & Leutner, 2007; Mietzel, 2007; Skinner, 1953; Shuell, 1986). Wissen gilt dabei als objektive Größe. Hinter der wahrgenommenen Welt gibt es eine 'wahre' Welt, die der Lernende erfahren kann (vgl. Mietzel, 2007). Behavioristen betrachten den Lernenden als *tabula rasa* und als *black box*, d.h. der Lernende wird als leere 'Wachstafel', also im übertragenen Sinne als unbeschriebenes Blatt gesehen wird, Vorkenntnisse spielen keine Rolle. Zudem gehen behavioristische Psychologen davon aus, dass innerpsychische Vorgänge der Beobachtung nicht zugänglich sind und daher keine Bedeutung zur Erklärung von Verhalten haben (vgl. Leutner, 2006; Klauer & Leutner, 2007; Greeno, Collins & Resnick, 1996). Das Verhalten gilt als die einzige Möglichkeit, objektive Evidenz zu erhalten, daher beschränkt sich das Forschungsinteresse auf die Erforschung von Verhalten. Behavioristische Instruktionstheorien wurden alleine durch objektive Verhaltensbeobachtungen und experimentell kontrollierte Verhaltensänderungen abgesichert (vgl. Klauer & Leutner, 2007; Miller, 2003; Schaub & Zenke, 1995).

1.2.1.2 Lernen als Erwerb von Wissen

Im Zuge der kognitiven Wende in den 1950er und 1960er Jahren verdrängten kognitionspsychologische Theorien den Behaviorismus. Im Kognitivismus rückt die Erforschung des 'Geistes' ins Zentrum des Interesses (vgl. Miller, 2003). Inspiriert durch die Erkenntnis, dass bei Computern Programme die Informationsverarbeitungsprozesse zwischen In- und Output steuern, steht die Erforschung dessen, was zwischen dem Reiz und der Reaktion geschieht, also die Erforschung von Kognition, d.h. von Prozessen der Wahrnehmung, des Denkens und der Erinnerung, im Vordergrund. Der Lernende wird nunmehr als *glass-box* gesehen. Zentral sind Fragen, wie ein Lernender die richtigen Informationen auswählt, wie er diese verarbeitet und wie er sie aus dem Gedächtnis abrufen (vgl. Leutner, 2006; Mietzel, 2007). Mit dem veränderten Forschungsinteresse ändern sich auch die Bilder vom Lernen und vom Lernenden. Kognitionspsychologen verstehen Lernen nicht mehr als mechanisch ablaufenden rezeptiven Prozess des Erwerbs von Reaktionen, sondern als Erwerb von Wissen, d.h. als Aufbau bzw. Veränderung (bereits vorhandener) kognitiver Strukturen (vgl. Leutner, 2006; Klauer & Leutner, 2007). Der Lernende wird nicht mehr als ein reaktiver und passiver Stimuli-Empfänger aufgefasst, der Reiz-Reaktions-Beziehungen zu lernen vermag, sondern dem Lernenden wird die Fähigkeit zum Wissenserwerb zugeschrieben (vgl. Mietzel, 2007). Dies bedeutet, dass der Lernende aktiv Informationen verarbeitet und dadurch Kenntnisse erwirbt, welche die weitere Informationsverarbeitung beeinflussen (vgl. Klauer & Leutner, 2007). Auch wenn der Lernende nun zu einer aktiven Instanz geworden ist und ihm die Fähigkeit zum Wissenserwerb zugesprochen wird, wird Wissen immer noch als objektive Größe betrachtet (vgl. Mietzel, 2007). Dies ändert sich erst im Übergang von der kognitiven zur konstruktivistischen Perspektive, weshalb Terhart (1999) den Informationsverarbeitungsansatz als liberalisierten Behaviorismus bezeichnet. Mietzel (2007) spricht von kognitiven Behavioristen.

1.2.1.3 Lernen als Konstruktion von Wissen

Konstruktivistische Lerntheorien fanden in den 1980er Jahren Eingang in die Instruktionspsychologie. In seinen Ursprüngen ist der Konstruktivismus eine Erkenntnistheorie mit philosophisch-psychologischen Grundlagen ('radikaler' Konstruktivismus). Dieser betrachtet Lernen als Wissenskonstruktion: Die Lernenden konstruieren ihr Wissen selbst auf Basis ihres Vorwissens. Damit wird der Beitrag des Lernenden an seinem Lernprozess stark betont; der Lernende gilt als Konstrukteur seines Wissens. Wird Lernen als individuelle Bedeutungskonstruktion verstanden, so kann es gemäß einer radikal konstruktivistischen Auffassung keine objektive Wirklichkeit, keine 'wahrheitsgetreue' Abbildung der Realität geben (vgl. Hoppe-Graff, 2000).

Als Grundlage für die didaktische Forschung gilt der radikale Konstruktivismus als ungeeignet, da ein gemeinsamer Unterricht gemäß der Vorstellung, es gäbe keine objektive Erkenntnis, nicht möglich ist (vgl. Einsiedler, 2005). In der aktuellen Didaktik haben sich so genannte *gemäßigte* bzw. *moderat-konstruktivistische Sichtweisen vom Lernen* durchgesetzt (vgl. Duit, 1995; Reinmann & Mandl, 2006; Gerstenmeier & Mandl, 1995; Möller, 2001b). Diese übernehmen zwar die konstruktivistische Grundidee, dass Wissen aktiv und individuell konstruiert wird; sie distanzieren sich aber von der radikalen Überbetonung der Annahme, dass menschliches Wissen menschliche Konstruktion und kein echtes Abbild der Realität ist (vgl. Duit & Häußler, 1997). Somit verzichtet der moderate Konstruktivismus auf die für den Lernprozess unwichtige Debatte der Frage, „in welchem Verhältnis die Konstrukte des Verstandes zur 'realen Welt', zur Wirklichkeit 'an sich' stehen“ (Stork 1995, S. 17), und konzentriert sich auf die Fragen, wie Konzepte durch den Lernenden aufgebaut werden und wie Wissensaufbau durch instruktionale Maßnahmen seitens der Lehrenden effektiv unterstützt werden kann.

Dabei berücksichtigen moderne konstruktivistisch orientierte Lerntheorien Sichtweisen des kognitiven Konstruktivismus (Piagets Theorie der Äquilibration), sozial-konstruktivistische Ansätze und Ansätze der situierten Kognition und beschreiben Lernen als einen aktiven Konstruktionsprozess auf der Basis von vorhandenen Wissensstrukturen, der zu einem gewissen Grad selbstgesteuert ist und der von sozialen und situativen Merkmalen beeinflusst wird (vgl. z.B. Gerstenmaier & Mandl, 1995; Reinmann-Rothmeier & Mandl, 1998).

Gemein ist allen Positionen, dass sie Lernen als aktiven und konstruktiven Prozess bezeichnen. Diese beiden Prozessmerkmale von Lernen können deshalb auch als gemeinsame Elemente konstruktivistischer Lerntheorien bezeichnet werden. Was genau unter den einzelnen Prozessmerkmalen zu verstehen ist, wird in den beiden folgenden Abschnitten herausgearbeitet.

1.2.2 Die theoretische Basis zum Bereich Conceptual Change

Im kognitiven Konstruktivismus wird Lernen als Conceptual Change beschrieben (vgl. Möller, 2001b). Conceptual Change-Ansätze stellen die Abhängigkeit der Wissenskonstruktion vom bestehen-

den Vorwissen der Lernenden heraus und stehen damit in Zusammenhang mit konstruktivistisch orientierten Sichtweisen von Lernen. Konstruktivistisch orientierte Lerntheorien betrachten Lernen gemäß den obigen Ausführungen als aktiven und konstruktiven Prozess auf der Basis von Vorerfahrungen. Effektives Lernen erfordert die aktive Beteiligung der Lernenden. Voraussetzung dafür ist, dass die Lernenden zum Lernen motiviert sind. Konstruktion bedeutet, dass Lernen auf der Basis vorhandener Kenntnisse und Fähigkeiten und vor dem Hintergrund individueller Erfahrungen geschieht. Demnach stellen der vorhandene Wissens- und Erfahrungshintergrund die 'Bausteine' dar, aus denen neues Wissen konstruiert wird (vgl. z.B. Reinmann-Rothmeier & Mandl, 1998; Reinmann & Mandl, 2006). Die bestehenden Wissensstrukturen sind demnach entscheidend für Konstruktionsprozesse.

Conceptual Change-Ansätze haben sich speziell zur Beschreibung und Erklärung von naturwissenschaftlichen Lernprozessen als geeignet erwiesen (z.B. Reinmann-Rothmeier & Mandl, 1998; Duit, 1999; Duit & Teagust, 2003; Wandersee, Mintzes & Novak, 1994).

1.2.2.1 Präkonzeptforschung

Ansätze zum Conceptual Change sind auf der Basis von Untersuchungsbefunden zu Schülervorstellungen und Lernschwierigkeiten entstanden.

Hintergründe zum Forschungsinteresse. Forschungen zu Schülervorstellungen haben sich im Bereich der Naturwissenschaftsdidaktik seit den 1970er Jahren wegen der Enttäuschung über den mangelnden Erfolg von wissenschaftsorientierten Curricula, die in den 1960er und 1970er Jahren vor allem in den USA entwickelt wurden, international etabliert (vgl. Duit, 1997). Die Orientierung an den Strukturen der Wissenschaften führte durch die Vernachlässigung individueller Lernvoraussetzungen und kindlicher Denkweisen dazu, dass Unterrichtsinhalte nicht verstanden wurden, so dass das Erlernen naturwissenschaftlicher Konzepte häufig erfolglos blieb (Duit, 1997; Möller, 2007; Wodzinski, 1996)²¹. Vor allem im Bereich der Naturwissenschaften (speziell zu physikalischen Inhalten) wurden (und werden immer noch) zahlreiche Untersuchungen zu Schülervorstellungen durchgeführt, wobei der Schwerpunkt zuerst auf der Untersuchung von Vorstellungen über naturwissenschaftliche Phänomene, Begriffe und Prinzipien lag (vgl. Duit, 1999). Zu Beginn der 1980er Jahre erfolgte durch die zunehmende Einflussnahme konstruktivistischer Sichtweisen des Lernens eine Ausweitung des Forschungsgebietes. Nun wurden nicht nur Schülervorstellungen auf inhaltlicher Ebene, sondern eine Vielzahl von Vorstellungen von Lernenden und Lehrenden (inhaltliche, meta-kognitive und wissenschaftstheoretische Vorstellungen) untersucht (vgl. Duit, 1999; Duit & Häußler, 1997), worauf im Rahmen dieser Arbeit jedoch nicht näher eingegangen wird.

Ein Überfluss an Begrifflichkeiten zur Beschreibung von Schülervorstellungen. Zur Benennung von vorunterrichtlichen Vorstellungen der Lernenden über naturwissenschaftliche Phänomene und Begriffe werden in der Literatur verschiedene Termini verwendet. Um die Vielschichtigkeit der in der Lite-

²¹Vgl. dazu auch Teilkapitel 1.1.4 zu den Ausführungen kognitiv anspruchsvollen Sachunterrichts.

ratur auftretenden Bezeichnungen zu verdeutlichen, sollen in Anlehnung an Wandersee, Mintzes und Novak (1994, S. 178) und Wodzinski (1996, S. 15) einige Begriffe zusammenfasst werden: Neben Schülervorstellungen wird auch von *Alltagsvorstellungen*, *Fehlvorstellungen (misconceptions)* oder *Präkonzepten (preconceptions)* (z.B. Nachtigall, 1986; Hashweh, 1986) gesprochen. Schecker (1985) spricht von *Schülerverständnis*. Im anglo-amerikanischen Raum finden sich weitere Bezeichnungen, wie *alternative frameworks* (Driver), *childrens' science* (Gilbert, Watts & Osborne, 1985), *intuitive science* (Preece, 1984), *facets of knowledge* (Minstrell, 1992), *naive beliefs* (Caramazza, McCloskey & Green, 1981) oder *spontaneous reasoning* (Viennot, 1979). Die Begriffe sind zum Teil in unterschiedliche Theorien eingebunden (vgl. Wodzinski, 1996), die sich in ihrer Bedeutung geringfügig bis teilweise deutlich unterscheiden (vgl. Wandersee, Mintzes & Novak, 1994). Eine umfassende Darstellung zu den unterschiedlichen theoretischen Ansätzen der Begriffe kann an dieser Stelle nicht gegeben werden²². In der vorliegenden Arbeit werden die Begriffe *Schülervorstellungen* und *Präkonzepte* verwendet. Vorstellungen bezeichnen, so Duit (1997, S. 234), ganz allgemein geistige Entwürfe, „die sich ein Mensch von der ihn umgebenden und durch Sinneseindrücke auf ihn wirkenden Welt macht“. Der Begriff der Präkonzepte ist nach Möller (2007, S. 260) am weitesten gefasst und verzichtet auf implizite theoretische Konnotationen. Präkonzepte bezeichnen ganz allgemein die Konzepte der Lernenden bzw. ihr Verständnis *vor* dem Unterricht (vgl. Wodzinski, 1996; Möller, 2007).

Charakteristika von Präkonzepten. Präkonzepte bzw. vorunterrichtliche Schülervorstellungen lassen sich nach deren Stabilität oder Flexibilität unterscheiden: *Deep structures* sind tief in bestehenden Strukturen verankerte, stabile Vorstellungen, die erfolgreich in Alltagsschemata eingebettet und daher gegen Veränderungen sehr resistent sind. Chinn und Brewer (1993) sprechen auch von *entrenched beliefs* (eingewurzelte Überzeugungen), wozu nicht nur Vorstellungen zu Phänomenen und Begriffen, sondern auch Vorstellungen auf der Meta-Ebene (z.B. Überzeugungen aus der Natur und zum Wesen der Naturwissenschaften) gehören (vgl. Duit, 1996). *Current constructions* oder Ad-hoc-Vorstellungen sind oft in einer Befragungssituation spontan erzeugte 'Verlegenheitsäußerungen'; Ad-hoc-Konstruktionen sind keine stabilen Strukturen und leichter zu verändern (vgl. Duit, 1993; Wodzinski, 1996; Niedderer & Schecker, 1992; Möller, 1999). Viennot (1979) spricht in diesem Zusammenhang von *spontaneous reasoning*.

Quelle von Präkonzepten. Schülervorstellungen können sowohl auf der Basis von Alltagserfahrungen als auch durch Unterricht entstehen (vgl. Duit & Häußler, 1997; Häußler et al., 1998). Entstehen Schülervorstellungen durch Alltagserfahrungen, geschieht dies zum Beispiel im Umgang mit Naturphänomenen, durch Sinneserfahrungen, Spracherfahrungen (Alltagssprache), durch Gespräche mit Mitmenschen oder durch Medien. Solche Vorstellungen werden auch als Alltagsvorstellungen bezeichnet (vgl. Duit & Häußler, 1997), in Alltagssituationen erlauben sie durchaus erfolgreiches Handeln. Oft stimmen Alltagsvorstellungen jedoch nicht mit naturwissenschaftlichen Sichtweisen überein und füh-

²²Der interessierte Leser sei z.B. verwiesen auf Niedderer und Schecker (1992).

ren zu Lernschwierigkeiten im Unterricht (Linn & Eylon, 2006). Vor allem in der Alltagssprache gibt es eine Reihe von Weltbildern, die in der Wissenschaft schon lange überholt sind (vgl. Duit & Häußler, 1997). Doch auch Vorstellungen, die durch Unterricht in der Schule entstanden sind, stehen gemäß Untersuchungsergebnissen oft nicht in Übereinstimmung mit wissenschaftlichen Sichtweisen. Bei Linn und Eylon (2006) findet sich ein Überblick über Forschungen zur Entstehung von Schülervorstellungen aus entwicklungspsychologischer Perspektive (z.B. Inhelder & Piaget, 1958), aus soziokultureller Sicht (basierend auf Arbeiten von Vygotsky (z.B. Vygotsky, 1978) und aus kognitiven und konstruktivistischen Perspektiven.

Empirische Befunde zu Schülervorstellungen. Die Befundlage zu Schülervorstellungen ist sehr umfassend. Einen Überblick über internationale Studien zu vorunterrichtlichen Schülervorstellungen geben Pfund und Duit in ihrer Bibliographie „Students' and Teachers' Conceptions and Science Education“ (Stand: März 2007). An dieser Stelle sollen die wichtigsten Ergebnisse aus Untersuchungen zu Vorstellungen von Schülern über Phänomene und naturwissenschaftliche Begriffe berichtet werden.

- *Lernende kommen nicht als unbeschriebene Blätter in den Unterricht.* Sehr bedeutend ist der Befund, dass Lernende (auch Grundschüler) bereits vor dem naturwissenschaftlichen Unterricht über zum Teil tief verwurzelte Vorstellungen und Überzeugungen über naturwissenschaftliche Phänomene und Begriffe verfügen, die oft nicht in Übereinstimmung mit den wissenschaftlichen Sichtweisen, teilweise sogar in starkem Kontrast zu ihnen stehen und zu Lernschwierigkeiten führen (vgl. Wandersee, Mintzes & Novak, 1994; Duit & Treagust, 2003; Duit, 1999).
- *Präkonzepte können stabil sein.* Außerdem ist vielfach belegt, dass Schülervorstellungen gegenüber Veränderung sehr resistent sind (vgl. Wandersee, Mintzes & Novak, 1994; Duit, 1999), durch Unterricht nicht hinreichend verändert werden und auch nach dem Unterricht bestehen bleiben können – vor allem dann, wenn sich die vorunterrichtlichen Vorstellungen im Alltag bewährt haben und in bestimmten Kontexten immer wieder erfolgreich angewendet werden (Duit, 1995, 1996, 1999). Der Überzeugungsgehalt solcher Vorstellungen kann so stark sein, dass er die Wahrnehmung der Lernenden verzerrt (*'confirmation bias'*, Hashweh, 1986): In vielen Studien hat sich herausgestellt, dass sich Schüler häufig nicht durch empirische Evidenz (z.B. Experimente) von der Unzulänglichkeit ihrer (tief-verankerten) Vorstellungen überzeugen lassen (vgl. Duit, 1996). Sie sehen also das, was sie gemäß ihren Präkonzepten sehen wollen (Möller, 2007, S. 260) und erfinden teilweise ad-hoc eine Reihe von Argumenten, weshalb der Ausgang eines Experiments nicht mit den von ihnen vorher genannten Vorstellungen übereinstimmt (vgl. Tiberghien, 1980). Chinn und Brewer (1993) untersuchten die Bedeutung von sich widersprechenden Daten auf den Wissenserwerb und konnten empirische Belege dafür finden, dass Schüler Widersprüche teilweise schlicht ignorieren.

- *Präkonzepte können flexibel sein.* Des Weiteren ist durch viele Untersuchungen belegt, dass Vorstellungen bei Lernenden nebeneinander existieren und für die Erklärung desselben Phänomens bzw. für Situationen, die für den Physiker äquivalent sind, parallel genutzt werden. Zudem scheinen Vorstellungen oft bruchstückhaft, fragmentiert und an Kontexte gebunden zu sein (vgl. diSessa, 1988; Linn & Eylon, 2006, zit. n. Kleickmann, 2008), d.h. das Antwortverhalten des Lernenden kann sich je nach Kontext, in dem ein Phänomen präsentiert wird, verändern (vgl. Wodzinski, 1996).

Die Ergebnisse zeigen, dass aus fachlicher Sicht 'falsche' Alltagsvorstellungen nicht einfach durch die 'richtigen' wissenschaftlichen Vorstellungen ersetzt werden können (vgl. Häußler et al., 1998).

1.2.2.2 Lernen als Conceptual Change

Die Frage, ob und unter welchen Umständen Lernenden der Übergang von einer (Alltags-)Vorstellung zu einem wissenschaftlicheren Konzept gelingt, ist Gegenstand der Conceptual Change-Forschung (vgl. Duit, 1997; Möller, 2007; Duit, 1999).

Zum Begriff. Der Terminus Conceptual Change (zu Deutsch: Konzeptwechsel) steht für Restrukturierungen der Wissensstrukturen (vgl. z.B. Duit & Treagust, 2003; Duit, 1999) bzw. für die Veränderung und Entwicklung von Vorstellungen und Begriffen (vgl. Möller, 2007). Sowohl der deutsche Begriff 'Konzeptwechsel' als auch die englische Bezeichnung 'conceptual change' sind verwirrend, da die Termini implizieren, dass beim Lernen ein Wechsel von nicht belastbaren Konzepten zu adäquateren Konzepten stattfindet bzw. dass Konzepte gegeneinander ausgetauscht werden (vgl. z.B. Duit, 1996, 1997; Möller, 2007). Begrifflichkeiten wie z.B. Konzeptveränderungen scheinen angemessener.

Arten von Conceptual Change. In Anlehnung an die Äquilibrationstheorie Piagets werden in der Literatur zwei Arten von Conceptual Change unterschieden:

- Ist es möglich, neue Konzepte in die bisherigen Wissensstrukturen aufzunehmen, ohne dass weitreichende Veränderungen der kognitiven Strukturen notwendig sind, spricht man von *Assimilation* (vgl. Posner et al., 1982). Der Lernende kommt mit Erweiterungen und kleineren Revisionen seiner Vorstellungen, also mit Wissensausdifferenzierungen (weak knowledge restructuring), aus (vgl. Harrison & Treagust, 2000; Duit & Treagust, 2003). Andere Bezeichnungen für diese Anpassungsprozesse sind in der Literatur z.B. *conceptual capture* (Hewson, 1982), *conceptual enrichment* (Vosniadou, 1994; Carey, 1991) oder *kontinuierliche Lernwege* (Duit, 1995).
- Bei der *Akkommodation* kommt es zu grundlegenden Umstrukturierungen der bestehenden Wissensstrukturen bzw. zum Neuaufbau von Wissen, indem nicht belastbare Konzepte aufgegeben und durch neue adäquatere Konzepte ersetzt werden (vgl. Posner et al., 1982; Duit, 1995). Akkomodative Wissensumstrukturierungen entsprechen im wörtlichen Sinne dem Begriff Conceptual Change (Carey, 1991) und werden auch als *strong/radical restructuring/reconstructi-*

on (Carey, 1985), *conceptual change* (Carey, 1991), *conceptual exchange* (Hewson, 1982), *conceptual revision* (Vosniadou, 1994) oder *diskontinuierliche Lernwege* (Duit, 1995) bezeichnet.

Conceptual Change-Theorien – Klassische Theorien, die Kritik an ihnen und ihre Weiterentwicklung

Seit den 1980er Jahren sind eine Reihe von Conceptual Change-Theorien entwickelt worden, die auf die Veränderung der Präkonzepte der Lernenden abzielen (vgl. Duit & Treagust, 2003). Die meisten von den in den 1980er Jahren entwickelten Ansätzen gehören zu den 'klassischen' Theorien, die ihren Schwerpunkt auf individuelle Konstruktionsprozesse legen (vgl. Duit, 1997). Sehr großen Einfluss auf die Entwicklung von Conceptual Change-Ansätzen hatte der klassische Ansatz von Posner, Strike, Hewson und Gertzog (1982). Gemäß dem Modell sind vier Bedingungen für Conceptual Change erforderlich: Als wesentliche Bedingung für Konzeptwechsel beschreiben Posner et al. (1982, S. 214) in Anlehnung an die Äquilibrationstheorie Piagets, dass die Lernenden mit ihren vorhandenen Vorstellungen unzufrieden sein müssen (*dissatisfactions with existing conceptions*). Des Weiteren müssen die neuen Konzepte dem Lernenden minimal verständlich sein (*intelligibility of a new conception*), die neuen Konzepte müssen ihm glaubwürdig (*initial plausibility of a new conception*) und fruchtbar erscheinen, d.h. sie müssen sich in der Anwendung auf andere Phänomene bewähren (*fruitfulness of a new conception*). Sind diese vier Bedingungen erfüllt, kommt es im Optimalfall zum Wechsel bzw. Austausch von Konzepten. Dabei konzentriert sich das Modell auf Akkomodationsprozesse, also harte Umstrukturierungen bestehender kognitiver Strukturen: „Our work has focused on the kinds of radical conceptual change which we describe as accommodations. How do accommodations take place?“ (Posner et al., 1982, S. 213). Dieser klassische Conceptual Change-Ansatz von Posner und Kollegen (1982) ist in den Folgejahren in verschiedener Hinsicht kritisiert und weiterentwickelt worden (vgl. Duit & Treagust, 2003).

Kritik an der radikalen Konzeptwechselannahme. Stark kritisiert wird an diesem Modell die Konzentration auf akkommodative Prozesse, da die Conceptual Change-Forschung gezeigt hat, dass eher von graduellen und häufig langwierigen Umstrukturierungsprozessen als von harten Konzeptwechseln auszugehen ist (vgl. Möller, 2007). Hewson (1982, S. 64) ergänzt deshalb den Ansatz von Posner und Strike (1982), indem er Konzeptveränderungen als Statuswechsel beschreibt. Ein Konzept kann (1) keinen Status, (2) den Status P (plausible), (3) den Status IP (intelligible, plausible) oder (4) den Status IPF (intelligible, plausible, fruitful) besitzen (Hewson, 1982). Somit wird der graduelle Charakter von Umstrukturierungsprozessen betont; der Status von wissenschaftlichen Konzepten wird Schritt für Schritt erhöht, der Status von Alltagsansichten wird entsprechend reduziert. Dabei können verschiedene Vorstellungen nebeneinander existieren (vgl. Hewson, 1982; Hewson & Hewson, 1992; Duit & Treagust, 2003).

Kritik an der Beschränkung des Modells auf kognitive Aspekte. Ein weiterer zentraler Kritikpunkt am Modell ist, dass der Schwerpunkt zu sehr auf den individuellen Konstruktionsprozessen liegt. Strike und Posner selbst bemängeln ihre ursprüngliche Theorie dahingehend, dass sich die Theorie vorwiegend auf kognitiv-rationale Aspekte bezieht und sozial-affektive Faktoren nicht berücksichtigt (vgl. Strike & Posner, 1992; Duit, 1996, 1999). Pintrich, Marx und Boyle schließen sich in ihrem Artikel „Beyond cold and conceptual change: the role of motivational beliefs and classroom contextual factors in the process of conceptual change“ aus dem Jahr 1993 der Kritik von Strike und Posner (1992) an. Mit Bezug auf zahlreiche empirische Evidenzen und theoretische Beweise, die belegen, dass Lernprozesse nicht kalt und isoliert vollzogen werden (Pintrich, Marx & Boyle, 1993, S. 172), ergänzen sie die von ihnen als 'kalt' bezeichnete klassische Theorie um sozial-affektive Faktoren (motivationale Faktoren und Faktoren des Klassenraumkontextes) und sprechen in diesem Zusammenhang von der 'heißen' Theorie:

„..., we take the constructivist position, that the process of conceptual change is influenced by personal, motivational, social, and historical processes, thereby advocating a hot model of individual conceptual change. ... Accordingly, we believe that the actual content of students' theories and models is influenced by personal, motivational, social, and historical factors, as shown by the existence and persistence of students' misconceptions in science“ (Pintrich, Marx & Boile, 1993, S. 170).

Weiter führen sie aus:

„There needs to be an integration of motivational constructs and an attention to classroom contextual factors in elaborating the model“ (S. 174).

Allgemein haben sozial-konstruktivistische Positionen dazu beigetragen, dass aktuelle Conceptual Change-Ansätze sowohl Aspekte der individuellen als auch der sozialen Konstruktion miteinander verbinden. Zudem spielt die Idee des 'situiereten Lernens' eine zentrale Rolle in neuen Conceptual-Change-Ansätzen.

1.2.3 Die theoretische Basis zum sozialen Konstruktivismus und zur situiereten Kognition

Sozial-konstruktivistische Theorien und Theorien der situiereten Kognition (engl. *situated cognition*; Clancey, 1993) gehören dem so genannten 'neuen' Konstruktivismus an. Die Positionen haben sich aus der Kritik am Kognitivismus entwickelt. Sowohl der Sozialkonstruktivismus als auch die situierete Kognition greifen den konstruktivistischen Grundgedanken auf, demzufolge Lernen als aktiver Konstruktionsprozess der Lernenden auf der Basis ihrer Vorerfahrungen verstanden und Wissen dementsprechend konstruiert wird. Dabei sind diese Ansätze jedoch weniger individuumszentriert und betonen die sozialen Interaktionen des Lernenden und die Situiertheit der Kognition im Kontext (vgl. Klauer, 2006; Mandl, Kopp & Dvorak, 2004; Gerstenmaier & Mandl, 1995; Mandl, Gruber & Renkl, 1997; Möller, 1997).

1.2.3.1 Lernen aus sozial-konstruktivistischer Sicht

Sozial-konstruktivistische Sichtweisen stellen heraus, dass jedes Individuum in einen bestimmten sozialen oder kulturellen Hintergrund eingebunden ist (vgl. Häußler et al., 1998). Diese Eingebundenheit hat zur Folge, dass Wissen aus Wechselbeziehungen zwischen sozialen und individuellen Prozessen resultiert. Interne (kognitive) und externe (Umwelt- und soziale) Faktoren steuern die Wissenskonstruktion (vgl. z.B. Palincsar, 1998; Duschl & Hamilton, 1998). Lernen und andere kognitive Prozesse sind somit keine rein individuellen Vorgänge, sondern beziehen immer auch soziale Prozesse mit ein (vgl. Resnick, 1991; Reinmann-Rothmeier & Mandl, 1998). Lernen erfolgt ko-konstruktiv in sozialen Aushandlungsprozessen. Unterstützt wird der soziale Aufbau von Wissensstrukturen durch kognitive und kulturelle Werkzeuge und Symbolsysteme (vgl. Reusser, 2006). Somit lässt sich mit den Worten Palincsars (1998, S. 348) sagen: „Learning and understanding are inherently social; and cultural activities and tools (ranging from systems to artifacts to language) are regarded as integral to conceptual development“.

Sozial-konstruktivistische Sichtweisen betonen also die Bedeutung sozialer Interaktion und den kulturellen Kontext für das Lernen und knüpfen damit an die sozio-kulturelle Theorie des russischen Psychologen Lev Vygotsky (1978) an.

Die sozio-kulturelle Theorie Vygotskys – grundlegende Annahmen

Vygotsky vertrat die Annahme, dass geistige Prozesse ihre Quelle in sozialen und kulturellen Prozessen haben (vgl. Mietzel, 2007). Kognitive Strukturen und geistige Prozesse werden auf der Grundlage sozialer Kontakte, d.h. durch Interaktionen mit anderen Menschen (Kinder, Erwachsene, Gleichaltrige) von frühester Kindheit an entwickelt. Diese sozialen Interaktionen – davon ging Vygotsky aus – beeinflussen die kognitive Entwicklung nicht nur, sie bringen die Entwicklung kognitiver Strukturen und Denkprozesse erst hervor (vgl. Palincsar, 1998; Mietzel, 2007). Wissen gilt also als sozial-konstruiert und wird von Generation zu Generation weitergegeben (vgl. Bliss, 1996).

Lernen auf interpsychischer und intrapsychischer Ebene. Nach Vygotsky ist die individuelle Entwicklung inklusive höherer mentaler Prozesse sozialen Ursprungs (vgl. Palincsar, 1998). „Every function in the child's cultural development“ – so Vygotsky (1978, S. 57) – „appears twice: first, on the social level, and later, on the individual level; first between people (interpsychological) and then inside the child (intrapsychological)“. Höhere geistige Fähigkeiten und Prozesse sind also zuerst soziale Beziehungen, sie werden zwischen Menschen in gemeinsamen Tätigkeiten ko-konstruiert. Die Lernenden tauschen Ideen, Begriffe oder Denkformen in Interaktionen mit weiterentwickelten Personen aus. Diese sozialen psychologischen Funktionen bzw. die ko-konstruierten Konzepte werden vom Lernenden internalisiert und dann zu individuellen psychologischen Funktionen (vgl. Woolfolk, 2008; Rieber & Carton, 1997, zit. n. Mietzel, 2007). Anders formuliert sind alle höheren geistigen Prozesse verinnerlichte soziale Beziehungen (Valsiner, 1987, S. 67, zit. n. Palincsar, 1998). Lernen kann „als Teil von

Veränderungen in den sozial-kulturellen Beziehungen und Kontexten, an denen der einzelne teil hat“ verstanden werden (Weinert, 1996, S. 9). Vorstellungen, Einstellungen, Begriffe und Werte entstehen also, indem der Mensch sich aneignet, in der Art und Weise zu denken und zu handeln, die die Kultur, in der er lebt, ihm vorgibt und die durch fähigere Menschen übertragen werden (vgl. Woolfolk, 2008). Denken und Lernen sind aus der Sicht Vygotskys somit hochgradig sozial; will man die kognitive Entwicklung verstehen, reicht es also nicht, nur auf den Lernenden zu schauen; ebenso muss man sein Augenmerk auf die soziale Umwelt des Lernenden richten (vgl. Mietzel, 2007).

Kulturelle Entwicklung und kulturelle Werkzeuge. Entscheidend für die kognitive Entwicklung sind nach Vygotsky 'kulturelle Werkzeuge'; sie vermitteln geistige Prozesse (vgl. Vygotsky, 1978; Palincsar, 1998; Mietzel, 2007). Zu kulturellen Werkzeugen gehören sowohl materielle Werkzeuge (z.B. Rechenschieber, Druckmaschinen) als auch symbolische und psychologische Mittel wie die Sprache, mathematische Systeme, Zahlen, Landkarten oder auch Kunstwerke (vgl. Wertsch, 1991). Kulturtechniken erlauben es den Menschen zu kommunizieren, gemeinsam zu denken, Probleme zu lösen und Wissensbestände auszubauen (vgl. Woolfolk, 2008). Der Lernende setzt technische Werkzeuge ein, um auf Objekte der Umgebung einzuwirken; psychologische und symbolische Werkzeuge dienen dazu, auf Vorstellungen und Einstellungen einzuwirken (vgl. Mietzel, 2007). Ein besonders bedeutsames kulturelles Werkzeug stellt nach Vygotsky die Sprache dar. Durch Sprache erfährt der Lernende, wie die Menschen seiner Kultur denken, welche Begriffe sie benutzen und wie sie Wahrgenommenes interpretieren. Die Befähigung des Menschen zur Sprache ermöglicht den Lernenden, Fragen zu stellen und Ideen zu äußern. Dadurch werden sie mit den Werkzeugen ihrer Kultur vertraut und können die Werkzeuge selbst zur Problemlösung nutzen. Die Sprache kann also als eine Kulturtechnik betrachtet werden, die hilft, weitere kulturelle Werkzeuge zu erwerben (vgl. Mietzel, 2007; Woolfolk, 2008).

Die Zone der nächstmöglichen Entwicklung. Um die Beziehung zwischen Entwicklung und Lernen verstehen zu können, muss man gemäß Vygotsky in der kognitiven Entwicklung eines Kindes zwei Bezugsebenen unterscheiden: Die tatsächliche und die potentielle Entwicklungsstufe. Die tatsächliche Entwicklungsstufe bezeichnet, was das Kind alleine leisten kann; die potentielle Entwicklungsstufe beschreibt, was das Kind unter Anleitung eines Erwachsenen oder eines weiter entwickelten Lernenden erreichen kann. Der Bereich der potentiellen Entwicklung stellt die 'Zone der nächstmöglichen Entwicklung' ('zone of proximal development', ZPD) dar. Anders formuliert charakterisiert die Zone der nächstmöglichen Entwicklung den Bereich zwischen der aktuellen Entwicklung, bestimmt durch die Fähigkeit zum selbstständigen Problemlösen, und dem potentiellen Entwicklungsstand, bestimmt durch die Problemlösefähigkeit mittels Unterstützung durch Erwachsene oder durch die Zusammenarbeit mit fähigeren Peers (vgl. Vygotsky, 1978; Rogoff & Wertsch, 1984; Palincsar, 1998, Reiser, 2004; Rogoff, 1990). Kann der Lernende ein Problem nicht ohne Hilfe lösen, befindet er sich in der 'Zone' (vgl. Mietzel, 2007), in dem Bereich, in dem Lernen durch Unterweisung erfolgreich umgesetzt werden kann und zu deutlichen Lernfortschritten führt (vgl. Woolfolk, 2008).

Vygotskys Ausführungen zur Zone der nächstmöglichen Entwicklung basieren auf seiner Kritik an psychologischen Tests und Unterricht (vgl. Rogoff & Wertsch, 1984). Er kritisierte die Fokussierung darauf, wohin sich ein Kind bisher entwickelt hat. Vygotsky vertrat die Meinung, dass die wichtigste Frage in der Entwicklung die nach der *möglichen* Entwicklung ist: „The only good learning is that which is in advance of development“, so Vygotsky (1978, S. 89).

Indem Vygotsky das Entwicklungspotential gegenüber dem erreichten Potential hervorhebt, unterstreicht er die Notwendigkeit von Hilfen für den Entwicklungsprozess (vgl. Bliss, 1996). Damit betont er die Bedeutung von Lernen und Unterricht für die Entwicklung eines Kindes bzw. eines Lernenden. Lernen und Unterricht stellen in den Augen Vygotskys ein wichtiges Werkzeug für die Entwicklung dar; soziale Interaktionen sind dabei wichtige Faktoren des Lernens. Dies ist besonders dann der Fall, wenn der Lernende durch die Hilfestellung anderer gefördert werden kann, indem er Kontakt mit Partnern hat, die in bestimmten Bereichen befähigter sind als der Lernende (vgl. Mietzel, 2007).

Vygotskys sozio-kulturelle Sichtweisen, vor allem seine Erkenntnisse über die Bedeutung der Unterstützung des Lernenden in seinem Erkenntnisprozess, finden sich in verschiedenen Ansätzen und Modellen von Vertretern der sozialen Konstruktion des Wissens und des situierten Lernens wieder. So entwickelte z.B. Barbara Rogoff das Konzept der gelenkten Beteiligung (engl. *guided participation*, Rogoff, 1990). Auch das Modell des kognitiven Lehrlingsystems (engl. *cognitive apprenticeship*) von Collins, Brown und Newman (1989) basiert auf Vygotskys Ideen. Das Konzept der Guided Participation und das Cognitive Apprenticeship-Modell werden zu den bedeutendsten Modellen der Situated Cognition-Bewegung gezählt und werden deshalb im nächsten Abschnitt aufgegriffen.

1.2.3.2 Lernen aus der Sicht der situierten Kognition

Theorien der situierten Kognition weisen deutliche Überschneidungen mit sozial-konstruktivistischen Theorien auf, indem sie Lernen ebenfalls als einen Prozess betrachten, in dem personeninterne und -externe Komponenten in Wechselbeziehung stehen (z.B. Mandl, Gruber & Renkl, 1995). Darüber hinaus betonen Ansätze der situierten Kognition die Kontextgebundenheit von Wissen. Collins, Brown und Duguid (1989, S. 32) beschreiben die Idee dieser Position wie folgt:

„Knowledge is situated, being in part a product of the activity, context, and culture in which it is developed and used. [...] The activity in which knowledge is developed and deployed, it is now argued, is not separable from or ancillary to learning and cognition. Nor is it neutral. Rather, it is an integral part of what is learned. Situations might be said to co-produce knowledge through activity. Learning and cognition, it is now possible to argue, are fundamentally situated“.

Wissen entsteht also als Relation zwischen Personen und Situation und konstituiert sich damit in der Koordination einer Person mit ihren Erfahrungen und einer Situation mit 'begünstigenden' und 'einschränkenden' Einflüssen (vgl. Greeno, Smith & Moore, 1993; auch Renkl, 1996, 2006; siehe dazu auch die folgende Seite).

Zum Situationsbegriff. Was unter dem Situationsbegriff verstanden wird, ist zwar nicht einheitlich definiert; es besteht aber Einigkeit darüber, dass unter 'Situation' sowohl materielle Aspekte (z.B. Bücher, Medien, Räumlichkeiten, 'Werkzeuge') als auch die soziale Umwelt (Interaktionen mit anderen Personen (Community-Gedanke) und historische und kulturelle Kontexte) verstanden werden können (vgl. Mandl, Gruber & Renkl, 1995). Der Community-Gedanke ist insbesondere im Kontext der situierten Kognition zentral. Communities bezeichnen einen Zusammenschluss von Personen mit bestimmten Zielvereinbarungen je nach Art der Community (vgl. Mandl, Kopp & Dvorak, 2004).

Transfer von Wissen. Vertreter der situierten Kognition gehen also davon aus, dass Wissen bzw. Gelerntes spezifisch für die Situation ist, in der es gelernt wird (Anderson, Reger & Simon, 1996, S. 5). Folglich kann es Wissen nicht als abgespeicherte, dekontextualisierte, d.h. allgemeine und anwendbare Repräsentationen geben (vgl. Renkel, 2006). Vielmehr ist Wissen durch die generelle Verbundenheit mit der Situation, in der es erlernt wurde, nur situationsspezifisch aktivierbar und kann allenfalls auf bekannte Kontexte übertragen werden. Ein Transfer ist kaum oder nur in geringem Maße möglich (vgl. Klauer, 2006). Allgemein gilt der Situietheitsgedanke als ein Erklärungsansatz für das *Phänomen des 'trägen Wissens'* (engl. *inert knowledge*). Der Begriff geht auf Whitehead (1929) zurück und beschreibt theoretisches Wissen, welches scheinbar vorhanden ist, in der Praxis bzw. zur Problemlösung jedoch nicht angewendet wird. Es besteht also eine Diskrepanz zwischen Wissen und Handeln. Insbesondere im schulischen Kontext ist zu beobachten, dass theoretisches Wissen, welches im Unterricht erworben wird, nicht in späteren Anwendungssituationen oder außerschulischen Kontexten eingesetzt oder angewandt wird (vgl. Renkl, 1996, 2006).

Instruktionsansätze

Die hier dargestellten Grundgedanken zur Situietheit und zum Transfer von Wissen werden bei den verschiedenen Vertretern der situierten Kognition teilweise sehr unterschiedlich konzipiert. Diese fehlende Einheitlichkeit lässt sich dadurch begründen, dass der Situated Cognition-Bewegung unterschiedliche Strömungen angehören, wie z.B. die kognitive Anthropologie (Rogoff, Lave), die ökologische Psychologie (Greeno) oder die sozio-kognitive Richtung (Resnick) (vgl. Reinmann & Mandl, 2006; Mandl, Gruber & Renkl, 1995; Mandl, Kopp & Dvorak, 2004).

So gilt z.B. das Interesse von James Greeno in seinem Konzept der Situietheit vor allem den Fragen, wie kognitive Prozesse durch Situationscharakteristika beeinflusst werden und wie Wissenstransfer gelingen kann. Situationscharakteristika unterteilt er in 'constraints' (Handlungseinschränkungen) und 'affordances' (Handlungsangebote), die sich auf den Transfer von Wissen auswirken: Wissenstransfer auf eine neue Situation gelingt nur, wenn entweder die Handlungsbedingungen der neuen Situation mit der ursprünglichen Situation übereinstimmen und erforderliche Aktivitäten auch im veränderten Kontext ausgeführt werden können, oder wenn das Handlungsrepertoire auf veränderte Handlungsangebote angepasst und entsprechend geändert werden kann (vgl. Greeno et al., 1993; Reinmann & Mandl, 2006; Mandl, Gruber & Renkl, 1995).

Eine weitere bedeutende Vertreterin der situierten Kognition, Jean Lave, greift insbesondere den Community-Gedanken auf. Auf sie geht der Ansatz 'community of practice' zurück (Lave, 1991). In dieser Community-Form wird „jedes Individuum als Mitglied einer Gemeinschaft betrachtet“, das „in einer 'Legitimate Peripheral Participation' bedeutende Werte, Normen und Inhalte dieser Gemeinschaft“ erwirbt (Mandl, Copp & Dvorak, 2004, S. 40) und sich damit 'enkulturiert' (vgl. Reinmann-Rothmeier & Mandl, 1998). Damit betont Lave, dass lebensnahes Lernen grundsätzlich im alltäglichen Tun und in der Gemeinschaft praktisch tätiger Menschen, also im ständigen sozialen Austausch, stattfindet (vgl. Reinmann & Mandl, 2006). Lave bezeichnet ihre Sichtweise als 'situierte soziale Praxis' (1991, S. 67).

Barbara Rogoff (1990) greift insbesondere Vygotskys Erkenntnisse über die Bedeutung der Unterstützung eines Lernenden in seiner kognitiven Entwicklung durch die Interaktion mit kompetenteren Sozialpartnern auf und entwickelt den Begriff der gelenkten Beteiligung. Der Begriff „impliziert die Hilfestellung anderer Personen beim Lernprozess durch Strukturierung der Problemstellung oder durch Unterstützung der Lernenden beim Erarbeiten von Problemlösungen“ (Mandl, Kopp & Dvorak, 2004, S. 12 in Anlehnung an Rogoff, 1991). „Lernen ist dann besonders erfolgreich, wenn der Lernende, angeleitet und unterstützt von kompetenten und von ihm anerkannten Sozialpartnern, aktiv und eigenverantwortlich an der Bewältigung von zunehmend komplexeren Aufgaben teilnimmt“ (Reinmann & Mandl, 2006, S. 627).

Im Detail soll auf die unterschiedlichen Akzentuierungen und Ausprägungen in den Ansätzen nicht weiter eingegangen werden. Gemein ist allen Ansätzen, dass sie das Wissen einer Gesellschaft als 'geteiltes Wissen' ansehen, welches vom Individuum durch soziale Interaktionen gemeinsam entwickelt und ausgetauscht wird, dass sich das konkrete Denken und Handeln eines Individuums nur vor dem Hintergrund eines konkreten sozialen Kontextes verstehen lässt, dass Wissen und somit Lernen stets situiert, also an inhaltliche und soziale Erfahrungen gebunden ist, und dass Wissen aktiv konstruiert wird (vgl. Reinmann & Mandl, 2006). Gemein ist allen Ansätzen auch, dass sie in Bezug auf die Gestaltung von Lernumgebungen fordern, die Situietheit des Lernens zu berücksichtigen (vgl. Mandl, Gruber & Renkl, 1995).

Instruktionsmodelle

Auf der Basis der dargelegten theoretischen Ansätze zum situierten Lernen sind in der Instruktionspsychologie seit Ende der 1980er Jahre Instruktionsmodelle entwickelt worden, die konkrete Hinweise zur Gestaltung von Lernumgebungen geben. Aus dem angloamerikanischen Raum sind insbesondere drei Ansätze bekannt geworden:

Cognitive Apprenticeship (Collins, Brown & Newman). Der Cognitive Apprenticeship-Ansatz, der 1989 von Collins, Brown und Newman entwickelt wurde, ist ein Beispiel für Lernumgebungen, welche die Wissenskonstruktion mit konkreten Hilfen fördern. Mit Bezug auf grundlegende Gestaltungsprinzipien aus der traditionellen Handwerkslehre (*apprenticeship*), fordern sie, bei der Gestaltung von

Lernumgebung die besonderen Qualitäten einer praxisnahen Anleitung bzw. der äußeren Anleitung zu berücksichtigen (Reinmann-Rothmeier & Mandl, 1998; Reinmann & Mandl, 2006). Grundprinzip der Methode ist, dass die Lernenden über authentische Aktivitäten und soziale Interaktionen in eine Expertenkultur eingeführt werden und dabei inhaltliches und strategisches Wissen erwerben.

Im Mittelpunkt des Unterrichts stehen dementsprechend praxisnahe und in Anwendungskontexte eingebettete Problemstellungen, die von den Lernenden mit Hilfe geeigneter Anleitungen seitens der Lehrenden gelöst werden (Collins et al., 1989; Mandl, Koop & Dvorak, 2004). Der Lernprozess kann dabei durch vier Phasen beschrieben werden (vgl. Collins, 1989, S. 467; Hartinger & Mörtl-Hafizović, 2004, S. 65; Mandl, Kopp & Dvorak, 2004; Reinmann & Mandl, 2006):

- *Modeling (modellhaftes Vorführen)*. Kann der Lernende eine Aufgabe auf der Basis seines Wissensstands nicht lösen, demonstriert der Lehrende als Experte, wie er mit einem Problem oder einer Aufgabe umgeht und löst dies bzw. diese exemplarisch; er dient sozusagen als kognitives Modell.
- *Coaching (Anleiten)*. Nach dem Modellieren beschäftigt sich der Lernende selbst mit einem Problem und wird dabei vom Lehrenden bei Bedarf betreut und gezielt unterstützt.
- *Scaffolding (strukturierende Unterstützung)*. Im Verlauf des Lernprozesses zieht sich der Lehrende aus seiner anleitenden Rolle zurück und bietet den Lernenden mentale Gerüste (scaffolds) als Hilfestellung an, wenn der Lernende eine Aufgabe nicht lösen kann. Insgesamt agiert der Lehrende mit größtmöglicher Zurückhaltung²³.
- *Fading (Zurücknehmen der Unterstützung)*. Der letzte Schritt ist eng verbunden mit dem Scaffolding. Gewinnt der Lernende im Verlauf des Lernprozesses zunehmend Selbstvertrauen und Kontrolle und kann selbstständiger arbeiten, blendet der Lehrende seine Hilfestellungen allmählich gänzlich aus.

Ein Umsetzungsbeispiel des Cognitive Apprenticeship-Ansatzes für die Schule ist die Methode des *Reciprocal Teaching* von Brown und Palincsar (1989). Dieser Ansatz soll insbesondere das Textverstehen leseschwacher Kinder fördern und ihre Lesekompetenz verbessern, indem das Modell den interaktiven und kommunikativen Aspekt des Lernens durch geleitete Diskussionen über einen Text betont (vgl. Mandl, Kopp & Dvorak, 2004). Die dabei verwendeten Strategien sind an das Prinzip des Cognitive Apprenticeship angelehnt und beinhalten das '*Questioning*' (gemeinsames Beantworten einer zentralen, von der Lehrperson gestellten Frage), das '*Clarifying*' (Überprüfen des Textverständnisses und Klären von Fragen), das '*Summarizing*' (Zusammenfassen des Textes unter Fokussierung zentraler Inhaltsaspekte) und das '*Predicting*' (Aufstellen von Vorhersagen über nachfolgende Inhaltsabschnitte zur Verdeutlichung des Textaufbaus). Rosenshine und Meister (1994) konnten die Effektivität des *Reciprocal Teaching* im Rahmen einer Metastudie (Berücksichtigung von 16 Studien) im Ver-

²³Siehe dazu die Ausführungen in Abschnitt 1.4.4.1.

gleich zu traditionellem Unterricht sowohl für untersuchungsspezifische als auch standardisierte Tests nachweisen (vgl. Mandl, Kopp & Dvorak, 2004).

Anchored Instruction (Cognition and Technology Group at Vanderbilt). Der Anchored Instruction-Ansatz der Cognition and Technology Group at Vanderbilt (1997) geht vom Problem des trägen Wissens aus und zielt darauf ab, das Problem fehlender Wissensnutzung zu überwinden, indem den Lernenden so genannte 'Anker' in Form von Erzählungen und Beschreibungen authentischer Problemsituationen geboten werden, welche das Interesse der Lernenden wecken und die den Lernenden die Möglichkeit bieten, eigenständig Probleme zu erkennen, zu definieren und zu lösen:

„Anchored instruction represents our attempt to overcome the inert-knowledge problem by creating meaningful problem-solving environments that permit sustained exploration by students and teachers. ... A related goal is to help students integrate their knowledge by exploring the same situation (anchor) from multiple points of view” (Cognition and Technology Group at Vanderbilt, CTGV, 1997, S. 24).

Besonders charakteristisch für die Gestaltung von 'Ankergeschichten' sind folgende Gestaltungsprinzipien (vgl. Cognition and Technology Group at Vanderbilt, CTGV, 1997; Mandl, Kopp & Dvorak, 2004, Reinmann & Mandl, 2006):

- *Videobasiertes Format* (Präsentation der Problemsituationen per Video zur Veranschaulichung des Problems und zur Förderung eines (situativen) Interesses)
- *Narratives Format* (Integration eines Problems in eine Geschichte bzw. in einen Kontext zur Aktivierung des Vorwissens und zur Unterstützung der Fähigkeit, Zielsetzungen mit allgemein einsetzbaren Strategien zu erreichen)
- *Generatives Lernformat* (Konstruktion der Geschichten oder Probleme in der Art, dass sie anhand bestimmter Regeln gelöst werden können und dadurch die Identifizierung und Differenzierung fördern)
- *Eingebettete Daten* (In die Geschichten bzw. in das Video sind alle notwendigen Daten, welche die Lernenden zur Problemlösung brauchen, eingebaut)
- *Problemkomplexität* (Darstellung der Problemsituation in einer gewissen Komplexität (weder über- noch unterfordernd) zur Förderung der Kompetenz der Lernenden, mit Komplexität umzugehen und zur Förderung des Selbstvertrauens in die eigenen Fähigkeiten)
- *Paare verwandter Abenteuer* (Präsentation ähnlicher bzw. analoger Geschichten zur Erleichterung bzw. zur Unterstützung des Transfers)
- *Fächerübergreifende Elemente* (Herstellen von Verknüpfungen mit anderen Disziplinen zur Erhöhung der Vernetztheit des Wissens und zur Erleichterung des Transfers auf andere Problemsituationen)
- *Kooperation* (Bearbeitung von Problemen in der Gruppe zur Schulung sozialer Kompetenzen und zur Anregung kognitiver Prozesse, die eine tiefere Verarbeitung des Wissens fördern)

Der erste Versuch der Umsetzung des Anchored Instruction-Modells sind die 'Jasper Woodbury Problem Solving Series', zwölf Abenteuergeschichten für den Mathematikunterricht (vgl. z.B. Cognition and Technology Group at Vanderbilt, CTGV, 1994). Die Kernidee ist, mathematische Aufgaben in realistische Geschichten mit 'spektakulären' Aufhängern zu kleiden, um das Interesse der Kinder zu wecken (vgl. Hartinger & Mörtl-Hafizović, 2004, S. 65).

Im Rahmen von Studien zur Verbesserung des mathematischen Verständnisses konnte gezeigt werden, dass der Einsatz von videobasierten Ankergeschichten mit deutlichen Fortschritten beim Lösen mathematischer Probleme und beim Transfer auf andere Kontexte einhergeht (vgl. Bransford et al., 1990; Mandl, Kopp & Dvorak, 2004). Zudem schätzten die Lernenden die Alltagsrelevanz und Nützlichkeit von Mathematik besser ein und äußerten sich gegenüber mathematischen Herausforderungen deutlich positiver als die traditionell unterrichteten Lernenden (vgl. Cognition and Technology Group at Vanderbilt, CTGV, 1993; Mandl, Kopp & Dvorak, 2004).

Cognitive Flexibility (Spiro, Feltovich, Coulson & Anderson). Spiro und Mitarbeiter (1989) betonen in ihrem Cognitive Flexibility-Ansatz vor allem den Aspekt der Situiertheit. Sie fordern für die Gestaltung von Lernumgebungen, dass zu starke Vereinfachungen bzw. Übersimplifizierung (Einschränkungen auf charakteristische Aspekte eines Konzepts) jeglicher Art vermieden werden müssen. Stattdessen sollten die Lernenden von Anfang an mit verschiedenen Situationen konfrontiert werden, damit erworbenes Wissen auf andere Situationen angewendet werden kann (vgl. Reinmann & Mandl, 2006; Gerstenmaier & Mandl, 1995; Mandl, Kopp & Dvorak, 2004). Zur Unterstützung des Aufbaus multipler und flexibel anwendbarer Wissensstrukturen bietet sich das Prinzip 'Landscape Criss-Crossing' (Spiro & Jehng, 1990) an, wobei dasselbe Konzept zu verschiedenen Zeiten in unterschiedlichen Kontexten aus unterschiedlichen Perspektiven unter verschiedenen Zielsetzungen betrachtet wird (vgl. Reinmann & Mandl, 2006; Mandl, Kopp & Dvorak, 2004). Diese umfassende Erarbeitung von Einzelaspekten eines Inhaltsgebietes, also multidirektionales und multiperspektivisches Lernen, soll den flexiblen Transfer von erworbenem Wissen vereinfachen.

Studien zur Überprüfung der Wirksamkeit dieses Ansatzes sind gemäß Mandl, Kopp und Dvorak (2004) bisher nur wenige vorhanden. Die Autoren nennen eine Untersuchung aus der Betriebswirtschaftslehre, in der bei einem computergestützten Lernprogramm multiple Perspektiven und instruktionale Unterstützung zur Selbstreflexion experimentell variiert wurden. Gemäß den Ergebnissen der Studie haben multiple Perspektiven nur dann eine positive Wirkung, wenn sie mit einer instruktionalen Unterstützung einhergehen (S. 22).

Hinweise auf die Notwendigkeit einer instruktionalen Unterstützung finden sich in verschiedenen Untersuchungen zur situierten Kognition. So zeigen Untersuchungsergebnisse, dass es vor allem in komplexen Lernsituationen ohne eine passende Unterstützung zu kognitiver Überforderung und motivationalen Passungsproblemen kommen kann (vgl. Stark, Gruber & Mandl, 1998; Renkl, Gruber & Mandl,

1999). Ein Aufgriff der Diskussion um instruktionale Unterstützungsformen erfolgt in 1.3.3.3 im Rahmen der Darstellung sachunterrichtsdidaktischer Konzeptionen).

1.3 Hinweise zur Gestaltung von kognitiv anregenden Lernumgebungen aus didaktischen Konzeptionen des Sachunterrichts

Nachdem im vorausgegangenen Teilkapitel wesentlichen (allgemeinen und naturwissenschaftsspezifischen) Merkmalen für ein auf Verstehen ausgerichtetes Lernen nachgegangen wurde, werden in diesem Teilkapitel Hinweise zur Gestaltung kognitiv anregender Lernumgebungen aus Konzeptionen des Sachunterrichts herausgearbeitet.

Entsprechend der Forderung nach einem kognitiv anspruchsvollen, das Denken fördernden naturwissenschaftlichen Sachunterricht (vgl. 1.1), gibt es seit den 1970er Jahren verschiedene sachunterrichtsdidaktische Methoden, die ein derartiges Lernen beschreiben und aus denen sich Vorschläge für die Gestaltung eines kognitiv anregenden Unterrichts ableiten lassen. Dazu gehören das genetische Lernen, das problemorientierte Lernen, das entdeckende Lernen und ein anspruchsvolles, Konzeptwechsel-förderndes Lernen. Im Folgenden werden die Charakteristika dieser Lehrformen in der genannten Reihenfolge herausgearbeitet. Dabei liegt der Schwerpunkt auf der Bedeutung der Konzeptionen für den naturwissenschaftlichen Sachunterricht. Das Kapitel abschließend erfolgt eine Zusammenfassung der Inhalte aus diesem und dem vorausgegangenen Kapitel 1.2.

1.3.1 Genetisches Lernen

In der aktuellen deutschsprachigen Mathematik- und Naturwissenschaftsdidaktik erhält das 'genetische Lehren und Lernen' erneut Aufschwung. So wurden genetisch orientierte Lehrformen nach den Diskussionen um TIMSS 1995 Video (vgl. 1.1) z.B. in der Mathematikdidaktik als Methoden zur Förderung kognitiv anregenden Unterrichts vorgeschlagen (z.B. Clausen, 2002).

Im Bereich der Naturwissenschaften gilt der Physikdidaktiker Martin Wagenschein als der bedeutendste Vertreter einer genetischen Didaktik (vgl. z.B. Möller, 2007). Wagenschein entwickelte seine Methode des 'genetischen Lehrens und Lernens' in den 1960er und 1970er Jahren für den Physikunterricht der Sekundarstufe (vgl. Reinmann & Mandl, 2006). Obwohl er sich mit seiner Konzeption nur auf den Unterricht in der Sekundarstufe bezogen hat, ist seine Pädagogik auch für den Sachunterricht in der Grundschule von Bedeutung (vgl. Köhnlein, 1998b). So gelang es Siegfried Thiel bereits Ende der 1960er und Anfang der 1970er Jahre, Wagenscheins Konzeption in praktischen Unterrichtsversuchen auf den Unterricht in der Grundschule zu übertragen²⁴. Bis heute sind Thiels Unterrichtsprotokolle, so Köhnlein, „eine Quelle der Anregungen für die Grundschule geblieben“ (1998b, S. 67). Auch Köhnlein selbst beschäftigt sich mit der Frage nach den Möglichkeiten und der Bedeutung des Prin-

²⁴Die von Thiel veröffentlichten Unterrichtsbeispiele sind z.B. zu finden in Thiel (1984).

zips in der Grundschule. Thiels Untersuchungen sind für ihn ein Beweis dafür, dass sich eine genetisch orientierte Unterrichtskonzeption bereits in der Grundschule als tragfähig erweist und verstehendes Lernen fördert. Somit überträgt er das Prinzip in den 1980er Jahren unter Berücksichtigung kognitionspsychologischer Gedanken auf den naturwissenschaftsbezogenen Sachunterricht in der Grundschule (Köhnlein, 1984; vgl. auch Möller, 2001a; 2007).

1.3.1.1 Genetisches Lehren und Lernen bei Martin Wagenschein

Unter dem Begriff des genetischen Unterrichts versteht Wagenschein eine Lehrform, welche die Dreiheit genetisch – sokratisch – exemplarisch beinhaltet. Zusammen bilden die einzelnen Grundbegriffe die Einheit des 'genetischen Lehrens und Lernens' (vgl. Köhnlein, 1973).

Den *Begriff des Genetischen* benutzt Wagenschein auf zwei Ebenen: Zum einen steht er für die Dreiheit 'genetisch – exemplarisch – sokratisch', also für die didaktische Unterrichtskonzeption. Zum anderen kommt das Element in der Dreiheit selbst vor (vgl. Köhnlein, 1973). Der Begriff 'genetisch' ist griechischen Ursprungs und bedeutet 'ursprünglich werdend, entstehend' (*gignomai*) (Berg, 1995, S. 349) und betrifft also die Entstehung und Entwicklung (*genesis*) (Schaub & Zenke, 1995, S. 156). Wagenschein bezieht sich auf diese Definition; er meint das Werden des menschlichen Geistes, die Entwicklung des Kindes. Die genetische Komponente gehört zur Grundbestimmung des Pädagogischen, denn „Pädagogik hat mit dem Werdenden zu tun: Mit dem werdenden Menschen und – im Unterricht als Didaktik – mit dem Werden des Wissens in ihm“ (Wagenschein, 1992, S. 75). Das Hauptanliegen Wagenscheins ist, Lernende langsam über alltägliche Erfahrungen in die Welt der Wissenschaft zu führen. Diesen Prozess bezeichnet Wagenschein als genetisch, der nur durch einen genetischen Unterricht entstehen kann (vgl. Köhnlein, 1973). Wagenscheins Anliegen beruht auf seiner Sichtweise, dass das Kind bereits auf dem Wege zur Physik ist, man braucht dem Kind also „nur entgegenkommen und es abzuholen da, wo es von sich aus gerade steht“ (Wagenschein, 1976, S. 73). Eine Unterrichtskonzeption ist genetisch, wenn sie dem Unterrichtsgegenstand in seinem Werdegang nachgeht und sein Zustandekommen verfolgt (vgl. Köhnlein, 1973, S. 407). Genetischer Unterricht trennt demnach die Erkenntnisse nicht von ihren Entstehungsprozessen ab, d.h. die Art und Weise „wie man darauf kommen kann“ ist Teil des Wissens selbst (vgl. Köhnlein, 1998a, S. 15). So wissen die Lernenden, wie sie zum Wissen gekommen sind, ihr Verständnis wird kontinuierlich aufgebaut.

Im *Prinzip des Exemplarischen* drückt sich Wagenscheins Forderung nach Stoffbeschränkung sowie die Konzentration des Unterrichts auf das an Beispielen grundlegend erfahrbare Wesentliche aus (vgl. Köhnlein, 1998a, S. 12). Ein genetischer Unterricht sollte und kann sich nach Wagenschein auf exemplarische Themenkreise beschränken (vgl. Wagenschein, 1992, S. 75). Im naturwissenschaftlichen Unterricht wird an wenigen, exemplarisch ausgewählten Phänomenen²⁵ gründlich gearbeitet (vgl. Soostmeyer, 1986, S. 436). Ein ausgewähltes Phänomen steht als Beispiel für einen allgemeinen physikali-

²⁵Nach Wagenschein ist das Phänomen ein Ereignis, welches dem Kind merkwürdig erscheint und einer Erklärung bedarf (vgl. Köhnlein 1998b, S. 82).

schen Sachverhalt und hat als Einstieg die Funktion, den Schüler an den Unterrichtsgegenstand heranzuführen (vgl. Köhnlein, 1998a), und ihm den Bezug zu physikalischen Grundlagen zu eröffnen, so dass er Zugang zur Wissenschaft findet (vgl. Soostmeyer, 1986, S. 436). Das Phänomen enthüllt über den Einzelfall hinaus den Zugang zum Allgemeinen, wofür es als konkretes Beispiel und als das Besondere steht (S. 13). Hat der Schüler den allgemeingültigen Charakter eines Phänomens erkannt, kann er generelle Gesetze formulieren und diese auf andere Gegenstandsbereiche transferieren. Er gewinnt Einblick in den physikalischen Hintergrund und gelangt auf diese Weise zu wirklichem Verstehen. Die exemplarische Komponente des genetischen Lehrens ist also, so Köhnlein (1998a, S. 13), angelegt „auf den Aufstieg vom Besonderen zum Allgemeinen“, das Beispiel wird „zum 'Spiegel des Ganzen' (Wagenschein)“ (Köhnlein, 1996, S. 60). Das Exemplarische hat also entlastende, erschließende und fundierende Funktion zugleich (vgl. Köhnlein, 1973, S. 294).

Die genetische und exemplarische Komponente der Wagenschein'schen Unterrichtsmethode bilden den Rahmen für das *sokratische Element*. Nach der Wahrnehmung des Phänomens als Einstieg in das Unterrichtsthema erfolgt die weitere Erarbeitung wesentlich im Gespräch (vgl. Köhnlein, 1998a, S. 15), denn wirkliches Verstehen, „das Erwachen der geistigen Kräfte“ (Wagenschein, 1992, S. 75) vollzieht sich am wirksamsten in der sprachlichen Auseinandersetzung: Im Gespräch werden Experimente entworfen und durchgeführt, Vermutungen aufgestellt und überprüft, Vorstellungen aufgebaut oder korrigiert (vgl. Köhnlein, 1998a, S. 15). Wagenschein (1976) verweist darauf, dass Bildung zuerst ein individueller Prozess ist, sich aber nie zwischen der Sache und dem Lernenden alleine vollzieht. Jeder lernt in Gruppen, somit kann der Bildungsprozess in der Schule niemals ohne die anderen Schüler in Gang kommen, denn „nichts setzt unsere Gedenkmasse so sehr in Bewegung wie das Gespräch“ (Wagenschein, 1976, S. 124f). Zwar ist nicht das Gespräch alleine „der Ort zum Reifen der Erkenntnis“, denn „das Wesentliche geschieht [...] in der Einsamkeit und Stille“, es „pflügt und sät“ jedoch (Wagenschein, 1976, S. 125). Das Gespräch ist demnach bedeutend für den Unterricht, es hat Verstehensfördernde Funktion (vgl. Köhnlein, 1973, S. 526), indem das eigene Denken in der sprachlichen Auseinandersetzung gefestigt und somit nachvollziehbar wird (vgl. Köhnlein, 1984, S. 210).

Der Ausgangspunkt für das sokratische Gespräch ist die Umgangssprache. Durch den Erwerb der Alltagssprache eignet sich der Lernende bereits ein Vorverständnis über die Dinge und die Struktur der Welt an, denn die Umgangssprache ist untrennbar mit ihren Weltansichten verbunden. Die Schule baut auf dieses Vorverständnis korrigierend auf (vgl. Köhnlein, 1984, S. 209). In diesem Sinne mahnt Wagenschein (1995, S. 137): „Die Muttersprache ist die Sprache des Verstehens, die Fachsprache besiegelt es als die Sprache des Verstandenen“. Hat Physikunterricht die Aufgabe, zu lehren, wie Physik entsteht, hat er auch die Aufgabe, zu lehren, wie ihre Sprache entsteht (vgl. Wagenschein, 1995, S. 153). Somit ist nach Wagenschein die 'Sprache im Physikunterricht' von der 'Sprache der Physik' zu unterscheiden (vgl. Köhnlein, 1973, S. 500). Die Muttersprache führt als Fundament zur Fachsprache (vgl. Wagenschein, 1995, S. 137).

Die Aufgabe des Lehrenden im sokratischen Gespräch ist eine mæutische (vgl. Loska, 1995, S. 180), d.h. die Lehrperson verhält sich nach den Prinzipien des sokratischen Gesprächs (vgl. Einsiedler 1994, S. 206): Sie hält sich selbst in ihren Kommentaren zurück und konzentriert sich auf die im Mittelpunkt des Unterrichtsgeschehens stehenden Schülersaussagen (vgl. Loska, 1995, S. 172). Das Gebot der Zurückhaltung bezieht sich auf das inhaltliche Eingreifen des Lehrenden, nicht jedoch auf das Einsetzen seiner fachlichen Kompetenz (S. 170). Die Fragestellung ist nicht vom Lehrenden an den Lernenden gerichtet, sondern von den Lernenden an die Sache selbst. Der Lehrkraft kommt eine leitende Funktion zu, indem sie durch gezieltes Fragen das Gespräch im Hinblick auf den Gesprächsgegenstand steuert und das gegenseitige Verstehen, eine Grundbedingung für den erfolgreichen Verlauf einer solchen Stunde, fördert.

Diese Art des Unterrichtsgesprächs stellt auch Forderungen an den Lernenden. Für das Gelingen eines sokratischen Gesprächs muss der Lernende zum einen akzeptieren, dass auch seine eigenen, für ihn zweifelsfreien Überzeugungen einer Überprüfung unterzogen werden. Zum anderen muss er fähig sein, auch andere Ansichten, aus seiner Sicht nicht zutreffende Erklärungen, wahrzunehmen. Eine Gesprächsatmosphäre in der jeder Gesprächsteilnehmer ohne Angst seine Ansichten frei äußern kann, ist selbstverständlich (vgl. Loska, 1995, S. 184f).

Zu den Zielen genetischen Lehrens gehören nach Wagenschein die Schulung der Lernenden zum produktiven Denken, z.B. zum Hervorbringen eigener Ideen und Entdeckungen (*produktive Findigkeit*) und zum kritischen Umgang mit eigenen Ideen und Vorschlägen (*kritisches Vermögen*), die Einwurzelung bzw. Verankerung des Wissens im eigenen Verständnis und in der Erfahrung (*'enracinement'*) und den Gebrauch der angeborenen Denk- und Lernlust (Wagenschein, 1992; S. 76-79 und 113; Köhnlein, 1998b, S. 83; Möller, 2007, S. 258)

Das genetische Prinzip bezieht sich also auf eine spezifische Lehrweise in exemplarischen Unterrichtseinheiten (vgl. Köhnlein, 1996). Die beispielhaft stehenden Phänomene lösen beim Kind den Drang zum 'Verstehenwollen', Interesse aus. Das Entwerfen und Durchführen von Experimenten, das Aufstellen und Überprüfen von Vermutungen sowie das Aufbauen und Korrigieren von Vorstellungen zur Lösung des Problems erfolgt im Gespräch (vgl. Köhnlein, 1998a, S. 15). So führt ein Unterricht genetischer Auslegung „mit dem Kind von *der* Sache aus, die *für* das Kind die Sache *ist*“ (Wagenschein, 1984, S. 67) „ohne Bruch vom *Sehen* zum *Verstehen*, vom Nachdenken über alltägliche und auffällige Naturphänomene in die wissenschaftliche Erforschung der Natur“ (S. 65).

1.3.1.2 Die genetische Orientierung des Sachunterrichts bei Walter Köhnlein

Walter Köhnlein betrachtet Wagenscheins Konzeption des genetischen Lehrens als ein Modell, „das in besonderer Weise auf die *Nachhaltigkeit des Aufbaus von Wissen und der tragfähigen Vorstellungen* angelegt ist“ (1998b, S. 67). Wesentliche Elemente eines guten Unterrichts sind seiner Meinung nach in dieser Konzeption angelegt. Somit überträgt er das Prinzip des Genetischen in den 1980er Jahren

auf den naturwissenschaftlichen Bereich des Sachunterrichts in der Grundschule und später (1996) allgemein auf die Didaktik des Sachunterrichts (vgl. Möller, 2001a).

Auch bei Köhnlein gehört die genetische Orientierung zu den Prinzipien seiner Konzeption eines genetisch-exemplarischen Unterrichts (vgl. Köhnlein, 1996, S. 54), wobei eine genetische Konzeption „die Erfahrungen, Vorkenntnisse und Überlegungen der Lernenden konstruktiv aufnimmt und zusammen mit ihnen Wege des Entdeckens sucht, um gemeinsam zu gesichertem und verstandenem Wissen zu kommen“ (S. 61). Indem das genetische Prinzip in den Inhalten des Unterrichts für die Lernenden verständlich fortschreitet und die Herstellung von Zusammenhängen gegeben ist, bezieht sich diese Methode auf den Aufbau des Curriculums: Genetischer Unterricht erfüllt die curriculare Forderung, an geeignete Vorerfahrungen und Interessen der Kinder anzuknüpfen. Die Vorstellungen der Schüler werden also nicht als störend betrachtet und durch formales Schulwissen ersetzt (vgl. Köhnlein, 1998b, S. 81). Kontinuierlich werden die Lernenden von ihren ursprünglichen Interessen und Fragen zu wissenschaftlicherem Denken geführt (vgl. Köhnlein, 1996).

Der Sinn des genetischen Unterrichts besteht nach Köhnlein darin, dass den Lernenden die produktive Bewältigung eines Gegenstandsbereiches ermöglicht wird. Damit meint er das Hervorbringen von Ideen und Strukturen durch eigene Aktivität. Köhnlein ist mit Aebli der Ansicht, dass sich Verstehen aus dem praktischen Handeln und der Wahrnehmung heraus entwickelt. Lernen ist somit ein konstruktiver Aufbauprozess (vgl. Köhnlein, 1996, S. 62), der geleitet wird durch den sozialen Prozess des gemeinsamen Bemühens um ein Problem (vgl. Köhnlein, 1998b, S. 82). Aus diesem Grund bezeichnet Köhnlein seinen genetischen Ansatz als *konstruktiv-genetisch* (Köhnlein, 1996, S. 63; vgl. auch Möller, 2001a).

Folgende Orientierungspunkte fasst Köhnlein für einen genetisch orientierten Sachunterricht zusammen (vgl. Köhnlein 1996, S. 63; Möller, 2001):

- Genetischer Unterricht hat *fundierende Funktion*, indem er, anknüpfend an das Vorverständnis der Lernenden, auf den Aufbau grundlegender und verstandener Vorstellungen gerichtet ist. Diese bilden die Basis für weiteres Lernen.
- Genetischer Unterricht hat *eröffnende Funktion*, da er die alltagsweltlichen Erfahrungen der Kinder berücksichtigt und sich um einen kontinuierlichen Weg von den Vorerfahrungen hin zu wissenschaftlichen Kenntnissen bemüht.
- Da genetischer Unterricht auf das Prinzip der Sachgemäßheit ausgerichtet ist, d.h. auf produktives Suchen und kritisches Prüfen, wird Sach- und Sprachkompetenz zugleich aufgebaut. Somit hat genetischer Unterricht *erschließende Funktion*.
- Genetisches Lehren und Lernen ist auf Verstehen ausgerichtet und hat daher *erkenntnisleitende Funktion*.

- Da eine genetische Konzeption einen forschenden und auf lange Lernzeiten ausgerichteten Unterricht erfordert, hat genetisches Lehren und Lernen schließlich *unterrichtsgestaltende Funktion*.

„Genetischer Unterricht hat die Offenheit und Flexibilität, schon in der Grundschule die jeweils themenspezifischen Meinungen und sachbezogenen Erwartungen konstruktiv aufzunehmen“ (Köhnlein 1998b, S. 81). In der direkten Auseinandersetzung mit Sachverhalten von exemplarischer Bedeutung werden Irrtümer überwunden, weiterführende Ideen hervorgebracht, geprüft und in tragfähiges Wissen überführt. Genetischer Unterricht verbindet somit Sache und Kind im Prozess des Aufbaus von Wissen und Verstehen (S. 76).

1.3.1.3 Lernen in Phänomenkreisen

Spreckelsen entwickelt in Anlehnung an Wagenschein eine eigene Methode. Er prägt den Begriff der 'Phänomenkreise'. Phänomenkreise bestehen aus Simultanexperimenten, d.h. aus Experimenten, die dem gleichen physikalischen Funktionsprinzip angehören. Sie sind also „strukturell identisch“, ihre Interpretationen stützen sich gegenseitig (vgl. Spreckelsen, 1997, S. 125). Den Lernenden werden Möglichkeiten zum Vergleich als Prozess der Analogiebildung gegeben, so dass sie für Phänomene adäquate Erklärungen finden und auf eine physikalische Gesetzmäßigkeit schließen können. Simultanexperimente versuchen also Verstehen auf der Grundlage eines physikalischen Funktionsprinzips anzuregen.

Spreckelsen spricht sich im Gegensatz zu Wagenschein also nicht für die Betrachtung eines einzelnen physikalischen Phänomens aus, sondern für die Gruppierung von Phänomenen, denn das Lösen von Problemen wird nicht durch die Betrachtung eines einzelnen Phänomens angegangen, sondern basiert auf den im Vergleich entstehenden und sich stabilisierenden Interpretationen von mehreren Phänomenen (vgl. Spreckelsen, 1993, S. 328). „Inseln des Verstehens“ gründen sich auf „Einzelkristallen des Verstehens“ (Wagenschein, 1976, S. 195, zit. n. Spreckelsen, 1991, S. 390), d.h. Schüler verstehen (im Sinne Wagenscheins), indem sie „einen Fremden bei näherer Betrachtung als einen verkleideten Bekannten wiedererkennen“ (Wagenschein, 1995, S. 14). Wagenscheins 'Einzelkristalle des Verstehens' entstehen aber nach Spreckelsen nicht durch Einlagerung in eine bereits vorhandene geistige Struktur, sondern durch das Vergleichen mehrerer strukturell identischer Phänomene.

Das unterrichtliche Ziel ist bei Spreckelsen das gleiche wie bei Wagenschein: Den Lernenden soll durch 'Einwurzelung' von Phänomenerklärungen ursprüngliches Verstehen ermöglicht werden (vgl. Spreckelsen, 1993, S. 328). Die Lehrperson nimmt in einem solchen Unterricht die Rolle eines Organisators ein: Sie hält sich zurück, verzichtet auf das Einbringen von fachsprachlichen Korrekturen und Eigendefinitionen. Die Lernenden durchlaufen den Lernprozess selbst (vgl. Hagstedt & Spreckelsen 1986, S. 321).

Es lässt sich festhalten, dass Formen des genetischen Unterrichts, die Konzeption des genetischen Lehrens nach Wagenschein und Köhnlein und die Methode der Phänomenkreise nach Spreckelsen die Vorstellungen der Kinder aufgreifen, falsche Schülervorstellungen in der direkten Auseinandersetzung mit dem Sachverhalt überwinden, weiterführende Ideen aufbauen und in tragfähiges Wissen überführen. Im Zuge einer didaktischen Führung eignen sich die Kinder zunehmend Wissen und Methodenkompetenz an (vgl. Köhnlein, 1998b, S. 81f). Genetischer Unterricht versucht also dort, wo er auf falsche Vorstellungen stößt, Umbauprozesse anzustoßen und mit den Kindern in einer gemeinsamen Arbeit an ihren Vorstellungsstrukturen angemessene Wissensstrukturen aufzubauen. Das wird erreicht, indem die Wissenskonstruktion durch soziale Kontexte unterstützt wird. Durch den sozialen Prozess wird die Aufmerksamkeit immer wieder auf neue Gesichtspunkte gelenkt, es wird intensiv nachgedacht und es werden neue Ideen entwickelt. Die Lernenden durchlaufen den Verstehensprozess also selbst, genetisches Lernen beteiligt die Lernenden aktiv an der Entstehung und Entwicklung ihres Wissens.

1.3.2 Problemorientiertes Lernen im Sachunterricht

Auch dem 'problemorientierten Lernen' wird in der aktuellen deutschsprachigen mathematik- und naturwissenschaftsbezogenen Didaktik wieder vermehrt Aufmerksamkeit entgegengebracht. Insbesondere seit TIMSS und PISA gehört die Entwicklung von Problemlösefähigkeit in der gegenwärtigen Diskussion um die Bildungsstandards für den mathematischen und naturwissenschaftlichen Unterricht (der Sekundarstufe) nach japanischem Vorbild zu einer zentralen Forderung (vgl. Teilkapitel 1.1).

In die sachunterrichtsdidaktische Diskussion gelangte der Begriff des problemorientierten Lernens bereits in den 1970er Jahren. So betont z.B. der Deutsche Bildungsrat im Jahre 1970 (S. 133; vgl. Möller, 2005) die Bedeutung einer Schulung im Problemlösen für den Primarbereich. Seit dieser Zeit gehört die Förderung des Problemlösens zu den erklärten Zielen des Sachunterrichts (vgl. z.B. Einsiedler, 1985).

1.3.2.1 Zum Begriff

'Problemorientiertes Lernen' bezeichnet allgemein ein Lernen mittels realitätsnahen Problemsituationen (vgl. Lankes, 2005), wobei ein Problem für den Lernenden entsteht, wenn ihm die Bewältigung einer Situation oder das Verstehen eines Sachverhalts durch Lücken, Widersprüche, Verwirrungen oder Komplikationen erschwert wird oder nicht möglich ist (Bönsch, 1985; Aebli, 1981). Als Unterrichtskonzeption im Sachunterricht verlangt problemorientiertes Lernen eine Lernumgebung, welche intrinsisch motivierende Probleme aus der Lebenswelt der Lernenden aufbaut. Generell werden derartige Probleme zum Ausgangspunkt des Unterrichts gemacht, im weiteren Unterrichtsverlauf aufgegriffen, in Teilprobleme zerlegt und im Gespräch und in der handelnden Auseinandersetzung mit Materialien von den Lernenden gelöst (vgl. Beinbrech, 2007).

1.3.2.2 Ansätze und Ziele problemorientierten Lernens im Sachunterricht

Es existiert eine Vielzahl von Ansätzen zum Problemlösen. Für die Sachunterrichtsdidaktik sind insbesondere zwei Konzeptionen problemorientierten Lernens von Bedeutung, die Ansätze von Michael Soostmeyer (1978) und Wolfgang Einsiedler (1985) (vgl. Beinbrech, 2003).

'Problemorientierter Sachunterricht' nach Einsiedler

Einsiedler erarbeitet eine Konzeption problemorientierten Lernens sowohl für den naturwissenschaftlichen als auch für den gesellschaftsbezogenen Lernbereich des Sachunterrichts. Anregungen ergeben sich dabei für ihn aus dem problemlösenden Lernen nach Aebli, der in den 1950er und 1960er Jahren problem- und handlungsorientierte Unterrichtsmethoden untersuchte und auch Vorschläge für den Sachunterricht machte (Einsiedler, 2007). Nach Einsiedler war Aebli's Didaktik ein bedeutender Beitrag zur Denkförderung. Gemäß Aebli (1969, S. 68, zit. n. Einsiedler, 2007, S. 392) sind Probleme Antizipationen für die folgenden Handlungen und sind somit ziel- und richtungweisend. Ein Unterricht sollte deshalb durch interessante Problemstellungen geleitet werden und weniger durch die Lehrperson. Mit Bezug auf Aebli schlägt Einsiedler (1994) vor, *Probleme im Sachunterricht aufzugreifen*. Zur Verwirklichung eines problemorientierten Sachunterrichts empfiehlt er folgende hauptsächliche Maßnahmen:

Bei Problemen im Unterricht sollte es sich gemäß Einsiedler im Idealfall um spontan in den Unterricht eingebrachte *Probleme der Lernenden* handeln oder aber um von der Lehrperson für den Unterricht *ausgewählte* und aufbereitete Fragen und *Probleme*, die fachbezogen oder aus der Lebenswelt der Lernenden sein können (1994, S. 203 und 205).

Problemtypen. Für eine intensive inhaltliche Auseinandersetzung mit dem Lerngegenstand im Sachunterricht schlägt Einsiedler in Anlehnung an Aebli (1981) Probleme mit Lücken (Inhalte mit Ablaufstruktur, bei denen in der Problemstellung Teile weggelassen werden) und Probleme mit Widersprüchen (z.B. eine überraschende physikalische Erscheinung) vor (1985, S. 133 und 141; 1994, S. 200). Ebenfalls reizen Probleme mit Kontroversen, welche bestehen, wenn Vorstellungen, Erklärungen, Meinungen etc. von Personen unverträglich sind und versucht wird, eine Übereinstimmung zu erreichen (Einsiedler, 1985, 1994) (siehe dazu auch die Ausführungen in 1.4.1). Wagenschein spricht bei Fragen, welche Selbstverständliches in Frage stellen, vom „Frontalangriff auf das Scheinwissen“ (Wagenschein 1968, S. 74 f., zit. n. Einsiedler, 1994, S. 205).

Probleme über den gesamten Unterrichtsverlauf. Einsiedler legt zudem Wert darauf, dass Probleme nicht nur als isoliert als Einstieg in ein Unterrichtsthema eingesetzt werden. Vielmehr sollte ein Problem im Sinne Wagenscheins als 'weittragende Frage' zu einer intensiven Sachauseinandersetzung führen, indem das Ausgangsproblem weitere Teilprobleme und -fragen mit sich zieht, die wiederum geklärt werden müssen (Einsiedler, 1994, S. 204f.). Probleme „wirken sich [...] nur über einen länge-

ren Zeitraum des Unterrichtsverlaufs aus, wenn das Problem gründlich entwickelt wird“, so Einsiedler (1985, S. 142). Probleme mit Lücken und Kontroversen eignen sich nach Einsiedler besonders für den Einstieg (1985, S. 141). Für den weiteren Unterrichtsverlauf schlägt Einsiedler (1994, S. 204) (halbstrukturierte) Problemreihen vor, worunter er verschiedene Untersuchungen versteht, welche die Problemfragen der Lernenden aufgreifen. Zur Klärung dieser Fragen sollten Alltagsmaterialien eingesetzt werden.

Gespräche über Probleme. Für Gespräche über Probleme schlägt Einsiedler (1994, S. 206) die Methode des sokratischen Gesprächs nach Wagenschein (vgl. 1.3.1) vor, da diese den Lernenden Raum zum Nachdenken gibt und das inhaltliche Gespräch der Lernenden und damit das Verstehen von Zusammenhängen fördert.

Hilfestellungen durch die Lehrkraft. (Problemorientierter) Unterricht bewegt sich nach Einsiedler zwischen den Polen der Instruktion und der Konstruktion (vgl. Einsiedler, 1985, 2007). Das bedeutet, dass die Lernenden nach dem Aufwerfen eines Problems weder sich selbst überlassen werden noch dass „sie inhaltlich und methodisch Schritt für Schritt zur Problemlösung geführt werden“ (Einsiedler, 1985, S. 142). Somit nennt Einsiedler als weitere notwendige methodische Maßnahmen im problemorientierten Unterricht Lernhilfen (1985, S. 142f.). Diese bieten den Lernenden eine Orientierungsgrundlage „innerhalb der sich die Schüler relativ selbstständig bewegen können“ (S. 142). Derartige Lernhilfen fungieren, um es in den Worten Einsiedlers zu sagen, als „Wegmarken“, die *problemorientiert* (z.B. Aufforderung zum genauen Betrachten eines Problems, methodische Hinweisreize) oder *ergebnisorientiert* (inhaltliche Lösungshilfen, z.B. „Denk an die hohe Lage des Wasserbehälters“) sein können. Einsiedler hält vor allem die problemorientierten Lernhilfen für besonders geeignet, da diese die Problemlösefähigkeit der Lernenden fördern (1985, S. 142f.) (vgl. dazu auch Abschnitt 1.4.3).

'Problemorientiertes Lernen im Sachunterricht' nach Soostmeyer

Michael Soostmeyer entwickelt seinen Ansatz problemorientierten Lernens in den 1970er Jahren für den naturwissenschaftlichen Sachunterricht. Dabei richtet er sein Interesse besonders auf das Verfahren des Experimentierens (Soostmeyer, 1977, 1978, 2002; Beinbrech, 2007). Auch Soostmeyer nennt Kriterien für die Gestaltung problemorientierter Lernumgebungen. Die wesentlichen Aspekte sollen im Folgenden erläutert werden.

Einsatz von Problemen über den gesamten Unterrichtsverlauf. Genau wie Einsiedler schlägt Soostmeyer (1978) den Einsatz von Problemen im Unterricht sowohl für den Unterrichtseinstieg als auch für den weiteren Unterrichtsverlauf vor (vgl. Beinbrech, 2003). Dabei misst er Fragen und Problemen der Lernenden ein hohes Gewicht bei: So sollte die Lehrperson zu Beginn des Unterrichts – aber auch während des fortlaufenden Unterrichts – die Erfahrungen, Fragen und Vermutungen der Lernenden zum Lerngegenstand erfragen und mit den Lernenden besprechen und analysieren.

Eigene Handlungserfahrungen mit Alltagsmaterialien. In einem problemorientierten Unterricht sollten die Lernenden Gelegenheiten für selbstständige konkrete Handlungserfahrungen haben. Dabei hält es Soostmeyer für sinnvoll, nicht standardisiertes Material anzubieten (Soostmeyer, 1978, S. 211). Alltagsgegenstände ermöglichen es den Kindern zu erfahren, dass sie mit den Dingen, mit denen sie spielen und die sie aus dem Alltag kennen, Probehandlungen durchführen und Fragen selbst beantworten können (1977, S. 218). Zudem führt das Arbeiten mit nicht standardisiertem Material oft dazu, dass ein Phänomen mit unterschiedlichen Mitteln überprüft wird, was den Schülern vermittelt, dass mehrere experimentelle Lösungen für ein Problem möglich sind (Soostmeyer, 1977, S. 220).

Gemeinsame Gespräche. Neben den Möglichkeiten zu selbstständigen handelnden Auseinandersetzungen betrachtet es Soostmeyer (1978, S. 211) als ebenso wichtig, dass die Lernenden die Gelegenheit bekommen, ihre Beobachtungen, Eindrücke und Erklärungen zu berichten und mit denen der anderen Lernenden zu vergleichen und gemeinsam zu diskutieren. Für Gespräche über den Unterrichtsgegenstand schlägt Soostmeyer genau wie Einsiedler die sokratische Methode nach Wagenschein vor (vgl. Biester & Soostmeyer, 1997).

Phasen der Problemlösung. Soostmeyer unterscheidet beim Problemlösen zwei Phasen, die Problemfindungsphase und die Phase der Problemlösung. Die *Problemfindungsphase* basiert vorwiegend auf entdeckenden Aktivitäten und ist durch methodisch freie und offene Lern- und Denkprozesse gekennzeichnet. Diese Phase dient der Erkennung, Spezifizierung und Definition des Problems. Die sich anschließende *Phase der Problemlösung* ist hauptsächlich mit Aktivitäten des Forschens verbunden, wozu Soostmeyer die Hypothesenbildung und Hypothesenüberprüfung zählt (Soostmeyer, 1977, 1978; vgl. auch Beinbrech, 2003, 2007).

Bezüglich der Hypothesenbildung muss zwischen Vermutungen und Hypothesen unterschieden werden. Vermutungen sind – so Soostmeyer (1977, S. 253) – „solche Aussagen und Fragen der Kinder, die mit Hilfe anthropomorpher, animistischer und teleologischer Konstrukte formuliert sind und somit nicht durch naturwissenschaftliche Verfahren verifiziert oder falsifiziert werden können“. Hypothesen dagegen sind wissenschaftlich überprüfbare Aussagen oder Fragen (S. 254). Zur Überprüfung von Hypothesen schlägt Soostmeyer das Verfahren des *Probierens* vor, das eine Vorform des *Experimentierens* darstellt und zu den Formen eines sinnvollen Problemlösens zählt (vgl. Soostmeyer 1977, S. 275 und 1978, S. 217). Beide Methoden unterscheiden sich in der Hinsicht, dass beim Experimentieren der Grad des Bewusstseins, eine Hypothese entwickelt zu haben, die geprüft werden muss, relativ hoch ist, während dieser Bewusstseinsgrad beim Probieren relativ gering ist, da es um eine dezidierte Versuchsplanung geht. Somit stellt das Probieren eine kognitiv geringere Belastung für den Schüler dar. Eine Vermittlung des komplexen Algorithmus des Experimentierens im Sachunterricht hält Soostmeyer für nicht möglich, da diese eine zu hohe kognitive Belastung für die Grundschul Kinder bedeuten würde.

Lernhilfen. Als weiteres Kriterium für die Gestaltung von Lerngelegenheiten, in denen Probleme geübt und analysiert werden, empfiehlt Soostmeyer (1978) ebenso wie Einsiedler das Anbieten von pro-

blem- und prozessbezogenen Lernhilfen durch die Lehrperson. Auch der Algorithmus des Probierens kann nach Soostmeyer (1977) als Lernhilfe betrachtet werden, mittels der die Lernenden Gegebenes ordnen und strukturieren können und so zu neuen Einsichten kommen.

Ziele problemorientierten Lernens

Zu den Zielsetzungen solchen Lernens im Sachunterricht zählen nach Einsiedler (1985, 1994) und Soostmeyer (1978, 1998) sowohl kognitive Ziele als auch Ziele aus dem motivationalen Bereich: Kognitive Ziele beziehen sich nach Einsiedler auf die 'Problemlösefähigkeit', d.h. auf „... die grundlegende Fähigkeit zur methodischen Bewältigung von Sachproblemen“ (Einsiedler, 1985, S. 126). Auf dieser Zielebene geht es um die „Sicherung des Erwerbs grundlegenden Wissens und elementarer Verfahren“ (Soostmeyer, 2002, S. 250). Motivationale Zielsetzungen beziehen sich auf die Bereitschaft, sich auf Probleme einzulassen und sich mit diesen konstruktiv auseinander zu setzen (Einsiedler, 1985; Soostmeyer, 1998). Einsiedler bezeichnet diesen Zielbereich als 'Problemlösehaltung' (1985, S. 126). Über beide Zieldimensionen hinaus betonen sowohl Einsiedler (1994) als auch Soostmeyer (1998), dass die Förderung einer Fragehaltung bei den Lernenden zum Ziel problemorientierten Lernens gehört.

1.3.3 Entdeckendes Lernen

Wie genetisch orientierte und problemorientierte Lehrkonzeptionen wird auch Versionen des 'entdeckenden Lernens' seit den 1990er Jahren wieder vermehrt Aufmerksamkeit gewidmet. Vor allem im naturwissenschaftlichen Bereich zählen sie international unter Bezeichnungen wie 'inquiry' oder 'scientific discovery' wieder zu den methodischen Standards zur Förderung kognitiver Aktivität, aktiven Wissenserwerbs, selbstgesteuerten Lernens und zur Generierung nutzbaren Wissens (vgl. Neber, 2006).

In den deutschen Sachunterricht gelangte der Begriff des entdeckenden Lernens in den 1970er und 1980er Jahren. Anregungen erhielt die Diskussion um entdeckendes Lernen im deutschen Sachunterricht (vor allem für den naturwissenschaftlich-technischen Bereich) durch das in England als Gegenbewegung auf die amerikanischen geschlossenen Curricula ('Science Curriculum Improvement Study' (SCIS) und 'Science – A Process Approach' (S-APA)) entwickelte „Nuffield Junior Science Projekt“, dessen methodische Grundlage das entdeckende Lernen bildete (Klewitz & Mitzkat, 1973)²⁶.

²⁶Im Nuffield-Projekt wurde auf die Festlegung von Zielen und von Unterrichtsverläufen verzichtet. Es wurden so genannte 'classroom examples' (Unterrichtsprotokolle) vorgelegt, aus denen die Prinzipien des Projekts abgelesen werden können (siehe dazu z.B. Klewitz & Mitzkat, 1973). Kritisiert wurden am Nuffield-Projekt u.a., dass durch die fehlende Festlegung von Unterrichtszielen und -verläufen das erworbene Wissen als zufällig und unsystematisch erschien. Für den deutschen Sachunterricht konnte das Projekt zwar Anregungen geben, durchsetzen konnte es sich in deutschen Schulen jedoch nicht (vgl. Klewitz & Mitzkat, 1973).

1.3.3.1 Zum Begriff

Unter 'entdeckendem Lernen' (engl. *learning by discovery* oder *discovery learning*) werden allgemein Formen des Lernens verstanden, bei denen sich die Lernenden aktiv und (weitgehend) selbstständig mit Problemen und Inhalten auseinandersetzen, eigenständig Lösungswege suchen und finden, auf diese Weise Einsichten in komplexe Sachverhalte und Prinzipien erhalten und somit (durch kognitive Aktivitäten/eigenes Denken) kognitive Strukturen aufbauen (vgl. Reinmann & Mandl, 2006; Hartinger, 2005; Neber, 2006). Im Vordergrund eines so definierten Lernens steht demnach die Nutzung des „eigenen Verstandes“ (Bruner, 1981, S. 16). Problematisch für die Definition des Begriffs des entdeckenden Lernens ist, dass der Terminus zugleich für die Lernprozesse der Lernenden ('Entdeckung als Lernakt'), für eine bestimmte Art von Lehrmethoden und für eine bestimmte Gruppe von Lernzielen des Unterrichts benutzt wird (vgl. Soostmeyer, 1977, S. 133; Hartinger, 2005).

Versteht man entdeckendes Lernen als Lehrmethode, so werden darunter Methoden verstanden, die den Lernenden gestatten, eigenständig Lernerfahrungen zu machen (z.B. selbstständig Probleme lösen, Begriffe erarbeiten, Zusammenhänge erkennen und neue Einsichten gewinnen). Der Lehrende fungiert als Berater und verzichtet auf die Vermittlung von Faktenwissen.

Historischer Rückblick

Als Lehrmethode geht das entdeckende Lernen auf Jerome S. Bruner zurück, der den Begriff des entdeckenden Lernens in den 1960er Jahren prägte. Seine Begründungen für 'entdeckendes Lernen' leitet er hauptsächlich aus kognitionspsychologischen Forschungen ab (Neber, 1981). Bruner selbst führt Untersuchungen zum Erlernen und Verstehen von Begriffen durch²⁷. Aus den Ergebnissen folgert er, dass derartige Lernprozesse aktive Prozesse der lernenden Person sind: Der Erwerb von Begriffen ist durch Hypothesenbildung und -überprüfung gekennzeichnet; dabei sucht der Lernende neue Informationen über die zu erlernenden Inhalte und verarbeitet diese mit Hilfe seines bestehenden Wissens (vgl. Neber, 2006; Hartinger, 2005; Bruner, 1981). Gestützt durch diese Befunde fordert Bruner die Entwicklung von wirksamen Unterrichtsmethoden zur Förderung entdeckenden Lernens und wendet sich gegen die zu der Zeit dominanten behavioristischen Lerntheorien und deren Auswirkungen in geschlossenen Curricula. Damit beeinflusste er die pädagogisch-psychologische Diskussion der 1960er Jahre stark (Reinmann & Mandl, 2006; Neber, 2006; Schaub & Zenke, 1995). Historisch betrachtet stellt das 'entdeckende Lernen' – so Neber (2006, S. 115) – „die erste breite Anwendung kognitions-wissenschaftlicher Forschung zum Lernen dar“. Bruner (1981, S. 17) sieht vier Vorteile, die sich durch entdeckendes Lernen für die Lernenden ergeben können: Den Zuwachs an intellektueller Potenz, den Übergang von extrinsischen zu intrinsischen Belohnungen, das Erlernen heuristischer Methoden des

²⁷Die Ergebnisse seiner Untersuchungen zeigen, dass der Erwerb von Begriffen strategiegesteuert abläuft und dass Strategien von der Entwicklung, der Lernumgebung und vom Vorwissen der Lernenden abhängig sind. Zudem zeigen die Ergebnisse, dass eine selbstständige strategiebasierte Ableitung begrifflichen Wissens zu längerfristigem Behalten und zu einer besseren Nutzbarkeit des erworbenen Wissens führt. Außerdem führt selbstständiges Lernen zu effektiveren Lernstrategien und fördert die Entwicklung intrinsischer Motivation (vgl. Neber, 2006).

Entdeckens, die Hilfe für die Verarbeitung im Gedächtnis. Bruner selbst erarbeitet jedoch keine für schulische Lernumgebungen umsetzbaren Unterrichtsmethoden zur Förderung entdeckenden Lernens. Sein Konzept wurde von Neber (1981) und Klewitz und Mitzkat (1973) weiter entwickelt.

1.3.3.2 Formen und Ziele entdeckenden Lernens

Formen entdeckenden Lernens. In der sachunterrichtsdidaktischen Literatur gibt es das entdeckende Lernen nicht als 'die' Methode, d.h. es gibt m. E. keinen Ansatz, der explizit als 'entdeckend' bezeichnet wird. Vielmehr gibt es viele sachunterrichtsspezifische Ansätze, die Prinzipien entdeckenden Lernens, wie sie im Nuffield Junior Science Projekt beschrieben werden, beinhalten. Folgt man Klewitz et al. (1977) und betrachtet das entdeckende Lernen als Unterrichtsmethode, „die dem Kind gestattet, weitgehend selbstständig Erfahrungen zu machen, Probleme zu lösen und Begriffe zu erarbeiten“ (S. 8), gehören dazu sowohl die bereits vorgestellten genetisch orientierten Methoden und problemorientierte Lehrformen, als auch weitere bisher nicht genannte Ansätze, wie z.B. handlungsbezogenes Lernen oder situiertes Lernen. Insbesondere mit dem problemorientierten Lernen bestehen starke Überschneidungen. Soostmeyer beschreibt das entdeckende Lernen als eine Variante problemorientierten Lernens, wobei der Unterschied zwischen beiden Verfahren darin liegt, dass das entdeckende Lernen auch andere Tätigkeiten als solche, die auf das Problemlösen gerichtet sind, umfassen kann. Spontane Prozesse des Entdeckens sind intuitive, also freie und offene Lern- und Denkprozesse, die auch zufällig auftreten können (vgl. Soostmeyer, 1978, S. 172; Beinbrech, 2003, S. 33). Somit unterscheiden sich das entdeckende und problemorientierte Lernen bezüglich „der Zielgerichtetheit und dem Auffinden von Problemlösungen“ (Beinbrech, 2003, S. 33).

An dieser Stelle soll nicht weiter auf Überschneidungsbereiche zwischen dem entdeckenden Lernen und anderen Unterrichtskonzeptionen eingegangen werden. Im Wesentlichen lässt sich sagen, dass es verschiedene Formen entdeckenden Lernens gibt. Neber (2006) nennt drei Grundformen des entdeckenden Lernens, die einzeln oder kombiniert kognitive Aktivitäten der Lernenden anregen können. Die genannten Varianten sind auch für die Grundschule relevant.

- *Lernen durch Beispiele und Erklärungen:* Ursprünglich ist entdeckendes Lernen als Lernen durch Beispiele oder Fälle bzw. als beispielgestützter Erwerb von Begriffen definiert worden (vgl. Neber, 2006). Dadurch betont diese Form entdeckenden Lernens ein induktives Verfahren des Lernens (vgl. Hartinger, 2005), d.h. die Lernenden finden allgemeingültige Aussagen durch Beispiele selbst (vgl. Soostmeyer, 1998). Nach Soostmeyer stellt ein induktiv-entdeckendes Lernen insbesondere für jüngere Lernende mit wenig Fachwissen eine geeignete Form dar. Diese Version entdeckenden Lernens findet sich insbesondere in Spreckelsens Methode des Lernens durch Phänomenkreise wieder (1.3.1.3). Aber auch im genetischen Lernen nach Wagenschein wird die Bedeutung des Lernens durch Beispiele durch das Moment des Exemplarischen betont (1.3.1.1).

- *Lernen durch Experimentieren.* Diese Form entdeckenden Lernens stellt nach Neber die wohl komplexeste Version entdeckenden Lernens im naturwissenschaftsbezogenen Bereich dar. Ziel ist die Generierung und der Erwerb kausalen Wissens. Dies stellt hohe Anforderungen an die Lernenden (vgl. Neber, 2002, 2006). Vor allem Soostmeyer betont die Bedeutung des Experimentierens in einem problemorientierten Sachunterricht, wie in Kapitel 1.3.2.2 ausführlich dargestellt wurde. Weitere Beispiele für diese Form entdeckenden Lernens für den Sachunterricht finden sich zum Beispiel im Curriculum AKTIF (vgl. Hameyer et al., 1986; Hartinger, 2005).
- *Lernen durch Konfliktinduktion und Konfliktlösung.* Bei dieser Art entdeckenden Lernens stehen Probleme im Zentrum des Interesses, welche bei den Lernenden kognitive Konflikte auslösen und dadurch die Neugierde der Lernenden als motivationale Voraussetzung für weiteres Lernen wecken (vgl. Neber, 2006). Einsiedler geht in Anlehnung an Aebli ausführlich auf Problemtypen ein (vgl. 1.3.2.2), siehe hierzu auch die Ausführungen in Kapitel 1.4.1.

Ziele entdeckenden Lernens. Zu den Zielen entdeckenden Lernens im naturwissenschaftlichen Sachunterricht gehören sowohl der Erwerb von anwendbarem (transferierbarem) Wissen, Kompetenzen und Fertigkeiten anstelle eines Erwerbs bloßen Fachwissens als auch der Erwerb von Lernmotivation, positiven Einstellungen und Haltungen gegenüber naturwissenschaftlich-technischen Sachverhalten (vgl. Soostmeyer, 1977 und 1978, Klewitz & Mitzkat, 1973; Neber, 2002).

1.3.3.3 'Pure discovery' versus 'guided discovery'

Im Rahmen der Diskussion um entdeckendes Lernen ist die Frage nach dem Ausmaß an Unterstützung durch die Lehrkraft (vor allem auch für den Primarbereich) seit jeher von zentraler Bedeutung und trifft auf alle Formen entdeckenden Lernens gleichermaßen zu.

Die 'Lenkungsfrage' geht bis auf die Anfänge des entdeckenden Lernens im pädagogischen Kontext in den 1960er Jahren zurück. So löst Bruner mit seinen Forderungen nach der Entwicklung von wirksamen Unterrichtskonzepten für entdeckendes Lernen in den frühen 1960er Jahren Kontroversen vor allem mit seinem Kontrahenten David P. Ausubel aus (vgl. Neber, 1981). Zwar vertritt Ausubel eine ähnliche Lerntheorie wie Bruner und grenzt sich vom mechanischen Lern- und Verhaltensbegriff des Behaviorismus ab, es geht ihm ebenfalls um eine Verbesserung schulischen Lernens und einer Befähigung der Lernenden zum selbstständigen Denken; Ausubel sieht aber im 'rezeptiven Lernen' – im darstellenden Lehren durch Anleitung und Erklärung – eine Antwort auf die Frage nach der geeigneten Lehrmethode zum Aufbau kognitiver Strukturen (vgl. Terhart, 1989; Edelmann, 2000), da sie „die einzige praktikierbare und effiziente Methode ist, mit der große Wissenssysteme übermittelt werden können. [...] Würde man von den Schülern verlangen, jeden Lehrsatz, der ihnen von ihrem Lehrer übermittelt wird, erst selbst zu prüfen, bevor sie ihn annehmen, könnten sie in keinem Fach jemals über die ersten Grundlagen hinauskommen“ (Ausubel et al., 1981, S. 39).

Die damaligen Auseinandersetzungen über das entdeckende Lernen zwischen Bruner und Ausubel werden bis heute in ähnlicher Weise Form weitergeführt. In der aktuellen Diskussion um entdeckendes Lernen geht es um die Frage nach dem Ausmaß an Unterstützung durch die Lehrkraft, d.h. um die Frage, ob und wenn ja wie viel eine Lehrkraft im Unterricht strukturieren darf. Dabei wird in der Literatur oft zwischen frei-entdeckenden (*'pure discovery learning'*) und gelenkt-entdeckenden Lernformen (*'guided discovery learning'*) unterschieden (vgl. z.B. Mayer, 2004). Neber (2006, S. 119) kritisiert die Beschränkung auf diese 'Entweder-Oder Frage'. Da es in seinen Augen in der Schule keine frei-entdeckenden Lernformen gibt, stellt sich vielmehr die Frage, welche internen Prozesse und Phasen und welche Merkmale der Lernumgebung extern strukturiert werden sollen. In seinen Augen können Formen entdeckenden Lernens „von weitgehend offen und projektähnlich bis hin zu extrem gelenkt realisiert werden“ (S. 118f.).

Inzwischen hat die Unterrichtsforschung gezeigt, dass un gelenkte Entdeckungsprozesse oft ein zu hohes Maß an Selbststeuerung für die Lernenden bedeuten und dementsprechend Verfahren des un gelenkten Entdeckens nicht zu optimalen Lernergebnissen führen (vgl. Schaub & Zenke, 1995). So gibt z.B. Mayer (2004) einen Überblick über Forschungen zu entdeckendem Lernen in den 1960er, 1970er und 1980er Jahren. In allen Fällen waren Formen gelenkt-entdeckenden Lernens erfolgreicher als frei-entdeckendes Lernen.

Speziell für die Grundschule können gemäß Hartinger (2005, S. 389) u.a. die folgenden Befunde als relativ gesichert gelten: Formen gelenkt-entdeckenden Lernens sind in Bezug auf Erwerb und Transfer von Wissen erfolgreicher als vollkommen offene Formen entdeckenden Lernens. Insbesondere leistungsschwächere Lernende profitieren von gelenkt-entdeckenden Unterrichtsformen stärker als von frei-entdeckendem Lernen. Zudem können gezielte Unterstützungen durch die Lehrkraft dazu beitragen, dass verschiedene, für entdeckende Lernformen notwendige Fähigkeiten trainiert werden können (vgl. dazu z.B. auch Einsiedler, 1976; Neber, 2002).

1.3.4 Anspruchsvolles, Konzeptwechsel-förderndes Lernen

Wie in Teilkapitel 1.2 herausgearbeitet wurde, haben auf die Frage nach der Gestaltung kognitiv anregender Lernumgebungen zum Ende des 20. Jahrhunderts vor allem konstruktivistisch-orientierte Lerntheorien großen Einfluss gehabt. Auch für die Grundschule werden Lernumgebungen, die auf der Basis eines moderat-konstruktivistischen Lehrverständnisses gestaltet werden, als geeignet angesehen, um ein anspruchsvolles, Konzeptwechsel-förderndes Lernen zu ermöglichen und den Aufbau von verstandenem und anwendbarem Wissen zu fördern. Im naturwissenschaftsbezogenen Sachunterricht werden konstruktivistisch orientierte Lerntheorien zur Gestaltung kognitiv anregender Lernumgebungen vor allem von den Arbeitsgruppen um Möller und Stern umgesetzt. Gemäß Möller (1999, 2000) umfasst diese Unterrichtskonzeption Formen offenen Unterrichts, Formen gelenkt-entdeckenden Lernens, problemorientierte Unterrichtsgespräche und auch lehrgesteuerte Unterrichtsformen. Wesent-

lich ist, dass die Formen einen individuellen und aktiven Wissenserwerb fördern. Im Zentrum des Interesses steht die „Aktivität des Lernenden im Lernprozess, die subjektive geistige Verarbeitung, die Integration des neu aufgebauten Wissens in bestehendes Wissen und nicht zuletzt die Anwendung und Überprüfung des neu konzipierten Wissens“ (Möller, 1999, S. 134). Dazu werden in Anlehnung an Reinmann und Mandl (2006), Dubs (1995), Reusser und Reusser-Weyeneth (1994) und Bliss (1996) die folgenden Merkmale für einen kognitiv anregenden und Konzeptwechsel-fördernden Sachunterricht vorgeschlagen (vgl. Möller, 1999, 2000; Möller et al., 2006):

Aktives Lernen. Entsprechend der Vorstellung, dass Lernen ein aktiver Konstruktionsprozess ist, sollte die Lernumgebung den Lernenden Handlungsmöglichkeiten bieten, die das Interesse der Lernenden berücksichtigen und fördern. Wie auch in den anderen sachunterrichtsdidaktischen Konzeptionen werden dazu z.B. problemhaltige Aufgaben, forschendes Experimentieren, selbsttätiges Handeln oder offene Fragestellungen empfohlen. Zudem sollten die Lehrperson den Lernenden Möglichkeiten zur Selbst- und Mitbestimmung geben.

Konstruktives Lernen. Die Lernenden sollten die Möglichkeit haben, eigene Wissenskonstruktionen vornehmen und eigenen Denkwege gehen zu können. Die Adäquatheit der Konzepte der Lernenden sollte in einer intensiven sachlichen Auseinandersetzung mit den Lerninhalten überprüft werden. Dabei sollte die Lehrkraft auf reine Instruktion verzichten. Sollte dies nicht möglich sein, müssen die Instruktionen für die Schüler nachvollziehbar sein und aktiv konstruiert werden.

Situatives Lernen. Um Lernen als situativen Prozess umsetzen zu können, werden authentische Aufgaben in relevanten und sinnvollen Kontexten empfohlen.

Soziales und kooperatives Lernen. Neben Gelegenheiten zur individuellen Konstruktion betont diese Konzeption auch die Bedeutung der Förderung der Kommunikation und Kooperation unter der Schülern.

Selbstgesteuertes und unterstütztes Lernen. Selbstbestimmtheit des Lernens ist ein langfristiges Ziel von Unterricht. Oft müssen die Voraussetzungen für selbstbestimmtes Lernen beim Lernenden aber erst geschaffen werden. Um einen aktiven Aufbau von Schemata gewährleisten zu können, sind Steuerungs- und Strukturierungshilfen bzw. allgemeine und individuelle Hilfen durch die Lehrkraft notwendig.

Ein derartiger Unterricht trägt dazu bei, dass neue Wissensinhalte unter Berücksichtigung des Vorwissens und auf der Basis eigener Erfahrungen entwickelt und somit verstanden werden. Neben kognitiven Lernzielen werden auch Ziele aus dem sozial-affektiven Bereich verfolgt.

1.3.5 Zusammenfassung der Inhalte zu Hinweisen zur Gestaltung kognitiv anregender Lernumgebungen aus der Instruktionspsychologie und aus Konzeptionen des Sachunterrichts

In Teilkapitel 1.2 wurde herausgearbeitet, dass sich in der aktuellen Instruktionspsychologie konstruktivistisch orientierte Lerntheorien durchgesetzt haben. Dabei gibt es nicht eine einzige konstruktivistische Lerntheorie. So genannte moderat-konstruktivistische Lerntheorien berücksichtigen Sichtweisen des kognitiven Konstruktivismus, sozial-konstruktivistische Ansätze und Ansätze der situierten Kognition. In Teilkapitel 1.3 wurden bedeutsame sachunterrichtsdidaktische Konzeptionen vorgestellt, die zum Ziel haben, ein kognitiv anspruchsvolles, auf Verstehen ausgerichtetes Lernen zu fördern. Im Rahmen dieser Zusammenfassung werden die wesentlichen Aspekte der Inhalte der Kapitel nochmals zusammengefasst und Hinweise für die Gestaltung kognitiv anregenden Unterrichts herausgearbeitet.

Lernen als aktiver Konstruktionsprozess auf der Basis von Vorerfahrungen

Instruktionspsychologie. Konstruktivistisch orientierten Lerntheorien ist der Grundgedanke gemein, dass Wissen vom Lernenden aktiv unter Rückgriff auf seine bestehenden Wissensstrukturen konstruiert wird. Voraussetzung dafür ist, dass die Lernenden für das Lernen motiviert sind. Die Präkonzeptforschung hat herausgestellt, dass Lernende speziell im Bereich der Naturwissenschaften über Vorstellungen von Phänomenen und Begriffen verfügen, die oft nicht mit den im Unterricht zu erlernenden Konzepten übereinstimmen und Restrukturierungen (Conceptual Change) bedürfen (vgl. Duit & Treagust, 2003; Vosniadou et al., 2001). Auf der Basis dieser Befunde sind Conceptual Change-Theorien entwickelt worden, die auf die Veränderung der Präkonzepte abzielen. Bezüglich der Veränderung von Wissensstrukturen werden in Anlehnung an Piaget Wissensumstrukturierungen und Wissensausdifferenzierungen unterschieden. Eine direkte Vermittlung im Sinne eines behavioristischen Lernverständnisses ist nicht möglich (vgl. z.B. Reinmann & Mandl, 2006).

Konzeptionen des Sachunterrichts. Auch Konzeptionen des Sachunterrichts gehen von einem aktiven und konstruktiven Lernen aus und betonen die Berücksichtigung von motivierenden und interessanten Inhalten. Explizit Bezug wird auf eigenständige Denkprozesse als Voraussetzung für wirkliches Verstehen im konstruktiv-genetischen Ansatz nach Köhnlein genommen, wie auch im entdeckenden Lernen in der Tradition Bruners und im Rahmen eines anspruchsvollen Konzeptwechsel-fördernden Lernens nach Möller et al. (2006). Ebenfalls ist in sachunterrichtsdidaktischen Konzeptionen ein konstruktiver Umgang mit Schülervorstellungen fest verankert. Insbesondere Wagenschein und Köhnlein haben auf die Bedeutung vorunterrichtlicher Vorstellungen für ein auf Verstehen ausgerichtetes Lernen hingewiesen. Im genetischen Lernen werden die Vorstellungen der Schüler in der Alltagssprache aufgenommen und zum Gegenstand gemeinsamen Nachdenkens gemacht, denn „der Lernprozess ist dem Verstehen des Schülers verpflichtet“, so formuliert Köhnlein (1973, S. 182). Er knüpft an das ursprüngliche Vorverständnis der Lernenden an und somit auch an ihre Umgangssprache (vgl. auch

Möller, 2001a, 2007). Wissen ist auch in Konzeptionen des Sachunterrichts nicht direkt vermittelbar (Möller, 1999, 2000).

Hinweise für das Lehren. Zur Unterstützung eines aktiven und konstruktiven Lernens ergeben sich für das Lehren in der Schule die folgenden Notwendigkeiten: (1) die Lernenden für das Lernen zu motivieren, (2) die kognitive Aktivität der Lernenden zu fördern und (3) die Vorerfahrungen der Lernenden zu berücksichtigen.

Soziales, kooperatives und unterstütztes Lernen

Instruktionspsychologie. Sozial-konstruktivistische Theorien unterstreichen die Bedeutung sozialer Aushandlungsprozesse für das Lernen, d.h. sie betonen, dass Lernen nicht nur eine individuelle Konstruktion ist und berücksichtigen die gegenseitige Beeinflussung der Mitglieder einer Gesellschaft beim Aufbau von Wissen (vgl. z.B. Palincsar, 1998; Duschl & Hamilton, 1998). Zudem greifen sozial-konstruktivistische Ansätze Vygotskys Ansicht, dass der Lernende kognitive Aufbauprozesse unter Anregung und Anleitung von 'unterstützenden anderen' vollzieht, auf und betonen die Bedeutung von Unterstützungshilfen für die Lernenden (vgl. Palincsar, 1998; Reiser, 2004; Bliss et al., 1996). Damit gestehen sie dem Lehrenden die Möglichkeit zu, die Lernenden anzuleiten und deren kognitiven Prozesse zu beeinflussen (vgl. Mietzel, 2007).

Konzeptionen des Sachunterrichts. Konzeptionen des Sachunterrichts greifen vor allem das sokratische Gespräch nach Wagenschein auf, welches gemäß Köhnlein hinsichtlich seiner Verstehen fördernden Funktion durch die Interaktionen untereinander ein wichtiges methodisches Instrument im Unterricht darstellt (vgl. Köhnlein 1973, S. 526). Die Bedeutung von Lernhilfen bzw. Unterstützung durch die Lehrkraft wird in allen sachunterrichtsdidaktischen Konzeptionen gefordert. So betonen beispielsweise Wagenschein und Köhnlein die Bedeutung der sokratischen Gesprächsführung; im sokratischen Gespräch unterstützt der Lehrende die Schüler in ihrem Erkenntnisprozess durch sparsame Impulse und Nachfragen. Einsiedler und Soostmeyer verweisen auf die Bedeutung prozessbezogener Lernhilfen.

Die Notwendigkeit von Strukturierungshilfen im Unterricht wird auch durch Studienbefunde unterstützt, die belegen, dass Lernen ohne instruktionale Unterstützung ineffektiv ist und zu viel Selbststeuerung dazu führen kann, dass die Lernenden ihre Lernprozesse nicht reflektieren und die angestrebten wissenschaftlichen Konzepte nicht erreichen (Mayer, 2004).

Hinweise für das Lehren. Für das Lehren in der Schule ergibt sich daraus also die Aufgabe, (1) soziales Lernen bzw. soziale Austauschprozesse zu ermöglichen. Zudem sollten den Lernenden (2) bei Bedarf Lernhilfen angeboten werden.

Situiertes Lernen

Instruktionspsychologie. Ansätze der situierten Kognition heben die Situationsbindung von Wissen hervor. Sie betonen die Kontextgebundenheit von Wissen und weisen auf das Problem des Transfers von Wissen hin. Instruktionsmodelle der situierten Kognition fordern in Bezug auf die Gestaltung von Lernumgebungen, die Situietheit des Lernens zu berücksichtigen (vgl. z.B. Brown, Collins & Duguid, 1989; Greeno et al., 1993).

Konzeptionen des Sachunterrichts. Eine Entsprechung findet sich in sachunterrichtsdidaktischen Konzeptionen. Wagenschein geht es um die 'Einwurzelung' des Erlernten in die Alltagswelt zwecks Aufbau eines tiefgehenden und anwendungsorientierten Verständnisses (vgl. Spreckelsen, 1991a, S. 257; vgl. auch Möller, 2007).

Hinweise für das Lehren. Für das Lehren in der Schule ergibt sich daraus, dass Lernen in situativen Kontexten erfolgen muss. Der Anwendungsbezug des Gelernten muss gesichert werden (vgl. Reinmann & Mandl, 2006).

Insgesamt liefern die beiden Theoriebereiche also ähnliche Hinweise für die Gestaltung kognitiv anregender Lernumgebungen.

1.4 Folgerungen für das Lehren kognitiv anregenden Sachunterrichts

Auf der Basis der Inhalte aus den Teilkapiteln 1.2. und 1.3 konnten die folgenden Hinweise zur Unterstützung eines aktiven, konstruktiven, sozialen und in situative Kontexte eingebetteten Lernens herausgearbeitet werden: Die Lehrperson hat die Aufgabe, Lernumgebungen zu gestalten, in denen die Lernenden für das Lernen motiviert werden, in denen die Lernenden Freiräume für individuelle Lernprozesse und für soziale Austauschprozesse haben, in denen sie aber auch Unterstützung durch die Lehrperson erfahren und in denen kontextbezogen gelernt wird. Es stellt sich die Frage, wie Lehrkräfte diesen Anforderungen gerecht werden können. Dieses Teilkapitel fokussiert die Umsetzung der in den letzten beiden Kapiteln herausgearbeiteten Gestaltungsprinzipien. Wie kann die Lehrperson die Lernenden motivieren? Welche Möglichkeiten bieten sich der Lehrperson, mit den Vorstellungen der Lernenden umzugehen? Wie kann die Lehrperson in der Zone der nächstmöglichen Entwicklung unterrichten, d.h. wie kann die Lehrperson den Lernenden adäquat Hilfestellung geben, ohne die Lernenden dabei zu belehren und ohne sie 'allein' zu lassen? Welche Verhaltensweisen gilt es zu vermeiden? Diesen Fragen soll im Folgenden nachgegangen werden.

1.4.1 Verhaltensweisen von Lehrkräften zur Anregung aktiven und konstruktiven Lernens

Motivation zur Auseinandersetzung mit den Lerngegenständen schaffen

Die Lernenden zum Lernen zu motivieren heißt für die Lehrperson konkret, ausreichend Anreizwerte zu schaffen, damit sich die Lernenden mit den Lerninhalten beschäftigen. Umso mehr es der Lehrkraft gelingt, den Lernenden die Unterrichtsinhalte als etwas Interessantes zu vermitteln, für das es sich lohnt, sich zu engagieren, desto höher wird die Motivation der Lernenden sein (vgl. Helmke, 2004). Was kann die Lehrperson im Unterricht tun, um die Lernmotivation anzuregen?

Förderung intrinsischer Motivation. Im Idealfall sollte der Lernende intrinsisch motiviert werden (persönliches Interesse am Lerngegenstand, Lust am Lernen). Zur Förderung intrinsischer Motivation werden in der Literatur verschiedene Ansatzpunkte gegeben. Einige lassen sich aus der Selbstbestimmungstheorie von Deci und Ryan (1993) ableiten. Unter Berücksichtigung dieser Theorie ist es notwendig, dass die Lehrperson mit ihrem Verhalten im Unterricht die Bedürfnisse der Lernenden nach Zugehörigkeit, Autonomie und Kompetenz berücksichtigt:

- Dem *Bedürfnis* der Lernenden *nach sozialer Eingebundenheit* kann die Lehrperson nachkommen, indem sie den Selbstwert der Lernenden durch Zuneigung, Anerkennung und Wertschätzung für persönliche Leistungen (nicht nur kognitive Leistung, z.B. auch Ausdauer oder Kreativität) bewahrt. Zudem sollte die Lehrperson sich bemühen, ein positives Klima im Unterricht zu entwickeln, im Sinne der 'Norm zur Gegenseitigkeit', d.h. die Schüler erfahren untereinander Anerkennung, lernen gemeinsam und helfen sich gegenseitig. Ein negatives Unterrichtsklima (ein Unterricht, in dem die Lernenden Angst haben) entspricht den Bedürfnissen der Lernenden nicht (vgl. Maehr & Anderman, 1993; Mietzel, 2007 in Anlehnung an Brophy, 2004; McCombs, 1998; Goodenow, 1993; Rakoczy & Pauli, 2006; Kunter, 2005).
- Dem *Bedürfnis nach Autonomie* nachzukommen, gestaltet sich schon schwieriger, da ein zu hohes Maß an Selbstbestimmung vor allem bei jüngeren Lernenden Überforderung bedeuten kann (vgl. Abschnitt 1.4.3). Das Erleben von Selbstbestimmung sollte sich nicht durch äußere Anleitung ausschließen. Unterstützen kann die Lehrperson die Selbstständigkeit der Lernenden z.B., indem sie ihnen Entscheidungsfreiheiten in Form von eingeschränkter Wahlfreiheit einräumt (zwischen verschiedenen Aufgaben auswählen können, Aufgabenmenge selbst bestimmen können) und ihnen Verantwortung überträgt (vgl. Maehr & Anderman, 1993; Woolfolk, 2008; Kunter, 2005; Rakoczy & Pauli, 2006).
- Für das *Kompetenzempfinden* ist es notwendig, dass die Lehrperson Aufgaben mit passendem Anforderungsniveau (Prinzip 'nicht zu schwer und nicht zu leicht') auswählt (vgl. Deci & Ryan, 1993).

Förderung von Interesse. Intrinsisch motiviertes Verhalten wird auch als interessenbestimmt beschrieben, d.h. intrinsische Lernmotivation steht in engem Zusammenhang zu den Interessen der Lernenden (Reinmann-Rothmeier & Mandl, 1998). Gemäß der Interessentheorie nach Krapp (z.B. Krapp, 1992, 2001, 2002) wird Interesse als Beziehung zwischen Person und Gegenstand verstanden. Aus dieser Kenntnis ergibt sich für den Unterricht die Notwendigkeit, auf Interessen der Lernenden einzugehen, an ihre Interessen anzuknüpfen und ihr Interesse am Lerngegenstand zu wecken (Reinmann-Rothmeier & Mandl, 1998). Neben einem persönlichen (individuellen) Interesse des Lernenden am Unterrichtsgegenstand kommt aber auch einem situationsspezifischen Interesse eine wesentliche Bedeutung im Unterricht zu. Es kann nicht immer auf die individuellen Interessen eines jeden Schülers eingegangen werden, situationsspezifisches Interesse kann durch Umgebungsfaktoren bei allen Lernenden geschaffen werden (vgl. Hidi & Anderson, 1992; Krapp, 1992; Reinmann-Rothmeier & Mandl, 1998). Zudem stehen beide Formen von Interesse in Beziehung zueinander; ein situationales Interesse kann sich im Verlauf des Unterrichts zu einem persönlichen Interesse entwickeln (Krapp, 1998; Mietzel, 2007). Wie kann die Lehrperson nun auf das Interesse der Lernenden eingehen und daran anknüpfen?

- Über die oben genannten Möglichkeiten zur Förderung intrinsischer Motivation werden auch Aspekte des persönlichen (individuellen) Interesses berücksichtigt (vgl. Reinmann-Rothmeier & Mandl, 1998).
- Aus dem Blickwinkel der Interessentheorie besteht eine weitere Möglichkeit zur Förderung des individuellen Interesses der Lernenden am Lerngegenstand darin, ihnen die Bedeutung oder (persönliche) Sinnhaftigkeit bzw. Relevanz des Unterrichtsgegenstands durch alltagsnahe Aufgaben zu vermitteln (vgl. Maehr & Anderman, 1993; Rakoczy & Pauli, 2006). Alltagsnahe Aufgaben werden in Teilkapitel 1.3 genannten Konzeptionen des Sachunterrichts erwähnt. Weitere Ausführungen dazu folgen unter 1.4.4.
- Zumindest situationales Interesse lässt sich durch die Interessantheit von Texten oder Geschichten erreichen (vgl. Hidi, 1990; Mietzel, 2007). Für den naturwissenschaftsbezogenen Sachunterricht der Grundschule eignet sich z.B. der Umgang mit (Sach-)Problemen und Phänomenen, welche die Lernenden neugierig machen. Auch die im Rahmen der Darstellung der Konzeptionen des Sachunterrichts genannten Handlungserfahrungen (z.B. Experimentieren) und die offenen Fragestellungen lassen sich hier ansiedeln (Einsiedler, 1994; Soostmeyer, 1978; Möller, 1999, 2000).

Wichtig bei der Förderung von Interesse ist, dass das Interesse der Lernenden nicht nur geweckt, sondern auch aufrecht erhalten wird (vgl. Printrich, 2003).

Gelegenheiten zu eigenen Wissenskonstruktionen schaffen und individuelle Lernwege zulassen

Um den konstruktiven Charakter des Lernens zu berücksichtigen, ist es, wie bereits erwähnt, wichtig, dass die Lernumgebung auf die Vorerfahrungen der Lernenden ausgerichtet ist und die Lernenden die

Gelegenheit zu eigenen Wissenskonstruktionen bekommen und dabei individuelle Lernwege gehen können (vgl. z.B. Dubs, 1995; Reinmann-Rothmeier & Mandl, 1998; Möller, 2000). Dabei wird gleichzeitig dem Aspekt der Selbststeuerung Rechnung getragen. Welche Möglichkeiten ergeben sich für die Lehrperson, eigene Wissenskonstruktionen und Denkwege im Unterricht zu fördern?

Wissenskonstruktionen der Lernenden fördern. Dazu gehört, dass die Lehrperson den Schülern Spielraum eröffnet und Zeit lässt, ihre eigenen Ideen, Vorstellungen und Fragen verbal zu äußern oder schriftlich bzw. zeichnerisch darzustellen (vgl. Tytler, 2002b; Möller et al., 2006; Einsiedler, 1994). Umsetzen kann die Lehrperson dies z.B., indem sie die Lernenden aufmuntert und ermutigt, Ideen zu äußern, indem sie Ideen oder Fragen der Lernenden akzeptiert, aufgreift und weiter verfolgt. Auf den Umgang speziell mit Präkonzepten wird in Abschnitt 1.4.1.4 eingegangen.

Individuelle Lernwege ermöglichen. Damit die Schüler wirklich eigene Denkwege gehen können, ist es wichtig, dass sie für die Darstellung ihrer Vorstellungen und Fragen ihre eigenen Formulierungen anstelle von naturwissenschaftlicher Fachsprache benutzen (vgl. Tytler, 2002b; Ausführungen zum genetischen Lernen in 1.3.1), denn die Alltagssprache kann oft eine Quelle für (falsche) Vorstellungen sein (vgl. Duit, 1999). Die Lehrperson kann das Nutzen eigener Formulierungen durch die Lernenden fördern, indem sie die Lernenden auffordert, ihre eigenen Formulierungen zu nutzen, und indem sie selbst die Formulierungen der Kinder aufgreift („Kannst du das mit deinen eigenen Worten sagen?“; ein Schüler äußert die folgende Idee „Das Wasser geht in die Luft“; die Lehrperson wiederholt die Äußerung wie folgt „Peter sagt, das Wasser geht in die Luft“ anstelle von „Peter sagt, das Wasser verdunstet“).

Zu eigenen Denkwegen gehören auch (Denk-)Umwege und Fehler (vgl. Möller, 1999). Die Auseinandersetzung mit Fehlern wirkt Verständnis fördernd und trägt zu einer besseren Konstruktion von verstandenem Wissen bei (vgl. Dubs, 1995 in Anlehnung an Smith et al., 1993). Will die Lehrperson fördern, dass die Schüler Fragen stellen, Probleme finden und auch lösen, sollte die Lehrperson einen positiven Umgang mit Fehlern pflegen, d.h. sie sollte Fehler nicht vertuschen, a priori vermeiden oder negativ bewerten. Vielmehr sollte sie Fehler als integralen Bestandteil des Lernprozesses sehen und diese im Unterricht zulassen, gedankliche Umwege der Schüler akzeptieren und auch falsche Ideen aufnehmen (Reusser, 1999; Oser et al., 1999; Spychiger et al., 1999; Heinze, 2004; Möller, 1999; Einsiedler, 1994; Soostmeyer, 1978). Zudem sollte sie den Lernenden vermitteln, dass Fehler nicht schlimm sind, sondern im Gegenteil auch positive und wichtige Funktionen im Lernprozess haben (vgl. Rakoczy & Pauli, 2006; Clausen, Reusser & Klieme, 2003). Letzteres findet sich z.B. in Lehreräußerungen, wie: „Jede Idee ist wichtig, auch wenn sie falsch ist“; „Es ist wichtig, dass jeder seine Idee äußert, auch wenn sie vielleicht falsch ist“, ...).

Transmission von Wissen vermeiden

Da traditionell behavioristische Positionen vom Lernen aus heutiger Sicht als nicht sehr lernwirksam gelten (vgl. Unterkapitel 1.2.1), gilt die Gestaltung von lehrerzentrierten Lerngelegenheiten, die auf einem traditionellen Lernverständnis basieren, als nicht angemessen. Derartiger Unterricht zeichnet sich durch ein systematisch geplantes und kleinschrittiges Vorgehen aus. Die Lernenden reproduzieren die Wissensinhalte und besitzen am Ende des 'Wissenstransports' die vermittelten Wissensinhalte in ähnlicher Form wie die Lehrkraft (vgl. Reinmann & Mandl, 2006). Es herrschen die Prinzipien des Instruierens und des fehlerlosen Lernens. Fehler werden der Reproduktionsleistung als abträglich betrachtet (vgl. Spychiger et al., 1999; Landwehr, 1995). Wie zeichnet sich Lehrerverhalten in einem derart eng geführten Unterricht konkret aus?

- Die Lehrkraft übernimmt im Unterricht die Rolle des 'didactic leaders', dessen Aufgabe darin besteht, Wissensinhalte als 'Fertigprodukte' zu präsentieren und zu erklären (vgl. Reinmann & Mandl, 2006), d.h. die Lehrperson informiert mehr als dass die Schüler selbst Lösungen finden können. Umsetzen kann die Lehrperson dies, indem sie z.B. selbst Erklärungen für offene Fragen, Probleme oder durchgeführte Experimente gibt. Typisch ist auch, dass die Lehrperson Merksätze formuliert, die von den Kindern gelernt werden. An dieser Stelle sei angemerkt, dass das Aufschreiben von Merksätzen nicht allgemein schlecht ist. Sie müssen allerdings gemeinsam mit den Kindern entwickelt und formuliert und nicht von der Lehrperson vorgegeben werden („*Wie können wir das für alle aufschreiben?*“ anstelle von „*Ich schreibe euch mal auf, wie ihr euch das merken könnt*“).
- Zudem zeichnet sich transmissives Verhalten der Lehrperson dadurch aus, dass sie die Lernenden anleitet und ihre Lernprozesse überwacht (vgl. Reinmann & Mandl, 2006), indem sie z.B. Denkwege und Formulierungen der Kinder unterbindet oder falsche Vermutungen per Autorität korrigiert. Dadurch nimmt die Lehrperson den Kindern Raum für individuelle Denk- und Lernwege im Unterricht.

Zum Umgang mit Präkonzepten und Anregung von Conceptual Change

Um einen auf Conceptual Change ausgerichteten Unterricht durchführen zu können, ist es, wie bereits erläutert, bedeutsam, den Einfluss, den die bestehenden Wissensstrukturen auf Lernprozesse haben, zu berücksichtigen (vgl. z.B. Duit, 1997). In einem naturwissenschaftsbezogenen Unterricht ist es somit unerlässlich, in einem ersten Schritt die Vorstellungen bzw. die Präkonzepte der Lernenden zu den Unterrichtsinhalten aufzugreifen (vgl. Möller et al., 2006). Dabei meint 'aufgreifen' nicht, dass die Lehrperson die Präkonzepte zu Beginn des Unterrichts oberflächlich (im Sinne eines Sammelns von Ideen) erfragt. Wesentlich ist, dass die Lehrperson über die bloße Abfrage der Konzepte hinaus während des gesamten Unterrichtsverlauf mit und an den Konzepten der Kinder arbeitet. Realisieren kann sie dies, indem sie z.B. im Verlauf der Unterrichtsstunde immer wieder auf die Präkonzepte der Kinder zurückkommt oder indem sie die Kinder auffordert, über ihre Präkonzepte nachzudenken.

Zur Unterstützung der Veränderung von Vorstellungen werden je nach Art der Konzepte Strategien zur Förderung von Wissensumstrukturierungen und Wissensausdifferenzierungen empfohlen:

- *Strategien zur Förderung von Wissensumstrukturierungen.* Sind tief greifende Wissensumstrukturierungen erforderlich, werden *Konfliktstrategien* zur Auslösung eines kognitiven Konflikts bei den Lernenden vorgeschlagen. Kognitive Konflikte kann die Lehrperson auslösen, indem sie die Lernenden mit objektiver Evidenz konfrontiert, z.B. mit einem Experiment zur Demonstration eines unerwarteten Phänomens („*Die kleine, leichte Stecknadel geht unter, der große, schwere Holzklötz schwimmt, wie kommt das?*“). Der Konflikt besteht in der Erwartung der Schüler, die sie aufgrund ihrer vorhandenen Konzepte haben, und der wahrgenommenen Realität. Die Kinder erfahren die Grenzen ihrer bisherigen Konzepte. Der Konflikt löst Unzufriedenheit mit den bestehenden Wissensstrukturen aus (vgl. Wodzinski, 1996; Scott, Asoko & Driver, 1992; Limón, 2001; Möller, 1999). Einsiedler (1985) spricht in diesem Zusammenhang in Anlehnung an Aebli von Problemen mit Widersprüchen (vgl. 1.3.2.2). Die Anwendung von Konfliktstrategien ist nicht problemlos. Empirisch belegt ist, dass das Erkennen eines kognitiven Konflikts nicht unbedingt zu den erwünschten physikalischen Vorstellungen führt. Insbesondere jüngere Schüler verteidigen ihre naiven Vorstellungen oft auch bei Widersprüchen (vgl. die Ausführungen zur Resistenz von Präkonzepten in 1.2.2.1). Der Erfolg der Konfrontationsstrategie hängt davon ab, ob die Kinder bereit sind, einen Konflikt wahrzunehmen (vgl. Wodzinski, 1996; Möller, 1999). Damit die Kinder den Widerspruch auch wahrnehmen bzw. annehmen, ist es wichtig, die Kinder den Ausgang eines Versuchs erst vermuten zu lassen bevor er durchgeführt wird (Newton, 2001; Tytler, 2002a) („*Was meinst du was passiert, wenn ich die Stecknadel/den Holzklötz in das Wasser tauche?*“).

Eine weitere Möglichkeit, kognitive Konflikte auszulösen, besteht darin, verschiedene nicht vereinbare Sichtweisen eines Problems darzustellen (vgl. z.B. Wodzinski, 1996; Scott, Asoko & Driver, 1992; Einsiedler, 1994; Soostmeyer, 1978). Möglich ist z.B., konträre Schülersagen gegenüber zu stellen („*Peter sagt, ein Schiff schwimmt, weil es einen Motor hat. Paul sagt ein Schiff schwimmt, weil es vom Wasser getragen wird. Was stimmt denn?*“; „*Lena sagt das Wasser verschwindet aus der Schale, weil es in die Schale eingesaugt wird. Laura sagt, das Wasser verschwindet in die Luft. Wer hat denn Recht?*“).

- *Strategien zur Förderung von Wissensausdifferenzierungen.* Zur Unterstützung von Wissensausdifferenzierungen werden Anknüpfungs- und Umdeutungsstrategien empfohlen. Bei der *Anknüpfungsstrategie* (auch Brückenstrategie) werden Vorstellungen von Schülern, die sich in einem gewissen Bereich mit physikalischen Konzepten überschneiden, als so genannte Anker- vorstellungen (anchoring conceptions) genutzt, von denen aus wissenschaftliche Konzepte aufgebaut werden (vgl. z.B. Wodzinski, 1996; Clement & Brown, 1993; Scott, Asoko & Driver, 1992; Newton 2001; Möller, 1999). Hier kann auch das Lernen über Analogien eingeordnet werden (z.B. Spreckelsens Phänomenkreise). *Umdeutungsstrategien* können eingesetzt wer-

den, um Schülern zu erklären, dass sie im Prinzip etwas richtig verstanden haben, dass es aber in der Wissenschaft anderes bezeichnet wird (vgl. Jung, 1986; Häußler et al., 1998) („*Wir sagen, das Wasser geht in die Luft; Wissenschaftler sagen, Wasser verdunstet*“).

Für die Veränderung von Wissensstrukturen (Conceptual Change) reicht es gemäß Posner et al. (1982) nicht aus, dass die Lernenden mit ihren bestehenden kognitiven Strukturen unzufrieden sind (vgl. dazu 1.2.2.2). Ebenso müssen die neuen Konzepte verständlich, plausibel und fruchtbar sein. Der Einsatz von geeigneten Experimenten kann dazu beitragen, dass den Lernenden die neuen Konzepte einsichtig gemacht werden und dass sie von den neuen Konzepten überzeugt sind (Möller, 2004b). Die 'Fruchtbarkeit' neuer Konzepte erweist sich für Kinder, wenn sie Situationen aus dem Alltag mit den neuen Konzepten erklären und verstehen können (vgl. z.B. Möller, 2004b). Nähere Ausführungen zum Aspekt der Fruchtbarkeit folgen bei den Erläuterungen zum anwendungsbezogenen Lernen in 1.4.4.

1.4.2 Verhaltensweisen von Lehrkräften zur Umsetzung sozialen und kooperativen Lernens

Gemeinsame Diskussionen

Auf der Basis eines Verständnisses von Lernen als Ergebnis sozialer Aushandlungsprozesse gewinnen Merkmale kooperativen Lernens an Bedeutung. Wie bereits erwähnt, sollte die Kommunikation unter den Schülern gefördert werden. Partner- und Gruppenarbeiten und gemeinsame Diskussionen bieten den Lernenden gute Möglichkeiten zu sozialen Aushandlungsprozessen mit dem Lerngegenstand. Vor allem dem gemeinsamen Diskutieren kommt ein besonderer Stellenwert zu. In gemeinsamen Diskussionen über den Unterrichtsinhalt sind die Lernenden einerseits dazu aufgefordert, ihre eigenen Vorstellungen zu Naturphänomenen und Begriffen zu versprachlichen, zu präsentieren und gegenüber den Mitlernenden zu vertreten, andererseits aber auch deren Standpunkte und Argumente aufzunehmen und Meinungsdivergenzen bzw. Widersprüche sachlich zu diskutieren. Diese sprachliche Interaktion fördert die Lernenden in ihrer kognitiven bzw. konzeptuellen Entwicklung (vgl. Palincsar, 1998; Duschl & Hamilton, 1998; Reinmann-Rothmeier & Mandl, 1998). Die Lehrperson sollte also im Unterricht die gemeinsame Diskussion fördern. Dies kann sie tun, indem sie die Kinder z.B. auffordert, ihre Ideen untereinander zu diskutieren, ihre Ideen gegenüber Mitschülern zu begründen oder auch die Ideen der Mitschüler kritisch zu überprüfen. Zudem sollte die Lehrperson selbst mit den Kindern über Probleme und Lösungswege diskutieren.

1.4.3 Verhaltensweisen von Lehrkräften zur Umsetzung unterstützten Lernens

Anleitung und Unterstützung der Lernenden

Gemäß den vorausgegangenen Ausführungen muss die Lehrperson den Lernenden insbesondere bei kognitiv anspruchsvollen (naturwissenschaftlichen) Unterrichtsinhalten nicht nur Raum für eigene (Denk-)Aktivitäten geben, sondern auch geeignete Unterstützungshilfen anbieten, um die Lernenden

beim Umstrukturieren ihres Wissens und beim Aufbau adäquater Konzepte zu unterstützen und um einer Überforderung der Lernenden durch ein zu hohes Maß an Selbststeuerung vorzubeugen (vgl. Möller et al., 2006; Mayer, 2004; Bliss et al., 1996).

Sozial-konstruktivistische Ansätze greifen zur Beschreibung notwendiger Denk- und Unterstützungshilfen vermehrt auf das Konzept des Scaffolding zurück. In Abschnitt 1.2.3.2 wurde Scaffolding bereits als ein zentrales Element des Cognitive Apprenticeship-Ansatzes (Collins, Brown & Newman, 1989) erwähnt. Ursprünglich geht das Konzept auf Bruner und Kollegen zurück (z.B. Wood, Bruner & Ross, 1976). In Anlehnung an Vygotskys 'Zone der nächstmöglichen Entwicklung' (siehe Abschnitt 1.2.3.1) beschreiben Wood et al. (1976) mit dem Konzept des Scaffolding die Notwendigkeit der Unterstützung und Anleitung der Lernenden durch die Lehrperson beim Lösen von Aufgaben oder Problemen, die ohne Hilfestellung außerhalb der Möglichkeiten bzw. kognitiven Kompetenzen der Lernenden liegen würden. Dabei ist es von Bedeutung, dass die Lehrkraft den Lernenden ein optimales Niveau an Denkkunterstützung bietet, d.h. die kognitive Aktivität der Lernenden darf durch die Anleitung und Unterstützung der Lehrkraft nicht eingeschränkt werden (vgl. Wood, Bruner & Ross, 1976; Reiser, 2004; Pea, 2004; McKenzie, 1999). Zur Förderung der kognitiven Aktivität muss die Lehrperson also „so viel Hilfe wie notwendig“ und gleichzeitig „so wenig Hilfe wie möglich“ anbieten (Möller, 2006, S. 124). Zu Recht wird diese Aufgabe in der Literatur als „delicate negotiation“ (Reiser, 2004, S. 275), „balancing act“ (McKenzie, 1999) oder „teaching dilemma“ (Edwards & Mercer, 1987, zit. n. Hardy, 2007, S. 52) beschrieben. Wie kann die Lehrperson diesen Balance-Akt bewerkstelligen?

Für McKenzie (1999) ist 'structure' das Schlüsselwort. Hardy übersetzt den Begriff 'Scaffolding' in ihrer Habilitationsschrift (2007) mit 'kognitive Strukturierung', wobei sie diese als prozessorientierte Unterstützung der individuellen Wissenskonstruktion versteht. Als Möglichkeiten prozessorientierter Lernhilfen im naturwissenschaftlichen Unterricht werden in der aktuellen Literatur zum Scaffolding eine diskursive Strukturierung, sequenzielle Strukturierungsmaßnahmen und der Einsatz von 'Denkwerkzeugen' empfohlen:

- *Diskursive Strukturierung*. Dies meint eine unterstützende Gesprächsführung durch strukturierende Lehreraußerungen und Impulse (vgl. Hogan & Pressley, 1997; Davis & Linn, 2000). Dazu gehören Fokussierungshilfen und Modellierungen. Wesentliche Initiativen seitens der Lehrkraft sind diesbezüglich z.B. das Zusammenfassen bzw. Hervorheben bedeutungsvoller Inhalte und Aussagen („Was Peter gesagt hat ist ganz wichtig“), Problematisierungshilfen (z.B. das Aufmerksammachen auf offene Probleme), aufmerksamkeitsregulierende Hinweise als Unterstützung bei der Wahrnehmung wesentlicher Aspekte (z.B. Aufforderungen zum genauen Betrachten, „Achte auf den Wasserstand“, „Schau genau, was mit dem Wasser passiert“), das Verlangen einer Überprüfung von Gesagtem („Kannst du das beweisen?“) und das Provozieren (vgl. Reiser, 2004; Pea, 2004; Davis & Miyake, 2004; Einsiedler, 2007; Möller et

al., 2006; Hardy, 2007). Letzteres weist gewisse Überschneidungen zur bereits erläuterten Konfliktstrategie auf.

- *Sequenzielle Strukturierungsmaßnahmen*. Unter Sequenzierungen werden Gliederungsmaßnahmen verstanden, die zur Reduzierung der Komplexität von Lerninhalten eingesetzt werden (vgl. Reiser, 2004; Davis & Miyake, 2004). Unterrichtsinhalte werden z.B. entsprechend den Zielen in Teileinheiten oder Phasen gegliedert, wodurch die Anzahl der Freiheitsgrade eingeschränkt wird (vgl. Hardy, 2007; Möller et al., 2006; Einsiedler, 2007) (siehe dazu das Beispiel 2 in 3.4.1.1).
- *'Denkwerkzeuge'*. Kognitive Strukturierung muss sich nicht nur auf Interaktionen zwischen Lehrenden und Lernenden beziehen (vgl. Hardy, 2007). Auch der Einsatz von 'Denkwerkzeugen', d.h. von Repräsentationen, Artefakten und Modellen wird zur Strukturierung vorgeschlagen (vgl. Brown, Collins & Duguid, 1989; Palincsar, 1998). Vygotsky spricht in diesem Zusammenhang von materiellen Werkzeugen und symbolischen Mitteln (Zeichensysteme, Sprache) (vgl. dazu die Ausführungen in 1.2.3.1). Im Mathematikunterricht der Grundschule stellt beispielsweise die Darstellung von 'Zehnern' als Striche und 'Einern' als Punkte eine visuelle Unterstützungsform dar. Im naturwissenschaftlichen Sachunterricht können visuelle Darstellungsformen z.B. Grafen oder Diagramme sein, welche den Lernenden proportionale Beziehungen zwischen Variablen (z.B. zwischen 'Masse' und 'Volumen' für das Konzept 'Dichte') verdeutlichen (Clement, 2000; Möller et al., 2006; Hardy et al., 2004). Visuelle Darstellungsformen können von den Lehrern vorgegeben, aber auch von Schülern selbst entwickelt werden (siehe dazu Möller et al., 2006).

Laissez-faire vermeiden

Gemäß den obigen Ausführungen zum unterstützenden Lernen wäre es also – um in den Worten Reusers (2006, S. 159) zu sprechen – ein „instruktionsdidaktischer Fehlschluss“, den Unterricht gemäß der Ansicht zu gestalten, nach der sich Lernen ohne Anleitung durch die Lehrperson am besten vollziehe und eine direkte Instruktion kein Lernen im konstruktivistischen Sinne auslösen könne. Vermeiden sollte die Lehrperson demnach eine starke Zurückhaltung in Unterrichtsgesprächen. Ebenso sollte sie nicht darauf verzichten, den Kindern beim Entwickeln und Auswerten von Experimenten Hilfe anzubieten.

Ein „aktionistischer Fehlschluss“ wäre es gemäß Reusser (2006, S. 158), Konstruktivität mit äußerer Aktivität gleichzusetzen. Mayer (2004) spricht in diesem Zusammenhang davon, dass reine 'hands-on'-Aktivitäten ohne Reflexion nicht zu Lernerfolgen führen.

1.4.4 Verhaltensweisen von Lehrkräften zur Umsetzung situierten Lernens

Authentische Kontexte und Anwendungsbezüge

Instruktionsmodellen, die auf der Basis von Theorien zur situierten Kognition entwickelt worden sind, ist gemein, dass sie in Bezug auf die Gestaltung von Lernumgebungen fordern, die Situiertheit des Lernens zu berücksichtigen. Wie kann die Lehrperson dieser Forderung nachkommen? Zur Sicherung des Anwendungsbezugs von Wissen und zur Vermeidung trägen Wissens werden die folgenden Möglichkeiten vorgeschlagen:

- *Kontextualisierung.* Die Annahme, dass Lernen ein weitgehend kontextgebundener und situierter Prozess ist, impliziert, Lern- und Anwendungssituationen möglichst ähnlich zu gestalten (vgl. Resnick, 1987). Die Einbettung der Unterrichtsinhalte in realistische, authentische und sinnvolle Kontexte soll zur Sicherung des Anwendungsbezugs beitragen. Erst dadurch, dass die Lernenden die Anwendungsbezogenheit der Lerninhalte erfahren, werden die zu erwerbenden bzw. die neu erworbenen Konzepte für die Lernenden bedeutungsvoll (vgl. Reinmann-Rothmeier & Mandl, 1998; Reinmann & Mandl, 2006; Mandl, Kopp & Dvorak, 2004). Ähnlich wie im Anchored Instruction-Ansatz (vgl. 1.2.3.2) sollte die Lehrperson im naturwissenschaftsbezogenen Sachunterricht Probleme aus der Lebenswelt der Kinder in den Unterricht aufnehmen (zum Thema „Schwimmen und Sinken“: „*Warum schwimmt ein Schiff?*“; zum Thema „Wasserkreislauf“: „*Wie kommt es eigentlich, dass es regnet?*“). Des Weiteren sollte sie den Lebensweltbezug nicht nur auf den Einstieg der Stunde beschränken, sondern im Unterrichtsverlauf immer wieder Bezüge zum Alltag der Kinder herstellen.
- *Dekontextualisierung.* Ebenfalls ist es notwendig, Unterrichtsinhalte in verschiedenen (multiplen) Kontexten und aus verschiedenen Perspektiven zu sehen, damit Wissen nicht auf einen Kontext beschränkt bleibt (vgl. z.B. Collins, Brown & Newman, 1989; Reinmann-Rothmeier & Mandl, 1998; Mandl, Gruber & Renkl, 1995).

1.4.5 Zusammenfassung

Aus den vorausgegangenen Erläuterungen können nun folgende Folgerungen für das Lehren eines kognitiv anregenden naturwissenschaftlichen Unterrichts festgehalten werden: Um einen kognitiv anregenden Unterricht umzusetzen, sollte die Lehrperson:

- die Lernenden zum Lernen motivieren, indem sie genügend Anreizwerte im Unterricht schafft, die den Lernenden die Unterrichtsinhalte als etwas Interessantes vermitteln (Handlungserfahrungen, offene Fragestellungen, Inhalte, die das Interesse der Lernenden wecken);
- die kognitive Aktivität fördern, indem sie den Lernenden Spielraum und Zeit eröffnet, eigene Ideen zu äußern, wobei sie die eigenen Formulierungen der Lernenden und auch (Denk-)Fehler

zulässt und auf eine direkte Vermittlung von Wissen (z.B. in Form von Erklärungen) verzichtet;

- an die vorunterrichtlichen Vorstellungen der Schüler anknüpfen, indem sie die Vorstellungen der Schüler aktiviert (erfragt) und im Unterricht immer wieder aufgreift bzw. in den Unterrichtsverlauf integriert, d.h. mit und an den Schülervorstellungen arbeitet;
- sich um Veränderung (falscher) Vorstellungen bemühen, indem sie die Lernenden mit objektiven Evidenzen konfrontiert;
- soziales Lernen ermöglichen, indem sie gemeinsame Diskussion fördert;
- Unterstützung in Form von sequenziellen Strukturierungsmaßnahmen (Gliederung komplexer Aufgaben) und in Form einer diskursiven Unterstützung (Hervorhebungen bzw. Fokussierungen, Problematisierungshilfen, Provokationen) anbieten;
- Fragen und Aufgaben in einen sinnvollen Kontext einbinden, indem sie Aufgaben mit lebensweltlicher Einbettung thematisiert.

Eine ähnliche Auflistung von Konzeptwechsel-fördernden Gestaltungsmöglichkeiten für den Unterricht findet sich in Möller et al. (2006).

1.5 Forschungsstand zur Erfassung von Prozessmerkmalen des Unterrichts per Videoanalyse

Dieses Teilkapitel befasst sich mit dem Forschungsstand zur Erfassung von Prozessmerkmalen des Unterrichts per Videoanalyse. Es dient als Grundlage für die Entwicklung eines Videoanalyse-Instruments zur Erfassung kognitiv anregenden Lehrerverhaltens im naturwissenschaftlichen Sachunterricht in dieser Arbeit. Videoanalysen werden im Rahmen der Unterrichtsforschung zur Beschreibung von Unterricht und zur Untersuchung der Wirksamkeit von Unterricht verstärkt genutzt, seitdem das Interesse an der Beschreibung von Unterrichtsmerkmalen und -prozessen nach der TIMS 1995 Videostudie wieder gestiegen ist (vgl. 1.1). Hinzu kommt, dass der rasante technische Fortschritt im Bereich der elektronischen Datenspeicherung und -verarbeitung und im Bereich der Entwicklung digitaler Videokameras den Umgang mit videographierten Datenquellen preiswerter und einfacher macht (vgl. Petko et al., 2003; Wild, 2003; Hugener, 2008, Pauli & Reusser, 2006). Welche Vorteile bietet die Videoanalyse gegenüber anderen Erhebungsmethoden?

Im Unterkapitel 1.1.3 wurde bereits erläutert, dass sich zur Erfassung und Bewertung von Qualitätsmerkmalen unterrichtlichen Lehrerhandelns der Einsatz externer Beobachter als Informationsquelle empfiehlt. Videoanalysen bieten (neben der direkten Beobachtung im Unterricht) die Möglichkeit, eine Beurteilung von Unterricht aus der Sicht des externen Beobachters erheben zu können. Dadurch kann der sichtbare Unterricht frei von Vorerfahrungen, Wahrnehmungen, Interpretationen und subjek-

tiven Erlebnissen der in den Unterricht involvierten Lernenden und Lehrenden erfasst werden. Zwar kann auch durch externe Beobachter keine rein objektive Erfassung der Geschehnisse im Unterricht gesichert werden, aber ihnen gelingt es besser, eine möglichst neutrale Bewertung des Unterrichts zu treffen (Clausen, 2002; Hugener, 2008; siehe dazu auch die Erläuterungen in 1.1.3). Zudem haben externe Beobachter gegenüber Lehrkräften oder Lernenden, die an dem zu bewertenden Unterricht beteiligt sind, den Vorteil, ihre Einschätzung der zu bewertenden Unterrichtsstunde auf der Basis eines Vergleichs mit anderen Unterrichtsstunden aus der Studie treffen zu können (vgl. Clausen, 2002; Hugener, 2008).

Neben dem Vorteil einer möglichst objektiven Erfassung der Geschehnisse im Unterricht durch externe Beobachter, bieten Videoanalysen als Datenmaterial in Rohform die Möglichkeit, Unterricht beliebig oft, unabhängig vom Zeitpunkt der Aufnahme, von unterschiedlichen Personen, unter verschiedenen Fragestellungen und theoretischen Perspektiven und unterschiedlichen Methoden zu untersuchen. Dabei können qualitative und quantitative Analysemethoden integriert werden (vgl. Stigler et al., 2000; Jacobs, Kawanaka & Stigler, 1999). So ist die Verwendung sowohl der qualitativen Videodaten als auch anderer quantitativ erhobener Daten möglich, um z.B. Wirkungszusammenhängen nachgehen zu können. Dadurch kann der im Rahmen der Unterrichtsforschung oft gestellten Forderung nach Methodentriangulation nachgekommen werden (vgl. Klieme & Bos, 2000).

Um Unterricht beschreiben und Fragen bezüglich der Unterrichtswirksamkeit nachgehen zu können, sind in der videogestützten Forschung in den letzten Jahren eine Reihe von Instrumenten der Videoanalyse entwickelt worden, die sich hinsichtlich ihrer Inhalte, Ziele und Verfahren unterscheiden. Das Teilkapitel 1.5.1 beinhaltet eine ausführliche Auseinandersetzung mit den Methoden der Videoanalyse. Daran anschließend wird in 1.5.2 kurz auf bisher bestehende Instrumente zur Erfassung kognitiv aktivierender Verhaltensweisen von Lehrkräften eingegangen.

1.5.1 Methoden der Videoanalyse

Im Rahmen der Auseinandersetzung mit den Methoden der Videoanalyse interessiert, was unter hoch- bis niedrig-inferenten Instrumenten der Videoanalyse verstanden wird bzw. was diese Verfahren kennzeichnet (1.5.1.1), wie Videoanalyse-Instrumente entwickelt werden (1.5.1.2), welche Möglichkeiten der Güteprüfung sich in der Videoanalyse anbieten und welche Güteprüfungen in Videostudien umgesetzt werden (1.5.1.3).

1.5.1.1 Niedrig- bis hoch-inferente Instrumente der Videoanalyse

Unterscheidungsmerkmal 'Inferenz': Die in der Unterrichtsforschung derzeit entwickelten Instrumente der Videoanalyse lassen sich „auf einem Kontinuum von 'niedrig-inferent' bis 'hoch-inferent' anordnen“ (Clausen, Reusser & Klieme, 2003, S. 124 in Anlehnung an Rosenshine, 1970). Darunter ist zu verstehen, dass die Instrumente im Grad der Inferenz variieren, wobei unter Inferenz interpretative Schlussfolgerungen zu verstehen sind (vgl. Seidel, 2003; Hugener, 2006, 2008). „Inference [...] refers

to the process intervening between the objective data seen or heard and to the coding of those data on an observational instrument” (Rosenshine, 1970, S. 281). Dementsprechend ist unter dem Grad der Inferenz das Maß an interpretativen Schlussfolgerungen gemeint, welches in das Urteil eines Betrachters einfließt (vgl. Hugener, 2006). Je höher die Inferenz, d.h. je höher das Maß an interpretativen Schlussfolgerungen, desto größer ist die Rolle von Vorwissen, Erwartungen, Einstellungen oder Denkweisen des jeweiligen Betrachters (Reyer, 2004). Dies bedeutet für die Unterrichtsbeobachtung, dass die reine 'objektive' Beobachtung von Merkmalen, Verhalten etc. mit steigendem Grad der Inferenz immer mehr zu einer Beurteilung wird. Clausen, Reusser und Klieme (2003) sprechen deshalb in Anlehnung an Rosenshine (1970) von niedrig-inferenten Beobachtungen und hoch-inferenten Beurteilungen. *Niedrig-inferente Beobachtungen* orientieren sich an konkret beobachtbarem Verhalten und ermöglichen eine objektive Wahrnehmung (Seidel, 2003; Clausen, Reusser & Klieme, 2003; Reyer, 2004). Sie benötigen somit ein eher geringes Maß an schlussfolgernden Interpretationen auf dem Weg zu einem Urteil (Hugener, 2006) und stellen nach Clausen et al. (2003) vergleichsweise wenig pädagogisch-didaktische Anforderungen an den Beobachter. Im Gegensatz zu niedrig-inferenten Beobachtungen gehen *hoch-inferente Beurteilungen* über konkret beobachtbares Verhalten hinaus und beziehen sich auf nicht sichtbare, latente Strukturen, d.h. auf abstraktere, globalere Sachverhalte oder Verhaltensmerkmale, die erst unter Nachvollzug des gesamten oder bisherigen Unterrichts bewertet werden können (vgl. Clausen, 2003; Hugener, 2006; Reyer, 2004). Dementsprechend erfordern hoch-inferente Beurteilungen ein höheres Maß an interpretativen Prozessen beim Beurteiler. Die folgenden Beispiele sollen den unterschiedlichen Interpretationsgehalt verschiedener Unterrichtsmerkmale und -prozesse verdeutlichen (siehe auch die Darstellungen in Tabelle 1.5-1):

- *Beispiel 1, 'Gesprächsanteile', niedrig-inferent.* Unter 'Gesprächsanteile' werden in diesem Beispiel die Redeanteile von Lehrern und Schülern verstanden („Lehrer spricht“, „Schüler spricht“, „Lehrer- und Schüler sprechen gemeinsam“, ...). Sprecherwechsel sind eindeutig identifizierbar. Somit bedarf das Setzen einer Variable (z.B. „Schüler spricht“) keinerlei Interpretationen. Auch spezielles Vorwissen ist zur Beobachtung nicht notwendig.
- *Beispiel 2, 'Unterrichtsphasen', niedrig- bis mittel-inferent.* In vielen Videostudien werden Unterrichtsphasen (Wiederholung, Einführung, Arbeitsphase, Reflexion etc.) erfasst (vgl. z.B. Seidel et al., 2003). Für die Beobachtung von Unterrichtsphasen lassen sich objektive Indikatoren als erwartete Beobachtungsmerkmale gut beschreiben, wie z.B. für die Bestimmung von Reflexionsphasen: „Die Lehrperson bespricht gemeinsam mit den Schülern die in der Stunde behandelten oder erarbeiteten Inhalte. Dazu kann gehören, dass die Schüler ihre Arbeitsergebnisse vorstellen, dass Ergebnisse gemeinsam herausgearbeitet und aufgeschrieben werden, dass Ergebnisse transferiert werden, ...“. Somit sind Unterrichtsphasen für Außenstehende gut beobachtbar; für die Beobachtung von Unterrichtsphasen ist die Berücksichtigung der Unterrichtsstunde als Ganzes nicht oder nur in einem sehr geringen Maße notwendig. Dennoch stellt die Bestimmung von Unterrichtsphasen mehr Anforderungen an den Beobachter als das Setzen

von Sprechercodes aus Beispiel 1, da der Beobachter über ein gewisses Maß an Verständnis über Unterrichtsphasierungen (Merkmale von Unterrichtsphasen und Phasenwechsel), also über ein allgemein pädagogisch-didaktisches Verständnis, verfügen muss. Oft sind Wechsel zwischen Phasen nicht immer eindeutig. Ebenso kann es vorkommen, dass sich z.B. eine Reflexionsphase nicht eindeutig von Übungen unterscheiden lässt. Zu einem großen Anteil kann derartigen Problemen durch spezifische Kodierregelungen begegnet werden. Dennoch kann es aber Situationen geben, in denen der Beobachter eine Kodierung nur vornehmen kann, indem er auf sein eigenes Verständnis von Unterrichtsphasen zurückgreift. Somit gehen in gewissem (wenn auch eher geringem) Maße individuelle Sichtweisen in die Bewertung mit ein.

- *Beispiel 3, 'Lehrer- und Schülerverhalten im Klassengespräch', mittel- bis hoch-inferent.* Verhaltensweisen von Lehrkräften im Unterricht, wie z.B. das Auffordern von Schülern, Ideen zu äußern, Begründungen zu geben, Hypothesen aufzustellen oder das Schülerverhalten, wie „Schüler äußern Ideen“; „Schüler begründen Ideen“; „Schüler formulieren Hypothesen“, wurden von B. Gais (2009) operationalisiert. Für derartige Verhaltensweisen lassen sich verhaltensnahe und für außenstehende Beobachter gut beobachtbare Indikatoren und Bestimmungshilfen definieren. Dennoch ist eine Beurteilung derartigen Verhaltens oft nur unter Berücksichtigung des vorausgegangenen Unterrichts möglich. So wird z.B. für die Bewertung eines bestimmten Schülerverhaltens das vorangegangene Lehrerverhalten (bestimmte Reaktionen oder Impulse) berücksichtigt. Die Erfassung von Schüler- und Lehrerverhalten aus diesem Beispiel verlangt vom Beobachter/Beurteiler also trotz der Definition verhaltensnaher Indikatoren deutlich mehr (fach-)didaktische Fähigkeiten bzw. ein höheres Maß an pädagogisch-didaktischem Verständnis als das vorausgegangene Beispiel 2. Insbesondere die Erfassung von Schülerverhalten kann Schwierigkeiten mit sich bringen. Wann handelt es sich um eine Begründung? Wenn die Schüler Worte wie „weil“, „darum“ benutzen? Eine Antwort aus Sicht eines Schülers kann eine Begründung für eine Idee oder ein Konzept darstellen. Aus der Sicht der Lehrperson oder des externen Beobachters aber handelt es sich ggf. eher um eine neue Behauptung als um eine Begründung. Es ist nicht zu vermeiden, dass ein gewisses Maß an Interpretationen in die Beobachtung bzw. Beurteilung einfließt.
- *Beispiel 4: 'Fehleroffenheit', hoch-inferent.* Ein Beispiel mit noch höherem Interpretationsgehalt stellt das Konstrukt „Fehleroffenheit“ (siehe Abschnitt 3.4.2; zum Umgang mit Fehlern vgl. auch Heinze, 2004) dar. Hier geht es darum, einzuschätzen, inwiefern für die Lehrperson Fehler nichts 'Schlimmes' sind. Zwar lassen sich verhaltensnahe Indikatoren finden, wie z.B. „Die Lehrperson betont, dass falsche Ideen wichtig sind, um zum Ziel zu kommen“. Diese stellen aber keine ausreichende Bewertungsgrundlage dar. Denn: Gilt eine Lehrkraft als nicht fehleroffen, wenn sie dies nicht explizit formuliert? Es muss also auch passives Lehrerverhalten, das auf Fehleroffenheit hindeutet, berücksichtigt werden, um die Dimension vollständig abzudecken. Hier geht es dann um eine Einschätzung dessen, was die Lehrperson denkt, was für

einen Eindruck sie bei den Beurteilern hinterlässt. Ob eine Lehrkraft im Unterricht offen für Fehler ist oder nicht, ist somit weniger leicht oder ggf. nicht direkt zu beobachten. Fehleroffenheit wird nicht aufgrund eines konkreten Verhaltens, sondern vor dem Hintergrund mehrerer Situationen und Interaktionen mit Schülern, demnach rückblickend auf den gesamten Unterricht oder größere Sinneinheiten als ganzheitliche Wahrnehmung beurteilt. Dementsprechend interpretativ ist dieses Bewertungsmerkmal.

Zu den Inhalten und Zielen niedrig- bis hoch-inferenter Instrumente der Videoanalyse. In der videogestützten Unterrichtsforschung werden niedrig- bis mittel-inferente Beobachtungen häufig zur Erfassung von Häufigkeiten und der Dauer von Ereignissen eingesetzt, wie z.B. zur Erfassung von Gesprächsanteilen (siehe Beispiel 1), von Arbeits- und Sozialformen (Arbeitsformen: z.B. Lehrervortrag, Klassengespräch; Sozialformen: Gruppen-, Partner-, Einzelarbeit) oder von Unterrichtsphasen (siehe Beispiel 2) (vgl. z.B. Seidel, Prenzel, Duit & Lehrke, 2003; Reyer, 2004; Hugener, Pauli & Reusser, 2006). Derartige Unterrichtsmerkmale werden als 'Sichtstruktur' oder 'Oberflächenstruktur' von Unterricht bezeichnet. Entsprechend ihrer Orientierung an Oberflächenmerkmalen haben niedrig-inferente Beobachtungen eine Beschreibung der Unterrichtsgestaltung und -organisation zum Ziel. Es geht, so Hugener (2006, S. 46), um eine „möglichst präzise Beschreibung der Unterrichtsgestaltung, Ereignisse und Organisationsformen usw.“ Hoch-inferente Beurteilungen werden zur Erfassung von Merkmalen der 'Tiefenstruktur' von Unterricht eingesetzt. Dazu gehören allgemeine und fachspezifische Unterrichtsqualitätsmerkmale, wie z.B. „Disziplin“, „Classroom Management“ oder „kognitive Aktivierung“ (z.B. Clausen, Reusser & Klieme, 2003; Hugener, Pauli & Reusser, 2006), also Merkmale des Lehrer- und Schülerhandelns im Unterricht. Das Ziel hoch-inferenter Beurteilungen ist die Einschätzung der Qualität von Unterricht und Unterrichtsereignissen (vgl. Hugener, 2006).

Kodierung versus Rating. Die Ausrichtung niedrig- bis mittel-inferenter Beobachtungsinstrumente an direkten, verhaltensnahen, gewissermaßen 'objektiven' Beobachtungen erzeugte in den Videostudien der letzten Jahre eine Fokussierung auf kurze Sequenzen, wodurch die videographierten Unterrichtsstunden oder -ausschnitte in überblicksartige Einheiten eingeteilt und auf das Auftreten bestimmter Aspekte hin untersucht worden sind (vgl. Hugener, 2006; Clausen, Reusser & Klieme, 2003). Dieses Vorgehen der Datenerhebung wird als *Kodierung* bezeichnet, wobei zwischen 'time-sampling' und 'event-sampling' unterschieden wird. Bei einem time-sampling-Verfahren (Zeitstichprobenplan) wird das gesamte Videomaterial in einheitliche Abschnitte, in so genannte slots, unterteilt (z.B. in 10-Sekunden-Intervalle). Für jeden slot wird ein entsprechendes Kodierurteil gefällt. Bei einem 'event-sampling' (Ereignisstichprobenplan) – auch als Turn-Kodierung bezeichnet – werden keine künstlichen Zeiteinheiten gesetzt, sondern es werden Sinnabschnitte ('turns') von einzelnen Ereignissen in Echtzeit kodiert. Anfang und Endpunkt eines Sinnabschnitts müssen für jedes Ereignis neu gesetzt werden. Ereignisse können z.B. Sprecherwechsel sein (vgl. Seidel, 2003; Hugener, Rakoczy, Pauli & Reusser, 2006). Kodierungen dienen oft als Basis für weiterführende/vertiefende Analysen, z.B. in Form von hoch-inferenten Beurteilungen.

Da es bei hoch-inferenten Beurteilungen um die Beurteilung von abstrakten Konstrukten geht, die erst im verstehenden Nachvollzug des Unterrichtsfortgangs beurteilt werden können, wobei es um die Einschätzung der Qualität von Unterricht geht, ist der Gesamteindruck vom beobachteten Unterricht durch den Beurteiler entscheidend (vgl. Reyer, 2004; Hugener, Rakoczy, Pauli & Reusser, 2006). Das Ausmaß eines Merkmals oder einer Verhaltensweise wird relativ ganzheitlich eingeschätzt (vgl. Clausen, Reusser & Klieme, 2003). Aus diesem Grund bilden entweder längere Unterrichtssequenzen oder aber die gesamte Unterrichtsstunde die Analyseeinheit. Dieses Verfahren wird als *Rating* bezeichnet.

Kategorisierungen erfolgen zumeist computergestützt, z.B. mit einer speziellen Software. Derartige Programme vereinbaren verschiedene Grundfunktionen, wie z.B. das am IPN entwickelte Programm Videograph (®Rimmele, 2003). In Videograph ist es möglich, digitalisiertes Video- bzw. Audiomaterial abzuspielen, zu transkribieren und mittels selbst konstruierter Beobachungskategorien zu kodieren. Die kodierten Rohdaten können zudem zur weiteren Datenverarbeitung in das Programm SPSS exportiert werden. Auch ein Datenimport ist möglich, um z.B. Daten zwischen Videographdateien auszutauschen (vgl. Seidel, Kobarg & Rimmele, 2003). Andere gebräuchliche Programme sind z.B. CatMovie (Wild, 1999) oder Vprism (Knoll & Stigler, 1999). Bei Ratings ist es üblich, dass das Videomaterial von ausgewählten Personen betrachtet wird, die ihre Urteile auf einem Fragebogeninventar notieren (vgl. Seidel, 2003) ('paper and pencil' oder Online-Bogen).

Inferenz und Auswirkungen auf Beobachterübereinstimmung bzw. Beurteilerreliabilität²⁸ und Validität. Wie bereits erläutert, sind niedrig-inferente Beobachtungen an konkret beobachtbaren Situationen und Verhaltensweisen ausgerichtet, während sich hoch-inferente Beurteilungen auf abstraktere und globalere Sachverhalte und Verhaltensweisen bzw. -tendenzen beziehen und dementsprechend mehr interpretative Schlussfolgerungen vom Beurteiler verlangen. Niedrig-inferente Beobachtungen gelten deshalb als die objektivere und neutralere Variante (vgl. Clausen, 2002, S. 48). Die Anleitungen zur Beobachtung niedrig-inferenter Merkmale sind für die Beobachter klar definierbar, woraus sich eine hohe Übereinstimmung der Beobachtungen ergibt (vgl. Seidel, 2003, S. 106). Entsprechend hat das höhere Maß an Inferenz bei hoch-inferenten Beurteilungen zur Folge, dass die Beurteilerreliabilität (auch Interrater-Reliabilität) meist geringer ist; die Beurteiler werden in einem höheren Maß als Fehlerquelle angesehen als dies bei niedrig-inferenten Beobachtungsverfahren der Fall ist (vgl. Clausen, Reusser & Klieme, 2003). Der Schluss, auf hoch-inferente Daten zu Gunsten einer möglichst hohen Reliabilität zu verzichten, ist allerdings nicht zulässig. Insbesondere die latenten, 'unsichtbaren' Konstrukte, wie z.B. spezielles Lehrerverhalten, das sich im Kontext von Lehrer-Schüler-Interaktionen äußert (und deren Beurteilung ein hohes Maß an pädagogisch-didaktischem Verständnis erfordert), sind für die Unterrichtsqualitätsforschung meist die interessanteren bzw. relevanteren (vgl. Clausen, Reusser & Klieme, 2003; Clausen, 2002). Bezüglich der Validität haben hoch-inferente Beurteilungen gegenüber niedrig-inferenten Beobachtungen den Vorteil, dass erstere oft einen stärkeren Bezug zur

²⁸Zum Gebrauch bzw. zur Unterscheidung der Begrifflichkeiten 'Übereinstimmung' und 'Reliabilität' siehe die Ausführungen in Abschnitt 1.5.1.3.

Theorie bzw. zur theoretischen Beschreibungsebene haben, wodurch die Qualität der Aussagen höher zu bewerten ist (vgl. Clausen, 2002; Reyer, 2004). Niedrig-inferente Verfahren büßen meist an Validität ein, da sie durch die detaillierte Aufteilung einzelner Unterrichtsereignisse und die Zuordnung in eine Einzelkategorie globale Unterrichtsereignisse oder -komplexe oft vernachlässigen (vgl. Seidel, 2003). Dementsprechend weisen Clausen et al. (2003) mit Bezug auf Rosenshine (1970) und Fraser und Walberg (1981) darauf hin, dass Prozess-Produkt-Studien zeigen, dass hoch-inferente Beurteilungen meist höhere Zusammenhänge zu schulischen Lernerfolgskriterien aufweisen als niedrig-inferente Kodierverfahren. So berichtet Rosenshine (1970) von sieben Untersuchungen, in denen Lehrerverhalten mit Hilfe von Kategorien- und Ratingsystemen erhoben wurde. Ergebnisse von sechs der sieben Studien zeigen höhere Zusammenhänge (bivariate Korrelationen) zwischen hoch-inferent geratetem Verhalten von Lehrkräften und Schülerleistungen als zwischen niedrig-inferent kodiertem Lehrerverhalten und Schülerleistungen.

Tab. 1.5-1 Überblick über Verfahren der Videoanalyse zur Erfassung von Unterrichtsmerkmalen

Überblick über Verfahren der Videoanalyse zur Erfassung von Unterrichtsmerkmalen			
Inferenz	niedrig	mittel	hoch
	wenig Interpretation		viel Interpretation
Was wird beobachtet?	Oberflächenstrukturen		Tiefenstrukturen
	z.B. Gesprächsanteile	Arbeitsphasen	Lehrer-/ Schülerverhalten
Wozu wird beobachtet?	Beschreibung der Unterrichtsgestaltung		Einschätzung der Qualität von Unterricht
Wie wird beobachtet?	kategorienbasiertes Verfahren (Kodierung)		Schätzverfahren (Rating)
Güte	höhere Beobachterübereinstimmung, geringe Validität		geringere Interrater-Reliabilität, höhere Validität

Die Ausführungen bezüglich der Güte von Videoverfahren zeigen also, dass niedrig-inferentere Beobachtungen meist die reliableren Messungen liefern und hoch-inferente Ratings als die validere Methode gelten (Clausen, Reusser & Klieme, 2003). Clausen (2002) und Clausen, Reusser und Klieme (2003) sprechen in diesem Zusammenhang von einem Übereinstimmungs-Bedeutsamkeits-Dilemma.

1.5.1.2 Zum Entwicklungsprozess von Videoanalyse-Instrumenten

Die Entwicklung von Kodier- und Ratingsystemen erfolgt in Videostudien anhand von zwei Methodentexten, und zwar nach den Standards inhaltsanalytischer Vorgehen nach Bos und Tarnai (1999) und auf der Basis des zyklischen Prozesses zur Kodierung und Analyse von Videoaufzeichnungen von Jacobs, Kawanaka und Stigler (1999). Entsprechend basieren die Ausführungen dieses Unterkapitels auf diesen beiden Methodentexten sowie auf Erläuterungen von Hugener (2006, 2008), Rakoczy und Pauli (2006), Seidel (2003) und Reyer (2004), die sich in ihren Darstellungen ebenfalls auf Bos und Tarnai (1999) und Jacobs et al. (1999) beziehen.

Theoretische Ebene (Schritt 1). Vor dem Hintergrund einer Theorie werden Untersuchungsbereiche, Forschungsfragen und Hypothesen abgeleitet.

Entwicklung der Videoverfahren (Schritt 2). An den ersten theoretischen Schritt schließt sich die Sichtung des Datenmaterials an, um Beobachtungsinhalte zu identifizieren, die mit der Theorie korrespondieren. Die beobachteten Inhalte werden dann – je nach Fragestellung und Eigenschaft der Inhalte – vor dem Hintergrund der theoretischen Annahmen in Form von Kategorien (Kodierung) oder Dimen-

sionen bzw. Konstrukten (Rating) inhaltlich beschrieben. Die Kategorien und die Konstrukte werden sodann durch die Beschreibung von Indikatoren (als erwartete Beobachtungsmerkmale) operationalisiert. Bei Kategoriensystemen beinhaltet die Operationalisierung eine genaue und umfangreiche Beschreibung der Kategorien auf der Beobachtungsebene (z.B. Sammlung musterhaften Verhaltens von Lehrkräften). Zudem sollten konkrete Stellen aus Unterrichtsvideos als Ankerbeispiele aufgeführt werden. Bei Ratingsystemen entsprechen meist die Items der Konstrukte den Indikatoren, die direkt beurteilt werden. Des Weiteren werden in diesem zweiten Schritt die Analyseeinheiten festgelegt (time-sampling oder event-sampling bei Kodierungen; ganze Unterrichtseinheiten, Unterrichtsphasen, -situationen etc. bei Ratings) und Regeln für die Kodierung bzw. für das Rating bestimmt. Für Ratingverfahren müssen zudem die Abstufungen der Einschätzungen der Items (z.B. zwei-, drei-, vier- oder fünfstufiges Antwortformat) festgelegt werden.

Pretest und Training (Schritt 3). Nachdem das Videoverfahren vorläufig steht, erfolgt die Testung (Erprobung und Überarbeitung) des Instruments innerhalb der Entwicklungsgruppe; möglichst an Videomaterial außerhalb der zu untersuchenden Stichprobe. Mittels Probebeobachtungen bzw. -beurteilungen werden die Beschreibungen der Kategorien und Konstrukte ausgeschärft und perfektioniert; ggf. werden weitere Indikatoren herausgearbeitet. Nach der definitiven Operationalisierung des Video-Instruments erfolgt ein Training für mindestens zwei Kodierer bzw. Rater. Allgemein beinhaltet das Training gemäß Bortz und Döring (2006) eine Einführung in das Konzept der Untersuchung und die Darstellung des theoretischen Ansatzes und schließt mit der Prüfung der Übereinstimmung der Beobachter ab. Bei einem kategorienbasierten Vorgehen in Form eines Time- oder Event-Samplings wird die Beobachterübereinstimmung im Anschluss an das Training an mindestens 30% des Datenumfanges anhand von Doppeltkodierungen überprüft (siehe z.B. Seidel, Prenzel, Duit & Lehrke, 2003; Hugener, Pauli & Reusser, 2006; Reyer, 2004). Bei Ratingverfahren ergibt sich durch die in Videostudien oft erfolgte Festlegung der Analyseeinheit auf das gesamte Unterrichtsvideo das Problem, dass die Anzahl der dem Training zugrunde liegenden Analyseeinheiten sehr gering ist, was keine ausreichende Basis für Reliabilitätsberechnungen darstellt (z.B. $N=1$ bei einer Analyseeinheit eines 45minütigen Unterrichtsvideos im Vergleich zu einem $N=270$ bei einer Analyseeinheit von 10 Sekunden bei einem 45minütigen Unterrichtsvideo). Aus diesem Grund hat sich in Videostudien das folgende Verfahren durchgesetzt: Die Schulung wird als abgeschlossen betrachtet, wenn die Rater sich einig sind, die Unterrichtsvideos auf der Basis eines gleichen theoretischen Verständnisses zu beurteilen und die Urteile auf ein reliables Rating deuten. Die Interrater-Reliabilität wird durchgeführt, nachdem 50% und 100% des zu analysierenden Datenmaterials unabhängig von mindestens zwei Personen geratet worden sind (Seidel, 2003 in Anlehnung an Jacobs et al., 2003; Clausen, Reusser & Klieme, 2003 und Rakoczy & Pauli, 2006). Bei Schätzverfahren ist eine Berechnung der endgültigen Interrater-Reliabilität somit erst zum Ende der Analyse (Schritt 4) möglich.

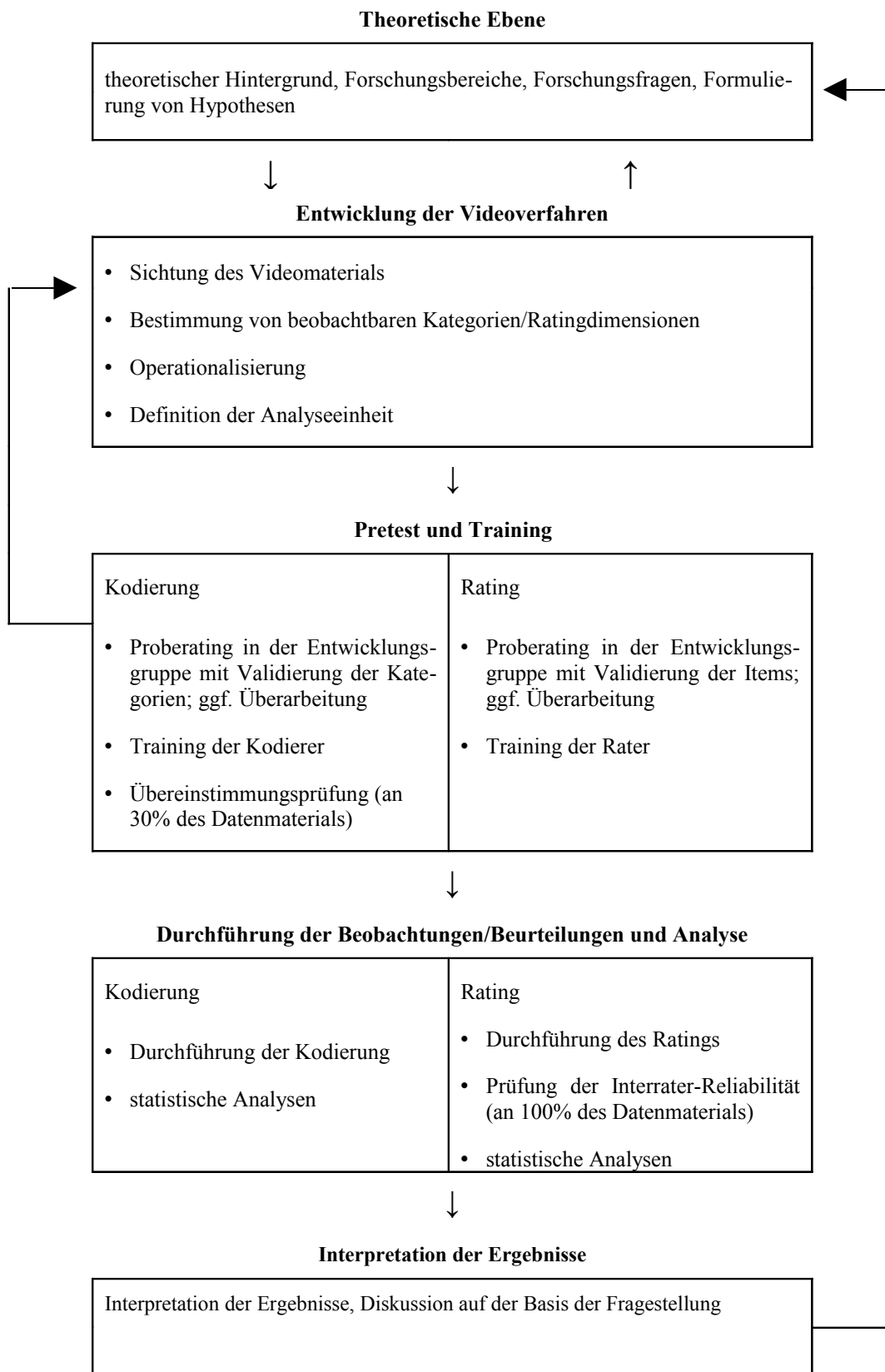
Analyse (Schritt 4). Dieser Schritt beinhaltet die Anwendung des Videoanalyse-Instruments auf die zu untersuchende Stichprobe. Mit diesem Schritt beginnt die Quantifizierung des Datenmaterials. Bei Ra-

tingverfahren schließt sich (wie bereits erwähnt) die endgültige Berechnung der Interrater-Reliabilität an.

Interpretation der Ergebnisse (Schritt 5). In diesem letzten Schritt werden die kategorisierten bzw. gerateten Daten aufbereitet. Das beurteilte Datenmaterial kann nun für weitere statistische Analysen eingesetzt werden. Dazu können je nach Fragestellung qualitative oder quantitative Verfahren genutzt werden. Schließlich werden die Ergebnisse der Analysen in Bezug auf die Hypothesen interpretiert und diskutiert. Dabei kann wiederum auf die Videodaten als qualitatives Datenmaterial zurückgegriffen werden. Aus der Diskussion der Ergebnisse ergeben sich möglicherweise neue Fragestellungen, die ggf. zur Entwicklung weiterer Analyseinstrumente führen.

Im Kern erfolgt die Entwicklung von Videoanalyse-Instrumenten also sowohl deduktiv als auch induktiv, d.h. theorie- und datengeleitet. Das Ziel der Entwicklung ist die Typisierung von Beobachtungsmerkmalen (qualitativ) und die Quantifizierung des Datenmaterials. Ein solches Verfahren der qualitativen Inhaltsanalyse wurde von Mayring (2003) ursprünglich zum nachvollziehbaren Durcharbeiten umfangreichen Textmaterials (qualitative Inhaltsanalyse) entwickelt (vgl. Bortz & Döring, 2006; auch Reyer 2004). Die Inhaltsanalyse gilt nach Atteslander (2000) und Bos und Tarnai (1999) aber auch für die Analyse von Videodaten als grundsätzlich geeignet.

Tab.1.5-2 Verfahren zur Entwicklung von Videoverfahren
 (angelehnt an Jacobs et al., 1999 und Bos & Tarnai, 1999; vgl. auch Hugener, 2008 und Seidel, 2003)



1.5.1.3 Zu den Gütekriterien bei Videoanalysen

Die Videoanalyse als Methode der Datengewinnung bedarf – wie andere Analyseinstrumente auch – verschiedener Qualitätskontrollen, um forschungsmethodischen Ansprüchen zu genügen. „Gütekontrollen liefern mit der Prüfung der Forschungsmethode Anhaltspunkte für den Wahrheitsgehalt und die Haltbarkeit von Aussagen“ (Lamnek, 1995, S. 153). Dabei stellt sich die Frage, welche Güteprüfungen für die Videoanalyse als angemessen gelten können. Die vorausgegangenen Ausführungen haben gezeigt, dass die Videoanalyse zugleich als qualitatives und quantitatives Verfahren betrachtet werden kann: Die Bildung der Beobachtungskategorien bzw. Beurteilungskonstrukte ist ein vorwiegend qualitatives Verfahren, während die Datenauswertung quantitativ erfolgen kann. Im Kontext der klassischen Testtheorie werden zur Güteprüfung von psychometrischen Tests die drei zentralen Gütekriterien *Objektivität*, *Reliabilität* und *Validität* genannt (vgl. Bortz & Döring, 2006). Diese werden nach Bortz & Döring (2006) in modifizierter Form auch in der qualitativen Forschung verwendet, wobei die Autoren darauf hinweisen (S. 326), dass im Rahmen von qualitativer Forschung die Begriffe Objektivität und Reliabilität eher ungebräuchlich sind und stattdessen von unterschiedlichen Kriterien der Validität gesprochen wird. Für Rost (2005) sind die Gütekriterien der klassischen Testtheorie sinngemäß auch auf andere Datenerhebungsmethoden, wie z.B. Interview, Beobachtung etc. übertragbar. Auch Lamnek (1995) hält die klassischen Testgütekriterien prinzipiell auf alle Beobachtungsinstrumente im weiteren Sinne anwendbar. In Bezug auf Inhaltsanalysen halten Mayring und Brunner (2007) eine Orientierung an den drei Gütekriterien der quantitativen Forschung prinzipiell für sinnvoll. Im Folgenden werden Möglichkeiten der Güteprüfung erläutert. Dabei wird darauf eingegangen, welche Möglichkeiten sich für die Güteprüfung in der Videoanalyse bieten.

Objektivität

Im Rahmen der *quantitativen Forschung* definieren Bortz und Döring (2006, S. 195) Objektivität wie folgt: „Ein Test oder Fragebogen ist objektiv, wenn verschiedene Testanwender bei denselben Personen zu den gleichen Resultaten gelangen, d.h. ein objektiver Test ist vom konkreten Testanwender unabhängig. Ein Test wäre also nicht objektiv, wenn in die Durchführung oder Auswertung z.B. besonderes Expertenwissen oder individuelle Deutungen des Anwenders einfließen, die intersubjektiv reproduzierbar sind“. Unterschieden werden Durchführungs-, Auswertungs- und Interpretationsobjektivität:

- *Durchführungsobjektivität*. Die Objektivität der Durchführung wird meist durch schriftlich festgelegte Durchführungsanweisungen in Form eines Begleithefts gesichert. So soll eine Variation in der Testdurchführung von Untersuchung zu Untersuchung verhindert werden (vgl. Bortz & Döring, 2006).
- *Auswertungsobjektivität*. Die Objektivität der Auswertung hängt von der Art der Itembeantwortung und Antwortbewertung ab (Bortz & Döring, 2006). Items mit geschlossenem Antwortformat (ja/nein-Antworten, multiple choice, mehrstufiges Antwortformat) begünstigen

eine hohe Auswertungsobjektivität. Bei offenen Fragen ohne festgelegte Antwortformate ist es notwendig, exakte Regeln zu definieren, wie ein Ergebnis zu bewerten ist. Je mehr die Bewertung einer Aufgabe von der Entscheidung eines Auswerters abhängt, desto geringer ist die Auswertungsobjektivität (vgl. Rost, 2005).

- *Interpretationsobjektivität*. Die Objektivität der Interpretation verlangt, dass keine individuellen Deutungen in die Interpretation eines Tests einfließen dürfen (vgl. Bortz & Döring, 2006). Verschiedene Personen müssen aus denselben Testergebnissen unabhängig voneinander die gleichen oder ähnliche Schlüsse ziehen (vgl. Rost, 2005).

Bei standardisierten Tests sind diese Objektivitätsmaßnahmen relativ leicht realisierbar. Bei qualitativ-deutenden Verfahren, wie z.B. projektiven Tests (der Psychologe interpretiert die Antworten eines Patienten und ordnet sie bestimmten Kategorien zu), muss die Objektivität empirisch überprüft werden (vgl. Bortz & Döring, 2006; Bühner, 2006). Gleiches gilt z.B. auch für quantifizierende Beobachtungs- und Beurteilungsmethoden. Dazu wird auf das Verfahren der Beurteilerübereinstimmung zurückgegriffen, das in der Literatur jedoch nicht einheitlich der Objektivität zugeordnet wird. Lienert und Raatz (1994) bewerten interpersonelle Übereinstimmungen genau wie Bortz und Döring (2006) als Merkmal der Objektivität. Wirtz und Caspar (2002) hingegen ordnen Beurteilerübereinstimmungen dem Zuverlässigkeitskriterium, also der Reliabilität, zu. Ihre Zuordnung wird auf S. 84f. ausgeführt.

Unabhängig von der Frage, welchem Gütekriterium das Konsenskriterium zugeordnet werden soll, eignen sich zur empirischen Überprüfung der Beurteilerübereinstimmung bzw. -reliabilität verschiedene Indizes (vgl. z.B. Wirtz & Caspar, 2002). Weitere Ausführungen dazu folgen auf den Seiten 84f.).

Im Rahmen von *qualitativen Daten* meint Objektivität nach Bortz und Döring (2006, S. 326) nicht 'höhere Wahrheit', sondern *interpersonalen Konsens*, d.h. bei einer Untersuchung desselben Sachverhalts mit denselben Methoden müssen unterschiedliche Forscher zu vergleichbaren Resultaten gelangen. Damit dies gewährleistet werden kann, sind eine klare Beschreibung des methodischen Vorgehens (Transparenz) und eine gewisse Standardisierung notwendig. Beruft sich ein Untersucher nur auf seine langjährigen Erfahrungen und verzichtet auf eine genaue Beschreibung seines Vorgehens, ist die Objektivität verletzt (Bortz & Döring, 2006). Mayring (2003) spricht in diesem Zusammenhang bei qualitativen Inhaltsanalysen von *Verfahrensdokumentation*, worunter er eine sehr detaillierte und weitgehende Darstellung des Vorgehens versteht, damit der Forschungsprozess für andere nachprüfbar wird (vgl. Mayring, 2003; Lamnek, 1995).

Auswertungs- und Interpretationsobjektivität können auch im qualitativen Ansatz mittels Übereinstimmungskoeffizienten abgeschätzt werden (vgl. Bortz & Döring, 2006). Ob Auswerter bzw. Interpreten übereinstimmen, wird in der qualitativen Forschung allerdings gemäß Bortz und Döring (2006) eher als Validitätsproblem aufgefasst. „Nur wenn intersubjektiver Konsens zwischen Auswertern besteht, kann eine Interpretation als gültig (valide) und wissenschaftlich abgesichert angesehen werden“ (Bortz

& Döring, 2006, S. 327). Im Rahmen von Inhaltsanalysen ordnen Mayring und Brunner (2007) die Prüfung der Übereinstimmung von Kodierern dem Kriterium der Objektivität zu.

Welche der genannten Möglichkeiten der Objektivitätsprüfungen lassen sich nun auf die Videoanalyse übertragen?

'Durchführungsobjektivität'/Verfahrensdokumentation. In den Publikationen zu den einzelnen Videostudien finden sich kaum Hinweise zur Sicherung der Objektivität von Video-Instrumenten. Im Rahmen von Videoanalysen kann eine differenzierte, detaillierte und transparente Darstellung des gewählten Verfahrens als Voraussetzung dafür betrachtet werden, dass verschiedene Testanwender zwar nicht zu denselben, aber nach einer Schulung zumindest zu vergleichbaren Ergebnissen kommen. Bei der Beschreibung ihrer Videoverfahren greift die Arbeitsgruppe um Reusser (vgl. z.B. Hugener, 2006; Hugener, Pauli & Reusser, 2006) auf spezifische inhaltsanalytische Qualitätsansprüche zurück. Dazu gehört entsprechend den Ausführungen zum Entwicklungsprozess von Videoverfahren in Kapitel 1.5.1.2 als zentraler Schritt der inhaltlichen und datengeleiteten Kategorien- bzw. Konstruktentwicklung eine klare und präzise Definition und Operationalisierung der Kategorien und Konstrukte sowie die Aufstellung von Regeln zum Kodieren bzw. Raten.

Auswertungs- und Interpretationsobjektivität. Bei Videoanalysen ist trotz einer genauen Beschreibung des Verfahrens nicht zu vermeiden, dass individuelle Deutungen in die Interpretationen eines Testwertes einfließen. Gemäß den obigen Ausführungen kann die Frage, ob Beobachter bzw. Beurteiler in ihren Urteilen übereinstimmen, mit dem Verfahren der Beurteilerübereinstimmung eingeschätzt werden. Beurteilerüberstimmungen werden im Rahmen der Videoanalyse jedoch als Aspekt der Reliabilität betrachtet (nähere Ausführungen siehe weiter unten).

Reliabilität

Im Rahmen der klassischen Testtheorie gibt die Reliabilität (Zuverlässigkeit) „den Grad der Messgenauigkeit (Präzision) eines Instruments an“ (Bortz & Döring, 2006), d.h., „den Grad der Genauigkeit, mit dem ein Test ein bestimmtes Merkmal misst, unabhängig davon, ob er dieses Merkmal auch zu messen beansprucht“ (Bühner, 2006, S. 35). Jeder erhobene Wert (X) setzt sich wie bereits erwähnt aus einem wahren Wert (W) und einem Fehleranteil (F) zusammen: $X=W+F$ (Rost, 2005, S. 130). Je kleiner der Messfehler eines Messwertes, desto höher ist die Reliabilität, also die Messgenauigkeit. Ist der Test in der Lage, den wahren Wert ohne Messfehler zu erfassen, würde das eine perfekte Reliabilität bedeuten ($X=W$). Dies tritt in der Realität nicht auf, da sich Fehlereinflüsse (z.B. situative Störungen) nicht ausschließen lassen (vgl. Bortz & Döringer, 2006). Reliabilität lässt sich allgemein als Anteil der wahren Varianz an der beobachteten Varianz beschreiben. Je größer der Anteil der 'wahren' Varianz, umso geringer ist der Fehleranteil in den Testwerten (vgl. Bortz & Döring, 2006).

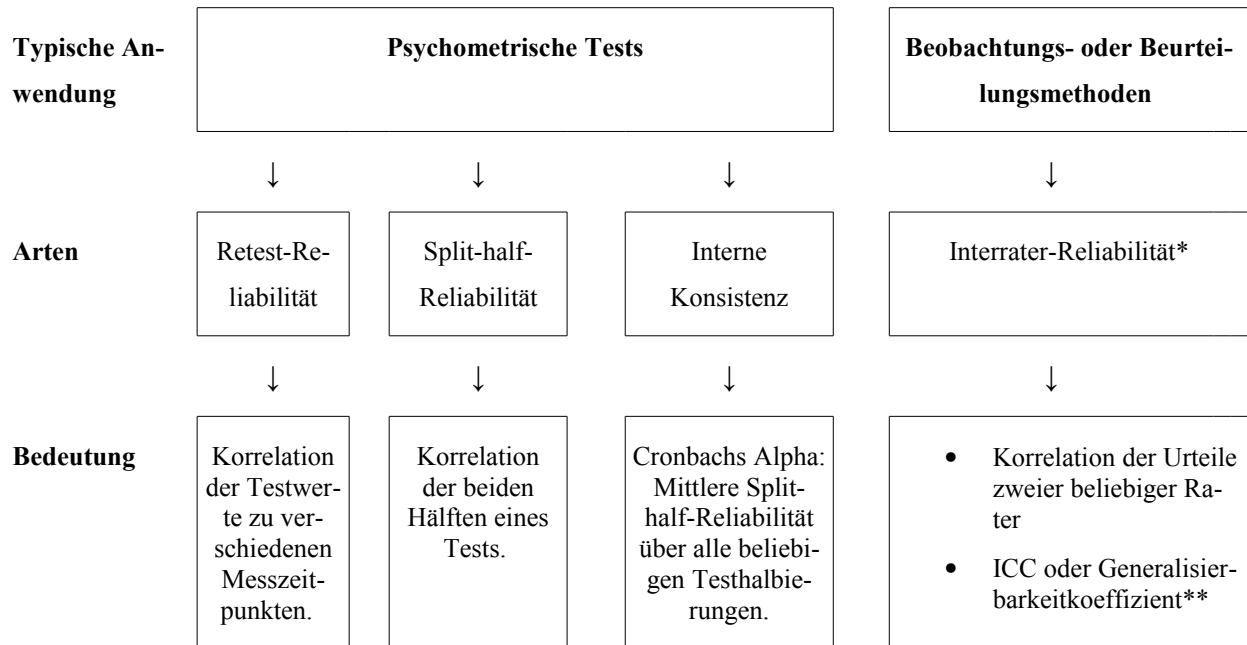
Die Zuverlässigkeit eines Tests kann mit verschiedenen Methoden ermittelt und mit unterschiedlichen Maßen quantifiziert werden. Für *psychometrische Tests* gehören die Retest-Reliabilität, die Split-half-

Reliabilität und die interne Konsistenz zu den wichtigsten Reliabilitätsarten (vgl. Bortz & Döring, 2006; Lienert & Raatz, 1994; Rost, 2005; Wirtz & Caspar, 2002; Bühner, 2006; vgl. auch Abbildung 1.5-1).

- *Retest-Reliabilität (Stabilität)*. Der Test wird zu zwei verschiedenen Testzeitpunkten durchgeführt und dann die Korrelation zwischen den beiden Testleistungen ermittelt. Zu beachten ist, dass die Korrelation in Abhängigkeit vom Zeitintervall zwischen den beiden Testleistungen variieren kann (vgl. Bühner, 2006); es ist also unerlässlich, die Größe des zeitlichen Abstands zwischen den Testungen anzugeben (vgl. Rost, 2005). Nachteile der Retest-Reliabilität sind u.a. die Überschätzung der Reliabilität, wenn die Lösungen von Testaufgaben erinnert werden, und der hohe zeitliche und untersuchungstechnische Aufwand (vgl. Bortz & Döring, 2006).
- *Split-half-Reliabilität (Testhalbierungsreliabilität)*. Der Test wird in gleiche Testhälften geteilt, die miteinander korreliert werden. Im Gegensatz zu der zuvor genannten Methode erfordert die Testhalbierungsmethode keinen untersuchungstechnischen Mehraufwand, sie hängt aber stark von der Art der zufälligen Testhalbierung ab (vgl. Bortz & Döring, 2006).
- *Interne Konsistenz*. Die Einschätzung der internen Konsistenz ist eine Erweiterung der Testhalbierungsmethode, die zu stabileren Schätzungen der Reliabilität führt. Interne Konsistenzschätzungen beruhen auf der Überlegung, „dass sich ein Test nicht nur in Testhälften, sondern in so viele 'kleinste' Teile zerlegen lässt, wie er Items hat. [...]. Die Korrelation zwischen den Items spiegelt dann die 'wahre Varianz' wider“ (Bortz & Döring, 2006, S. 198). Cronbachs Alpha (Cronbach, 1951) stellt das gebräuchlichste Maß der internen Konsistenz dar (vgl. Bortz & Döring, 2006).

Für *Beobachtungs- und Beurteilungsmethoden* stellt nach Wirtz und Caspar (2002) die Interrater-Reliabilität eine geeignete Reliabilitätsart dar. Die Abbildung 1.5-1 verdeutlicht die Einordnung der Interrater-Reliabilität. Gemäß den Autoren kann die Verbindung zur Interrater-Reliabilität dadurch hergestellt werden, „dass die Merkmalsausprägung einer Person nicht – wie im Falle psychometrischer Tests – durch die Werte auf mehreren Fragebogenitems geschätzt wird, sondern dass stattdessen eine Person Werte von mehreren Ratern zugewiesen bekommt“. Hier interessiert nicht, „ob die *Summe* aller Einschätzungen (Itemantworten) reliabel ist, sondern ob die Einschätzung durch *jeden einzelnen Rater* zuverlässig ist“ (S. 17f.).

Abb. 1.5-1 Überblick über die verschiedenen Arten von Reliabilitätsdefinitionen
(in Anlehnung an Wirtz & Caspar, 2002; vgl. auch Bortz & Döring, 2006)



* Reliabilität steht hier als übergeordneter Terminus für Übereinstimmung und Reliabilität.

** Generalisierbarkeitskoeffizient nach Cronbach et al. (1972); wird nicht bei Wirtz und Caspar (2002) genannt.

Der Terminus 'Reliabilität' steht in diesem Zusammenhang sowohl für Übereinstimmung als auch Reliabilität. Entsprechend den Begriffen werden zwei Arten von Zuverlässigkeitsmaßen unterschieden, Übereinstimmungs- und Reliabilitätskoeffizienten (vgl. Wirtz & Caspar, 2002):

- Maße der Übereinstimmungen eignen sich für die Berechnung der Übereinstimmung von Kategorien (nominalskaliert). Sie machen Aussagen darüber, inwiefern verschiedene Beobachter verschiedene Objekte exakt gleich beurteilen. Werden Objekte/Personen mittels nominalskalierter Kategoriensystemen (sich gegenseitig ausschließende, exakt definierte Kategorien) beurteilt, können die Häufigkeiten der Zuordnungen verglichen werden. Koeffizienten geben dann Informationen über die Gleichheit bzw. Ungleichheit der Kategorienzuordnung, also über den Grad der Übereinstimmung. Geeignete Maße sind die prozentuale Übereinstimmung (PÜ) und die Übereinstimmungsmaße Cohens κ (Kappa) und Scotts π (Pi), die auf der prozentualen Übereinstimmung basieren.
- Reliabilitätskoeffizienten hingegen werden für Einschätzungen auf Ratingskalen verwendet. Hier ist eine exakte Übereinstimmung nicht notwendig. Es werden Unterschiede zwischen verschiedenen Urteilern zugelassen; sind diese Unterschiede gering, kann von einer hohen Zuverlässigkeit ausgegangen werden. Das Ausmaß der Unterschiede ist also von Bedeutung; es wird die Intensität oder die graduelle Abstufung eines Merkmals eingeschätzt. Bei Reliabilitätsmaßen muss die relative Lage der Werte zum Mittelwert der Rater für die jeweiligen Personen ähnlich sein. Liegen intervallskalierte Ratingwerte vor, schlagen Wirtz und Caspar (2002) die

Intraklassenkorrelation (ICC), welche auf einem varianzanalytischen Modell beruht, als geeignetes Reliabilitätsmaß vor. Ein anderes Reliabilitätsmaß stellt der so genannte Generalisierbarkeitskoeffizient von Cronbach und Mitarbeitern (Cronbach, Gleser, Nanda & Rajaratnam, 1972) dar, die ebenfalls auf einem varianzanalytischen Modell beruht (siehe Abschnitt 3.5.2.1).

In der *qualitativen Forschung* werden die referierten Methoden der Reliabilitätsprüfungen abgelehnt. Es werden aber auch keine eigenen Methoden der Zuverlässigkeitsprüfung entwickelt (Bortz & Döring, 2006; Lamnek, 1995). Qualitative Forscher heben die Einzigartigkeit, Individualität und den situativen Kontext hervor, der eine Wiederholbarkeit der Messungen generell verbietet. Für qualitative Inhaltsanalysen ist gemäß Mayring und Brunner (2007) die so genannte Intrakoder-Reliabilität ein wichtiges Kriterium. Der Inhaltsanalytiker nimmt sich nach Beendigung der Auswertung Passagen des Materials noch einmal vor und vergleicht seine Kodierungen mit den ursprünglichen.

Welche der genannten Möglichkeiten der Reliabilitätsprüfungen lassen sich nun auf die Videoanalyse übertragen?

Interrater-Reliabilität. Entsprechend den o.g. Zuverlässigkeitsmaßen, ist es für die Prüfung der Beurteilerübereinstimmung bei niedrig-inferenten Kodierungen üblich, auf Cohens κ (Kappa) und die prozentuale Übereinstimmung (PÜ) zurückzugreifen (vgl. z.B. Clausen, Reusser & Klieme, 2003). Für die Berechnung der Qualität der Beurteilungen (Reliabilität der Beurteilungen durch die Rater) halten Clausen, Reusser und Klieme (2003) einen Rückgriff auf die Generalisierbarkeitstheorie von Cronbach und Mitarbeitern (Cronbach, Gleser, Nanda & Rajaratnam, 1972) für angemessen, da dieser Ansatz eine sehr komplexe Reliabilitätsanalyse erlaubt. Weitere Ausführungen zu diesem Ansatz folgen in 3.5.2.1. In vielen Videostudien folgt man Clausen und Mitarbeitern und berichtet für Ratingdaten Generalisierbarkeitskoeffizienten (z.B. Hugener, 2008; Knierim, 2008).

Interne Konsistenz. Bei Ratingverfahren, in denen die zu beobachtenden Dimensionen als Skalen angelegt werden und bei denen die Items likert-skaliert im mehrstufigen Antwortformat beurteilt werden, ist es möglich, die internen Konsistenzen mittels Cronbachs Alpha zu berechnen. Dies wird auch in Arbeitsgruppen, die ein derartiges Verfahren im Rahmen ihrer Videoanalysen gewählt haben, durchgeführt, wie z.B. bei Clausen (2002) und Clausen, Reusser und Klieme (2003).

Retest-Reliabilität (Stabilität). Nach Wirtz und Caspar (2002) ist es auch bei Beurteilungsverfahren (Ratingkalen) möglich, die Retest-Reliabilität zu berechnen. Dies entspricht dem Vorschlag von Mayring und Brunner (2007). Die Autoren halten dies jedoch aus den folgenden Gründen für problematisch: Retest-Reliabilitätsmaße haben das Problem, dass sie das Ausmaß angeben, in dem derselbe Rater bei wiederholten Messungen zu gleichen oder ähnlichen Ergebnissen kommt. Diese Maße beziehen sich nur darauf, inwiefern dieselbe Person die Skala stabil oder konsistent misst. Das Problem besteht darin, dass verschiedene Rater die Skala immer in gleicher Art und Weise unterschiedlich verste-

hen, so dass sie für sich im Urteil konsistent sein können ohne dass Reliabilität zwischen den Ratern besteht. Es kann also bei einer wiederholten Messung keine Aussage darüber gemacht werden, ob andere Rater zu ähnlichen Resultaten kommen. Zudem werden derartige Stabilitätsmaße leicht durch Erinnerungseffekte und bei zu großen Zeitintervallen zwischen den Messungen durch eventuelle Veränderungen der Merkmalsausprägungen in den Objekten verfälscht. Wirtz und Caspar (2002) halten deshalb die Berechnung von Koeffizienten, die sich auf die Reliabilität zwischen verschiedenen Ratern beziehen, für ausreichend. Die Berechnung der Retest-Reliabilität wird m. E. in Videoanalysen nicht praktiziert – vermutlich aus den oben genannten Gründen. Zumindest wird darüber in Veröffentlichungen nicht berichtet.

Validität

Die Validität (Gültigkeit) gibt das Ausmaß an, in dem ein Test das misst, was er messen soll. Trotz hoher Reliabilität kann ein Test nicht brauchbar sein, wenn er etwas anderes misst, als er zu messen vorgibt. Validität gilt somit als das wichtigste Gütekriterium. In der klassischen Testtheorie werden drei Hauptarten von Validität unterschieden (vgl. Bortz & Döring, 2006; Rost, 2005; Bühner, 2006), Inhalts-, Kriteriums- und Konstruktvalidität.

- *Inhaltsvalidität (Face-Validity)*. Inhaltsvalidität (auch Augenscheinvalidität oder logische Validität) ist gegeben, „wenn der Inhalt des Testitems das zu messende Konstrukt in seinen wichtigsten Aspekten erschöpfend erfasst“ (Bortz & Döring, 2006, S. 200). Die Inhaltsvalidität eines Tests kann nicht numerisch mittels eines Kennwerts bestimmt werden; sie beruht auf subjektiven Einschätzungen (vgl. Bortz & Döring, 2006; Bühner, 2006). Daher werden die Messwerte der Stichprobe eines Tests benutzt, um dessen Validität indirekt zu erfassen (Kriteriums- und Konstruktvalidität) (vgl. Bühner, 2006). „Streng genommen bestimmt man mit der Kriteriums- und Konstruktvalidität nicht die Validität eines Tests im eigentlichen Sinne, sondern die Validität der abgeleiteten Aussagen, welche mit Hilfe der Testkennwerte getroffen werden“ (Bühner, 2006, S. 36).
- *Kriteriumsvalidität*. Kriteriumsbezogene Validität „bezieht sich auf den Zusammenhang zwischen den empirisch gemessenen Ergebnissen des Messinstrumentes und einem anders gemessenen empirischen („externen“) Kriterium“ (Schnell, Hill & Esser, 2005, S. 155). Kriteriumsvalidität ist also nur bei einem geeigneten Außenkriterium sinnvoll (vgl. Bortz & Döring, 2006). Unterschieden werden hauptsächlich 'predictive validity' und 'concurrent validity'. Ein Test besitzt prädiktive Validität, wenn „Voraussagen, die auf einer ersten Messung mit dem Instrument beruhen, durch spätere Messungen mit einem anderen Instrument bestätigt werden können“ (Schnell, Hill & Esser, 2005, S. 155). Der Unterschied zur 'concurrent validity' (Übereinstimmungsvalidität) liegt darin, dass hier die Messungen zum selben Zeitpunkt erfolgen. Eine Möglichkeit der Bestimmung der 'concurrent validity' stellt nach Schnell, Hill und Esser (2005, S. 156) die 'Methode der bekannten Gruppen' ('known groups') dar. Die Zugehörigkeit zu

Gruppen ist das Kriterium. Für die Gruppen werden Unterschiede in der Ausprägung des zu messenden Konstrukts erwartet (vgl. Schnell, Hill & Esser, 2005; Bortz & Döring, 2006).

- *Konstruktvalidität.* Da die Inhaltsvalidität wenig aussagekräftig ist (kein objektivierbarer Kennwert) und die Kriteriumsvalidität selten anwendbar ist (Außenkriterium fehlt), kommt der Konstruktvalidität besondere Bedeutung zu (vgl. Bortz & Döring, 2006; Schnell, Hill & Esser, 2005). „Unter dem Begriff „Konstrukt“ werden theoretische Eigenschaftsdimensionen (latente Variablen) verstanden. Konstruktvalidität liegt dann vor, wenn aus dem Konstrukt empirisch überprüfbare Aussagen über Zusammenhänge dieses Konstrukts mit anderen Konstrukten theoretisch hergeleitet werden können und sich diese Zusammenhänge empirisch nachweisen lassen“ (Schnell, Hill & Esser, 2005, S. 156). Bortz und Döring (2006) sowie Bühner (2006) merken an, dass einige Autoren unter Konstruktvalidität auch inhaltsbezogene und kriterienbezogene Aspekte der Validität fassen. Im engeren Sinne gehören zur Konstruktvalidität nur die faktorielle (per Faktorenanalysen), konvergente und diskriminante Validität (z.B. per Multi-trait-Multimethod-Ansatz nach Campbell und Fiske (1959)). Eine weitere Möglichkeit der Validierung stellt die Expertenvalidierung (expert validity) dar. Das Erhebungsinstrument wird einem Experten vorgelegt, der den Grad der Gültigkeit des Messverfahrens auf der Basis seines Wissens, seiner wissenschaftlichen Erkenntnisse und aufgrund von Plausibilität prüft (vgl. Lamnek, 1995, S. 160).

Im Rahmen der *qualitativen Forschung* stellt die *konsensuelle Validierung* (interpersonale Konsensbildung) das bedeutendste Validitätskriterium dar. Als Indiz für Validität gilt, ob sich mehrere Personen auf den Bedeutungsgehalt des Materials einigen können. Dabei kann die Konsensfindung zwischen verschiedenen Personengruppen stattfinden: zwischen den Forschern eines Projektes (interpersonale Konsensbildung), zwischen Forschern und Beforschten (kommunikative Validierung), zwischen außenstehenden Laien und Experten (argumentative Validierung) (vgl. Bortz & Döring, 2006). Letzteres ähnelt der Expertenvalidierung (vgl. Lamnek, 1995).

In Bezug auf qualitative Inhaltsanalysen halten Mayring und Brunner (2007) für die Validität „vor allem (im Sinne der Konstruktvalidität) die theoriegeleitete Entwicklung der Kategoriensysteme und Kodierleitfäden (bei Strukturierungen) bzw. die theoriegeleitete Bestimmung von inhaltsanalytischen Regeln [für] ausschlaggebend“ (S. 678). Früh (1991, zit. n. Hugener, 2006, S. 52) erwähnt auch die Möglichkeit, inhaltsanalytisch erhobene Ergebnisse dadurch zu validieren, dass in die Berechnung der Beurteilerübereinstimmung nicht nur die geschulten Kodierer einbezogen werden, sondern auch der Forscher, welcher das Instrument entwickelt hat.

Welche der genannten Möglichkeiten der Validitätsprüfung lassen sich nun auf die Videoanalyse übertragen?

Kriteriumsvalidität. Kriteriumsbezogene Validität kann in Videoanalysen bei geeignetem Außenkriterium berechnet werden. In Videostudien wird darüber jedoch m. E. nicht berichtet.

Konstruktvalidität. Faktorielle, konvergente und diskriminante Validität können bei entsprechendem Datenformat berechnet werden. Dies geschieht in den einzelnen Videostudien z.T. auch. Eine Expertenvalidierung ist in jedem Fall möglich, wird aber kaum berichtet.

Konsensuelle Validierung. Eine interpersonale Konsensbildung muss in jeder Videostudie Berücksichtigung finden; Konsens zwischen den am Projekt beteiligten Personen ist ein wichtiger Bestandteil des Entwicklungsprozesses von Videoverfahren (vgl. Tabelle 1.5-2). Auch die anderen Arten der Konsensbildung sind denkbar, wobei Lamnek in Bezug auf die kommunikative Validierung anmerkt, dass ungeklärt bleibe, welche Kriterien für die Stimmigkeit zwischen Forschern und Beforschten gelten können (vgl. Lamnek, 1995).

Mayring und Brunner (2007) heben die Bedeutung der theoriegeleiteten Entwicklung von inhaltsanalytischen Instrumenten hervor. Dieser Aspekt ist ein wichtiger Bestandteil des in 1.5.1.2 dargestellten Verfahrens zur Entwicklung von Videoanalyse-Instrumenten und kann als Voraussetzung für die Entwicklung von Verfahren der Videoanalyse betrachtet werden. Dem Vorschlag von Früh (1991), die Validität der Kodierungen durch den Einbezug der Entwickler des Instruments in die Berechnung der Übereinstimmung zu erhöhen, kann leicht nachgekommen werden. Im Rahmen der schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“ wurde der Forderung nachgekommen, indem die Beurteilerübereinstimmung nach Abschluss des Trainings mit einer von der Entwicklergruppe im Konsensverfahren gebildeten Modellkodierung verglichen wurde (vgl. Hugener, 2006, S. 52).

1.5.2 Bisher bestehende Instrumente zur Erfassung kognitiv anregender Verhaltensweisen von Lehrkräften

Entsprechend den Ausführungen zu den Wirkungen eines kognitiv aktivierenden Mathematikunterrichts in Teilkapitel 1.1.3 sind Videoanalyse-Instrumente für die Sekundarstufe entwickelt worden, die kognitiv aktivierendes Lehrerverhalten im Mathematikunterricht zu erfassen versuchen. Eine Durchsicht der Instrumente zeigt, dass diese sich nur sehr bedingt für eine Anwendung auf die Grundschule bzw. den naturwissenschaftlichen Sachunterricht eignen. Zurückzuführen ist dies zum einen darauf, dass es sich bei kognitiv aktivierenden bzw. kognitiv anregenden Elementen des Unterrichts um fachspezifische Unterrichtsqualitätsmerkmale handelt, die entsprechend für jedes Fach und auch für jede Schulstufe neu zu bestimmen sind (vgl. dazu die Ausführungen in Teilkapitel 1.1). So erfassen z.B. Klieme, Schümer und Knoll (2001, S. 51) das Konstrukt der kognitiven Aktivierung u.a. mittels der Skalen „Anspruchsvolles Üben“, und „Repetitives Üben“, die Übungsformen im Mathematikunterricht beschreiben und somit speziell auf diesen ausgerichtet sind und sich nicht auf den Sachunterricht übertragen lassen. Es werden aber auch Konstrukte in der Literatur verwendet, die Anregungen für das in dieser Arbeit zu entwickelnde Instrument liefern können bzw. die für den Sachunterricht adaptierbar sind. Dazu zählen Konstrukte wie „Motivierungsfähigkeit des Lehrers“, „Fehlerkultur“ oder „Argumentatives Aushandeln von Ideen“, wie sie von Clausen, Reusser und Klieme (2003) beschrieben wer-

den. Eine ausführliche Darstellung der für die Sekundarstufe entwickelten Instrumente wird an dieser Stelle nicht vorgenommen. Stattdessen werden die Instrumente anderer Arbeitsgruppen, die Anregungen für die Entwicklung des im Rahmen dieser Arbeit vorgestellten Instruments geben konnten, bei der Darstellung des entwickelten Instruments im Methodenteil genannt.

Für die Grundschule bzw. für den naturwissenschaftlichen Sachunterricht existierte m. E. zum Zeitpunkt der Entwicklung des hier vorgestellten Instruments nur ein Instrument, welches speziell Lehrhandlungen im naturwissenschaftlichen Sachunterricht erfasst (Gais, 2009).

Dabei handelt es sich um ein kategorienbasiertes Vorgehen, welches Lehrerverhalten, wie z.B. das Auffordern der Schüler zum Äußern von Ideen oder Hypothesen und zum Begründen von Aussagen durch die Lehrperson erfasst. Auch dieses Instrument konnte Anregungen für die Entwicklung eines hoch-inferentes Schätzverfahrens, um das es sich bei dem vorgestellten Instrument handelt (siehe dazu die Ausführungen im Rahmen der Fragestellung und in den Methoden) geben.

1.5.3 Zusammenfassung

Die Ausführungen zu Methoden der Videoanalyse haben gezeigt, dass das hauptsächliche Unterscheidungsmerkmal bei Videoverfahren der Grad der Inferenz ist, d.h. der Grad der vom Beobachter geforderten interpretativen Schlussfolgerungen (vgl. Hugener, 2006, 2008; Clausen, Reusser & Klieme, 2003; Rosenshine, 1970). Unterschieden werden niedrig-inferente Beobachtungen bis hoch-inferente Beurteilungen (vgl. Clausen, Reusser & Klieme, 2003). Niedrig-inferente Beobachtungen ermöglichen die quantitative Erfassung einzelner Ereignisse und die Beschreibung konkret beobachteter Unterrichtsprozesse. Sie werden meist kategorienbasiert im event- oder time-sampling erhoben. Qualitätsaussagen zu Unterrichtsprozessen sind bei dieser Methode allerdings nur begrenzt möglich. Hoch-inferente Beurteilungen hingegen eignen sich besonders dafür, komplexe zusammenhängende Merkmale oder Prozesse zu erheben und gleichzeitig zu bewerten. Sie werden als Rating erfasst und ermöglichen somit eine qualitative Einschätzung im Sinne eines Gesamteindrucks (vgl. Hugener, 2006; Rakoczy & Pauli, 2006). Die Wahl für eine Methode orientiert sich also an der Fragestellung. In Bezug auf die Güte der Videoverfahren liefern niedrig-inferentere Kodierungen meist die reliableren Messungen, hoch-inferente Ratings gelten dahingegen als die validere Methode.

Die Entwicklung von Videoanalyse-Instrumenten erfolgt theorie- und datengeleitet nach den Bedingungen qualitativer Inhaltsanalysen (vgl. Bos & Tarnai, 1999). Das Ziel der Entwicklung von Videoverfahren ist die Typisierung von Beobachtungsmerkmalen (qualitativ) und die Quantifizierung des Datenmaterials.

Videoanalyse-Instrumente bedürfen verschiedener Qualitätskontrollen, um forschungsmethodischen Ansprüchen zu genügen. Zur Güteprüfung von Videoanalyse-Instrumenten gibt es allerdings keine fest vorgeschriebenen Konventionen der Testprüfung, wie sie z.B. im Rahmen der klassischen Testtheorie für psychometrische Tests üblich sind. Die in diesem Kontext genannten Methoden zur Prüfung der

Objektivität, Reliabilität und Validität können aber in modifizierter Form auch für die Videoanalyse verwendet werden.

Abschließend lässt sich bemerken, dass es für die Grundschule kein Videoanalyse-Instrument gibt, welches die in 1.4 herausgearbeiteten Lehrerverhaltensweisen hoch-inferent erfasst.

2 Zielsetzung und Fragestellung

Vor dem Hintergrund der Forderung nach einem anspruchsvollen Sachunterricht vor allem im Vorfeld eines naturwissenschaftsbezogenen Lernens und in Anbetracht der Bedeutung, die fachspezifische Unterrichtsqualitätsmerkmale für den Lernerfolg der Lernenden bedeuten können, ist das Ziel der Arbeit, kognitiv anregende Lehrerverhaltensweisen im naturwissenschaftsbezogenen Sachunterricht zu konzeptualisieren und zu erfassen.

Konzeptualisierung kognitiv anregenden Lehrerverhaltens. Auf der Basis von Hinweisen zur Gestaltung kognitiv anregenden Unterrichts aus dem Bereich der Instruktionspsychologie, die für das naturwissenschaftliche Lernen von Bedeutung sind, und aus Konzeptionen des Sachunterrichts sind Folgerungen für das Lehren eines anspruchsvollen, kognitiv anregenden naturwissenschaftsbezogenen Sachunterrichts herausgearbeitet worden. Dazu gehören Lehrerverhaltensweisen (1) zur Anregung aktiven und konstruktiven Lernens (Motivation zur Auseinandersetzung mit dem Lerngegenstand schaffen, Gelegenheiten zu eigenen Wissenskonstruktionen schaffen und dabei individuelle Lernwege zulassen, Transmission von Wissen vermeiden, Anknüpfen an Präkonzepte und Anregung von Conceptual Change), (2) zur Umsetzung sozialen und kooperativen Lernens (gemeinsame Diskussionen), (3) zur Umsetzung unterstützten Lernens (Anleitung und Unterstützung der Lernenden) und zur Umsetzung situierten Lernens (authentische Kontexte und Anwendungsbezüge).

Erfassung kognitiv anregenden Lehrerverhaltens. Um die herausgearbeiteten Verhaltensweisen von Lehrkräften in Folgestudien auf ihre Wirksamkeit hin untersuchen zu können, und da es bisher für die Grundschule kein Instrument gibt, welches derartige Lehrerverhaltensweisen umfassend misst, ist es das Hauptziel dieser Arbeit, ein solches Videoanalyse-Instrument zu entwickeln und hinsichtlich test-theoretischer Gütekriterien zu analysieren.

Für die Entwicklung eines Videoanalyse-Instruments stehen niedrig- bis mittel-inferente kategorienbasierte Vorgehen und hoch-inferente Schätzverfahren zur Verfügung. Die Wahl für ein Verfahren der Videoanalyse hängt gemäß den Ausführungen in Abschnitt 1.5.1 im Wesentlichen von den Inhaltsbereichen, die mit dem Instrument beobachtet bzw. beurteilt werden sollen, und von den Forschungsfragen ab.

Inhaltsbereiche. Das Ziel dieser Arbeit ist, kognitiv anregende Verhaltensweisen von Lehrkräften mit Hilfe eines Videoanalyse-Instruments zu erfassen. Es sollen Aussagen über die Qualität des Lehrerhandelns im Unterricht gemacht werden. Dabei handelt es sich um Verhaltensweisen von Lehrkräften, die oft erst durch das komplexe Zusammenwirken struktureller und inhaltlicher Merkmale des Unterrichts sowie durch soziale Interaktionen der Lehrkräfte und Schüler entstehen. Es müssen dabei verschiedene Kriterien gleichzeitig analysiert und bewertet werden. Aus diesen Gründen bietet es sich an, auf hoch-inferente Ratings zurückzugreifen.

Forschungsfrage. In dieser Arbeit steht die Entwicklung des Instruments im Vordergrund. Das Instrument wurde bzw. wird natürlich auch zur Untersuchung von Fragestellungen bzw. zur Hypothesenprüfung eingesetzt. Wie bereits erwähnt wurde es im Rahmen des Münster'schen DFG-Projekts zur Untersuchung von Zusammenhängen zwischen Vorstellungen von Lehrkräften zum Lehren und Lernen und deren Handlungen im Unterricht genutzt (vgl. Vehmeyer, Kleickmann & Möller, 2007; Kleickmann, Vehmeyer & Möller, i. D.). Des Weiteren ist geplant, das Instrument zur Erfassung von Zusammenhängen zwischen unterrichtlichen Lehrerhandlungen und Schülerleistungen einzusetzen. Studien zeigen, dass hoch-inferente Beurteilungen meist höhere Zusammenhänge zu schulischen Erfolgs- und Entwicklungskriterien aufweisen als niedrig-inferente Kodierverfahren (vgl. die Ausführungen in Abschnitt 1.5.1.1).

Insgesamt sprechen sowohl die mit dem Instrument zu beurteilenden Inhaltsbereiche als auch die Fragestellungen aus dem Projekt für die Entwicklung eines hoch-inferenten Rating. Um das Instrument für Hypothesenprüfungen einsetzen zu können, muss es forschungsmethodischen Ansprüchen genügen. Dabei steht im Rahmen der Testkonstruktion vor allem die Reliabilität und Validität im Zentrum des Interesses. Somit kann die folgende Fragestellung formuliert werden:

Frage: Ist es möglich, ein hoch-inferentes Videoanalyse-Instrument zu entwickeln, welches kognitiv anregende Verhaltensweisen von Lehrkräften objektiv, reliabel und valide erfasst?

Die Ausführungen in Teilkapitel 1.5.1.1 haben gezeigt, dass hoch-inferente Beurteilungen, die über konkret beobachtbares Verhalten hinausgehen, besonders anfällig für systematische und unsystematische Beurteilungsfehler sind. Andererseits haben hoch-inferente Schätzverfahren oft einen stärkeren Bezug zur Theorie und liefern durch die ganzheitliche Beobachtungen latenter Konstrukte recht valide Ergebnisse.

In Bezug auf die Analyse der Güte des Instruments ist somit zu erwarten, dass die Ergebnisse bezüglich der Untersuchung der Qualität der Beurteilungen, also der Interrater-Reliabilität, tendenziell eher niedrig ausfallen. Dagegen steht zu vermuten, dass das Instrument die Konstrukte valide misst.

3 Methoden

Aufgabe dieser Arbeit ist, ein Instrument zur Erfassung kognitiv anregenden Verhaltens von Grundschullehrkräften im naturwissenschaftlichen Sachunterricht zu entwickeln. Dies geschieht auf der Basis von Unterrichtsvideos zu den Themen „Wasserkreislauf“ und „Schwimmen und Sinken“. In diesem dritten Kapitel wird in einem ersten Schritt die Herkunft der Daten beschrieben (Teilkapitel 3.1). Daran anschließend folgen die Erläuterung der Untersuchungsanlage (Teilkapitel 3.2), die Beschreibung der Stichprobe (Teilkapitel 3.3), die Darstellung des entwickelten Videoanalyse-Instruments (Teilkapitel 3.4) und die Beschreibung der in der Arbeit genutzten Methoden zur Prüfung der Güte des Instruments (Teilkapitel 3.5).

3.1 Herkunft der verwendeten Daten

Die in dieser Arbeit verwendeten Daten stammen aus der in der Einleitung genannten Münster'schen DFG-Interventionsstudie aus dem Schwerpunktprogramm Biqua. Die Studie untersuchte im dritten Antragszeitraum des Schwerpunktprogramms (2004-2006) die Wirksamkeit von Lehrerfortbildungen im Bereich des naturwissenschaftlichen Sachunterrichts. In der Studie wurde zum einen der Frage nach Auswirkungen verschiedener Fortbildungskonzepte im naturwissenschaftsbezogenen Sachunterrichts auf Aspekte des professionellen Wissens von Lehrenden und auf motivationale und selbstbezogene Variablen seitens der Lehrenden nachgegangen (Möller et al., 2006; Kleickmann, Möller & Jonen, 2006). Zum anderen wurden Zusammenhänge des Wissens der Lehrenden mit Zielkriterien auf Ebene der Lernenden (Kleickmann, 2008) und mit deren Handlungen im Unterricht untersucht (Vehmeier, Kleickmann & Möller, 2007). Für letzteres wurde das entwickelte Instrument eingesetzt²⁹.

Die Stichprobe der vorliegenden Arbeit stellt einen Teil der Lehrkräfte dar, die im Rahmen der Studie an Fortbildungen zum naturwissenschaftlichen Sachunterricht (16 ganztägige Sitzungen) partizipierten. Es nahmen insgesamt 54 Lehrkräfte an der Intervention teil, die sich zu je 18 Lehrkräften auf drei verschiedene Fortbildungsmodule verteilten (zwei Experimentalgruppen, die an einer Präsenzfortbildung mit tutorieller Unterstützung teilnahmen; eine Kontrollgruppe erarbeitete sich die Inhalte der Fortbildung im Selbststudium). Auf die Inhalte und die methodische Gestaltung der Fortbildung, welche ausführlich in Möller et al. (2006) beschrieben sind, wird an dieser Stelle nicht näher eingegangen, da dies im Rahmen dieser Arbeit nicht notwendig ist.

Die Tabelle 3.1-1 zeigt die Untersuchungsanlage des Projekts in vereinfachter Form, welche die Einbettung der im Rahmen dieser Arbeit verwendeten Daten in die DFG-Studie veranschaulicht. Die Untersuchungsanlage des Projekts wurde ebenfalls an anderer Stelle ausführlich erläutert (vgl. z.B. Mül-

²⁹Vgl. dazu auch die Ausführungen in Abschnitt 5.4.

ler et al., 2006). An dieser Stelle wird deshalb nur auf die Zusammenstellung der Videostichprobe eingegangen.

Tab. 3.1-1 Vereinfachte Untersuchungsanlage der DFG-Interventionsstudie (vgl. Möller et al., 2006; Kleickmann, 2008)

09/03 - 02/04			03/04 - 07/04	09/04 - 11/05				
Parallelisierungsfragebogen	Unterrichtsbeobachtung „Wasserkreislauf“ Videoanalyse & Fragebogen 'constraints'	Lehrerfragebogen	Intervention N=54	Lehrerfragebogen	Schülerfragebogen zum konzeptuellen Verständnis von „Schwimmen und Sinken“	Unterrichtsbeobachtung „Schwimmen und Sinken“ Videoanalyse & Fragebogen 'constraints'	Schülerfragebogen zum konzeptuellen Verständnis von „Schwimmen und Sinken“	Lehrerfragebogen
	MZP 1	MZP 1		MZP 2	MZP 1	MZP 2	MZP 2	MZP 3
			Baselinegruppe N=19					

Für die Unterrichtsbeobachtungen wurden je 10 Lehrkräfte der drei fortgebildeten Gruppen ausgewählt. Die Auswahl der Lehrkräfte geschah nach Extremgruppenkriterien: Aus jeder der drei Gruppen wurden die Teilnehmer mit den höchsten und niedrigsten Ausprägungen in Skalen zu Vorstellungen zum Lehren und Lernen sowie zur Selbstwirksamkeitsüberzeugung bezüglich des Unterrichtens naturwissenschaftlicher Inhalte ausgewählt. Die Skalen und die den Skalen zugrunde liegenden Konstrukte sind erläutert in Kleickmann (2008). Die 30 Lehrkräfte wurden gebeten, selbst Unterricht in einer dritten oder vierten Jahrgangsklasse durchzuführen; jeweils einmal vor der Fortbildung zum Thema „Wasserkreislauf“ und einmal nach der Fortbildung zum Thema „Schwimmen und Sinken“.

Die Fortbildungsteilnehmer wurden über eine Ausschreibung der Fortbildung durch die Bezirksregierung Münster gewonnen. Aus insgesamt 96 interessierten Lehrkräften wurden 54 Personen für die Teilnahme an den Fortbildungen per Parallelisierungsverfahren ausgewählt.

3.2 Untersuchungsanlage

Für die Entwicklung des Videoanalyse-Instruments standen insgesamt 59 Unterrichtsvideos zur Verfügung. Von allen 30 oben genannten Lehrkräften liegt jeweils eine videographierte Unterrichtseinheit zum „Wasserkreislauf“ vor. Einer Lehrkraft war es aus organisatorischen Gründen nicht möglich, Unterricht zum „Schwimmen und Sinken“ durchzuführen, so dass zu diesem Thema 29 Videoaufzeichnungen vorliegen.

Datenerhebung. Die Unterrichtsaufzeichnungen zum Thema „Wasserkreislauf“ fanden von September 2003 bis Februar 2004 statt, die zum Thema „Schwimmen und Sinken“ von September 2004 bis Februar 2005. Die Aufzeichnung der Videos erfolgte nach standardisierten Richtlinien, um eine Vergleichbarkeit der Unterrichtsstunden zu gewährleisten. In einem Manual wurden in Anlehnung an das Kameraskript der TIMSS 1995 Videostudie alle für das Filmen wichtigen Regeln festgehalten, wie z.B. die technische Vorbereitung der Aufnahmen oder Kamerapositionen. Nur so können Unterrichtsvideos gemäß Petko (2006) für wissenschaftliche Zwecke genutzt werden³⁰.

Themenauswahl der Videoaufzeichnungen. Die Themen „Wasserkreislauf“ und „Schwimmen und Sinken“ wurden gewählt, da es sich hierbei um sehr verbreitete Themen im Sachunterricht handelt, die außerdem in den Richtlinien stehen. Zudem können diese Themen auf vielfältige Art und Weise umgesetzt werden. Eine intensive Auseinandersetzung mit den Inhalten in Form von Schülerexperimenten, in Form von anspruchsvollen, das Denken der Kinder anregenden Gesprächen oder in Form eines konstruktiven Arbeitens mit und an den (Fehl-)Konzepten der Schüler ist möglich, genau wie oberflächliche oder belehrende Abhandlungen. In jedem Fall stellen beide Themen hohe Anforderungen an das Handeln der Lehrkräfte im Unterricht.

Durchführung des Unterrichts. Um eine gewisse Vergleichbarkeit der Unterrichtsstunden sicherzustellen, erhielten die Probanden einige Instruktionen zur Durchführung ihres Unterrichts. Die Instruktionen waren bei beiden Unterrichtsthemen dieselben: Die Lehrkräfte wurden gebeten, den Unterricht in einem dritten oder vierten Schuljahr durchzuführen. Es wurde jeweils die erste Unterrichtseinheit der Themen „Wasserkreislauf“ (Regen, Wolken, Verdunstung, Kondensation) und „Schwimmen und Sinken“ (Schwimmverhalten von Vollkörpern) erhoben. Den Lehrkräften wurde vorgegeben, dass die gezeigte Unterrichtseinheit einen zeitlichen Umfang von 60 bis 90 Minuten haben und mindestens die Anteile Hinführung, Erarbeitung und Auswertung beinhalten sollte. Zusätzlich fertigten die Lehrkräfte beim Thema „Wasserkreislauf“ schriftliche Unterrichtsplanungen an. Beim Thema „Schwimmen und Sinken“ wurden den Lehrkräften Materialboxen mit einer Lehrerhandreichung bereitgestellt, um die materielle Ausstattung für den Unterricht konstant zu halten. Die Materialboxen enthielten sämtliche für das Unterrichten notwendige Experimentier- und Demonstrationsmaterialien. Das Handbuch beinhaltete Informationen über den fachlichen Hintergrund, Vorschläge für die Unterrichtsgestaltung, methodisch-didaktische und organisatorische Hilfen für den Unterricht und Hinweise auf typische Schülerleistungen und Lernschwierigkeiten (vgl. Jonen & Möller, 2005). Zudem wurden die Lehrkräfte direkt im Anschluss an die Videoaufzeichnung mittels Fragebogen um eine Einschätzung über wahrgenommene situative Besonderheiten der gezeigten Stunde, allgemeine Klassenmerkmale und Lernvoraussetzungen der Schüler (so genannte 'constraints') befragt.

³⁰Zur Datenaufbereitung: Die aufgezeichneten Unterrichtsstunden wurden digitalisiert im Format mpeg1 auf einem Videoserver und auf CD gespeichert. Zudem wurden die Gesprächsphasen aller Unterrichtsvideos auf der Basis eines Transkriptionsmanuals transkribiert.

3.3 Stichprobenbeschreibung

Die Videostichprobe der Lehrkräfte wird im Folgenden näher beschrieben.

Geschlechterverteilung. Der Anteil der weiblichen Probanden beträgt 90.0% (N=27), der Anteil der männlichen Probanden 10.0% (N=3). Dies entspricht der durchschnittlichen Geschlechterverteilung in Grundschulen in Nordrhein-Westfalen.

Altersverteilung. Das Alter der 30 Probanden liegt im Durchschnitt bei 46.67 Jahren (SD=8.12 Jahre; Range von 27 bis 56 Jahre). Am stärksten vertreten ist die Gruppe der 45- bis 54-jährigen mit über der Hälfte der Lehrkräfte. Insbesondere die jüngste und älteste Gruppe sind nur in geringer Stärke vertreten. Dagegen zeigt sich bei den landesweiten Zahlen (NRW-Population für das Schuljahr 2008/2009) eine relativ gleichmäßige Verteilung über die Altersgruppen ab 25 Jahren bei leicht zunehmender Tendenz mit aufsteigendem Alter. Auf Landesebene liegt der Altersdurchschnitt bei geschätzten 46.21 Jahren³¹.

Abb. 3.3-1 Alter Lehrkräfte NRW-Population

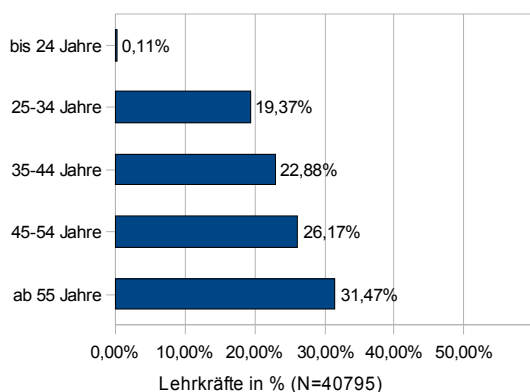
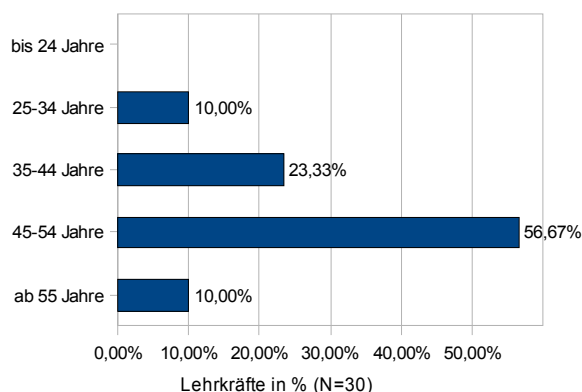


Abb. 3.3-2 Alter der Lehrkräfte Untersuchungsstichprobe



Berufserfahrung. Bei den Lehrkräften der Videostichprobe reicht die Berufserfahrung (in Dienstjahren) von zwei bis 34 Jahren. Dabei handelt es sich überwiegend um erfahrene Lehrkräfte, mit einer durchschnittlichen Berufserfahrung von 22.10 Jahren (SD=8.86 Jahre).

Bezüglich der Geschlechterverteilung und des Durchschnittsalters kann die Videostichprobe als repräsentativ gelten, wobei die Lehrkräfte ab 45 Jahre und älter in der Untersuchungsstichprobe mit 66.67% etwas stärker vertreten sind als in der NRW-Population (57.64%). Um weitere Anhaltspunkte zur Re-

³¹Da für die NRW-Population die Anzahl der Lehrkräfte lediglich pro Altersgruppe (5-Jahres-Intervall) vorlag, wurde das Durchschnittsalter approximiert, indem angenommen wurde, dass das Durchschnittsalter in jeder Altersgruppe dem Mittelpunkt der 5-Jahres-Intervalle entspricht (vgl. Anhang C.1), URL: <http://www.it.nrw.de/statistik/d/daten/eckdaten/r313-schul2.html>, letzter Aufruf am 04.05.09.

präsentativität der Videostichprobe der 30 Lehrkräfte zu erhalten, wurde diese mit einer größeren Stichprobe von Grundschullehrkräften (N=271) verglichen, welche für Grundschullehrkräfte in Nordrhein-Westfalen als weitgehend repräsentativ betrachtet werden kann (vgl. Möller, 2004a). Ein Vergleich erfolgte auf der Basis von Interessen und Selbsteinschätzungen zu Physik und physikbezogenem Unterricht³² (siehe Tabelle 3.3-1). Um die Abweichungen der Untersuchungsstichprobe von der NRW-Stichprobe bestimmen zu können, wurden Effektstärken berechnet. Diese entsprechen dem Verhältnis der Differenz aus den Mittelwerten der NRW-Stichprobe und der Untersuchungsstichprobe zur Standardabweichung der NRW-Stichprobe (vgl. Kleickmann, 2008). Gemäß Bortz und Döring (2006) sind derart berechnete Effektgrößen mit einem Wert von .20 als 'klein', von .50 als 'mittel' und von .80 als 'groß' zu bezeichnen.

Ein Vergleich der Stichproben erfolgt zu beiden Messzeitpunkten (MZP) der Videostichprobe (vor der Intervention 'MZP 1' und nach der Intervention 'MZP 2'; vgl. 3.1-1).

Tab. 3.3-1 Vergleich der Untersuchungsstichprobe (N=30) mit einer für NRW weitgehend repräsentativen Stichprobe von Grundschullehrkräften (N=271)

Kriterium	NRW-Stichprobe (N=271) M (SD)	Videostichprobe		Effektstärke	
		vor der Intervention (MZP 1) (N=30) M (SD)	nach der Intervention (MZP 2) (N=29) M (SD)	vor der Intervention (MZP 1)	nach der Intervention (MZP 2)
Fachinteresse Physik*	M=1.58 (SD=1.18)	M=1.43 (SD=1.16)	M=1.39 (SD=1.18)	.13	.16
Sachinteresse an Physik	M=2.53 (SD=.82)	M=2.58 (SD=.69)	M=2.80 (SD=.58)	-.06	-.34
Fähigkeitsselbstkonzept Physik	M=1.96 (SD=.82)	M=1.60 (SD=.67)	M=1.86 (SD=.69)	.44	.12
Selbstwirksamkeitserwartungen bezüglich des Unterrichtens physikbezogenen Sachunterrichts	M=2.58 (SD=.91)	M=2.42 (SD=.72)	M=2.90 (SD=.68)	.18	-.35
Interesse am Unterrichten physikbezogenen Sachunterrichts	M=2.88 (SD=.90)	M=3.08 (SD=.83)	M=3.41 (SD=.61)	-.22	-.59

Antwortformat von 0 bis 4 ((0) *stimmt gar nicht* bis (4) *stimmt völlig*).

Die Konstrukte und Skalen sind nachzulesen in Kleickmann (2008).

Zum Zeitpunkt der Unterrichtsbeobachtung „Wasserkreislauf“ (vor der Intervention) sind das Interesse am Schulfach Physik (Fachinteresse) und das Interessens an der Sache Physik (Sachinteresse) in beiden Stichproben ähnlich ausgeprägt (geringes Fachinteresse, mittleres Sachinteresse). Eine mittlere Abweichung der Untersuchungsstichprobe von der NRW-Stichprobe zeigt sich bei den Beurteilungen der eigenen Fähigkeiten bezüglich physikalischer Inhalte (Fähigkeitsselbstkonzept). Der Mittelwert der Untersuchungsstichprobe ist deutlich niedriger als der ohnehin schon relativ niedrige Mittelwert der

³²Zu den Skalen und den Konstrukten, die den Skalen zugrunde liegen, siehe Kleickmann (2008).

Vergleichsstichprobe. Hinsichtlich des Interesses am Unterrichten physikbezogenen Sachunterrichts und in Bezug auf die Selbstwirksamkeitserwartungen bezüglich des Unterrichts physikbezogenen Sachunterrichts ergeben sich kleine Abweichungen der Untersuchungsstichprobe von der repräsentativen Stichprobe, wobei in der Untersuchungsstichprobe das Mittel der Selbstwirksamkeitserwartungen etwas niedriger und der Mittelwert des Interesses am Unterrichten physikalischen Sachunterrichts leicht höher ist als in der NRW-Stichprobe. Das geringe Fähigkeitsselbstkonzept bezüglich Physik und das relativ hohe Interesse der Lehrkräfte am Unterrichten naturwissenschaftsbezogenen Sachunterrichts erklärt die freiwillige Teilnahme an den Fortbildungen.

Insgesamt kann die Untersuchungsstichprobe bis auf das Fähigkeitsselbstkonzept aber noch als der repräsentativen Stichprobe ähnlich betrachtet werden. Es kann davon ausgegangen werden, dass die Unterrichtsvideos zum „Wasserkreislauf“ in etwa der Qualität des naturwissenschaftlichen Sachunterrichts zu diesem Thema in Grundschulen entsprechen.

Zum Zeitpunkt der Unterrichtsbeobachtung „Schwimmen und Sinken“ (nach der Intervention) zeigen sich durch den Fortbildungseinfluss in fast allen Vergleichskriterien mittlere bis stärkere Abweichungen der Videostichprobe von der NRW-Stichprobe zu Gunsten der Videostichprobe. Vor dem Hintergrund des möglichen Einflusses der Fortbildung auf die Planung und Durchführung des Unterrichts sowie hinsichtlich der Tatsache, dass den Lehrkräften Materialien für den Unterricht zur Verfügung gestellt wurden, handelt es sich bei den Videos zum „Schwimmen und Sinken“ sicherlich um keine repräsentativen Unterrichtsbeobachtungen.

3.4 Rating-Instrument zur Erfassung kognitiv anregender Handlungen von Lehrkräften

Im Rahmen dieses Teilkapitels wird das entwickelte Videoanalyse-Instrument zur Erfassung kognitiv anregenden Lehrerverhaltens, ein hoch-inferentes Rating, beschrieben. In einem ersten Abschnitt (3.4.1) wird die Konstruktion des Ratingverfahrens (inhaltliche und methodische Entwicklung) beschrieben. Daran anschließend werden die abgeleiteten und operationalisierten Ratingdimensionen vorgestellt (3.4.3).

3.4.1 Konstruktion des Ratingverfahrens

Die inhaltliche und methodische Entwicklung des Videoanalyse-Instruments erfolgte in Anlehnung an die in Abschnitt 1.5.1.2 dargestellte Methode zur Entwicklung von Videoverfahren (Jacobs et al., 1999; Bos & Tarnai, 1999).

3.4.1.1 Theoretische Basis und Entwicklung des Videoverfahrens

Theoretische Basis. Die inhaltliche Entwicklung der Ratingdimensionen erfolgte auf dem Stand der berichteten Forschung zum Lernen und Lehren (Teilkapitel 1.2 und 1.3) und vor dem Hintergrund der o.g. Projektfragestellungen.

Inhaltliche Entwicklung der Ratingdimensionen. Es wurden insgesamt 14 Konstrukte theoretisch beschrieben und durch Items als Indikatoren operationalisiert. Zur Konstruktion der Items wurden vor allem konzeptwechsel- und denkfördernde Handlungen von Lehrkräften aus der Theorie abgeleitet und auf der Beobachtungsebene abgebildet. Dabei fanden im Sinne des in Abschnitt 1.5.1.2 dargestellten zyklischen Verfahrens zur Entwicklung von Kodier- und Ratingsystemen Rückkopplungen mit dem Datenmaterial statt. Zudem erfolgte die Entwicklung der Konstrukte weitgehend analog zu einem von T. Kleickmann erstellten Fragebogen, der Vorstellungen zum Lehren und Lernen im naturwissenschaftsbezogenen Sachunterricht erfasst. Die ausführlich beschriebene Entwicklung des Fragebogens ist nachzulesen in Kleickmann (2008). Außerdem wurden bei der Entwicklung des Rating-Instruments Anregungen aus bereits bestehenden Instrumenten zur Erfassung des Lehrerhandelns anderer Arbeitsgruppen berücksichtigt. Impulse kamen vor allem aus den Arbeiten von Clausen, Reusser und Klieme (2003) und teilweise auch von Kobarg und Seidel (2003), Rakoczy und Pauli (2006), Gais (2009) und Heinze (2004) (vgl. 1.5.2). Die Beschreibung der Beobachtungsdimensionen erfolgt in Abschnitt 3.4.3.

Festlegung der Analyseeinheit. Als Analyseeinheit wurde bei dem Rating der Unterrichtsvideos die ganze Unterrichtsstunde (49 bis 90 Minuten) gewählt. Der Grund dafür ist der folgende: Bei den mittels dieses Instruments zu bewertenden Konstrukten handelt es sich um Merkmale, die oftmals nicht ohne Berücksichtigung des gesamten Arrangements, d.h. ohne Nachvollzug des vollständigen Unterrichtsverlaufs beurteilt werden (vgl. Abschnitt 1.5.1.1; vgl. auch Rakoczy & Pauli, 2006, S. 210; Reyer, 2004, S. 122). Zudem geht es um die Einschätzung von Lehrerverhalten im Sinne eines Gesamteindrucks.

Antwortformat. Das Antwortformat des Instruments besteht aus einer vierstufigen Likert-Skala von (1) *trifft überhaupt nicht zu* über (2) *trifft eher nicht zu*, (3) *trifft eher zu* bis (4) *trifft voll und ganz zu*. Einzuschätzen sind alle Items eines Konstrukts. Zusätzlich gibt es zu jedem Konstrukt eine Gesamteinschätzung.

Bei der Einschätzung der Merkmale auf einer mehrstufigen Skala können gemäß Rakoczy und Pauli (2006) Unsicherheiten auftreten, da einem Urteil verschiedene Dimensionen zugrunde liegen können, wie z.B. die Häufigkeit eines einzelnen gezeigten Verhaltens oder auch die Intensität bzw. Ausprägung eines gezeigten Verhaltens oder Merkmals. In Anlehnung an Seidel (2003) wurde bei einzelnen Items die Vergabe bestimmter Abstufungen schriftlich festgelegt. Für andere Items erfolgten Konkretisierungen. Teilweise wurden auch Gegenbeispiele formuliert.

- *Beispiel 1, Item „Es werden Alltagsmaterialien benutzt“.* (1) *trifft überhaupt nicht zu:* Keine Alltagsmaterialien; (2) *trifft eher nicht zu:* Größtenteils Materialien aus der Physiksammlung; (3) *trifft eher zu:* Größtenteils Alltagsmaterialien; (4) *trifft voll und ganz zu:* Ausschließlich Alltagsmaterialien.
- *Beispiel 2, Item „Der Unterricht ist so sequenziert, dass die Schüler die angestrebten Konzepte/Ziele verstehen können“.* Die Abstufung für das Thema „Wasserkreislauf“ sah wie folgt aus: (1) *trifft überhaupt nicht zu:* In der gezeigten Unterrichtsstunde wird der komplette Wasserkreislauf, z.B. in Form eines komplexen Stationslaufs, behandelt. Die Bedeutung der einzelnen Teilaspekte wird nicht herausgearbeitet. (2) *trifft eher nicht zu:* In der gezeigten Unterrichtsstunde werden die Teilaspekte „Verdunstung“ und „Kondensation“ vermischt behandelt. Es findet also eine Beschränkung auf Teilaspekte des Wasserkreislaufs statt, wobei die Unterschiede bzw. die Bedeutungen der Phänomene nicht deutlich herausgearbeitet werden. (3) *trifft eher zu:* In der gezeigten Unterrichtsstunde werden die Teilaspekte „Verdunstung“ und „Kondensation“ nacheinander mit genügend Gesprächsphasen, in denen die Bedeutung der Teilaspekte herausgearbeitet wird, behandelt. (4) *trifft voll und ganz zu:* Es wird in der gezeigten Unterrichtsstunde nur ein Teilaspekt des Wasserkreislaufes behandelt.
- *Beispiel 3, Item „Die Lehrperson lässt die Sprache der Kinder zu, auch wenn diese noch ungenau ist“.* Dieses Item wurde wie folgt konkretisiert: 'Ungenau' wird im Sinne von physikalisch 'nicht ganz korrekt' verwendet; d.h. die Lehrperson lässt kinderspezifische Äußerungen zu Konzepten, wie z.B. „Das Wasser drückt das Schiff nach oben“ zu. Nicht gemeint ist mit diesem Item, dass die Lehrperson missverständliche oder unklare Aussagen stehen lässt. Das Item spiegelt passives Verhalten im Sinne von 'akzeptieren' wider.

Ratingdurchgänge. Wie bereits erwähnt, umfasst das Rating-Instrument 14 aus der Theorie abgeleitete Konstrukte. Jedes Konstrukt besteht wiederum aus einer Anzahl von Items als Indikatoren für das Konstrukt. Jedes Item wurde von den Ratern einzeln eingeschätzt. Es ergab sich also eine große Anzahl von Merkmalen, die von den Ratern beurteilt werden sollten, was große Ansprüche an die Rater setzte (inhaltliches Verständnis, Trennung der Konstrukte, Auffassungsgabe). Zu entscheiden war daher, ob alle Konstrukte von den Ratern in einem Durchgang beurteilt werden sollten, oder ob die Konstrukte im Sinne eines zyklischen Verfahrens auf mehrere Ratingdurchgänge aufgeteilt werden sollten. Das Rating aller Beobachtungsdimensionen in einem Durchgang ist einerseits ökonomisch, kann andererseits aber zur Überforderung der Rater führen (vgl. Wirtz & Caspar, 2002, S. 246). Eine Aufteilung des Ratings in mehrere Durchgänge würde eine Entlastung für die Rater bedeuten. Für niedrig-inferente Kodierungen, die sich auf sehr unterschiedliche Aspekte des Unterrichtsgeschehens beziehen, hält Reyer (2004) ein zyklisches Verfahren für sinnvoll. Hingegen scheint ihm für hoch-inferente Beurteilungen eine zyklische Vorgehensweise nicht günstig, da „für das verstehende Nachvollziehen des Unterrichtsfortgangs eine eher 'ganzheitliche' Wahrnehmung der betreffenden Unterrichtsebene erforderlich ist“ (Reyer, 2004, S. 127). Reyers Argumentation folgend und vor dem Hintergrund zeitlicher und

finanzieller Aspekte wurden im Rahmen dieses Projekts alle Konstrukte von vier Ratern in einem Durchgang beurteilt.

3.4.1.2 'Pretest' und Training der Rater

Nach einer ausführlichen Diskussion der entwickelten Ratingdimensionen bzw. der operationalisierten Items in der Arbeitsgruppe fand eine erste Testung des Instruments durch die Entwicklerin mit einer weiteren Person, die zu den ausgewählten Ratern gehörte, statt. Daran anschließend erfolgte die Schulung von insgesamt drei Ratern (zur Anzahl der Rater siehe die folgenden Ausführungen in diesem Abschnitt).

Zum Schulungsablauf und zu den Schulungsinhalten. Die Schulung beinhaltete in einem ersten Schritt die Vermittlung der Beobachtungsdimensionen des Instruments, um gewährleisten zu können, dass alle Rater ein gleiches inhaltliches Verständnis der zu bewertenden Konstrukte erreichen (in Anlehnung an Rakoczy & Pauli, 2006; Kobarg & Seidel, 2003; Reyer 2004).

Die weitere Schulung erfolgte an Extrembeispielen von Videos, welche die zu beobachtenden Unterrichtsmerkmale darstellten. Diese Videoextrembeispiele stammten aus der Stichprobe und dienten zur Verankerung der höchsten und niedrigsten Werte der Beobachtungsdimensionen. So konnte den Ratern ein Gefühl für die Spannbreite in den Unterrichtsstunden vermittelt werden. Bei der Auswahl der Extrembeispiele wurde darauf geachtet, dass Lehrkräfte für einzelne Beobachtungsdimensionen als Positiv- oder Negativbeispiele ausgewählt wurden und nicht für möglichst viele Konstrukte. So sollte vermieden werden, dass die Rater sich Urteile im Sinne von 'guten' oder 'schlechten' Lehrkräften machten. Daran anschließend beurteilten die Rater weitere Übungsvideos unabhängig voneinander und diskutierten anschließend ihre Bewertungen bis Konsens über die Beurteilungen erreicht wurde. Dieser letzte Teil der Schulung stellte den Schwerpunkt der Schulung dar und nahm mehrere Sitzungen in Anspruch. Das gemeinsame Diskutieren über die Verteilung der Ratingwerte der einzelnen Rater als eine Form der Validierung der Beurteilungsdimensionen (vgl. Ausführungen zur konsensuellen Validierung in Abschnitt 1.5.1.3) stellte die Basis zur Verbesserung der Raterleistung dar. Durch die gemeinsamen Gespräche konnten typische Beurteilungsfehler, wie Wahrnehmungsfehler (z.B. Konsistenzeffekte, verstanden als das Problem, bei der Bewertung konsistent zu bleiben) und Deutungs- und Interpretationsfehler (z.B. Milde-/Härtefehler) genau nachvollzogen und geklärt werden (vgl. Wirtz & Caspar, 2002). Wenn nötig wurden aus den Erfahrungen der Vergleiche der Übungsbeurteilungen Regeln für die weiteren Ratings abgeleitet.

In Anlehnung an das bei Seidel (2003, S. 108) vorgestellte und von Rakoczy und Pauli (2006) übernommene Verfahren wurde die Schulung als abgeschlossen betrachtet, als die Rater sich einig waren, die Unterrichtsvideos auf der Basis eines gleichen theoretischen Verständnisses zu beurteilen und die Urteile auf ein reliables Rating schließen ließen (vgl. 1.5.1.2). Durch die Festlegung der Analyseeinheit auf eine Unterrichtsstunde ergab sich das in Abschnitt 1.5.1.2 beschriebene Problem, dass die dem

Training zugrunde liegende Anzahl der Analyseeinheiten keine ausreichende Basis für die Berechnung der Beurteilerreliabilität darstellte, so dass die Prüfung der Interrater-Reliabilität erst erfolgen konnte, nachdem das gesamte Datenmaterial beurteilt war.

Für den Themenbereich „Schwimmen und Sinken“ konnte die als ideal angestrebte völlige Trennung zwischen Trainingsvideos und Auswertungsvideos wegen der relativ geringen Datenmenge nicht eingehalten werden. Es wurden jedoch so wenige Videos wie möglich für das Training verwendet. Für den Themenbereich „Wasserkreislauf“ konnte auf Videomaterial aus einer anderen am Seminar für Didaktik des Sachunterrichts durchgeführten Videostudie zurückgegriffen werden³³, so dass die Videos aus der Trainingsphase nicht Teil der zu untersuchenden Stichprobe waren. Die Videos wurden zeitgleich zu den Unterrichtsbeobachtungen „Wasserkreislauf“ der DFG-Studie und unter den gleichen Bedingungen erhoben.

Zu den Voraussetzungen und zur Auswahl der Rater. Im Allgemeinen herrscht Konsens darüber, dass jede geschulte Person in der Lage sein sollte, eine Kodierung bzw. ein Rating nach einem Training durchführen zu können. Bei den Beurteilungsdimensionen dieses Instruments handelt es sich jedoch um komplexe und inhaltlich anspruchsvolle Konstrukte, deren Anwendung auf Datenmaterial hohe Ansprüche an die Rater stellt. Die Rater müssen über einen speziellen fachwissenschaftlichen Hintergrund und über fachspezifisch-pädagogisches Wissen verfügen. Um den Umfang der ohnehin inhaltlich umfangreichen und zeitlich intensiven Schulung möglichst gering zu halten und um dem Problem von sehr ungleichen Raterurteilen (versursacht durch unterschiedliches Hintergrundwissen) entgegen zu wirken, wurden Rater ausgewählt, welche mit dem theoretischen Hintergrund der Konstrukte (oder zumindest mit einem Teil der Theorie) in den Grundzügen vertraut waren und die über einen annähernd homogenen fachlichen und fachdidaktischen Beurteilungshintergrund³⁴ verfügten.

Zudem wurde bei der Auswahl der Rater darauf geachtet, Rater auszusuchen, die in der Lage schienen, sowohl komplexe unterrichtliche Lehrerhandlungen im Kontext flexibel beurteilen als auch ihre Beurteilungen reflektieren, einordnen und begründen zu können. Auf die Bedeutung der Reflexionsfähigkeit als zentrales Merkmal eines Raters verweist Reyer (2004, S. 114f.). Vor allem hoch-inferente Ratings erfordern vom Rater ein bewusstes Anwenden des Ratingschemas. Ein blindes Nachmachen und Übernehmen von Musterurteilen ist bei dieser Form der Videoanalyse nicht möglich.

³³ Dabei handelt es sich um ein von der Landesregierung Nordrhein-Westfalen gefördertes Forschungsprojekt „Zum Einfluss von Lehrerausbildungskonzepten auf Lehrerkognition, Lehrerhandeln und Unterrichtserfolg im naturwissenschaftlichen Lernfeld der Grundschule“ – Teiluntersuchung Videostudie: „Gibt es einen Zusammenhang zwischen der Teilnahme an einem fachdidaktischen Studienmodul und konstruktivistischen Orientierungen im Unterrichtshandeln von Sachunterrichtslehrkräften?“. Nachzulesen in Gais et al., 2004.

³⁴ Selbstverständlich können die 'Eingangsvoraussetzungen' der Rater nur in einem begrenzten Umfang gleich gehalten werden. Bei jedem Rater ist der fachwissenschaftliche und fachdidaktische Hintergrund unterschiedlich ausgeprägt. Zudem verfügt jeder Rater über weitere persönliche Voraussetzungen, welche den Rater bei den Beurteilungen beeinflussen.

Zur Anzahl der Rater. Im Rahmen eines hoch-inferenten Ratings stellt sich die Frage, wie viele Rater in den Beurteilungsprozess mit einbezogen werden müssen, um möglichst reliable Werte zu erhalten und dabei die finanzielle Belastung³⁵ möglichst gering zu halten.

Die Anwendung von Ratingskalen erlaubt den Mittelwert mehrerer Rater als Informationsgrundlage für die Reliabilität eines Urteils zu nutzen. Je mehr Raterurteile in ein Gesamturteil einfließen, desto reliabler fällt das Urteil aus (vgl. Wirtz & Caspar, 2002). Im Rahmen dieser Arbeit wird die Interrater-Reliabilität varianzanalytisch mittels Generalisierbarkeitskoeffizienten berechnet (vgl. Abschnitt 3.5.2.1). Diese Generalisierbarkeitskoeffizienten sind mit abnehmender Steigung wachsende Funktionen in der Anzahl der Rater. Anhand der in Abschnitt 3.5.2.1 erläuterten Formeln zur Berechnung der Generalisierbarkeitskoeffizienten wird deutlich, dass diese neben der Anzahl der Rater nur noch vom Generalisierbarkeitskoeffizienten bei einem Rater abhängen. Die Interrater-Reliabilität für eine unterschiedliche Anzahl von Ratern kann also bestimmt werden, indem man die Reliabilitätsschätzung für einen Rater festlegt. Die Abbildung 3.4-1 zeigt den funktionalen Zusammenhang zwischen der Anzahl der Rater und der absoluten Generalisierbarkeitskoeffizienten³⁶. Die einzelnen Kurven basieren dabei jeweils auf der Annahme eines konkreten Wertes des Generalisierbarkeitskoeffizienten für einen Rater. Die Darstellung zeigt, dass der Reliabilitätswachstum unter Hinzunahme eines weiteren Raters zu drei oder mehr Ratern nur geringfügig ist. Nimmt man beispielsweise einen Generalisierbarkeitskoeffizienten Φ bei einem Rater von .50 an, so ergibt sich bei dem Übergang von drei auf vier Rater nur noch eine Erhöhung des Koeffizienten um .05 (vgl. Tabelle 3.4-1; die ausführlichen Rechenwege der in der Tabelle dargestellten Werte können dem Anhang B.1 entnommen werden). Somit ist es nicht notwendig, mehr als drei Rater für einen Beobachtungsprozess zu verwenden.

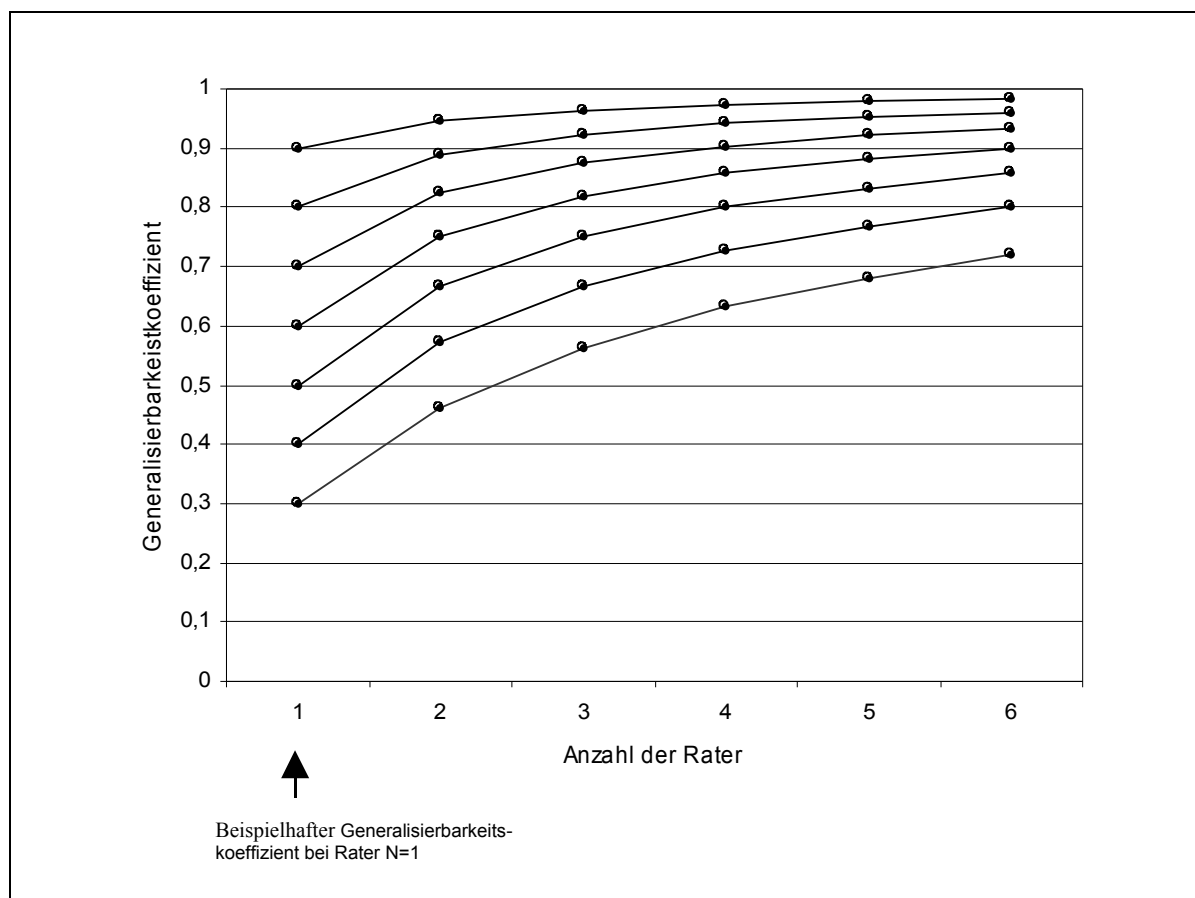
Ebenfalls berücksichtigt werden muss bei der Beantwortung der Frage nach der Anzahl der Rater, dass es erfahrungsgemäß vorkommen kann, dass sich Rater in ihren Werten von der Mehrheitsmeinung unterscheiden. In den meisten Fällen lassen sich die Gründe für die Unterschiede beheben. Es kann jedoch auch vorkommen, dass die Gründe nicht nachvollziehbar bzw. nicht behebbar sind. Deshalb wird in der Literatur empfohlen, eine ausreichende Anzahl von Ratern zu schulen, um notfalls von der Mehrheitsmeinung abweichende Rater von weiteren Ratings ausschließen zu können (Wirtz & Caspar, 2002).

Den Ausführungen oben folgend wurde im Rahmen dieser Studie entschieden, drei Rater zu trainieren, deren Urteile zusammen mit den Urteilen der Schulungsleiterin in die Generalisierbarkeitsanalyse eingehen. So konnte sichergestellt werden, dass auch im Falle eines oben skizzierten notwendigen Ausschlusses eines Raters ausreichend Raterurteile in die Analysen eingehen konnten, um reliable Wert zu erhalten.

³⁵Ratingprozesse (vor allem die Schulung) gehen mit einem hohen Zeitfaktor und daher mit einem hohen Kostenfaktor einher.

³⁶Für den absoluten Generalisierbarkeitskoeffizienten fallen die Steigungen der Funktionen analog zu den Funktionssteigungen der relativen Generalisierbarkeitskoeffizienten aus.

Abb. 3.4-1 Generalisierbarkeitskoeffizient in Abhängigkeit von der Anzahl der Rater



Tab. 3.4-1 Zuwachs des Generalisierbarkeitskoeffizienten (absolut) bei steigender Anzahl der Rater

Angenommener Generalisierbarkeitskoeffizient (absolut) für einen Rater (N=1)	Zuwachs des Koeffizienten von zwei Ratern zu drei Ratern (N=2 → N=3)	Zuwachs des Koeffizienten von drei Ratern zu vier Ratern (N=3 → N=4)	Zuwachs des Koeffizienten von vier Ratern zu fünf Ratern (N=4 → N=5)
.50	.0833	.0500	.0333
.60	.0682	.0390	.0252
.70	.0515	.0282	.0178

3.4.1.3 Durchführung des Ratings und anschließende Analyse

Rating des Datenmaterials und Interrater-Reliabilität. Nach Abschluss der Raterschulung wurde das gesamte Datenmaterial unabhängig von vier Personen (Entwicklerin des Instruments und drei geschulte Rater) beurteilt. Daran anschließend wurde die Interrater-Reliabilität mittels der Generalisierbarkeitsanalyse berechnet. Um abschätzen zu können, wie sich die Reliabilität der Raterurteile entwickelte bzw. ob überhaupt reliable Schätzungen erwartet werden konnten, wurden auch während des Ratingprozesses probeweise Reliabilitäten berechnet (z.B. nachdem 50% des Materials geratet waren). Der

Einbezug der Urteile der Instrumententwicklerin in die Prüfung der Qualität der Ratings kann gemäß Früh (1991) gleichzeitig als eine Form der Validitätsprüfung gesehen werden (vgl. die Ausführungen in 1.5.1.3).

Weitere Analysen. Die Raterurteile wurden für weiterführende Analysen gemittelt. Dazu gehörten zusätzliche Güteprüfungen (siehe dazu Teilkapitel 3.5 und Kapitel 4) und die Anwendung des Instruments auf Fragestellungen des DFG-Projekts.

3.4.2 Darstellung der Ratingdimensionen

Bei der Entwicklung des Instruments konnten wie bereits erwähnt Anregungen aus bereits bestehenden Instrumenten von Clausen, Reusser und Klieme (2003), Gais (2009), Kobarg und Seidel (2003), Rakoczy und Pauli (2006) und Heinze (2004) berücksichtigt werden. Aus anderen Instrumenten übernommene oder adaptierte Items sind markiert.

Im Folgenden werden die aus dem Entwicklungsprozess resultierenden Beobachtungsdimensionen inhaltlich erläutert. Dazu werden die Grundideen der Dimensionen beschrieben. Jede Beobachtungsdimension besteht aus einer Anzahl von Items, welche ihre Grundidee widerspiegeln. Aufgeführt sind alle Items, die geratet wurden. In grauer Schrift sind die Items gehalten, die nach den Reliabilitätsanalysen für weitere Analysen eliminiert wurden.

Motivation zur Auseinandersetzung mit dem Lerngegenstand schaffen

Konstrukt: „Motivation des Unterrichts“

Dieses Konstrukt erfasst, inwieweit es die Lehrperson schafft, die Lernenden für das unterrichtliche Handeln zu motivieren und inwieweit es der Lehrperson gelingt, den Schülern den Unterrichtsstoff als etwas Interessantes zu vermitteln. Dabei wird im Rahmen dieses Konstrukts nicht danach geschaut, *welche* Anreize die Lehrperson für die Beschäftigung mit dem Lerngegenstand schafft, sondern ganz allgemein, ob sie *ausreichend* Anreize schafft. Bewertungshintergrund dieses Konstrukts ist also die Frage, wie motivierend der Unterricht auf die Lernenden wirkt.

Items	Mt1	Die Lehrperson gestaltet den Unterricht spannend. (übernommen von Clausen, Reusser & Klieme, 2003)
	Mt2	Die Lehrperson kann auch trockenen Stoff interessant machen. (übernommen von Clausen, Reusser & Klieme, 2003)
	Mt3	Die Lehrperson kann die Schüler auch mal richtig begeistern. (übernommen von Clausen, Reusser & Klieme, 2003)
	Mt6	Die Lehrperson macht die Schüler neugierig.
	Mt7	Die Lehrperson ruft bei den Schülern Staunen hervor.
	Mt9	Die Lehrperson kann das Interesse der Kinder am Lerninhalt aufrecht erhalten.

Gelegenheiten zu eigenen Wissenskonstruktionen schaffen und individuelle Lernwege zulassen

Im Rahmen von drei Konstrukten werden Verhaltensweisen von Lehrkräften im Unterricht fokussiert, welche den Lernenden Raum und Zeit für eigene Denk- und Lernwege im Unterricht schaffen. Durch derartige Verhaltensweisen unterstützt die Lehrperson die Lernenden in ihren Wissenskonstruktionsprozessen.

Konstrukt „Eigene Ideen zulassen“

Dieses Konstrukt erfasst, inwiefern die Lehrperson den Lernenden im Unterricht die Gelegenheit gibt, eigene Ideen (Deutungen, Interpretationen und Erklärungen zu Naturphänomenen, Fragen oder Problemen; Lösungsvorschläge für Probleme) zu entwickeln.

- Items
- Ei_i1 Die Lehrperson ermuntert die Kinder, weiter nachzudenken/Ideen zu äußern.
 - Ei_i2 Die Lehrperson räumt den Kindern die Möglichkeit ein, selbst Deutungen/Erklärungen für ein Phänomen zu finden.
 - Ei_i3 Die Lehrperson verfolgt die Ideen der Kinder weiter.
 - Ei_i4 Die Lehrperson greift Schülerbeiträge im weiteren Unterrichtsverlauf auf.
 - Ei_i5 Die Lehrperson gibt den Kindern Zeit, Ideen zu entwickeln und Antworten zu finden.
 - Ei_i6 Die Lehrperson fragt – ohne zu bewerten – nach, wenn ein Schüler eine unvollständige oder unklare Idee formuliert.
(übernommen von Clausen, Reusser & Klieme, 2003)

Konstrukt „Eigene Formulierungen“

Dieses Konstrukt erfasst, inwiefern die Lehrperson eigenen Formulierungen der Lernenden (im Gegensatz zur naturwissenschaftlichen Fachsprache) Platz im Unterricht einräumt.

- Items
- Ei_e1 Die Lehrperson lässt die Sprache der Kinder zu, auch wenn diese noch ungenau ist.
 - Ei_e2 Die Lehrperson legt Wert darauf, dass die Kinder Sachverhalte in ihren eigenen Worten erklären.
 - Ei_e3 Die Lehrperson greift die Formulierungen der Kinder auf und arbeitet damit weiter.
 - Ei_e5 Die Lehrperson sorgt nicht dafür, dass 'Fachtermini' mit eigenen Worten erklärt werden.*

* umgepoltes Item

Konstrukt „Fehleroffenheit“

Dieses Konstrukt erfasst, inwiefern die Lehrkraft im Unterricht einen positiven Umgang mit Fehlern betreibt, Fehler als etwas Positives bzw. als 'Lernchancen' sieht und Schülerfehlern im Unterricht Platz einräumt.

- Items
- Ei_f1 Die Lehrperson ist geduldig, wenn ein Schüler im Unterricht ein Fehlkonzepnt nennt/hat.
(in Anlehnung an Clausen, Reusser & Kieme, 2003)
 - Ei_f2 Bei dieser Lehrperson ist das Nennen von falschen Vermutungen/Erklärungen nichts Schlimmes.
(in Anlehnung an Clausen, Reusser & Klieme, 2003)
 - Ei_f3 Die Lehrperson achtet darauf, dass keiner ausgelacht wird, der etwas Falsches sagt.
 - Ei_f4 Die Lehrperson betont, dass auch falsche Ideen wichtig sind, um zum Ziel zu kommen.
 - Ei_f7 Die Kinder dürfen auch falsche Vermutungen/ Erklärungen äußern.
 - Ei_f8 Die Lehrperson korrigiert nicht sofort jedes Fehlkonzepnt.
(in Anlehnung an Clausen, Reusser & Klieme, 2003)

Transmission von Wissen vermeiden

Die folgenden zwei Konstrukte bilden transmissive Lehrerverhaltensweisen ab. Diese Lehrerverhaltensweisen spiegeln eine sehr enge Unterrichtsführung wider. Im Unterricht gehen alle Aktivitäten von der Lehrkraft aus, die Rolle der Lernenden beschränkt sich auf das Rezipieren von Inhalten und Durchführen von Aufgaben. Dadurch nimmt die Lehrkraft den Lernenden die Gelegenheit zur eigenen Konstruktion von Wissen.

Konstrukt: Transmission von Wissen – Erklärungen

Mit Hilfe dieses Konstrukts wird bewertet, inwiefern die Lehrperson im Unterricht den passiven Lernenden fertige Wissensinhalte direkt vermittelt.

- Items
- I11 Die Lehrperson beantwortet aufkommende Fragen vorschnell selbst.
 - I12 Die Lehrperson gibt die Erklärungen für die zuvor durchgeführten Experimente vor.
 - I13 Die Lehrperson lässt die Kinder zuvor aufgeworfenen Problemen nicht selbst lösen.
 - I15 Den Kindern werden 'fertige Erklärungen' durch Abbildungen vermittelt.
 - I16 Die Lehrperson gibt den Schülern Merksätze vor, die nicht mit den Kindern entwickelt worden sind.
 - I18 Die Lehrperson fragt nach den Vermutungen der Kinder, arbeitet damit aber nicht weiter.

Konstrukt: Transmission von Wissen – eigene Denk- und Lernwege ablehnen

Dieses Konstrukt erfasst, inwiefern die Lehrkraft die Lernenden im Unterricht in ihren Lernprozessen und Leistungen überwacht, wobei die Lehrperson eng an ihrem eigenen Konzept festhält und den Lernenden dadurch keinen Platz für individuelle Denk- und Lernwege im Unterricht einräumt.

- Items
- II9 Die Lehrperson unterbindet eigene Denkwege der Kinder.
 - II13 Die Lehrperson berichtigt Fehlkonzepte ohne Begründung.
(in Anlehnung an Heinze, 2004)
 - II4 Die Lehrperson korrigiert falsche Vermutungen qua Autorität.
 - II7 Die Lehrperson unterdrückt eigene Formulierungen der Kinder.
 - II12 Die Lehrperson ignoriert Fehlkonzepte.
 - II10 Die Lehrperson geht nicht auf Vorschläge der Kinder zur Überprüfung ein.
 - II11 Die Lehrperson bricht Lösungsvorschläge der Kinder ab.

Umgang mit Präkonzepten und Conceptual Change anregen*Konstrukt: „Präkonzepte – Vorwissen aktivieren und aufgreifen“*

Einem konstruktivistisch orientierten Verständnis von Lernen folgend geht es bei diesem Konstrukt um das Aufgreifen der vorunterrichtlichen Vorstellungen bzw. um die Präkonzepte der Lernenden durch die Lehrkraft. Hier wird erfasst, inwieweit die Lehrperson die Präkonzepte der Schüler aktiviert und diese im weiteren Unterrichtsverlauf wieder aufgreift bzw. in den Unterrichtsverlauf einbezieht. Hier geht es nicht um ein flüchtiges Abfragen von Vorstellungen der Lernenden im Sinne eines Ideensammelns am Anfang der Unterrichtsstunde.

- Items
- Pk1 Die Lehrperson kommt im Verlauf der Stunde auf Präkonzepte der Kinder zurück.
 - Pk3 Die Lehrperson knüpft im Unterricht an Präkonzepte der Schüler an.
(in Anlehnung an Kobarg & Seidel, 2003)
 - Pk4 Der Lehrer regt die Schüler an, über ihre Präkonzepte nachzudenken.
 - Pk5 Die Lehrperson nimmt sich viel Zeit, Präkonzepte begründen zu lassen.

Konstrukt: „Conceptual Change – kognitive Konflikte erzeugen“

Dieses Konstrukt erfasst, inwiefern die Lehrperson mit Hilfe einer objektiven, überzeugenden Evidenz die Kinder auf Widersprüche in ihrem Denken aufmerksam macht und ihnen somit die Unzulänglichkeit und Grenzen ihrer Konzepte bewusst macht. Dadurch erzeugt die Lehrperson auf Seiten der Schüler eine Unzufriedenheit mit alten Konzepten und lässt neue Konzepte als nützlicher und plausibler erscheinen. Auf den Aspekt der Fruchtbarkeit neuer Konzepte wird im Rahmen des Konstrukts „Anwendungsbezogenes Lernen“ eingegangen.

Anmerkung: Von besonderer Bedeutung sind aus didaktischer Sicht die tief verwurzelten Präkonzepte, die aufgrund ihrer Stabilität schwer zu beeinflussen sind und im Unterricht Lernschwierigkeiten auslösen können (vgl.

z.B. Wodzinski, 1996). Aus diesem Grund erfasst das Video-Analyseinstrument nur das Erzeugen kognitiver Konflikte durch die Lehrkraft.

- | | |
|-------|--|
| Items | Cc1 Die Lehrperson hat Versuche ausgesucht, die für die Kinder überzeugend sind. |
| | Cc4 Die Lehrperson konfrontiert die Kinder mit Beobachtungen oder Phänomenen, die den Erwartungen der Kinder widersprechen. |
| | Cc6 Die Lehrperson widerlegt Fehlkonzepte der Kinder mit Hilfe ausgewählter Materialien. |
| | Cc8 Der Lehrperson ist es besonders wichtig, dass objektive Evidenzen intensiv mit Präkonzepten in Beziehung gesetzt werden. |

Umsetzung sozial-kooperativen Lernens

Konstrukt „Ideen gemeinsam diskutieren“

Dieses Konstrukt erfasst, inwiefern die Lehrperson mit den Kindern Unterrichtsinhalte gemeinsam diskutiert und inwiefern die Lehrkraft die gemeinsame Diskussion fördert. Dabei spielt es keine Rolle, ob es fachlich angemessene oder falsche Vorstellungen über Phänomene oder Begriffe sind.

- | | |
|-------|--|
| Items | Dk4 Missverständliche, unvollständige oder unklare Äußerungen werden nicht ignoriert oder bloß kommentiert, sondern es wird nachgefragt.
(übernommen von Clausen, Reusser & Klieme, 2003) |
| | Dk5 Die Lehrperson fordert die Kinder auf, ihre eigenen Vermutungen/Erklärungen untereinander zu diskutieren. |
| | Dk6 Die Lehrperson fordert die Kinder im Unterricht auf, ihre Deutungen zu einem Phänomen gegenüber Mitschülern zu begründen. |
| | Dk7 Die Lehrperson regt die Kinder auch dann an, ihre Vorstellungen untereinander zu diskutieren, wenn diese falsch sind. |
| | Dk8 Die Lehrperson diskutiert gemeinsam mit den Kindern Probleme, Ideen, Lösungswege, Fehlkonzepte. |

Anleitung und Unterstützung der Lernenden – Scaffolding

Im Rahmen von drei Konstrukten wird erfasst, inwiefern die Lehrperson die Lern- und Denktätigkeit der Schüler im naturwissenschaftlichen Sachunterricht durch prozessorientierte Strukturierungshilfen unterstützt. Berücksichtigt werden diskursive und sequenzielle Unterstützungsmaßnahmen.

Anmerkung: Eine Einschätzung zum Einsatz von materiellen Denkhilfen durch die Lehrkräfte (geeignete Experimente, Materialien, Repräsentationsformen, ...) konnte an den zur Verfügung stehenden Unterrichtsvideos aus Gründen der Kameraführung während der Videoaufzeichnung nicht vorgenommen werden, so dass diesbezüglich keine Beobachtungsdimension aus der Theorie abgeleitet wurde.

Konstrukt „Scaffolding – Hervorhebungen“

Hervorhebungen meinen eine diskursive Unterstützung durch die Lehrkraft in Form von *Fokussierungshilfen* zum Lenken der Aufmerksamkeit der Lernenden auf wesentliche Inhalte und Aspekte im Unterricht. Im Rahmen dieses Konstrukts wird erfasst, inwiefern die Lehrperson Hervorhebungen im Unterricht einsetzt und durch diese Art der Denkkunterstützung den Lernenden hilft, die für den Lernprozess relevanten Inhalte, Aspekte und Strukturen wahrzunehmen und weiter zu verfolgen (Reiser, 2004).

- Items Sc_h1 Die Lehrperson weist darauf hin, was sich die Schüler merken sollen.
(übernommen von Clausen, Reusser & Klieme, 2003)
- Sc_h2 Die Lehrperson hebt wichtige Fakten hervor.
(übernommen von Clausen, Reusser & Klieme, 2003)
- Sc_h4 Die Lehrperson fasst Ergebnisse oder bereits Gesagtes häufig wiederholend zusammen.
- Sc_h6 Die Lehrperson fasst das Wichtigste nochmals zusammen.
(übernommen von Clausen, Reusser & Klieme, 2003)
- Sc_h7 Die Lehrperson fokussiert die Aufmerksamkeit der Kinder auf wichtige inhaltliche Aspekte.
- Sc_h8 Die Lehrperson fordert die Kinder auf, verständlich zu wiederholen.

Konstrukt „Scaffolding – Perturbieren“

In Bezugnahme auf Maturana und Varela (1987) kann Lernen als selbstschaffender Prozess betrachtet werden, der in Kontakt mit der Umwelt erfolgt und der von 'außen' durch Perturbationen, d.h. durch wahrgenommene Störungen, durch Anregungen und durch Irritationen angestoßen werden kann (Siebert, 1999). Im pädagogischen Kontext bedeutet dies, dass Lehrkräfte die Lernenden 'perturbieren' können, das heißt sie können die Lernen durch entsprechende Handlungen dazu anregen, über ihre Wissensstrukturen, über ihre Konzepte und Sichtweisen nachzudenken und diese ggf. zu modifizieren (Siebert 1999).

Unter Perturbationen werden im Rahmen dieses Konstrukts diskursive Unterstützungsformen verstanden, die einerseits Hilfen beinhalten, welche den Argumentationsprozess der Lernenden unterstützen. Zudem fallen Problematisierungshilfen und Provokationen in dieses Konstrukt, um die Aufmerksamkeit der Lernenden auf Probleme, Aufgaben oder Fragen zu lenken, welche die Lernenden sonst möglicherweise ignoriert oder übersehen hätten (Reiser, 2004).

- Items Sc_p1 Die Lehrperson fordert Begründungen ein.
- Sc_p2 Die Lehrperson fordert Beweise ein.
- Sc_p3 Die Lehrperson macht auf Widersprüche in Aussagen/Vermutungen aufmerksam.
- Sc_p5 Die Lehrperson hinterfragt Aussagen.
- Sc_p4 Die Lehrperson macht auf ungelöste Probleme aufmerksam.

Konstrukt „Scaffolding – Sequenzierungen“

Hier geht es um sequenzielle Unterstützungsmaßnahmen durch die Lehrkraft. Darunter werden Gliederungsmaßnahmen zur Reduzierung der Komplexität von Lerninhalten verstanden (Reiser, 2004). Dieses Konstrukt erfasst, inwiefern die Lehrperson komplexe Inhalte, Aufgaben, Gespräche sinnvoll sequenziert bzw. gliedert (z.B. Zerlegung in Teilaspekte). Dadurch schränkt die Lehrkraft die Freiheitsgrade der Lernenden ein und gibt ihnen eine 'bestimmte Richtung' vor. Dies hilft den Lernenden, die wesentlichen Inhalte, Aufgaben und Ziele im Auge zu behalten und in einer sinnvollen Reihenfolge zu bearbeiten (Reiser, 2004, siehe auch Hardy, 2007).

- Items Sc_s1 Der Unterricht ist so sequenziert, dass die Schüler die angestrebten Konzepte/Ziele verstehen können.
- Sc_s2 Die Lehrperson achtet darauf, dass das Gespräch nicht zwischen unterschiedlichen Aspekten hin und her springt.

Laisser-faire vermeiden*Konstrukt „Laisser-faire“*

Dieses Konstrukt bildet Lehrerverhaltensweisen ab, bei denen die Lehrkraft den Lernenden kaum oder keine Strukturen anbietet. Die Lehrkraft folgt in ihrem Verhalten der Ansicht, dass Schüler im naturwissenschaftlichen Unterricht am besten ohne Eingreifen und Lenken durch die Lehrperson sowie weitgehend selbstgesteuert auf der Basis eigener Erfahrungen und Experimenten lernen sollten und Strukturierungshilfen nicht notwendig sind (Reusser, 2006).

- Items O13 Die Lehrperson gibt den Kindern kaum Hilfe beim Entwickeln und Auswerten von Experimenten.
- O14 Die Lehrperson begnügt sich vorschnell damit, wenn ein Kind einen erwünschten Begriff/Zusammenhang genannt hat.
- O15 Die Lehrperson beschränkt den Umgang mit Präkonzepten auf ein oberflächliches Abfragen.
- O11 In Gesprächen über Deutungen von Naturphänomenen oder der Suche nach geeigneten Lösungswegen hält sich die Lehrperson zurück.
- O12 Die Lehrperson lässt die Kinder einfach nur 'laufen'.

Authentische Kontexte und Anwendungsbezüge

Konstrukt „Anwendungsbezogenes Lernen“

Dieses Konstrukt erfasst, inwiefern die Lehrperson den Kindern die Möglichkeit gibt, die Anwendungsbezogenheit des Unterrichtsinhalts bzw. der neu erlernten Konzepte zu erfahren. Dadurch erfahren die Lernenden neue Konzepte als 'fruchtbar'.

- Items
- Cc3 Die Lehrperson thematisiert die Anwendung neuer Konzepte.
 - Aw2 Fragestellungen (Bezüge, Beobachtungen) aus der außerschulischen Welt werden angesprochen.
(übernommen von Clausen, Reusser & Klieme, 2003)
 - Aw4 Es werden Alltagsmaterialien benutzt.
 - Aw5 Die Lehrperson stellt Bezüge zum außerschulischen Alltag der Kinder her.
 - Aw6 Die Lehrperson stellt im Unterricht Bezüge zu Naturphänomenen her.
 - Aw7 Die Lehrperson fragt die Kinder nach Alltagsbeispielen/Anwendungsbezügen.

3.5 Methoden der Prüfung des Rating-Instruments

Das im Rahmen dieser Arbeit entwickelte Videoanalyse-Instrument soll unterschiedliches Verhalten von Lehrkräften in bestimmten, genau definierten Bereichen (aus der Theorie abgeleiteten Konstrukten) möglichst objektiv, zuverlässig und gültig erfassen. Auf der Basis der theoretischen Ausführungen zur Güteprüfung von Videoanalyse-Instrumenten wird im Rahmen dieses Teilkapitels beschrieben, auf welche Methoden der Qualitätsprüfung des Rating-Instruments zurückgegriffen wird.

3.5.1 Zur Objektivität

Gemäß den Ausführungen in 1.5.1.3 kann die Objektivität in Videostudien mit einer transparenten Darstellung des gewählten Vorgehens gesichert werden. Hugener (2006) berücksichtigt bei der Beschreibung der Verfahren spezielle inhaltsanalytische Ansprüche. Dazu gehört die inhaltliche und datengeleitete Entwicklung der zu beobachtenden Konstrukte, die klare und präzise Definition und Operationalisierung der Kategorien und Konstrukte sowie die Aufstellung von Regeln zum Kodieren bzw. Raten, damit Rater zu vergleichbaren Resultaten gelangen.

Bezüglich der Beurteiler- bzw. Interrater-Reliabilität ist eine Zuordnung sowohl zur Objektivität als auch zur Reliabilität möglich. Der Argumentation von Wirtz und Caspar (2002) folgend, wird die Interrater-Reliabilität in dieser Arbeit – wie in Videostudien üblich – dem Zuverlässigkeitskriterium zugeordnet (vgl. dazu 1.5.1.3).

3.5.2 Zur Reliabilität

Um dem Gütekriterium der Reliabilität Rechnung zu tragen, wurde zum einen die Interrater-Reliabilität zur Analyse der Qualität der Beurteilungen durch die vier Rater bestimmt. Dazu wurde auf Methoden aus der Generalisierbarkeitstheorie zurückgegriffen, die in 3.5.2.1 beschrieben werden. Zum anderen ist die Qualität eines Instruments von der Art und der Zusammensetzung der Items, aus denen das Instrument besteht, abhängig. Darum ist die Item- bzw. Skalenanalyse ein zentraler Aspekt bei der Konstruktion eines Tests (vgl. Bortz & Döring, 2006). In 3.5.2.2 wird kurz erläutert, welche Kennwerte zur Bestimmung der Merkmale der konstruierten Items und Skalen berechnet wurden.

3.5.2.1 Analyse der Interrater-Reliabilität mittels der Generalisierbarkeitstheorie

Da die Beobachtungsdimensionen als Skalen angelegt sind, bei denen die Items in einem Schätzverfahren auf einer Skala in einem mehrstufigen Antwortformat eingeschätzt werden, eignen sich zur Berechnung der Interrater-Reliabilität sowohl Intraklassenkorrelationen als auch Generalisierbarkeitskoeffizienten. Beide Verfahren basieren auf einem varianzanalytischen Verfahren. In dieser Arbeit wird auf den Generalisierbarkeitsansatz zurückgegriffen, welcher im Folgenden erläutert wird. An geeigne-

ter Stelle wird darauf eingegangen, was die Generalisierbarkeitsanalyse von der Berechnung von Intra-klassenkorrelationen unterscheidet.

Bei dem Generalisierbarkeitsansatz handelt es sich um eine Erweiterung der klassischen Testtheorie. Während bei letzterer nur eine Fehlerquelle betrachtet wird, kann bei der Generalisierbarkeitstheorie (GT) der Einfluss verschiedener Fehlerquellen (Facetten) simultan abgeschätzt werden (vgl. Shavelson, Webb & Rowley, 1989). Dazu bedient sich die Generalisierbarkeitstheorie eines varianzanalytischen Modells (vgl. Renkl & Helmke, 1993). Das Generalisierbarkeitsdesign der vorliegenden Analysen berücksichtigt zwei Faktoren, den Faktor 'Video' und den Faktor 'Rater'. In Anlehnung an die Ausführungen von Brennan (2001) ergibt sich ein 'Single-facet G-study v(video) x r(rater) design', wobei 'x' gelesen wird als '*crossed with*'. Beim v x r Design wird jedes Video von jedem Rater beurteilt.

Prinzipiell lässt die Generalisierbarkeitstheorie die Spezifizierung einer beliebig großen Zahl von den Messwert beeinflussenden Facetten zu. Renkl und Helmke (1993, S. 71) weisen allerdings darauf hin, dass sich mit wachsender Anzahl von Facetten das Problem ergibt, dass die Schätzungen der Varianzkomponenten dann immer unzuverlässiger werden. Dies liegt daran, dass mit steigender Komplexität des Designs auch die Anzahl stichprobenabhängiger Größen wächst, mit deren Hilfe man die Varianzkomponenten schätzt. Demnach ist es günstig, das Generalisierbarkeitsdesign möglichst einfach zu halten.

Im Fall zweier Faktoren, wie es in den vorliegenden Analysen mit den Facetten 'Video' und 'Rater' der Fall ist, ergibt sich ein varianzanalytisches Modell mit zwei Haupteffekten und einem Interaktionseffekt (vgl. Brennan, 2001; Helmke & Renkl, 1993). Formal bezeichnen wir nach Brennan (2001, S. 22, Formeln (2.1) bis (2.3)) mit X_{vr} die Beurteilung von Video v durch Rater r und mit

$$\mu_{v\cdot} \equiv E_r [X_{vr}]$$

bzw.

$$\mu_{\cdot r} \equiv E_v [X_{vr}]$$

die Erwartungswerte über die Rater bzw. über die Videos. Außerdem ist das Gesamtmittel ('grand mean') definiert durch

$$\mu \equiv E_v [E_r [X_{vr}]]$$

Die Beurteilung von Video v durch Rater r setzt sich additiv zusammen aus dem 'grand mean' μ , dem Einfluss $(\mu_{v\cdot} - \mu)$ des Videos v, den Einfluss $(\mu_{\cdot r} - \mu)$ des Raters r und der Interaktion $(X_{vr} - \mu_{v\cdot} - \mu_{\cdot r} + \mu)$ von Video und Rater (vgl. Tabelle 3.5-1).

Tab. 3.5-1 Varianzanalytische Zerlegung der beobachteten Werte und Varianzkomponenten, die auf die verschiedenen Effekte zurückgehen (in Anlehnung an Cronbach et al., 1972, S. 28)

	Varianzanalytische Zerlegung der beobachteten Werte	Varianzkomponenten
	$X_{vr} = \mu$ (grand mean)	$\sigma^2[X_{vr}] =$
<u>Haupteffekt</u> Video	$+ \mu_{v\bullet} - \mu$	$E[\mu_{v\bullet} - \mu]^2 = \sigma_v^2$
Rater	$+ \mu_{\bullet r} - \mu$	$+ E[\mu_{\bullet r} - \mu]^2 = \sigma_r^2$
<u>Interaktion</u> Video x Rater (Residualeffekt)	$+ X_{vr} - \mu_{v\bullet} - \mu_{\bullet r} + \mu$	$+ E E[X_{vr} - \mu_{v\bullet} - \mu_{\bullet r} + \mu]^2 = \sigma_{vr}^2$

Für jede Komponente gibt es eine zugehörige Varianz, die als Varianzkomponente bezeichnet wird (vgl. Brennan, 2001; vgl. Tabelle 3.5-1). So weist z.B. die Komponente 'Video' die Varianz $E[\mu_{v\bullet} - \mu]^2 = \sigma_v^2$ auf. In diesem Modell setzt sich die gesamte Varianz (Varianz der Einzelwerte $\sigma^2[X_{vr}]$ ³⁷), gemäß der Tabelle 3.5-1 (zur differenzierten Betrachtung siehe Brennan, 2001, S. 24f.) additiv aus den drei Varianzkomponenten der zwei Facetten und deren Interaktion zusammen (vgl. Brennan, 2001; Clausen, Reusser & Klieme, 2003):

- Die erste Varianzkomponente σ_v^2 wird als 'universe score'-Varianz bezeichnet und entspricht der 'wahren Varianz' in der klassischen Testtheorie. Hier stellt sie die 'systematische' Varianz zwischen den Unterrichtsstunden dar.
- Die zweite Varianzkomponente σ_r^2 drückt systematische Niveau-Unterschiede in den Beurteilungen der Rater aus (systematische Fehlervarianz).
- Die dritte Varianzkomponente σ_{vr}^2 beschreibt die unsystematische Variation zwischen Ratern und Videos (unsystematische Fehlervarianz).

Renkl und Helmke (1993, S. 67) erwähnen, dass die „Größe der einzelnen Fehlervarianzkomponenten [...] Aufschluss darüber [gibt], wie stark die jeweiligen Faktoren bzw. deren Wechselwirkung zum Messfehler beitragen.“

³⁷Die Varianz der Werte jeweils einzelner Videos, die von jeweils einem Rater vergeben wurden (vgl. Renkl & Helmke, 1993).

Die empirischen Varianzkomponenten berechnen sich gemäß der klassischen Varianzanalyse aus den bekannten mittleren Abweichungsquadraten (vgl. dazu Brennan, 2001, S. 27, Formel (2.19) bis (2.21)). Die genauen Terme zeigt die Tabelle 3.5-3.

Tab. 3.5-2 Überblick über die Struktur der vorliegenden Daten

Videos	Rater				R _{gesamt} = $\bar{X}_{i\bullet}$
	R 1	R 2	R 3	R 4	
V 1	X _{1,1}	X _{1,2}			$\bar{X}_{1\bullet}$
V 2					$\bar{X}_{2\bullet}$
V 3		X _{3,2}			$\bar{X}_{3\bullet}$
...					...
V i			X _{i,3}		$\bar{X}_{i\bullet}$
...					...
V 60				X _{60,4}	$\bar{X}_{60\bullet}$
V _{gesamt} = $\bar{X}_{\bullet j}$	$\bar{X}_{\bullet 1}$	$\bar{X}_{\bullet 2}$	$\bar{X}_{\bullet 3}$	$\bar{X}_{\bullet 4}$	\bar{X}

$\bar{X}_{\bullet r}$: empirischer Mittelwert für Rater r, analog zum theoretischen Mittelwert $\mu_{\bullet r}$

$\bar{X}_{v\bullet}$: empirischer Mittelwert für die Videos v, analog zum theoretischen Mittelwert $\mu_{v\bullet}$

\bar{X} : empirischer Gesamtmittelwert, analog zum theoretischen Gesamtmittelwert μ

Tab. 3.5-3 Formeln für die Schätzer der Varianzkomponenten

Geschätzte Varianzkomponenten	
Video	$\hat{\sigma}_v^2 = \frac{1}{n_v - 1} \sum_{i=1}^{n_v} (\bar{X}_{i\bullet} - \bar{X})^2 - \frac{1}{n_r} \hat{\sigma}_{vr}^2$
Rater	$\hat{\sigma}_r^2 = \frac{1}{n_r - 1} \sum_{j=1}^{n_r} (\bar{X}_{\bullet j} - \bar{X})^2 - \frac{1}{n_v} \hat{\sigma}_{vr}^2$
Video x Rater	$\hat{\sigma}_{vr}^2 = \frac{n_v \cdot n_r}{(n_v - 1)(n_r - 1)} \sum_{i=1}^{n_v} \sum_{j=1}^{n_r} (X_{ij} - \bar{X}_{i\bullet} - \bar{X}_{\bullet j} + \bar{X})^2$

$\hat{\sigma}_v^2$: empirische Varianz, analog zur theoretischen Varianz σ_v^2

$\hat{\sigma}_r^2$: empirische Varianz, analog zur theoretischen Varianz σ_r^2

$\hat{\sigma}_{vr}^2$: empirische Varianz, analog zur theoretischen Varianz σ_{vr}^2

n_v : Anzahl der Videos; n_r : Anzahl der Rater

Im Rahmen der Generalisierbarkeitstheorie werden zwei Arten von Fehlervarianzen unterschieden, die relative und die absolute Fehlervarianz (Brennan, 2001). In dem $v \times r$ -Design dieser Studie resultiert die Fehlervarianz für relative Entscheidungen $\sigma^2(\delta)$ aus dem Verhältnis des Interaktionseffekts von Video und Rater zur Anzahl der Rater (vgl. Brennan, 2001, Formel (2.36) auf S. 33):

$$\sigma^2(\delta) = \frac{\sigma_{vr}^2}{n_r}$$

Für die absolute Fehlervarianz $\sigma^2(\Delta)$, in die alle Varianzanteile bis auf die 'universe-score'-Varianz einfließen (Renkl & Helmke, 1993), ergibt sich gemäß Brennan (2001, Formel (2.32) auf S. 32) die folgende Formel:

$$\sigma^2(\Delta) = \frac{\sigma_{vr}^2}{n_r} + \frac{\sigma_r^2}{n_r}$$

Anhand der drei Varianzkomponenten und deren relativer Bedeutungen lässt sich die Güte der Messung abschätzen. Diese kann in Generalisierbarkeitskoeffizienten ausgedrückt werden (vgl. Clausen, Reusser & Klieme, 2003), welche mit dem Reliabilitätskennwert der klassischen Testtheorie vergleichbar sind (vgl. Renkl & Helmke, 1993). Die Generalisierbarkeitskoeffizienten (als Kennwerte für die Zuverlässigkeit) berechnen sich aus dem Verhältnis der 'universe-score'-Varianz σ_v^2 zur Gesamtvarianz σ_{vr}^2 .

Entsprechend den zwei unterschiedlichen Fehlervarianzen wird zwischen dem relativen und dem absoluten Generalisierbarkeitskoeffizienten unterschieden. Bei dem relativen Generalisierbarkeitskoeffizienten (*generalizability coefficient*) $E\rho^2$ wird der Haupteffekt der Facette 'Rater' nicht mit einbezogen, es wird nur die Rangreihe der Beurteilungsgegenstände (rank ordering of judged) berücksichtigt (Renkl & Helmke, 1993; Shavelson, Webb & Rowley, 1989; Brennan, 2001). Der Koeffizient entspricht der Fehlervarianz in der klassischen Testtheorie und wird wie folgt definiert (Brennan, 2001, S. 25):

$$E\rho^2 = \frac{\sigma_v^2}{\sigma_v^2 + \frac{\sigma_{vr}^2}{n_r}} = \frac{\sigma_v^2}{\sigma_v^2 + \sigma^2(\delta)}$$

Bei der Generalisierbarkeit absoluter Entscheidungen (*index of dependability*) Φ gehen alle Varianzanteile in den Term mit ein, es werden sowohl die Höhe als auch die Rangreihe der Beurteilungsgegenstände (level and rank ordering) berücksichtigt (Renkl & Helmke, 1993; Shavelson, Webb & Row-

ley, 1989; Brennan, 2001). Der absolute Generalisierbarkeitskoeffizient wird definiert als (Brennan, 2001, S. 35):

$$\Phi = \frac{\sigma_v^2}{\sigma_v^2 + \frac{\sigma_{vr}^2}{n_r} + \frac{\sigma_r^2}{n_r}} = \frac{\sigma_v^2}{\sigma_v^2 + \sigma^2(\Delta)}.$$

Ein Vergleich der Formeln für Generalisierbarkeitskoeffizienten und der Koeffizienten der Intraklassenkorrelation (vgl. dazu Wirtz & Caspar, 2002, S. 170) zeigt den Unterschied zwischen Generalisierbarkeits- und Intraklassenkoeffizienten: Generalisierbarkeitskoeffizienten berücksichtigen die Interaktion von Video und Rater. In Intraklassenkorrelationen hingegen wird angenommen, dass keine Interaktion zwischen gerateten Videos und Ratern stattfindet. Entsprechend wird diese Varianzkomponente gleich Null gesetzt (S. 169).

In der Literatur wird eine Intraklassenkorrelation von mindestens .70 als Indiz für 'gute' Reliabilität gesehen (Wirtz & Caspar, 2002, S. 160). Für Generalisierbarkeitskoeffizienten werden in der Literatur keine Konventionen genannt. In Anlehnung an Clausen, Reusser & Klieme (2003) wird in dieser Arbeit ein Generalisierbarkeitskoeffizient ab .70 für gut befunden.

3.5.2.2 Analyse der internen Konsistenzen

Die Ausführungen zur Testung der Item- und Skalenqualität fallen kürzer aus als die zur Generalisierbarkeitsanalyse, da es sich um sehr konventionelle und daher bekannte Methoden der Reliabilitätstestung handelt.

Im Rahmen der klassischen Testkonstruktion werden zur Itemanalyse die Berechnung der Itemschwierigkeit, Trennschärfe und sowie die Dimensionalitätsüberprüfung³⁸ gezählt (Bortz & Döring, 2006). Wie allgemein üblich werden als Maße der Schwierigkeit der Items deren Mittelwerte und Standardabweichungen angegeben. Die Berechnung der Trennschärfen der Items als Maß dafür, wie gut ein Item das Testergebnis repräsentiert, erfolgt mittels der korrigierten Item-Skalenwert-Korrelation (r_{it}). Zur Prüfung der Reliabilität der Skalen wurden die internen Konsistenzen mit dem dafür gebräuchlichsten Maß, dem Alpha-Koeffizienten von Cronbach (1951) ('Cronbachs Alpha'), berechnet (vgl. Bortz & Döring, 2006). Die Berechnung der internen Konsistenz basiert je Video und je Item auf dem jeweiligen Mittelwert über alle Rater.

Gemäß Bühner (2004, S. 129, vgl. auch Bortz & Döring, 2006) gelten folgende Beurteilungsrichtlinien für die korrigierte Trennschärfe: Ein empirischer Wert unterhalb von .30 gilt als 'niedrig', ein Wert von .30 bis .50 gilt als 'mittel' und ein Wert größer als .50 als 'hoch'. Bezüglich Cronbachs Alpha gilt in der Literatur meist ein empirischer Wert oberhalb von .80 als akzeptabel (Schnell, Hill & Esser, 2005; Bortz & Döring, 2006), wobei in der Praxis meist noch weit niedrigere Werte des Koeffizienten ak-

³⁸Zur Dimensionalitätsprüfung siehe 3.5.3.

zeptiert werden (vgl. Schnell, Hill & Esser, 2005, S. 153). In dieser Arbeit gelten für die korrigierte Trennschärfe ein Wert unterhalb von .30 und für die interne Konsistenz ein Wert für Cronbachs Alpha kleiner als .65 bei der Itemanalyse als Ausschlusskriterien.

3.5.3 Zur Validität

Im Rahmen der Beschreibung der Konstruktion des Rating-Verfahrens in 3.4.3 wurde bereits darauf eingegangen, dass zum einen der Einbezug der Urteile der Instrumententwicklerin in die Prüfung der Interrater-Reliabilität und zum anderen die interpersonale Konsensbildung als eine Form der konsensualen Validierung als Möglichkeiten der Validitätsprüfung speziell für inhaltsanalytische Vorgehen betrachtet werden können. In Folgenden wird auf weitere Validitätsprüfungen eingegangen.

3.5.3.1 Kriteriumsvalidität

Zur Überprüfung der 'concurrent validity' als einer Form der Kriteriumsvalidität kann auf die von Schnell, Hill und Esser (2005) vorgeschlagene 'Methode der bekannten Gruppen' zurückgegriffen werden (vgl. die Ausführungen in Abschnitt 1.5.1.3). Die Lehrkräfte der Videostichprobe verteilen sich gemäß den Darlegungen in 3.1 auf verschiedene Fortbildungsgruppen. 20 der 30 Lehrkräfte nahmen an einer Präsenzfortbildung mit tutorieller Unterstützung teil (Experimentalgruppen, EGs); 10 Lehrkräfte erarbeiteten sich die Inhalte der Fortbildung im Selbststudium (Kontrollgruppe, KG). Als Kriterium kann hier die Zugehörigkeit der Lehrkräfte zu den Fortbildungsgruppen (EGs und KG) gelten.

Geht man davon aus, dass sich die Fortbildungsgruppen nach der Intervention (zum Zeitpunkt der Unterrichtsaufzeichnung „Schwimmen und Sinken“) hinsichtlich ihres unterrichtlichen Handelns zu Gunsten der Experimentalgruppen (EGs) unterscheiden, so sollte das Instrument diese Differenzen erfassen können, um 'concurrent validity' zu besitzen. Diese Annahme lässt sich dadurch motivieren, dass die Fortbildungen Einfluss auf das Lehrerhandeln haben.

3.5.3.2 Konstruktvalidität

Faktorenanalysen. Zur Überprüfung der Dimensionalität der erfassten Daten ist es üblich, Faktorenanalysen durchzuführen. Für faktorenanalytische Testungen stehen zwei Verfahrenstypen zur Verfügung, explorative und konfirmatorische Faktorenanalysen. Während *explorative Faktorenanalysen* eine Datenreduktion und das Zurückführen von Zusammenhängen zwischen Items auf latente Dimensionen zum Ziel haben, dienen *konfirmatorische Faktorenanalysen* dazu, theoretisch oder empirisch gut fundierte Modelle auf ihre Modellgüte zu testen (Bühner, 2004). Bei konfirmatorischen Verfahren liegt also a priori ein Modell über hypothetische Faktoren zugrunde, d.h. aufgrund theoretischer Vorüberlegungen bestehen Annahmen über die Beziehungen zwischen den einzelnen Variablen und deren Zuordnung zu den latenten Dimensionen (vgl. Bühl & Zöfel, 2008; Döring, 2005). Da dem zu testenden Instrument eine theoretisch angenommene Struktur zugrunde liegt, würde sich die Durchführung konfirmatorischer Faktorenanalysen anbieten, um zu überprüfen, ob und inwiefern die empirisch erho-

benen Daten dem hypothetischen Modell entsprechen und um weiterhin zu überprüfen, ob die auf der Basis theoretischer Vorüberlegungen entstandenen idealtypischen Konstrukte haltbar sind. Auf die Durchführung dieses Verfahrens wird jedoch aus den folgenden Gründen verzichtet: In der Literatur wird darauf hingewiesen, dass für konfirmatorische Faktorenanalysen die Stichprobe relativ groß sein sollte. Genauer werden zur Vermeidung von Schätzproblemen, die bei kleinen Stichprobenumfängen ($N=100$) bei konfirmatorischen Faktorenanalysen auftreten können, Stichprobengrößen ab einem N von 200 empfohlen. Zudem sollte das Verhältnis von Stichprobengröße zu Item- bzw. Variablenanzahl nicht unter 5:1 liegen (bei 10 Variablen ergibt sich also eine Stichprobengröße von mindestens $N=50$) (vgl. Bühner, 2004). Bei der zur Testung des Instruments vorliegenden kleinen Stichprobe von $N=59$ und dem sehr ungünstigen Verhältnis von Unterrichtsvideos zu Items ($N=63$) sind keine aussagekräftigen Ergebnisse zu erwarten.

Multitrait-Multimethod-Verfahren. Gemäß den theoretischen Ausführungen zu Möglichkeiten der Güteprüfung bei Videoanalysen in Abschnitt 1.5.1.3, kann die Konstruktvalidität eines Instruments neben einer faktoriellen Testung bei entsprechender Datenlage auch mittels des Multitrait-Multimethod-Verfahrens nach Cambell und Fiske (1959) abgeschätzt werden. Bei diesem Verfahren werden mehrere Konstrukte (Multitrait) durch mehrere Methoden (Multimethod) erfasst. Die Höhe der Konstruktvalidität wird abgeschätzt, indem die wechselseitigen Beziehungen zwischen Konstrukten und Methoden analysiert werden. Das Verfahren unterscheidet zwei Bestandteile der Konstruktvalidität: konvergente und diskriminante Validität. Konvergente Validität liegt vor, wenn verschiedene Methoden (z.B. Tests) dasselbe Konstrukt mit großer Übereinstimmung erfassen; diskriminante Validität liegt vor, wenn verschiedene Konstrukte durch eine Methode differenziert werden. Beide Validitätsformen werden durch Zusammenhangsmaße abgeschätzt, wobei die wechselseitigen Zusammenhänge in einer speziellen Korrelationsmatrix dargestellt werden (vgl. Bortz & Döring, 2006, S. 203). Um das Multitrait-Multimethod-Verfahren als besonders sorgfältige Methode der Konstruktvalidierung auf das vorliegende Videoanalyse-Instrument anwenden zu können, bräuchte es eine weitere Methode, welche die entwickelten Unterrichtshandlungen der Lehrkräfte erfasst. Da jedoch kein weiteres Instrument vorliegt, welches derartige Lehrerverhaltensweisen misst, können im Rahmen dieser Arbeit keine Multitrait-Multimethod-Matrizen modelliert werden.

Expertenvalidierung. Da eine empirische, faktorielle Überprüfung der Konstruktvalidität aus den erläuterten Gründen nicht möglich ist, wurde die Zuordnung der Items zu den theoretischen Konstrukten in einem Experten-Verfahren validiert. Dazu erhielten die 'Experten' eine theoretische Beschreibung der Konstrukte und die Items in beliebiger Reihenfolge. Die Items waren dann von den Experten blind zurück zu sortieren.

An der Expertenvalidierung nahmen fünf Personen teil, drei Fachexperten und zwei Lehrpersonen. Die Lehrpersonen waren mit der Theorie, die den Konstrukten zugrunde liegt, vertraut.

4 Ergebnisse

In diesem Kapitel werden die Ergebnisse der Güteprüfungen des Video-Instruments dargestellt. In den ersten beiden Teilkapiteln werden Ergebnisse der Reliabilitätsprüfung vorgestellt. Dazu gehören die Ergebnisse der Generalisierbarkeitsanalyse zur Prüfung der Interrater-Reliabilität (4.1) und der Analyse zentraler Kennwerte der Items und Skalen (4.2). In Teilkapitel 4.3 werden Ergebnisse zu Unterschieden im Unterrichtshandeln der Lehrkräfte zur Analyse der 'concurrent validity' als einer Form der Kriteriumsvalidität dargestellt. Im letzten Teilkapitel (4.4) werden die Ergebnisse der Expertenvalidierung zur Prüfung der Konstruktvalidität berichtet.

4.1 Ergebnisse der Generalisierbarkeitsanalyse

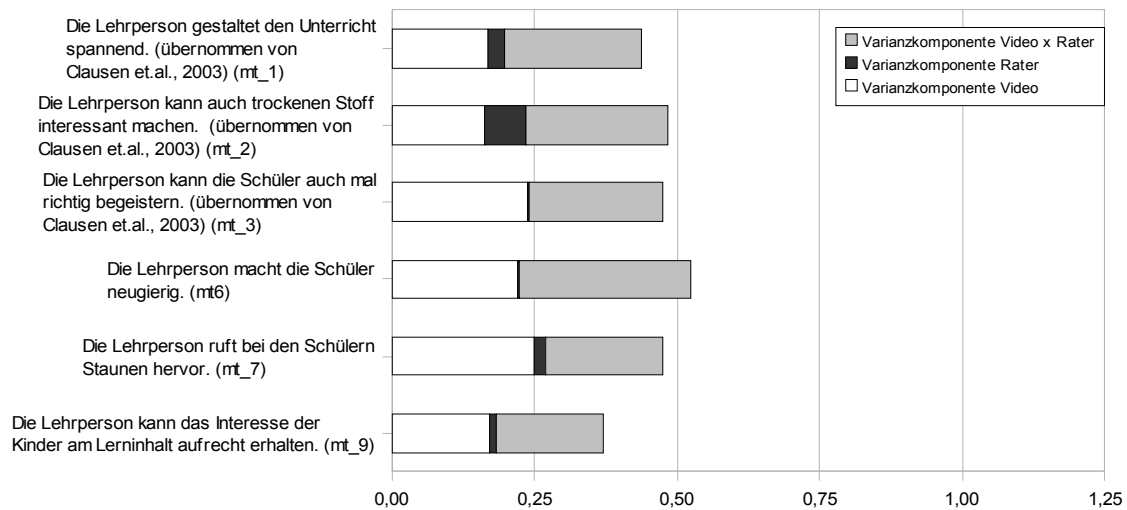
Mit Hilfe der Generalisierbarkeitsanalyse wurde die Qualität der Beurteilungen auf der Ebene der Items untersucht. Die folgenden Grafiken zeigen die Ergebnisse der Itemanalysen für die theoretisch angenommenen Konstrukte. Dargestellt ist jeweils die beobachtete Varianz der Ratings und die Zerlegung in die Varianzkomponenten für Video, Rater und die Interaktion Video x Rater³⁹. Zudem ist jeweils die Anzahl der Videos angegeben, die in die Analysen eingegangen sind. Aufgeführt sind auch die Generalisierbarkeitskoeffizienten für absolute Entscheidungen⁴⁰.

Die Grafiken können folgendermaßen interpretiert werden: Ist die Varianzkomponente der Videos gleich der Summe der Varianzkomponenten Rater und Video x Rater, so ergibt sich immer ein absoluter Generalisierbarkeitskoeffizient von .80. Ist sie größer (kleiner), so ist der Koeffizient stets größer (kleiner) als .80. Auf der Basis der Ausführungen in Abschnitt 3.3.2.1 werden Generalisierbarkeitskoeffizienten ab .70 für gut befunden. Dies ist bei einer Untersuchung mit vier Ratern dann erfüllt, wenn die Varianzkomponente der Videos mindestens knapp 60% der Summe der anderen beiden Varianzkomponenten beträgt (siehe dazu die Rechnung in Anhang B.2). Items mit einem Generalisierbarkeitskoeffizienten unterhalb des Schwellenwertes von .70 werden jeweils unter den einzelnen Grafiken aufgeführt.

³⁹Die den Grafiken zugrunde liegenden Werte der Varianzkomponenten sind nachzulesen in Tabelle A.1-1 in Anhang A.1.

⁴⁰Berichtet werden nur die absoluten Generalisierbarkeitskoeffizienten, da die Varianz der Rater auf den relativen Koeffizienten keinen Einfluss hat (vgl. Abschnitt 3.5.2.1).

Abb. 4.1-1 Motivation des Unterrichts

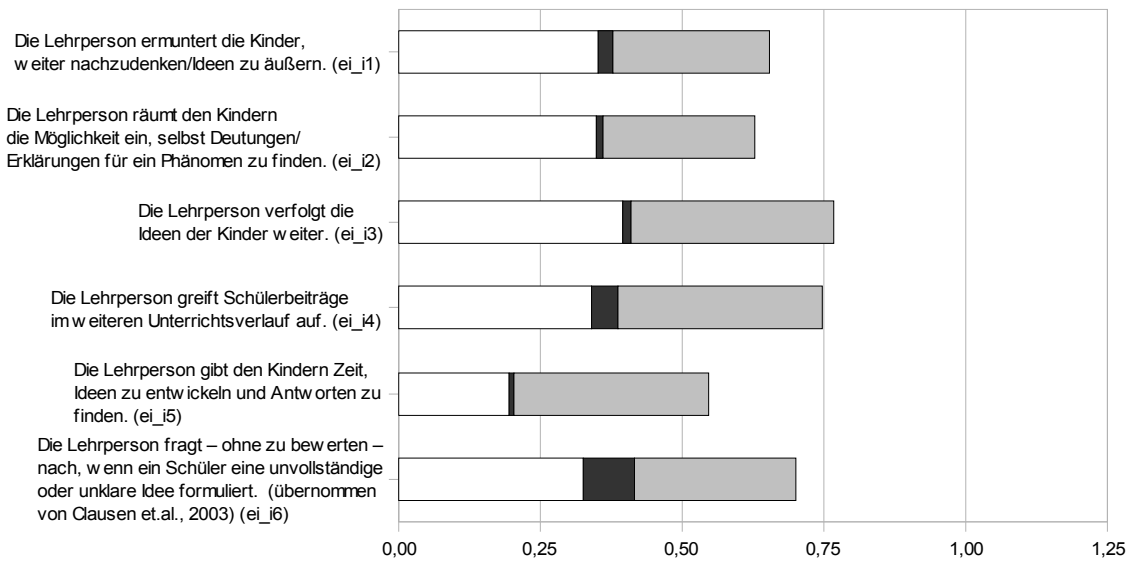


N=57

Generalisierbarkeitskoeffizient (absolut): mt_2: $\Phi = .67$; restliche Items: Φ von .71 bis .81.

Bei den Items, die dem theoretischen Konstrukt „Motivation des Unterrichts“ zugeordnet sind, wird für das Item mt_2 keine ausreichende Übereinstimmung erreicht (Koeffizient von .67). Bei allen anderen Items liegen die Koeffizienten zwischen .71 und .81. Die mangelnde Übereinstimmung bei Item mt_2 ergibt sich aus dem ungünstigen Verhältnis der Varianzkomponente Video als der im Sinne der klassischen Testtheorie 'wahren Varianz' zu der Summe der systematischen Raterunterschiede (Varianzkomponente Rater) und der systematischen Fehlervarianz (Varianzkomponente Video x Rater). Zurückzuführen ist dies auf die relativ geringe Varianz in den Unterrichtsvideos.

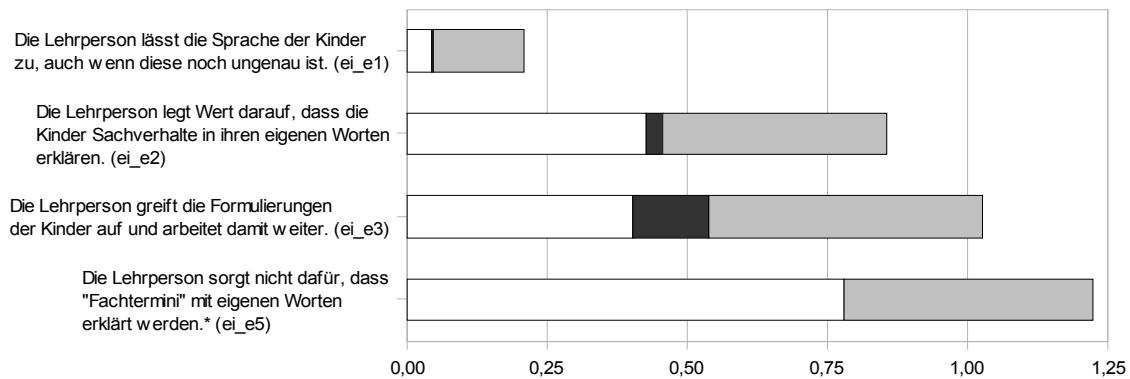
Abb. 4.1-2 Eigene Ideen zulassen



N=57

Generalisierbarkeitskoeffizient (absolut): ei_i5 $\Phi = .69$; restliche Items: Φ von .77 bis .83.

Abb. 4.1-3 Eigene Formulierungen

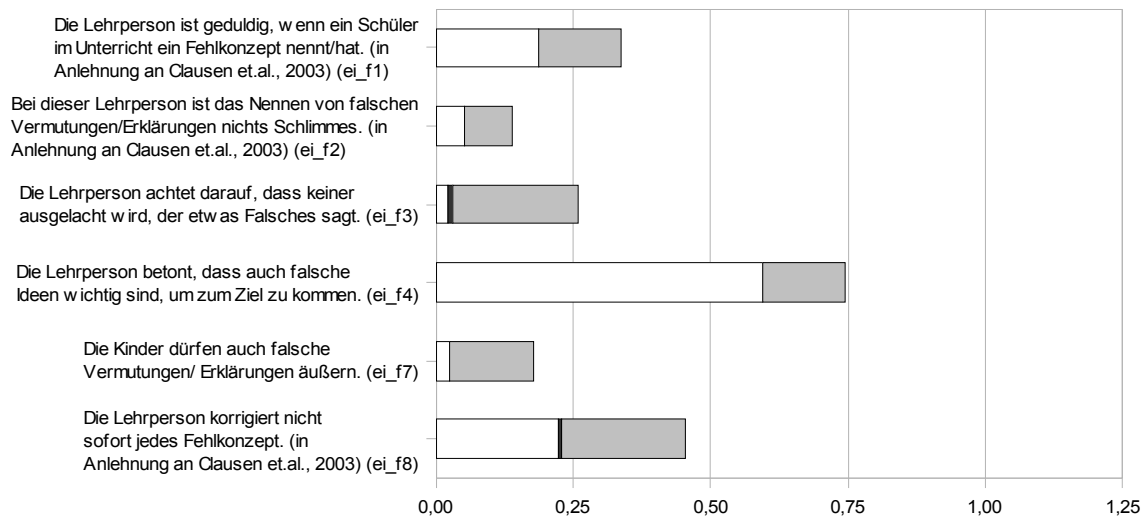


* umgepoltes Item

N=57

Generalisierbarkeitskoeffizient (absolut): ei_e1 $\Phi = .69$; restliche Items: Φ von .72 bis .88.

Abb. 4.1-4 Fehleroffenheit



N=57

Generalisierbarkeitskoeffizient (absolut): ei_f3: $\Phi = .26$; ei_f7: $\Phi = .40$; restliche Items: Φ von .71 bis .94.

Bei den Items der Konstrukte „Eigene Ideen zulassen“, „Eigene Formulierungen“ und „Fehleroffenheit“ können bei den folgenden Items keine ausreichenden Übereinstimmungen erreicht werden:

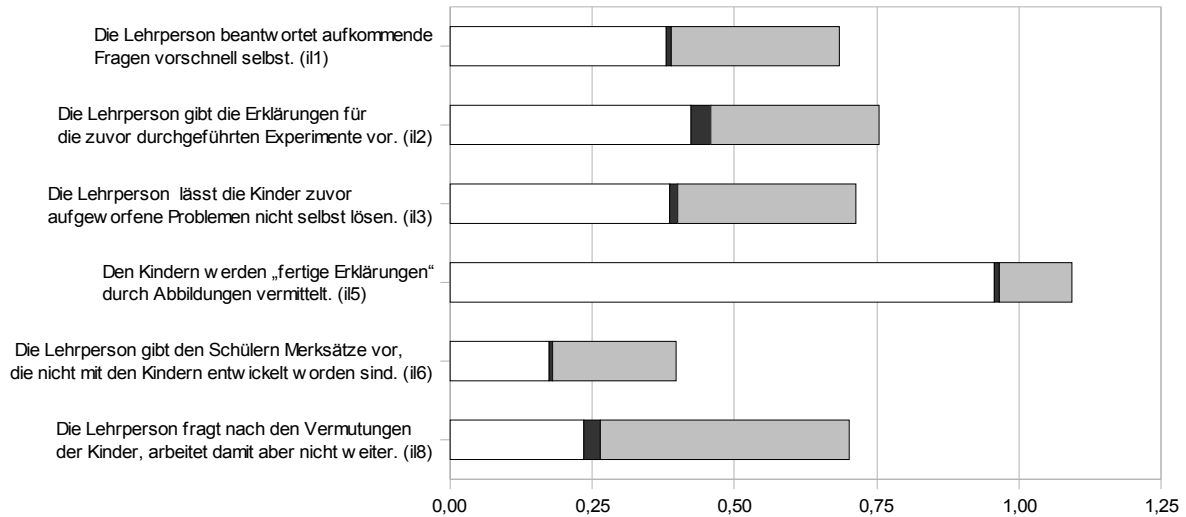
- „Die Lehrperson gibt den Kindern Zeit, Ideen zu entwickeln und Antworten zu finden“ (ei_i5) (Koeffizient von .69)
- „Die Lehrperson lässt die Sprache zu, auch wenn diese noch ungenau ist“ (ei_e1) (Koeffizient von .52)
- „Die Lehrperson achtet darauf, dass keiner ausgelacht wird, der etwas Falsches sagt“ (ei_f3) (Koeffizient von .26)
- „Die Kinder dürfen auch falsche Vermutungen/Erklärungen äußern“ (ei_f7) (Koeffizient von .40)

Für die übrigen Items ergeben sich absolute Generalisierbarkeitskoeffizienten von .71 bis .94.

Die unzureichende Übereinstimmung der genannten vier Items ist auch hier auf die geringe Varianz in den Videos zurückzuführen, die zwangsläufig dazu führt, dass die Varianzkomponente Video in einem ungünstigen Verhältnis zu systematischen und unsystematischen Fehleranteilen steht. Besonders deutlich wird dies bei den drei zuletzt genannten Items, für die sich Deckeneffekte ergeben, d.h.: In den Videos ist selten zu beobachten, dass die Lehrperson die Sprache der Kinder nicht zulässt ($M=3.74$, $SD=.39$), dass die Lehrperson nicht darauf achtet, dass keiner ausgelacht wird, der etwas Falsches sagt ($M=3.9$, $SD=.28$) und dass die Kinder keine falschen Vermutungen äußern dürfen ($M=3.84$; $SD=.44$).

Dies ist in Anbetracht dessen, dass die Lehrkräfte in Grundschulen meist schülerorientiert denken und arbeiten, nicht verwunderlich.

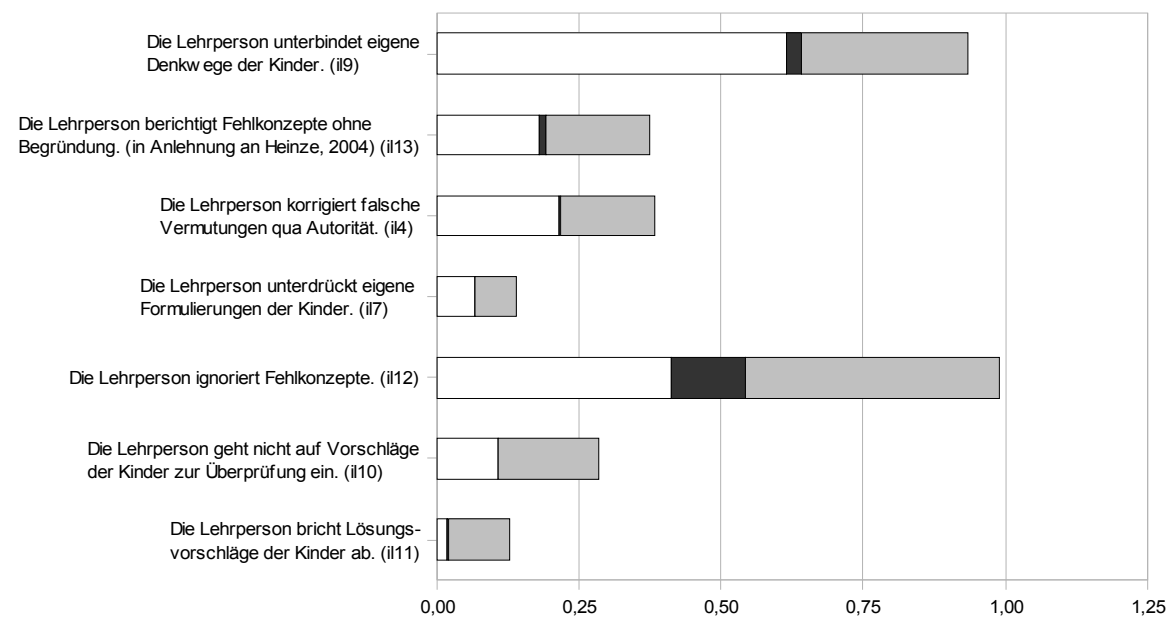
Abb. 4.1-5 Transmission von Wissen – Erklärungen



N=56

Generalisierbarkeitskoeffizient (absolut): il8 $\Phi = .67$; restliche Items: Φ von .76 bis .97.

Abb. 4.1-6 Transmission von Wissen – Denkwege ablehnen



N=56

Generalisierbarkeitskoeffizient (absolut): il11: $\Phi = .41$; restliche Items: Φ von .71 bis .84.

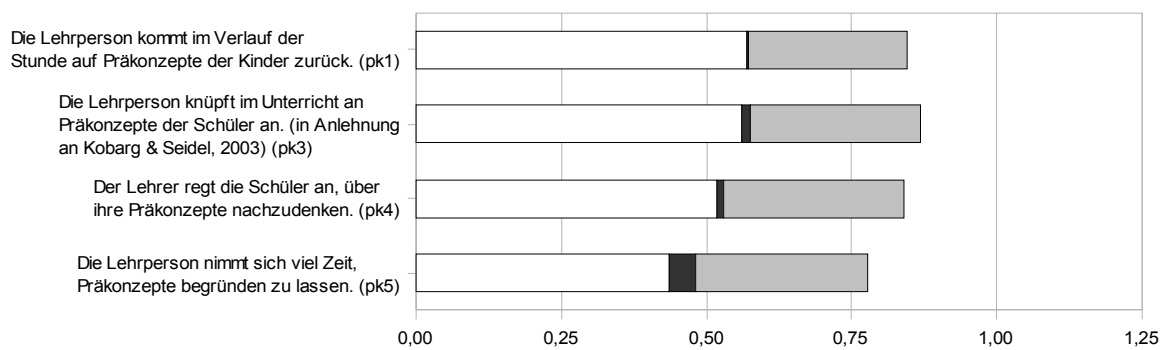
Für den Bereich „Transmission von Wissen“ (Erklärungen, Denkwege ablehnen) können für zwei Items keine ausreichenden absoluten Generalisierbarkeitskoeffizienten erreicht werden. Dies sind:

- „Die Lehrperson fragt nach den Vermutungen der Kinder, arbeitet damit aber nicht weiter“ (il8) (Koeffizient von .67)
- „Die Lehrperson bricht Vorschläge der Schüler ab“ (il11) (Koeffizient von .41)

Die übrigen Items erreichen gute Generalisierbarkeitskoeffizienten zwischen .71 und .81.

Bei Item il11 liegt mit einem Mittelwert von 1.12 (SD=.24) ein Bodeneffekt vor, d.h. es wird in den Videos – erfreulicherweise – kaum bis gar nicht beobachtet, dass die Lehrperson Lösungsvorschläge abbricht. Auch wird kaum beobachtet, dass die Lehrperson Formulierungen der Kinder unterdrückt (il7) (M=1.24; SD=.37). Durch die geringe systematische Fehlervarianz konnte bei diesem Item aber dennoch ein ausreichend hoher absoluter Generalisierbarkeitskoeffizient erreicht werden.

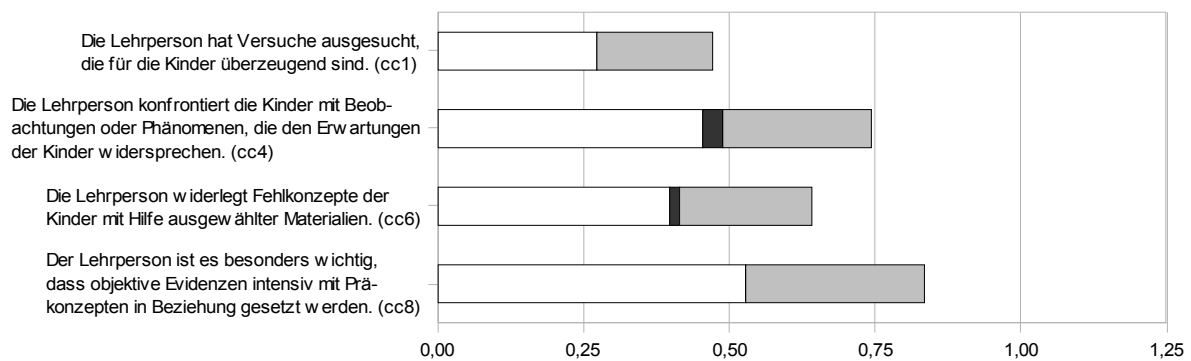
Abb. 4.1-7 Präkonzepte – Vorwissen aktivieren und aufgreifen



N=56

Generalisierbarkeitskoeffizient (absolut): Φ von .85 bis .89.

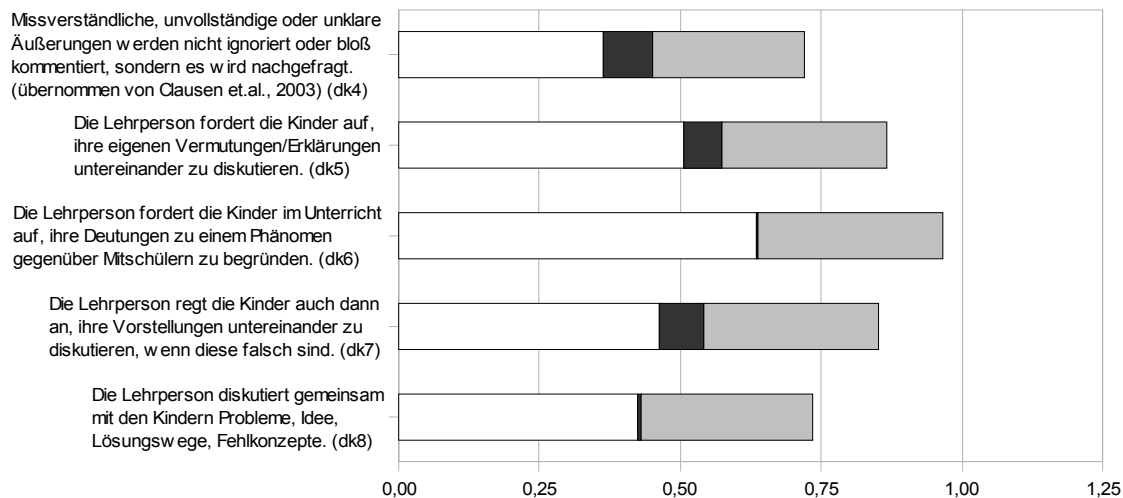
Abb. 4.1-8 Conceptual Change – kognitive Konflikte erzeugen



N=56

Generalisierbarkeitskoeffizient (absolut): Φ von .85 bis .87.

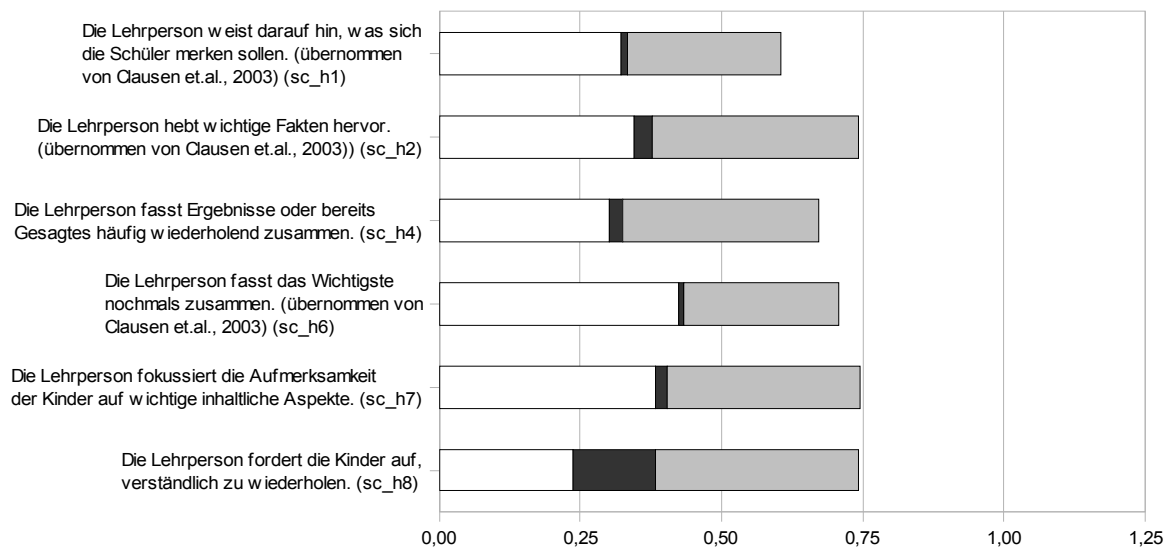
Abb. 4.1-9 Ideen gemeinsam diskutieren



N=56

Generalisierbarkeitskoeffizient (absolut): Φ von .80 bis .88.

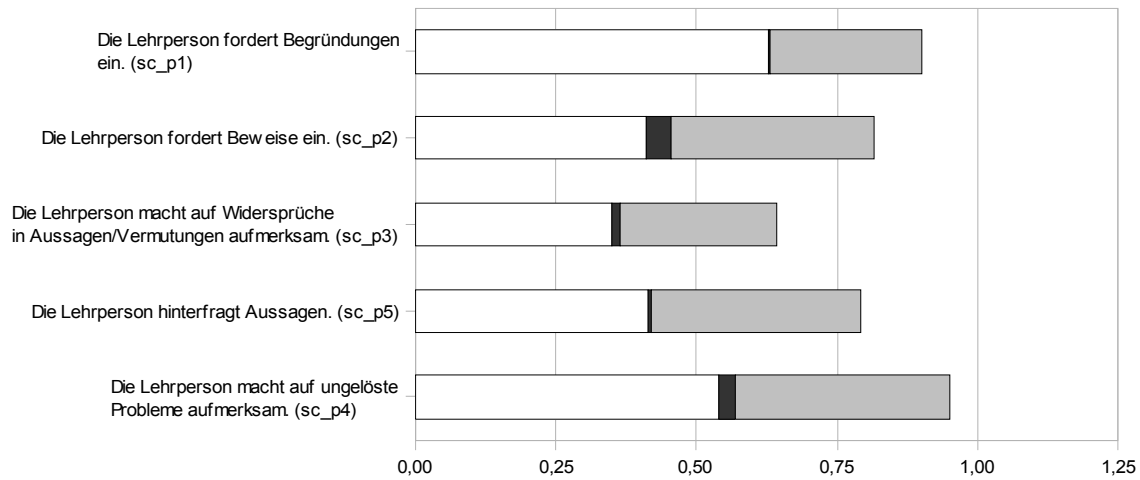
Abb. 4.1-10 Scaffolding – Hervorhebungen



N=56

Generalisierbarkeitskoeffizient (absolut): sc_h8 Φ =.65; restliche Items: Φ von .76 bis .86.

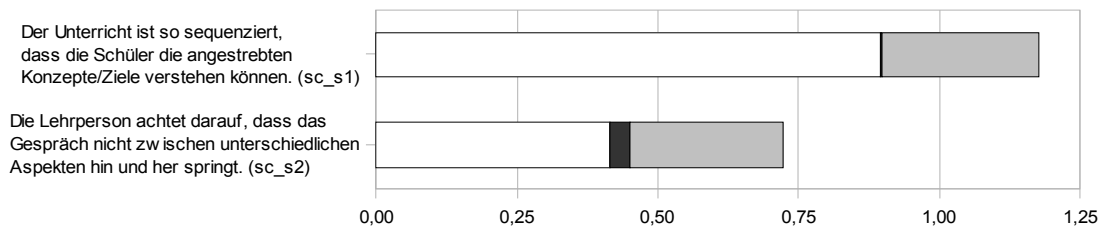
Abb. 4.1-11 Scaffolding – Perturbieren



N=56

Generalisierbarkeitskoeffizient (absolut): Φ von .80 bis .90.

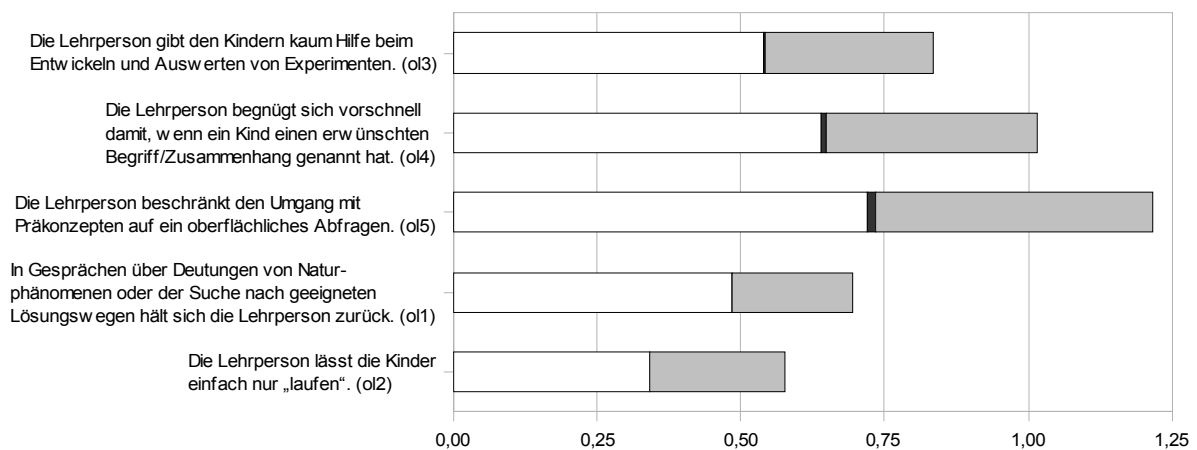
Abb. 4.1-12 Scaffolding – Sequenzierungen



N=57

Generalisierbarkeitskoeffizient (absolut): Φ =.93 und .94.

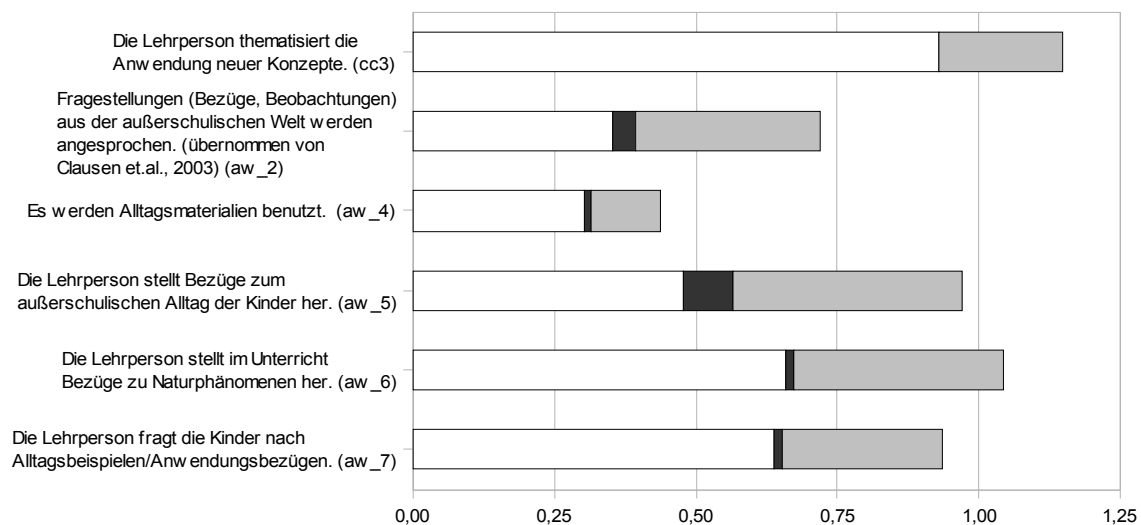
Abb. 4.1-13 Laisser-faire



N = 56

Generalisierbarkeitskoeffizient (absolut): Φ von .85 bis .90.

Abb. 4.1-14 Anwendungsbezogenes Lernen



N=57

Generalisierbarkeitskoeffizient (absolut): Φ von .79 bis .94.

Für die Items der theoretischen Dimensionen „Präkonzepte – Vorwissen aktivieren und aufgreifen“, „Conceptual Change – kognitive Konflikte“, „Ideen gemeinsam diskutieren“, „Laisser-faire“, „Anwendungsbezogenes Lernen“ und der drei Scaffolding-Konstrukte (Hervorhebungen, Perturbieren, Sequenzierungen) zeigen sich gute bis sehr gute Generalisierbarkeitskoeffizienten bei fast allen Items. Die einzige Ausnahme stellt das Item „Die Lehrperson fordert die Kinder auf, verständlich zu wiederholen“ (sc_h8) aus dem Konstrukt „Scaffolding – Hervorhebungen“ dar, für das sich ein niedriger absoluter Generalisierbarkeitskoeffizient von .65 ergibt. Zurückzuführen ist dies auf den relativ hohen Anteil an systematischen Raterunterschieden (Varianzkomponente Rater), d.h. die Rater unterscheiden sich hinsichtlich ihres mittleren Urteilsniveaus. Die Rater scheinen das Item unterschiedlich verstanden zu haben, was nicht überraschend ist. Für zukünftige Anwendungen müsste konkretisiert werden, was mit 'verständlich zu wiederholen' gemeint ist. Bei den übrigen Items stehen die Varianzkomponenten Video durchweg in einem günstigen Verhältnis zu den systematischen und unsystematischen Fehlervarianzen. Insgesamt zeigt sich in Bezug auf die letztgenannten Konstrukte, dass das bei diesen Dimensionen beobachtete Verhalten der Lehrkräfte deutlich mehr Varianz aufweist als in den zuvor beschriebenen.

4.2 Ergebnisse der Item- und Skalenanalyse

Die Ergebnisse des vorherigen Teilkapitels zeigen, für welche Items gute bis sehr gute Generalisierbarkeitskoeffizienten erreicht wurden. Diese Items werden nun einer weiteren Reliabilitätsanalyse unterzogen. Im Folgenden werden die üblichen Testkennwerte, d.h. Mittelwerte und Standardabweichungen für Items und Skalen, die korrigierte Item-Skalenwert-Korrelation als Maß für die Trennschärfe der Items und die internen Konsistenzen mittels Cronbachs Alpha berichtet. Die Kennwerte werden getrennt für die Themen „Wasserkreislauf“ und „Schwimmen und Sinken“ berechnet und sind den Tabellen 4.2-1 und 4.2-2 zu entnehmen. Anders als die Grafiken zur Generalisierbarkeitsanalyse beinhalten die Tabellen nur noch Items, die nach den Reliabilitätsanalysen bestehen geblieben sind⁴¹. Nach der in diesem Teilkapitel durchgeführten Reliabilitätsanalyse wurden weiterhin die folgenden Items aufgrund von niedrigen Trennschärfen in beiden Themengebieten und schlechten Werten von Cronbachs Alpha ausgeschlossen:⁴²

- Ei_f4 Die Lehrperson betont, dass auch falsche Ideen wichtig sind, um zum Ziel zu kommen.
- II5 Den Kindern werden 'fertige Erklärungen' durch Abbildungen vermittelt.
- II7 Die Lehrperson unterdrückt eigene Formulierungen der Kinder.
- II10 Die Lehrperson geht nicht auf Vorschläge der Kinder zur Überprüfung ein.
- Sc_p4 Die Lehrperson macht auf ungelöste Probleme aufmerksam.
- Aw4 Es werden Alltagsmaterialien benutzt.

In Bezug auf die Itemschwierigkeit, welche durch Mittelwert und Standardabweichung beschrieben wird, zeigt sich, dass bei einigen Items Decken- und Bodeneffekte vorliegen.

Deckeneffekte betreffen Items der Skala „Fehleroffenheit“, und zwar sowohl beim Thema „Wasserkreislauf“ als auch beim Thema „Schwimmen und Sinken“. Die Streuungen in den Items sind dementsprechend niedrig (SD von .29 bis .69 sowie von .28 bis .46). Die Deckeneffekte spiegeln sich auch in den Skalenmittelwerten wider (M=3.58 bzw. M=3.86). Diese Beobachtungen decken sich mit den Ergebnissen der Generalisierbarkeitsanalyse, welche ebenfalls gezeigt hat, dass bei diesem Konstrukt kaum Varianz in den Videos zu finden ist. Ein weiterer Deckeneffekt (M=3.58) tritt bei Item ei_e5 aus

⁴¹Eine detaillierte Übersicht über die Mittelwerte, Standardabweichungen und Trennschärfen aller Items befindet sich in Anhang A.2.

⁴²Ausgeschlossen wurden Items mit einem Trennschärfenwert unterhalb von .30. Eine Ausnahme stellt die Aufnahme zweier Items aus der Skala „Motivation des Unterrichts“ dar, dessen Trennschärfen Werte unterhalb von .30 für das Thema „Wasserkreislauf“ erreichen (.26 bzw. .24). Gleiches gilt für ein Item der Skala „Transmission von Wissen – Erklärungen“, dessen Trennschärfe mit einem Wert von .23 beim Thema „Schwimmen und Sinken“ ebenfalls unterhalb der gesetzten Grenze liegt. Da die Trennschärfen der genannten drei Items beim jeweils anderen Unterrichtsthema deutlich höher liegen, verbleiben die Items in den Skalen.

der Skala „Eigene Formulierungen“ beim Thema „Schwimmen und Sinken“ auf, wobei die Mittelwerte hier allgemein hoch sind (M von 2.97 bis 3.58).

Bodeneffekte zeigen sich in nahezu allen Items der beiden Transmissions-Skalen und in den zugehörigen Skalenmittelwerten der Stichprobe zum Thema „Schwimmen und Sinken“ sowie bei einzelnen Items zum Thema „Wasserkreislauf“ (il6 und il13, il4). Des Weiteren liegen beim Thema „Wasserkreislauf“ bei je zwei Items der Konstrukte „Ideen diskutieren“, „Präkonzepte“ und „Laisser-faire“ Bodeneffekte vor. Bei der Skala „Laisser-faire“ finden sich außerdem beim Thema „Schwimmen und Sinken“ in nahezu allen Items Bodeneffekte. Entsprechend wirkt sich diese Tatsache auf den Skalenmittelwert (M=1.57) aus.

Tab. 4.2-1 Statistische Kennwerte der Skalen und Items zum Thema „Wasserkreislauf“ (N=30)

Skala	N _{Items}	M _{Skala}	SD _{Skala}	M _{Items} (min.-max.)	SD _{Items} (min.-max.)	r _{it} (min.-max.)	Cronbachs Alpha
Motivation des Unterrichts	5	2.22	.33	1.80 - 2.58	.35 - .60	.24 - .65	.70
Eigene Ideen zulassen	5	2.43	.58	2.31 - 2.61	.63 - .73	.53 - .87	.90
Eigene Formulierungen	3	2.35	.64	2.22 - 2.61	.68 - .99	.42 - .72	.73
Fehleroffenheit	3	3.58	.44	3.49 - 3.80	.29 - .69	.49 - .71	.69
Transmission – Erklärungen	4	2.00	.65	1.52 - 2.17	.72 - .78	.49 - .86	.88
Transmission – eigene Denk- und Lernwege ablehnen	4	1.93	.59	1.54 - 2.31	.59 - .84	.54 - .80	.83
Präkonzepte – Vorwissen aktivieren und aufgreifen	4	1.66	.56	1.33 - 1.88	.39 - .75	.78 - .93	.92
Conceptual Change – kognitive Konflikte erzeugen	4	2.15	.56	1.75 - 2.55	.56 - .89	.56 - .78	.83
Ideen gemeinsam diskutieren	5	1.76	.47	1.40 - 2.45	.50 - .72	.59 - .70	.83
Scaffolding – Hervorhebungen	5	2.12	.62	1.65 - 2.37	.57 - .77	.87 - .93	.97
Scaffolding – Perturbieren	4	1.86	.49	1.60 - 2.22	.65 - .70	.35 - .75	.71
Scaffolding – Sequenzierungen	2	2.09	.75	1.88 - 2.29	.71 - .94		.77
Laisser-faire	5	2.01	.68	1.44 - 2.60	.73 - .88	.59 - .84	.89
Anwendungsbezogenes Lernen	5	2.54	.63	2.13 - 2.94	.65 - .90	.50 - .79	.68

Antwortformat von 1 bis 4 ((1) trifft überhaupt nicht zu, (2) trifft eher nicht zu, (3) trifft eher zu, (4) trifft voll und ganz zu)

r_{it}: Trennschärfe der Items

Tab. 4.2-2 Statistische Kennwerte der Skalen und Items zum Thema „Schwimmen u. Sinken“ (N=29)

Skala	N _{Items}	M _{Skala}	SD _{Skala}	M _{Items} (min.-max.)	SD _{Items} (min.-max.)	r _{it} (min.-max.)	Cronbachs Alpha
Motivation des Unterrichts	5	2.66	.48	2.54 - 2.83	.55 - .64	.52 - .79	.87
Eigene Ideen zulassen	5	3.16	.53	2.97 - 3.39	.52 - .66	.63 - .91	.93
Eigene Formulierungen	3	3.22	.58	2.97 - 3.58	.59 - .76	.62 - .82	.84
Fehleroffenheit	3	3.86	.36	3.82 - 3.93	.28 - .46	.80 - .92	.91
Transmission – Erklärungen	4	1.25	.39	1.12 - 1.35	.27 - .54	.23 - .88	.86
Transmission – eigene Denk- und Lernwege ablehnen	4	1.31	.42	1.14 - 1.74	.39 - .72	.45 - .85	.84
Präkonzepte – Vorwissen aktivieren und aufgreifen	4	2.68	.71	2.38 - 2.85	.71 - .81	.85 - .92	.95
Conceptual Change – kognitive Konflikte erzeugen	4	2.81	.53	2.63 - 3.12	.45 - .81	.82 - .91	.93
Ideen gemeinsam diskutieren	5	2.49	.77	2.11 - 3.04	.86 - 1.00	.69 - .89	.92
Scaffolding – Hervorhebungen	5	2.41	.75	1.89 - 2.70	.72 - .87	.88 - .95	.97
Scaffolding – Perturbieren	4	2.57	.67	2.04 - 2.90	.66 - .88	.64 - .90	.91
Scaffolding – Sequenzierungen	2	3.00	.62	2.83 - 3.18	.58 - .74		.86
Laisser-faire	5	1.57	.66	1.39 - 1.76	.61 - .89	.74 - .89	.94
Anwendungsbezogenes Lernen	5	1.67	.34	1.06 - 2.57	.14 - .69	.23 - .70	.78

Antwortformat von 1 bis 4 ((1) trifft überhaupt nicht zu, (2) trifft eher nicht zu, (3) trifft eher zu, (4) trifft voll und ganz zu)

r_{it}: Trennschärfe der Items

Die internen Konsistenzen sind insgesamt sehr zufriedenstellend, die Werte für Cronbachs Alpha liegen zwischen .69 und .97. Auffällig ist, dass die Werte für Cronbachs Alpha beim Thema „Wasserkreislauf“ durchweg höher ausfallen als beim Thema „Schwimmen und Sinken“. Insbesondere fallen stärkere Abweichungen der Alpha-Werte zwischen den Themen bei den Skalen „Fehleroffenheit“ (Wasserkreislauf: Cronbachs Alpha von .69; Schwimmen und Sinken: Cronbachs Alpha von .91) und „Scaffolding – Perturbieren“ (Wasserkreislauf: Cronbachs Alpha von .71; Schwimmen und Sinken: Cronbachs Alpha von .91) auf. Eine Ausnahme stellt die Skala „Scaffolding – Hervorhebungen“ dar; hier betragen die Alpha-Werte für beide Themen .97.

Die Item-Skalen-Korrelation liegt bei fünf Items unter .30, davon sind zwei beim Thema „Wasserkreislauf“ (Item mt_9 mit .24 und Item mt_6 mit .26) und drei beim Thema „Schwimmen und Sinken“ (Item il6 mit .23, Item cc_3 mit .23 und Item aw2 mit .29). Da eine niedrige Trennschärfe für ein Item

jeweils nur bei höchstens einem der Themen „Wasserkreislauf“ und „Schwimmen und Sinken“ vorkommt, wurde keines der Items eliminiert.

Insgesamt kann von einer hinreichenden Reliabilität der Messwerte ausgegangen werden.

Die Tabelle 4.2-3 zeigt die Korrelationen der Skalen untereinander (n. Pearson 2-seitig) zum Zeitpunkt der Unterrichtserhebung „Wasserkreislauf“ (N=30)⁴³. Vor dem Hintergrund der Konventionen von Cohen, wonach Korrelationen um +/- .10 als kleine Zusammenhänge, um +/- .30 als mittlere Zusammenhänge und um +/- .50 als starke Zusammenhänge interpretiert werden können (Sedlmeier & Renkewitz, 2008, S. 221), weisen die Interskalenkorrelationen teilweise auf mittlere bis starke Zusammenhänge hin. Die in dieser Arbeit konzeptualisierten Skalen stellen Teilbereiche kognitiv anregenden Lehrerverhaltens dar, gewisse Zusammenhänge liegen nahe. Somit wird aus theoretischen Erwägungen heraus nicht davon ausgegangen, dass die Skalen unkorreliert sind. Vielmehr wird ein Korrelationsmuster erwartet.

So ist zu erwarten, dass die drei Skalen „Eigene Ideen zulassen“, „Eigene Formulierungen“ und „Fehleroffenheit“, die aus dem Indikator *'Gelegenheiten zu eigenen Wissenskonstruktionen schaffen und dabei individuelle Lernwege zulassen'* operationalisiert worden sind, positiv korrelieren. Gleiches gilt sowohl für die zwei Transmissions-Skalen als auch für die drei Skalen aus dem Bereich Scaffolding, die jeweils innerhalb signifikant positiv korrelieren. Ebenso ist es nachvollziehbar, dass für die Skala „Laisser-faire“ signifikant negative Korrelationen mit allen Scaffolding-Skalen vorliegen, da Laisser-faire-Verhalten eine Negation des Scaffolding darstellt.

Auch sind Negativkorrelationen der Konstrukte zur Transmission von Wissen („Erklärungen“ und „Denk- und Lernwege ablehnen“) mit den Skalen „Eigene Ideen zulassen“, „Eigene Formulierungen“, „Fehleroffenheit“ und „Motivation des Unterrichts“ nachvollziehbar, da transmissives Lehrerverhalten gemäß der theoretischen Basis den Lernenden keinen Raum für individuelle Denk- und Lernwege einräumt. Umso weniger Zeit den Lernenden für individuelle Denk- und Lernwege im Unterricht eingeräumt wird, desto weniger motivierend wirkt wahrscheinlich der Unterricht auf die Schüler.

Die positiven Korrelationen zwischen der Skala „Präkonzepte – Vorwissen aktivieren und aufgreifen“ und den Skalen „Eigene Ideen zulassen“, „Eigene Formulierungen“, und „Fehleroffenheit“ liegen nahe, da es sich bei vorunterrichtlichen Schülervorstellungen im Prinzip um eigene Ideen der Kinder handelt, die umgangssprachlich formuliert sind und zudem fehlerbehaftet sind.

Auch der hoch signifikante starke Zusammenhang zwischen den beiden aus dem Indikator *'Umgang mit Präkonzepten und Anregung von Conceptual Change'* abgeleiteten Skalen „Conceptual Change – kognitive Konflikte erzeugen“ und „Präkonzepte – Vorwissen aktivieren und aufgreifen“ lässt sich erklären. Versucht die Lehrperson Konzeptwechsel bei den Lernenden anzuregen, d.h. Schülervorstellungen

⁴³Die Interskalenkorrelationen zum Messzeitpunkt „Schwimmen und Sinken“ werden aufgrund des möglichen Einflusses der Intervention auf das Unterrichtshandeln der Lehrkräfte nicht angegeben.

lungen zu verändern, bringt sie diese auch in Erfahrung und arbeitet an den Vorstellungen der Kinder im Unterricht.

Die Korrelationen der Skala „Conceptual Change – kognitive Konflikte erzeugen“ mit den Skalen aus dem Bereich des Scaffolding, insbesondere mit der Skala „Perturbieren“, verdeutlichen die starken theoretischen Bezüge; alle Konstrukte erfassten Lehrerverhaltensweisen zur Lern- und Denkkunterstützung (Maßnahmen zur kognitiven Strukturierung) (vgl. dazu auch die Ausführungen auf den Seiten 154f. und 162f).

Auch die Zusammenhänge der Skalen aus den Bereichen '*Gelegenheiten zu eigenen Wissenskonstruktionen schaffen und dabei individuelle Lernwege zulassen*' und '*Umgang mit Präkonzepten und Anregung von Conceptual Change*' mit dem Konstrukt „Motivation des Unterrichts“ sind nicht erstaunlich, da es nahe liegt, dass ein Unterricht, in dem Lernende Gelegenheiten erhalten eigene Ideen zu äußern und eigene Formulierungen zu nutzen, in dem Fehler begangen werden dürfen und in dem auch die vorunterrichtlichen (Fehl-)Vorstellungen der Lernenden aufgegriffen und verändert werden, motivierend auf die Schüler wirkt. Demgegenüber kann davon ausgegangen werden, dass ein tranmissiver Unterricht eine weniger motivierende Wirkung auf die Kinder hat (negative Korrelativen der Skalen „Erklärungen“ und „Denkwege ablehnen“ mit der Skala „Motivation des Unterrichts“).

An dieser Stelle werden nicht alle Korrelationen diskutiert. Insgesamt verdeutlichen die Ausführungen, dass sich für viele der berichteten Zusammenhänge inhaltliche Gründe finden lassen. Aufgrund der sorgfältigen theoretischen Ableitung der Konstrukte und in Anbetracht der hohen internen Konsistenzen der Skalen, wird die Trennung der Konstrukte in den nachfolgenden Analysen vorerst beibehalten. Aus methodischer Perspektive, vor allem vor dem Hintergrund der diskriminanten Validität der Skalen, stellen die Interskalenkorrelationen jedoch eine wichtige Information dar, aus denen sich wertvolle Hinweise für weiterführende Arbeiten in Folgestudien ziehen lassen. So ist z.B. über eine Reduktion der Beobachtungsdimensionen nachzudenken (siehe dazu auch die Ausführungen zu Folgerungen für Überarbeitungen des Instruments im Ausblick auf Seite 175f.).

Tab. 4.2-3 Interskalenkorrelationen zum Messzeitpunkt „Wasserkreislauf“ (N=30)

	Mot. Unt.	Eig Ideen	Eig Form.	Fehleroff.	Erklärungen	Denkw. ablehnen	Präkonzepte	Conceptual Change	Ideen diskutieren	Scaff. Hervorh.	Scaff. Perturb.	Scaff. Sequenzier.	Laisser-faire	Anwend. Lernen
Mot . Unt.	1.0													
Eig Ideen	.54**	1.0												
Eig Form.	.49**	.68**	1.0											
Fehleroff.	.38*	.55**	.36	1.0										
Erklärungen	-.43*	-.47*	-.33	-.63**	1.0									
Denkw. ablehnen	-.51**	-.51**	-.43*	-.69**	.85**	1.0								
Präkonzepte	.41*	.66**	.64**	.28	-.20	-.41*	1.0							
Conceptual Change	.55**	.73**	.72**	.26	-.21	-.33	.79**	1.0						
Ideen diskutieren	.51**	.82**	.60**	.36	-.35	-.37*	.62**	.65**	1.0					
Scaff. Hervorh.	.35	.56**	.55**	.05	.04	-.08	.72**	.73**	.44*	1.0				
Scaff. Perturb.	.49**	.70**	.62**	.11	-.11	-.19	.77**	.80**	.81**	.63**	1.0			
Scaff. Sequenzier.	.09	.44*	.37	-.01	.17	.04	.39*	.44*	.50**	.38*	.46*	1.0		
Laisser-faire	-.53*	-.79**	-.85**	.04	.08	.23	-.86**	-.79**	-.76**	-.69**	-.79**	-.76**	1.0	
Anwend. Lernen	.19	.44	.47*	.21	.02	-.25	.56**	.52**	.32	.71**	.46*	.03	-.45*	1.0

* p<.05, ** p<.01

4.3 'Concurrent validity': Ergebnisse der Analyse von Unterschieden im Unterrichtshandeln der Lehrkräfte in den Fortbildungsgruppen

Zur Einschätzung der 'concurrent validity' der Ratingskalen wurde auf die 'Methode der bekannten Gruppen' ('known groups') mit dem Kriterium 'Zugehörigkeit der Lehrkräfte zu den Fortbildungsgruppen' zurückgegriffen. Gemäß den Ausführungen in 3.1. nahmen die in dieser Studie untersuchten 30 Lehrkräfte im Rahmen des BIQUA-Projekts an verschiedenen Fortbildungen zum naturwissenschaftlichen Sachunterricht teil. Von den 30 untersuchten Lehrkräften gehörten 20 Lehrkräfte so genannten Experimentalgruppen (EGs) an, d.h. diese Lehrkräfte nahmen an Präsenzfortbildungen mit tutorieller Unterstützung teil. Die 20 Lehrkräfte verteilten sich zu je 10 Personen auf zwei Experimentalgruppen. Die Fortbildungen waren an Conceptual Change-Ansätzen und Theorien des Scaffolding orientiert; die Inhalte der Fortbildungen bezogen sich auf fachliches und fachdidaktisches Wissen zu Themen des naturwissenschaftlichen Sachunterrichts (z.B. Conceptual Change-Ansätze zum Lehren und Lernen, typische Schülervorstellungen und Lernschwierigkeiten, Umgang mit Fehlkonzepten) (vgl. Möller et al., 2006). Die übrigen 10 Lehrkräfte gehörten der Kontrollgruppe (KG) an; diese Lehrkräfte erarbeiteten sich die Fortbildungsinhalte im Selbststudium anhand von Handreichungen (vgl. 3.1). Zur Berechnung der Unterschiede wurden die beiden Experimentalgruppen zusammengefasst und der Kontrollgruppe gegenüber gestellt.

Unter den Annahmen, dass Fortbildungen, die gestaltet sind wie oben beschrieben, Einfluss auf das Unterrichtshandeln der Lehrkräfte haben und dass tutoriell unterstützte Fortbildungen einen größeren Einfluss auf die unterrichtlichen Handlungen der Lehrkräfte haben, als Fortbildungen ohne tutorielle Unterstützung, in denen sich die Lehrkräfte die Fortbildungsinhalte weitgehend selbstgesteuert erarbeiten, werden für die Gruppen nach der Intervention (Unterrichtsbeobachtung „Schwimmen und Sinken“) Unterschiede in den Ausprägungen der folgenden zu messenden Konstrukte zugunsten der Experimentalgruppen (EGs) erwartet: „Präkonzepte – Vorwissen aktivieren und aufgreifen“, „Conceptual-Change – kognitive Konflikte erzeugen“, „Ideen gemeinsam diskutieren“ und in allen drei Skalen zum Scaffolding. Zudem wird erwartet, dass die Lehrkräfte der Experimentalgruppe weniger Laissez-faire-Verhalten zeigen, als Lehrkräfte der Kontrollgruppe.

In den Ausprägungen der Skalen zur Transmission von Wissen und der Konstrukte „Motivation des Unterrichts“, „Eigene Ideen zulassen“, „Fehleroffenheit“ und „Anwendungsbezogenes Lernen“ werden keine signifikanten Unterschiede in den Gruppen erwartet, da derartige Inhalte in den tutoriell unterstützten Fortbildungsgruppen nicht intensiver thematisiert wurden als in den Lehrerhandreichungen beschrieben ist, die alle Fortbildungsteilnehmer, auch die der Selbststudiumsgruppe, erhalten haben.

Die Unterschiede wurden mittels U-Test (nach Mann und Whitney) berechnet. Die Ergebnisse der Berechnungen sind in Tabelle 4.3-1 aufgeführt, sie zeigen das folgende Bild: Bei den Skalen „Präkonzept-

te – Vorwissen aktivieren und aufgreifen“, „Conceptual-Change – kognitive Konflikte erzeugen“, „Ideen gemeinsam diskutieren“ und den Scaffolding-Skalen „Perturbieren“ und „Sequenzierungen“ zeigen sich wie erwartet signifikante Unterschiede zwischen den Gruppen; die Lehrkräfte, die an tutoriell unterstützten Fortbildungen teilgenommen haben (EGs, N=19⁴⁴), zeigen im Mittel signifikant häufiger entsprechendes Verhalten im Unterricht als ihre Kollegen der Selbststudiumsgruppe (KG, N=10). Ein signifikanter Unterschied zugunsten der Experimentalgruppen zeigt sich wie erwartet auch bei Unterrichtshandlungen, die dem Konstrukt „Laisser-faire“ zuzuordnen sind.

Tab. 4.3-1 Unterschiede im Unterrichtshandeln der Lehrkräfte in den Fortbildungsgruppen zum MZP 2 (nach der Intervention)

Kriterium	Nach der Intervention (MZP 2) Unterrichtsbeobachtung 'Schwimmen und Sinken'			
	KG (N=10) M (SD)	EGs (N=19) M (SD)	Signifikanz	Effektstärke
Motivation des Unterrichts	M=2.40 (SD=.53)	M=2.79 (SD=.41)	.021*	-.82
Eigene Ideen zulassen	M=2.88 (SD=.61)	M=3.31 (SD=.43)	.050*	-.81
Eigene Formulierungen	M=2.83 (SD=.71)	M=3.43 (SD=.38)	.040*	-1.03
Fehleroffenheit	M=3.88 (SD=.21)	M=3.86 (SD=.42)	.910 ns	.04
Transmission – Erklärungen	M=1.31 (SD=.42)	M=1.21 (SD=.37)	.330 ns	.26
Transmission – Denkwege ablehnen	M=1.39 (SD=.35)	M=1.26 (SD=.46)	.094 ns	.31
Präkonzepte – Vorwissen aktivieren und aufgreifen	M=2.21 (SD=.90)	M=2.92 (SD=.46)	.012*	-1.00
Conceptual Change – kognitive Konflikte erzeugen	M=2.51 (SD=.68)	M=2.96 (SD=.45)	.040*	-.80
Ideen gemeinsam diskutieren	M=2.13 (SD=.95)	M=2.67 (SD=.60)	.040*	-.72
Scaffolding – Hervorhebungen	M=2.20 (SD=.78)	M=2.52 (SD=.70)	.228 ns	-.44
Scaffolding – Perturbieren	M=2.19 (SD=.76)	M=2.77 (SD=.53)	.009**	-.86
Scaffolding – Sequenzierungen	M=2.65 (SD=.58)	M=3.19 (SD=.57)	.024*	-.87
Laisser-faire	M=2.09 (SD=.75)	M=1.30 (SD=.41)	.004**	1.19
Anwendungsbezogenes Lernen	M=1.51 (SD=.36)	M=1.76 (SD=.33)	.056 ns	-.71

Antwortformat von 1 bis 4 ((1) trifft überhaupt nicht zu, (2) trifft eher nicht zu, (3) trifft eher zu, (4) trifft voll und ganz zu). Grau markiert sind Skalen, zu denen Unterschiede in den Fortbildungsgruppen erwartet werden. * p<.05, ** p<.01

⁴⁴Da es zum Zeitpunkt der Unterrichtsbeobachtung „Schwimmen und Sinken“ einer Lehrperson der Experimentalgruppen nicht möglich war, Unterricht zum genannten Thema zu zeigen, basieren die Berechnungen für die Experimentalgruppen auf einem N=19.

Bei der Skala „Scaffolding – Hervorhebungen“ zeigt sich entgegen der Erwartung kein signifikanter Mittelwertunterschied. Dennoch weisen die Lehrkräfte der Experimentalgruppe einen etwas höheren Mittelwert auf als ihre Kollegen der Kontrollgruppe. Der mittlere Effekt weist auf eine tendenzielle Überlegenheit der Experimentalgruppen hin.

Mittlere Abweichungen der Experimentalgruppe von der Kontrollgruppe zeigen sich zudem in den Unterrichtshandlungen „Motivation des Unterrichts“, „Eigene Ideen zulassen“ und „eigene Formulierungen“. In allen Bereichen sind die Mittelwerte der Experimentalgruppe höher als die der Kontrollgruppe. Die Unterschiede werden entgegen den Erwartungen signifikant.

Bei Verhaltensweisen, die dem Bereich Transmission von Wissen und die der Skala „Fehleroffenheit“ zuzuordnen sind, erreichen beide Gruppen gemäß den Erwartungen ähnliche Ausprägungen. Beim Konstrukt „Fehleroffenheit“ liegen in beiden Gruppen Deckeneffekte und bei den Transmissions-Skalen Bodeneffekte vor. Diese haben sich bereits in den Mittelwerten über die gesamte Untersuchungsstichprobe gezeigt (vgl. 4.2). Hinsichtlich der deskriptiven Werte beider Untersuchungsgruppen fällt weiterhin auf, dass die Streuung der Werte in der Kontrollgruppe größer ist als bei den zusammengefassten Experimentalgruppen. Die Standardabweichungen sind bei den Lehrkräften der Selbststudiumsgruppe fast durchweg höher als bei den tutoriell unterstützten Kollegen. Dies deutet auf eine größere Heterogenität der Kontrollgruppe bezüglich der jeweiligen Merkmale.

Insgesamt zeigen die Ergebnisse, dass die angenommenen Veränderungsintensitäten für die Skalen „Präkonzepte – Vorwissen aktivieren und aufgreifen“, „Conceptual Change – kognitive Konflikte erzeugen“, „Ideen gemeinsam diskutieren“, „Laisser-faire“, „Anwendungsbezogenes Lernen“, „Fehleroffenheit“, für beide Transmissions-Skalen und für die zwei Scaffolding-Skalen „Hervorhebungen“ und „Sequenzierungen“ bestätigt werden konnten.

Auch für das Verhalten der Lehrkräfte im Unterricht zum Thema „Wasserkreislauf“, der vor der Intervention aufgezeichnet wurde (vgl. 3.1), wurden Unterschiede für die Experimental- und Kontrollgruppen berechnet (EGs: N=20; KG: N=10). Zu diesem Messzeitpunkt sollten im Unterrichtshandeln der Lehrkräfte keine bedeutsamen Unterschiede vorliegen. Gemäß den Ergebnissen unterscheiden sich die Gruppen in keinem Merkmal signifikant voneinander, auch auf dem Niveau von $\alpha=.10$ führen alle Tests nicht zur Ablehnung der Nullhypothese (vgl. Tabelle 4.3-2). Anzumerken sind jedoch mittlere Abweichungen der Fortbildungsgruppen voneinander bei den Konstrukten „Scaffolding – Perturbieren“ und „Laisser-faire“. Bei der genannten Scaffolding-Skala liegt ein leicht höherer Mittelwert für die Experimentalgruppen vor, bei der Skala „Laisser-faire“ liegt der Mittelwert der Kontrollgruppe über dem Mittelwert der Experimentalgruppe. Die Lehrkräfte der der tutoriell unterstützten Fortbildungsgruppen scheinen den Lehrkräften der Kontrollgruppe in diesen Verhaltensweisen bereits vor der Intervention leicht überlegen zu sein, wenn auch diese Unterschiede statistisch nicht bedeutsam sind.

Tab. 4.3-2 Unterschiede im Unterrichtshandeln der Lehrkräfte in den Fortbildungsgruppen zum MZP 1 (vor der Intervention)

Kriterium	Vor der Intervention (MZP 1) Unterrichtsbeobachtung 'Wasserkreislauf'			
	KG (N=10) M (SD)	EGs (N=20) M (SD)	Signifikanz (p)	Effektstärke
Motivation des Unterrichts	M=2.24 (SD=.40)	M=2.21 (SD=.30)	.895 ns	.01
Eigene Ideen zulassen	M=2.36 (SD=.89)	M=2.46 (SD=.36)	.930 ns	-.17
Eigene Formulierungen	M=2.29 (SD=.75)	M=2.38 (SD=.60)	.612 ns	-.14
Fehleroffenheit	M=3.51 (SD=.55)	M=3.62 (SD=.39)	.928 ns	-.26
Transmission – Erklärungen	M=1.80 (SD=.89)	M=2.04 (SD=.49)	.152 ns	-.37
Transmission – Denkwege ablehnen	M=1.89 (SD=.85)	M=1.95 (SD=.45)	.509 ns	-.09
Präkonzepte – Vorwissen aktivieren und aufgreifen	M=1.55 (SD=.65)	M=1.72 (SD=.51)	.388 ns	-.29
Conceptual Change – kognitive Konflikte erzeugen	M=2.09 (SD=.78)	M=2.18 (SD=.44)	.791 ns	-.16
Ideen gemeinsam diskutieren	M=1.65 (SD=.68)	M=1.81 (SD=.34)	.134 ns	-.36
Scaffolding – Hervorhebungen	M=2.03 (SD=.69)	M=2.16 (SD=.58)	.552 ns	-.22
Scaffolding – Perturbieren	M=1.70 (SD=.65)	M=1.94 (SD=.38)	.122 ns	-.49
Scaffolding – Sequenzierungen	M=2.18 (SD=.89)	M=2.04 (SD=.69)	.659 ns	.17
Laisser-faire	M=2.25 (SD=.86)	M=1.90 (SD=.55)	.280 ns	.51
Anwendungsbezogenes Lernen	M=2.38 (SD=.62)	M=2.62 (SD=.63)	.226 ns	-.38

Antwortformat von 1 bis 4 ((1) trifft überhaupt nicht zu, (2) trifft eher nicht zu, (3) trifft eher zu, (4) trifft voll und ganz zu).

* $p < .05$, ** $p < .01$

4.4 Konstruktvalidität: Ergebnisse der Expertenvalidierung

Im Rahmen der Validierung stand die folgende Frage an die Experten im Zentrum des Interesses: Repräsentieren die Items die Konstrukteinhalte wie sie auf Basis der Theorie beschrieben sind? Die Experten wurden um eine Zuordnung der Items zu den Konstrukten gebeten. Aufgenommen wurden nur die reliablen und trennscharfen Items. Eine Ausnahme stellen die Items ei_f4, il5 und sc_p4 dar, die in den Tabellen 4.4-1 und 4.4-2 in grau dargestellt sind. Diese Items werden in dieser Untersuchung zwar aufgrund mangelnder Reliabilität aus den Skalen ausgeschlossen (vgl. 4.2); eine Berücksichtigung der Items bei der Expertenvalidierung ist aber dennoch von Interesse, um zu sehen, wie eine Zuordnung durch die Experten ausfällt. Möglicherweise erreichen diese Items in einer anderen Stichprobe reliable Werte und können dann bei einer korrekten Zuordnung durch die Experten in die Skalen aufgenommen werden.

Die Tabelle 4.4-1 zeigt die Zuordnung der Items zu den Konstrukten für jede Person, die an der Expertenvalidierung teilgenommen hat. Mit ✓ wird eine Übereinstimmung der Zuordnung der Experten mit der theoretisch vorgenommenen Zuordnung der Items gekennzeichnet. In Fällen, bei denen ein Experte das Item zwar in erster Priorität einem anderen als dem theoretisch zugeordneten Konstrukt zuweist, letzteres aber auch für passend hält, ist dies mit (✓) gekennzeichnet. Die abweichende Zuordnung ist dann jeweils angegeben. Dies gilt auch für den Fall, dass das theoretisch zugeordnete Konstrukt nicht einmal mit niederer Priorität zugewiesen wurde.

Tab. 4.4-1 Expertenvalidierung (von Person 1 bis Person 5) durch Zuordnung der Items zu den theoretischen Konstrukten

Motivation des Unterrichts	Person 1	Person 2	Person 3	Person 4	Person 5
Die Lehrperson gestaltet den Unterricht spannend. (<i>in Anlehnung an Clausen, Reusser & Klieme, 2003</i>) (mt_1)	✓	✓	✓	✓	✓
Die Lehrperson kann die Schüler auch mal richtig begeistern. (<i>in Anlehnung an Clausen, Reusser & Klieme, 2003</i>) (mt3)	✓	✓	✓	✓	✓
Die Lehrperson macht die Schüler neugierig. (mt6)	✓	✓	✓	✓	✓
Die Lehrperson ruft bei den Schülern Staunen hervor. (mt7)	✓	✓	✓	✓	✓
Die Lehrperson kann das Interesse der Kinder am Lerngegenstand aufrecht erhalten. (mt9)	✓	✓	✓	✓	✓

Eigene Ideen zulassen	Person 1	Person 2	Person 3	Person 4	Person 5
Die Lehrperson ermuntert die Kinder, weiter nachzudenken/Ideen zu äußern. (ei_i1)	Perturbieren	✓	✓	✓	✓
Die Lehrperson räumt den Kindern die Möglichkeit ein, selbst Deutungen/Erklärungen für ein Phänomen zu finden. (ei_i2)	✓	Eigene Formulierungen	✓	✓	✓
Die Lehrperson greift Schülerbeiträge im weiteren Unterrichtsverlauf auf. (ei_i4)	Präkonzepte	Ideen diskutieren	✓	✓	✓
Die Lehrperson fragt – ohne zu bewerten – nach, wenn ein Schüler eine unvollständige oder unklare Idee formuliert. (<i>übernommen von Clausen, Reusser & Klieme, 2003</i>) (ei_i6)	✓ auch Eigene Formulierungen, Präkonzepte, Perturbieren	✓	✓	Perturbieren	✓
Die Lehrperson verfolgt Ideen der Kinder weiter. (ei_i3)	✓ auch Bezug zu Präkonzepten, Conceptual Change	✓	✓	✓	✓
Eigene Formulierungen	Person 1	Person 2	Person 3	Person 4	Person 5
Die Lehrperson legt Wert darauf, dass die Kinder Sachverhalte in ihren eigenen Worten erklären. (ei_e2)	✓	✓	✓	✓	✓
Die Lehrperson greift die Formulierungen der Kinder auf und arbeitet damit weiter. (ei_e3)	Präkonzepte, auch Eigene Ideen zulassen	Präkonzepte	Präkonzepte	✓	Präkonzepte
Die Lehrperson sorgt nicht dafür, dass 'Fachtermini' mit eigenen Worten erklärt werden. (ei_e5) (umgepoltes Item)	✓ auch Denkwege ablehnen	Denkwege ablehnen	Denkwege ablehnen	Denkwege ablehnen	✓ auch Denkwege ablehnen
Fehleroffenheit	Person 1	Person 2	Person 3	Person 4	Person 5
Die Lehrperson ist geduldig, wenn ein Schüler im Unterricht ein Fehlkonzent nennt/hat. (<i>in Anlehnung an Clausen, Reusser & Kieme, 2003</i>) (ei_f1)	✓	(✓) Präkonzepte	✓	✓	✓
Bei dieser Lehrperson ist das Nennen von falschen Vermutungen/Erklärungen nichts Schlimmes. (<i>in Anlehnung an Clausen, Reusser & Klieme, 2003</i>) (ei_f2)	✓	✓	✓	✓	✓
Die Lehrperson korrigiert nicht sofort jedes Fehlkonzent. (<i>in Anlehnung an Clausen, Reusser & Klieme, 2003</i>) (ei_f8)	✓ auch Conceptual Change	✓ auch Präkonzepte	✓	✓	✓
Die Lehrperson betont, dass auch falsche Ideen wichtig sind, um zum Ziel zu kommen. (ei_f4).	✓	✓	✓	✓	✓

Transmission von Wissen – Erklärungen	Person 1	Person 2	Person 3	Person 4	Person 5
Die Lehrperson beantwortet aufkommende Fragen vorschuell selbst. (il1)	Denkwege ablehnen	Denkwege ablehnen	Denkwege ablehnen	✓	Denkwege ablehnen
Die Lehrperson gibt die Erklärungen für die zuvor durchgeführten Experimente vor. (il2)	Denkwege ablehnen	✓	Denkwege ablehnen	✓	✓
Die Lehrperson lässt die Kinder zuvor aufgeworfene Problemen nicht selbst lösen. (il3)	Denkwege ablehnen	✓	Denkwege ablehnen	Denkwege ablehnen	Denkwege ablehnen
Die Lehrperson gibt den Schülern Merksätze vor, die nicht mit den Kindern entwickelt worden sind. (il6)	Denkwege ablehnen	✓	Denkwege ablehnen	✓	Denkwege ablehnen
Den Kindern werden 'fertige Erklärungen' durch Abbildungen vermittelt. (il5)	✓	✓	✓	✓	✓
Transmission von Wissen – eigene Denk- und Lernwege ablehnen	Person 1	Person 2	Person 3	Person 4	Person 5
Die Lehrperson unterbindet eigene Denkwege der Kinder. (il9)	✓ auch negativ Eigene Ideen zulassen	✓ auch Erklärungen denkbar	✓	✓	✓
Die Lehrperson berichtigt Fehlkonzepte ohne Begründung. (in <i>Anlehnung an Heinze, 2004</i>) (il13)	✓	✓	✓	✓	✓
Die Lehrperson ignoriert Fehlkonzepte. (il_12)	✓ auch negativ Präkonzepte	(✓) Erklärungen	✓	✓	✓
Die Lehrperson korrigiert falsche Vermutungen qua Autorität. (il4)	✓	✓	✓	✓	✓
Präkonzepte – Vorwissen aktivieren und aufgreifen	Person 1	Person 2	Person 3	Person 4	Person 5
Die Lehrperson kommt im Verlauf der Stunde auf Präkonzepte der Kinder zurück. (pk1)	✓	✓	✓	✓	✓
Die Lehrperson knüpft im Unterricht an Präkonzepte der Schüler an. (in <i>Anlehnung an Kobarg & Seidel, 2003</i>) (pk3)	✓	✓	✓	✓	✓
Die Lehrperson regt die Schüler an, über ihre Präkonzepte nachzudenken. (pk4)	✓	✓	✓	✓	✓
Die Lehrperson nimmt sich viel Zeit, Präkonzepte begründen zu lassen. (pk5)	✓	✓	✓	✓	✓

Conceptual Change – kognitive Konflikte erzeugen	Person 1	Person 2	Person 3	Person 4	Person 5
Die Lehrperson hat Versuche ausgesucht, die für die Kinder überzeugend sind. (cc1)	Hervorhebungen auch Motivation	(✓) Präkonzepte	Hervorhebungen	(✓) auch Motivation	(✓) Hervorhebungen
Die Lehrperson konfrontiert die Kinder mit Beobachtungen oder Phänomenen, die den Erwartungen der Kinder widersprechen. (cc4)	✓ auch Bezug zu Scaffolding	✓	✓ auch Perturbieren	✓ auch Perturbieren, aber schwächer	✓
Die Lehrperson widerlegt Fehlkonzepte der Kinder mit Hilfe ausgewählter Materialien. (cc6)	✓	✓	✓	✓	✓
Der Lehrperson ist es besonders wichtig, dass objektive Evidenzen intensiv mit Präkonzepten in Beziehung gesetzt werden. (cc8)	Präkonzepte, auch Scaffolds	✓	✓	✓	✓
Ideen gemeinsam diskutieren	Person 1	Person 2	Person 3	Person 4	Person 5
Missverständliche, unvollständige oder unklare Äußerungen werden nicht ignoriert oder bloß kommentiert, sondern es wird nachgefragt. (übernommen von Clausen, Reusser & Klieme, 2003) (dk_4)	Hervorhebungen, auch Fehleroffenheit, Präkonzepte	Fehleroffenheit, auch Präkonzepte	Eigene Formulierungen	Perturbieren	Fehleroffenheit
Die Lehrperson fordert die Kinder auf, ihre eigenen Vermutungen/Erklärungen untereinander zu diskutieren. (dk5)	✓	Eigene Ideen zulassen	✓	✓	✓
Die Lehrperson fordert die Kinder im Unterricht auf, ihre Deutungen zu einem Phänomen gegenüber Mitschülern zu begründen. (dk6)	✓ auch Präkonzepte	eigene Formulierungen	✓ auch eigene Ideen zulassen, eigene Formulierungen	✓	✓ auch Präkonzepte
Die Lehrperson regt die Kinder auch dann an, ihre Vorstellungen untereinander zu diskutieren, wenn diese falsch sind. (dk7)	(✓) Präkonzepte, auch Conceptual Change, Laisser-faire	Fehleroffenheit, auch eigene Ideen zulassen, eigene Formulierungen, Präkonzepte, Conceptual Change	✓	(✓) Eigene Formulierungen,	✓
Die Lehrperson diskutiert gemeinsam mit den Kindern Probleme, Idee, Lösungswege, Fehlkonzepte. (dk_8)	✓ auch Präkonzepte	✓ auch Präkonzepte, müsste eingeschränkt werden, über was diskutiert wird	✓	✓	✓

Scaffolding – Hervorhebungen	Person 1	Person 2	Person 3	Person 4	Person 5
Die Lehrperson weist darauf hin, was sich die Schüler merken sollen. (<i>übernommen von Clausen, Reusser & Klieme, 2003</i>) (sc_h1)	✓	✓	✓	Erklärungen	✓
Die Lehrperson hebt wichtige Fakten hervor. (<i>übernommen von Clausen, Reusser & Klieme, 2003</i>) (sc_h2)	✓	✓	Erklärungen	✓	✓
Die Lehrperson fasst Ergebnisse oder bereits Gesagtes häufig wiederholend zusammen. (sc_h4)	✓ auch Sequenzierungen	✓	✓	✓	✓
Die Lehrperson fasst das Wichtigste nochmals zusammen. (<i>übernommen von Clausen, Reusser & Klieme, 2003</i>) (sc_h6)	✓	✓	✓	✓	✓
Die Lehrperson fokussiert die Aufmerksamkeit der Kinder auf wichtige inhaltliche Aspekte. (sc_h7)	✓	✓	✓	✓	✓
Scaffolding – Perturbieren	Person 1	Person 2	Person 3	Person 4	Person 5
Die Lehrperson fordert Begründungen ein. (sc_p1)	Conceptual Change, auch Bezug zu Präkonzepten, Hervorhebungen	Hervorhebungen, auch Conceptual Change, Ideen diskutieren	Ideen diskutieren	✓	✓
Die Lehrperson fordert Beweise ein. (sc_p2)	Conceptual Change auch Präkonzepte	✓	Ideen diskutieren	✓	✓
Die Lehrperson macht auf Widersprüche in Aussagen/Vermutungen aufmerksam. (sc_p3)	Conceptual Change	✓	Conceptual Change	✓	Conceptual Change
Die Lehrperson hinterfragt Aussagen. (sc_p5)	Conceptual Change	✓ auch eigene Formulierungen, Conceptual Change	✓ auch Conceptual Change	✓	✓
Die Lehrperson macht auf ungelöste Probleme aufmerksam. (sc_p4)	Conceptual Change	✓	Conceptual Change	✓	Conceptual Change
Scaffolding – Sequenzierungen	Person 1	Person 2	Person 3	Person 4	Person 5
Der Unterricht ist so sequenziert, dass die Schüler die angestrebten Konzepte/Ziele verstehen können. (sc_s1)	✓ auch Präkonzepte	✓	✓	✓	✓
Die Lehrperson achtet darauf, dass das Gespräch nicht zwischen unterschiedlichen Aspekten hin und her springt. (sc_s2)	✓ auch Präkonzepte	✓	✓	✓	✓

Laisser-faire	Person 1	Person 2	Person 3	Person 4	Person 5
Die Lehrperson gibt den Kindern kaum Hilfe beim Entwickeln und Auswerten von Experimenten. (ol3)	✓	✓	✓	✓	✓
Die Lehrperson begnügt sich vorschnell damit, wenn ein Kind einen erwünschten Begriff/Zusammenhang genannt hat. (ol4)	Denkwege ablehnen	Denkwege ablehnen	Erklärungen	✓ auch Tansmission	Denkwege ablehnen
Die Lehrperson beschränkt den Umgang mit Präkonzepten auf ein oberflächliches Abfragen. (ol5)	Denkwege ablehnen	Denkwege ablehnen	<i>Item nicht zugeordnet</i>	Erklärungen	Denkwege ablehnen
In Gesprächen über die Deutung von Naturphänomenen oder der Suche nach geeigneten Lösungswegen hält sich die Lehrperson stark zurück. (ol 1)	(✓) Eigene Ideen zulassen	Eigene Ideen zulassen	✓	✓	✓
Die Lehrperson lässt die Kinder einfach nur 'laufen'. (ol2)	✓	✓	✓	✓	✓
Anwendungsbezogenes Lernen	Person 1	Person 2	Person 3	Person 4	Person 5
Die Lehrperson thematisiert die Anwendung neuer Konzepte. (cc3)	(✓) Perturbieren	✓	✓	✓	✓
Fragestellungen (Bezüge, Beobachtungen) aus der außerschulischen Welt werden angesprochen. (<i>übernommen von Clausen, Reusser & Klieme, 2003</i>) (aw2)	✓	(✓) Präkonzepte	✓	✓	✓
Die Lehrperson stellt Bezüge zum außerschulischen Alltag der Kinder her. (aw5)	✓	✓	✓	✓	✓
Die Lehrperson stellt im Unterricht Bezüge zu Naturphänomenen her. (aw6)	✓	✓	✓	✓	✓
Die Lehrperson fragt die Kinder nach Alltagsbeispielen/Anwendungsbezügen. (aw7)	✓	✓	✓	✓	✓

Die Tabelle 4.4-2 beinhaltet:

- Die Anzahl richtiger Zuordnungen pro Item durch die Beurteiler in erster oder zweiter Priorität.
- Die Bewertung der Zuordnung der Items; ein Item gilt als korrekt zugeordnet, wenn von mindestens drei Beurteilern (=Mehrheit) Zuordnungen in erster oder zweiter Priorität vorliegen; eine korrekte Zuordnung wird mit 'ja' gekennzeichnet, eine unzureichende Zuordnung wird mit 'nein' gekennzeichnet.
- Die Trennschärfen der Items sowohl für das Thema „Wasserkreislauf“ (pre) als auch für das Thema „Schwimmen und Sinken“ (post).
- Die Nennung einer Alternativskala bei mindestens zwei Nennungen eines Items.

Tab. 4.4-2 Bewertung der Zuordnung der Items zu den theoretischen Konstrukten

Motivation des Unterrichts	Anzahl korrekter Zuordnungen	korrekt zugeordnet: ja / nein	r_{it} pre / post	Alternativskala (mind. 2 Nennungen)
Die Lehrperson gestaltet den Unterricht spannend. (in Anlehnung an Clausen, Reusser & Klieme, 2003) (mt_1)	5	ja	.61 / .79	
Die Lehrperson kann die Schüler auch mal richtig begeistern. (in Anlehnung an Clausen, Reusser & Klieme, 2003) (mt3)	5	ja	.65 / .80	
Die Lehrperson macht die Schüler neugierig. (mt6)	5	ja	.26 / .58	
Die Lehrperson ruft bei den Schülern Staunen hervor. (mt7)	5	ja	.24 / .77	
Die Lehrperson kann das Interesse der Kinder am Lerngegenstand aufrecht erhalten. (mt9)	5	ja	.53 / .52	
Eigene Ideen zulassen	Anzahl korrekter Zuordnungen	korrekt zugeordnet: ja / nein	r_{it} pre / post	Alternativskala (mind. 2 Nennungen)
Die Lehrperson ermuntert die Kinder, weiter nachzudenken/Ideen zu äußern. (ei_i1)	4	ja	.56 / .87	
Die Lehrperson räumt den Kindern die Möglichkeit ein, selbst Deutungen/Erklärungen für ein Phänomen zu finden. (ei_i2)	4	ja	.77 / .63	
Die Lehrperson greift Schülerbeiträge im weiteren Unterrichtsverlauf auf. (ei_i4)	3	ja	.82 / .82	
Die Lehrperson fragt – ohne zu bewerten – nach, wenn ein Schüler eine unvollständige oder unklare Idee formuliert. (übernommen von Clausen, Reusser & Klieme, 2003) (ei_i6)	4	ja	.84 / .80	Perturbieren (2 Nennungen)
Die Lehrperson verfolgt Ideen der Kinder weiter. (ei_i3)	4	ja	.87 / .91	
Eigene Formulierungen	Anzahl korrekter Zuordnungen	korrekt zugeordnet: ja / nein	r_{it} pre / post	Alternativskala (mind. 2 Nennungen)
Die Lehrperson legt Wert darauf, dass die Kinder Sachverhalte in ihren eigenen Worten erklären. (ei_e2)	5	ja	.72 / .82	
Die Lehrperson greift die Formulierungen der Kinder auf und arbeitet damit weiter. (ei_e3)	1	nein	.61 / .69	Präkonzepte (4 Nennungen)
Die Lehrperson sorgt nicht dafür, dass 'Fachtermini' mit eigenen Worten erklärt werden. (ei_e5) (umgepoltes Item)	2	nein	.42 / .62	Denk- und Lernwege ablehnen (5 Nennungen)

Fehleroffenheit	Anzahl korrekter Zuordnungen	korrekt zugeordnet: ja / nein	r_{it} pre / post	Alternativskala (mind. 2 Nennungen)
Die Lehrperson ist geduldig, wenn ein Schüler im Unterricht ein Fehlkonzent nennt/hat. (in <i>Anlehnung an Clausen, Reusser & Kieme, 2003</i>) (ei_f1)	5	ja	.49 / .83	
Bei dieser Lehrperson ist das Nennen von falschen Vermutungen/Erklärungen nichts Schlimmes. (in <i>Anlehnung an Clausen, Reusser & Kieme, 2003</i>) (ei_f2)	5	ja	.71 / .92	
Die Lehrperson korrigiert nicht sofort jedes Fehlkonzent. (in <i>Anlehnung an Clausen, Reusser & Kieme, 2003</i>) (ei_f8)	5	ja	.57 / .80	
Die Lehrperson betont, dass auch falsche Ideen wichtig sind, um zum Ziel zu kommen. (ei_f4).	5	ja		
Transmission von Wissen – Erklärungen	Anzahl korrekter Zuordnungen	korrekt zugeordnet: ja / nein	r_{it} pre / post	Alternativskala (mind. 2 Nennungen)
Die Lehrperson beantwortet aufkommende Fragen vorschnell selbst. (il1)	1	nein	.84 / .87	Denk- und Lernwege ablehnen (4 Nennungen)
Die Lehrperson gibt die Erklärungen für die zuvor durchgeführten Experimente vor. (il2)	3	ja	.82 / .88	Denk- und Lernwege ablehnen (2 Nennungen)
Die Lehrperson lässt die Kinder zuvor aufgeworfene Problemen nicht selbst lösen. (il3)	1	nein	.86 / .87	Denk- und Lernwege ablehnen (4 Nennungen)
Die Lehrperson gibt den Schülern Merksätze vor, die nicht mit den Kindern entwickelt worden sind. (il6)	2	nein	.49 / .23	Denk- und Lernwege ablehnen (3 Nennungen)
Den Kindern werden 'fertige Erklärungen' durch Abbildungen vermittelt. (il5)	5	ja		
Transmission von Wissen – eigene Denk- und Lernwege ablehnen	Anzahl korrekter Zuordnungen	korrekt zugeordnet: ja / nein	r_{it} pre / post	Alternativskala (mind. 2 Nennungen)
Die Lehrperson unterbindet eigene Denkwege der Kinder. (il9)	5	ja	.67 / .83	
Die Lehrperson berichtigt Fehlkonzente ohne Begründung. (in <i>Anlehnung an Heinze, 2004</i>) (il13)	5	ja	.80 / .82	
Die Lehrperson ignoriert Fehlkonzente. (il_12)	5	ja	.54 / .45	
Die Lehrperson korrigiert falsche Vermutungen qua Autorität. (il4)	5	ja	.64 / .85	

Präkonzepte – Vorwissen aktivieren und aufgreifen	Anzahl korrekter Zuordnungen	korrekt zugeordnet: ja / nein	r_{it} pre / post	Alternativskala (mind. 2 Nennungen)
Die Lehrperson kommt im Verlauf der Stunde auf Präkonzepte der Kinder zurück. (pk1)	5	ja	.86 / .89	
Die Lehrperson knüpft im Unterricht an Präkonzepte der Schüler an. (<i>in Anlehnung an Kobarg & Seidel, 2003</i>) (pk3)	5	ja	.93 / .92	
Die Lehrperson regt die Schüler an, über ihre Präkonzepte nachzudenken. (pk4)	5	ja	.85 / .90	
Die Lehrperson nimmt sich viel Zeit, Präkonzepte begründen zu lassen. (pk5)	5	ja	.78 / .85	
Conceptual Change – kognitive Konflikte erzeugen	Anzahl korrekter Zuordnungen	korrekt zugeordnet: ja / nein	r_{it} pre / post	Alternativskala (mind. 2 Nennungen)
Die Lehrperson hat Versuche ausgesucht, die für die Kinder überzeugend sind. (cc1)	3	ja	.56 / .85	Hervorhebungen (3 Nennungen)
Die Lehrperson konfrontiert die Kinder mit Beobachtungen oder Phänomenen, die den Erwartungen der Kinder widersprechen. (cc4)	5	ja	.63 / .86	Perturbieren (2 Nennungen)
Die Lehrperson widerlegt Fehlkonzepte der Kinder mit Hilfe ausgewählter Materialien. (cc6)	5	ja	.78 / .90	
Der Lehrperson ist es besonders wichtig, dass objektive Evidenzen intensiv mit Präkonzepten in Beziehung gesetzt werden. (cc8)	4	ja	.77 / .82	
Ideen gemeinsam diskutieren	Anzahl korrekter Zuordnungen	korrekt zugeordnet: ja / nein	r_{it} pre / post	Alternativskala (mind. 2 Nennungen)
Missverständliche, unvollständige oder unklare Äußerungen werden nicht ignoriert oder bloß kommentiert, sondern es wird nachgefragt. (<i>übernommen von Clausen, Reusser & Klieme, 2003</i>) (dk_4)	0	nein	.59 / .69	Fehleroffenheit, Präkonzepte (jeweils 2 Nennungen)
Die Lehrperson fordert die Kinder auf, ihre eigenen Vermutungen/Erklärungen untereinander zu diskutieren. (dk5)	4	ja	.70 / .84	
Die Lehrperson fordert die Kinder im Unterricht auf, ihre Deutungen zu einem Phänomen gegenüber Mitschülern zu begründen. (dk6)	4	ja	.60 / .79	Präkonzepte, eigene Formulierungen (jeweils 2 Nennungen)
Die Lehrperson regt die Kinder auch dann an, ihre Vorstellungen untereinander zu diskutieren, wenn diese falsch sind. (dk7)	4	ja	.66 / .84	Präkonzepte, Concept. Ch., eig. Formulierung. (jeweils 2 Nennungen)

Die Lehrperson diskutiert gemeinsam mit den Kindern Probleme, Idee, Lösungswege, Fehlkonzepte. (dk_8)	5	ja	.62 / .89	Präkonzepte (2 Nennungen)
Scaffolding – Hervorhebungen	Anzahl korrekter Zuordnungen	korrekt zugeordnet: ja / nein	r_{it} pre / post	Alternativskala (mind. 2 Nennungen)
Die Lehrperson weist darauf hin, was sich die Schüler merken sollen. (<i>übernommen von Clausen, Reusser & Klieme, 2003</i>) (sc_h1)	4	ja	.91 / .88	
Die Lehrperson hebt wichtige Fakten hervor. (<i>übernommen von Clausen, Reusser & Klieme, 2003</i>) (sc_h2)	4	ja	.92 / .95	
Die Lehrperson fasst Ergebnisse oder bereits Gesagtes häufig wiederholend zusammen. (sc_h4)	5	ja	.87 / .91	
Die Lehrperson fasst das Wichtigste nochmals zusammen. (<i>übernommen von Clausen, Reusser & Klieme, 2003</i>) (sc_h6)	4	ja	.93 / .90	
Die Lehrperson fokussiert die Aufmerksamkeit der Kinder auf wichtige inhaltliche Aspekte. (sc_h7)	4	ja	.89 / .93	
Scaffolding – Perturbieren	Anzahl korrekter Zuordnungen	korrekt zugeordnet: ja / nein	r_{it} pre / post	Alternativskala (mind. 2 Nennungen)
Die Lehrperson fordert Begründungen ein. (sc_p1)	2	nein	.35 / .81	Conceptual Ch. Hervorhebungen (jeweils 2 Nennungen)
Die Lehrperson fordert Beweise ein. (sc_p2)	3	ja	.35 / .64	
Die Lehrperson macht auf Widersprüche in Aussagen/Vermutungen aufmerksam. (sc_p3)	2	nein	.59 / .86	Conceptual Ch. (3 Nennungen)
Die Lehrperson hinterfragt Aussagen. (sc_p5)	4	ja	.75 / .90	Conceptual Ch. (3 Nennungen)
Die Lehrperson macht auf ungelöste Probleme aufmerksam. (sc_p4)	2	nein		Conceptual Ch. (3 Nennungen)
Scaffolding – Sequenzierungen	Anzahl korrekter Zuordnungen	korrekt zugeordnet: ja / nein	r_{it} pre / post	Alternativskala (mind. 2 Nennungen)
Der Unterricht ist so sequenziert, dass die Schüler die angestrebten Konzepte/Ziele verstehen können. (sc_s1)	5	ja		
Die Lehrperson achtet darauf, dass das Gespräch nicht zwischen unterschiedlichen Aspekten hin und her springt. (sc_s2)	5	ja		

Laisser-faire	Anzahl korrekter Zuordnungen	korrekt zugeordnet: ja / nein	r_{it} pre / post	Alternativskala (mind. 2 Nennungen)
Die Lehrperson gibt den Kindern kaum Hilfe beim Entwickeln und Auswerten von Experimenten. (ol3)	5	ja	.79 / .88	
Die Lehrperson begnügt sich vorschnell damit, wenn ein Kind einen erwünschten Begriff/Zusammenhang genannt hat. (ol4)	1	nein	.67 / .74	Denkwege ablehnen (4 Nennungen)
Die Lehrperson beschränkt den Umgang mit Präkonzepten auf ein oberflächliches Abfragen. (ol5)	0	nein	.59 / .89	Denkwege ablehnen (3 Nennungen)
In Gesprächen über die Deutung von Naturphänomenen oder der Suche nach geeigneten Lösungswegen hält sich die Lehrperson stark zurück. (ol 1)	4	ja	.81 / .85	Eigene Ideen zulassen (3 Nennungen)
Die Lehrperson lässt die Kinder einfach nur 'laufen'. (ol2)	5	ja	.84 / .85	
Anwendungsbezogenes Lernen	Anzahl korrekter Zuordnungen	korrekt zugeordnet: ja / nein	r_{it} pre / post	Alternativskala (mind. 2 Nennungen)
Die Lehrperson thematisiert die Anwendung neuer Konzepte. (cc3)	5	ja	.77 / .23	
Fragestellungen (Bezüge, Beobachtungen) aus der außerschulischen Welt werden angesprochen. (übernommen von Clausen, Reusser & Klieme, 2003) (aw2)	5	ja	.72 / .29	
Die Lehrperson stellt Bezüge zum außerschulischen Alltag der Kinder her. (aw5)	5	ja	.79 / .66	
Die Lehrperson stellt im Unterricht Bezüge zu Naturphänomenen her. (aw6)	5	ja	.55 / .70	
Die Lehrperson fragt die Kinder nach Alltagsbeispielen/Anwendungsbezügen. (aw7)	5	ja	.50 / .46	

Die Ergebnisse der Zuordnung der Items zu den theoretischen Konstrukten durch die Experten zeigen gemäß den Tabellen 4.4-1 und 4.4-2 das im Folgenden beschriebene Bild.

Eine durchweg eindeutige Zuordnung der Items durch alle Experten ergibt sich für die Konstrukte „Motivation des Unterrichts“ und „Präkonzepte – Vorwissen aktivieren und aufgreifen“. Korrekte Zuordnungen liegen ebenfalls für die Items der Konstrukte „Scaffolding – Sequenzierungen“, „Anwendungsbezogenes Lernen“ und „Fehleroffenheit“ vor. Beim Konstrukt „Scaffolding – Hervorhebungen“ liegen für alle Items von mindestens vier der fünf Personen eindeutige Zuordnungen vor. Gleiches gilt für die Items des Konstrukts „Eigene Ideen zulassen“, mit Ausnahme des Items ei_i4, für das drei Zuweisungen vorliegen. Alle genannten Skalen sind durch ihre Items validiert.

Für das Konstrukt „Conceptual Change – kognitive Konflikte erzeugen“ liegt eine passende Zuordnung für alle Items vor. Das Item cc1 wird von drei Beurteilern in zweiter Priorität zugeordnet, zudem liegen für das Item drei Nennungen für das Konstrukt „Scaffolding – Hervorhebungen“ vor. Eine Berücksichtigung des Items bei dieser Skala führt nicht zu einer Verbesserung der internen Konsistenz der genannten Scaffolding-Skala. Insgesamt scheint beim Item cc1 für die Experten nicht eindeutig zu sein, was mit 'überzeugenden Versuchen' gemeint ist. Da für dieses Item keine eindeutig mehrheitliche Zuordnung zu einer Skala vorliegt, sollte es ausgeschlossen werden.

Beim Konstrukt „Conceptual Change – kognitive Konflikte erzeugen“ werden zudem vereinzelt Bezüge zur Skala „Scaffolding – Perturbieren“ gesehen. Umgekehrt stellen vier der fünf Experten bei Items des Konstrukts „Scaffolding – Perturbieren“ Bezüge zum Konstrukt „Conceptual Change – kognitive Konflikte“ her. Nur zwei Items werden gemäß der Tabelle 4.4-2 korrekt zugeordnet. Bei drei Items liegen Nennungen für die Conceptual Change-Skala vor. Die wechselseitigen Zuordnungen lassen sich durch den empirisch nachgewiesenen starken Zusammenhang der Skalen untereinander ($r=.80^{**}$, n. Pearson 2-seitig, $N=30$) erklären (vgl. Tabelle 4.2-3 auf Seite 138). Dies ist aus dem Blickwinkel der Theorie nachvollziehbar. Das Auslösen kognitiver Konflikte kann als Teilaspekt des Perturbierens aufgefasst werden. Diese beiden Konstrukte sollten zusammengefasst werden als „Konfrontation“ (Infra-gestellten/Erschüttern von Schülerkonzepten). Eine erneute Reliabilitätsanalyse führt zu einer sehr guten Reliabilität der Gesamtskala (vgl. Tabelle 4.4-3; vgl. auch Tabelle A.3-1 in Anhang A.3); die Werte für Cronbachs Alpha der Gesamtskala erhöhen sich bei beiden Themen im Vergleich zu den Werten der Einzelskalen.

Für das Item sc_p1, welches der Scaffolding-Skala „Perturbieren“ zugeordnet ist, liegen neben einer Sortierung zur Skala „Conceptual Change – kognitive Konflikte erzeugen“ auch zwei Nennungen für die Skala „Scaffolding – Hervorhebungen“ vor. Eine Zuordnung des genannten Items zur Skala „Hervorhebungen“ wird inhaltlich als nicht passend angesehen und führt zudem zu einer Verschlechterung der Reliabilität der Skala, wie eine erneute Reliabilitätsanalyse zeigt. Das Item geht in das Gesamtkonstrukt „Konfrontation“ ein.

Des Weiteren zeigt sich, dass die Items aus dem Konstrukt „Transmission von Wissen – Erklärungen“ von den Experten zu einem großen Teil dem Transmissions-Konstrukt „Denk- und Lernwege ablehnen“ zugeordnet werden. Beim Konstrukt „Denk- und Lernwege ablehnen“ liegt eine mehrheitlich eindeutige Zuordnung vor. Insgesamt zeigt sich, dass die beiden Transmissions-Konstrukte nicht trennscharf sind, aber zusammen eine große Dimension bilden, die als „Transmission von Wissen“ bezeichnet werden kann. Auch für dieses Gesamtkonstrukt ergeben sich gemäß der Tabelle 4.3-3 sehr gute Werte für Cronbachs Alpha; diese sind höher als die internen Konsistenzen der Einzelskalen (vgl. auch die Tabelle A.3-1 in Anhang A.3).

Beim Konstrukt „Laisser-faire“ werden drei der fünf Items eindeutig zugeordnet. Dagegen werden die zwei Items ol4 und ol5 mit vier bzw. drei Nennungen mehrheitlich der Skala „Transmission von Wis-

sen – Denk- und Lernwege ablehnen“ zugeordnet. Auch hier führt eine weitere Reliabilitätsanalyse zu schlechteren Trennschärfen und niedrigeren Werten für Cronbachs Alpha, so dass die beiden Items nicht in die Transmissions-Skala aufgenommen werden. Für das Item ol_1 liegen neben den vier korrekten Zuordnungen ebenfalls drei Nennungen für die Skala „Eigene Ideen zulassen“ vor. Aufgrund dieser drei Nennungen und in Anbetracht dessen, dass es sich bei den vier korrekten Zuordnungen des Items bei einer Nennung um eine Sortierung zweiter Priorität handelt, wird das Item ausgeschlossen. Die Skala „Laisser-faire“ wird als Kurzskala geführt. Durch den Ausschluss der drei Items aus der Skala „Laisser-faire“ erhöht sich die interne Konsistenz des Konstrukts, wie der Tabelle 4.4-3 zu entnehmen ist (vgl. auch die Tabelle A.3-1 in Anhang A.3).

Tab. 4.4-3 Statistische Kennwerte der Skalen und Items der Konstrukte „Konfrontation“, „Transmission von Wissen“ und „Laisser-faire“ (Kurzskala) zu den Themen „Wasserkreislauf“ (N=30) und „Schwimmen und Sinken“ (N=29)

MZP	Skala	N _{Items}	M _{Skala}	SD _{Skala}	r _{it} (min.-max.)	Cronbachs Alpha
Wasserkreislauf Schwimmen und Sinken	Konfrontation	7	1.95	.52	.32 - .85	.87
			2.63	.64	.66 - .93	.95
Wasserkreislauf Schwimmen und Sinken	Transmission von Wissen	8	1.94	.59	.51 - .89	.92
			1.28	.40	.17 - .93	.93
Wasserkreislauf Schwimmen und Sinken	Laisser-faire	2	1.65	.76		.92
			1.50	.65		.96

Antwortformat von 1 bis 4 ((1) trifft überhaupt nicht zu, (2) trifft eher nicht zu, (3) trifft eher zu, (4) trifft voll und ganz zu).

Beim Konstrukt „Ideen gemeinsam diskutieren“ werden vier von fünf Items dem angedachten Konstrukt mehrheitlich zugeordnet. Es muss jedoch beachtet werden, dass bei fast allen Items mehrere Bezüge zu anderen Konstrukten gesehen werden, was konform geht mit den mittleren bis hohen Korrelationen der Skala „Ideen gemeinsam diskutieren“ mit den anderen Skalen des Instruments (vgl. die Interskalenkorrelationen in Tabelle 4.2-3 auf Seite 138). Dieses Konstrukt wird als eingeschränkt valide betrachtet und sollte überarbeitet werden.

Problematisch gestaltet sich die Zuordnung der Items zum Konstrukt „Eigene Formulierungen“, bei dem lediglich ein Item eindeutig von allen Experten zugeordnet wurde. Das Items ei_e3 wird mehrheitlich der Skala „Präkonzepte – Vorwissen aktivieren und aufgreifen“ (vier Nennungen) und das Items ei_e5 der Transmissions-Skala „Denk- und Lernwege ablehnen“ zugeordnet (fünf Nennungen). Eine Aufnahme der Items in die jeweils genannten Skalen und erneute Reliabilitätsprüfungen führen jedoch zu geringeren Trennschärfen der Items und schlechteren internen Konsistenzen (Cronbachs Al-

pha). Somit verbleiben die Items in der vorgesehen Skala, wobei das Konstrukt vor weiterer Anwendung einer Überarbeitung bedarf (vgl. die Ausführungen auf den Seiten 163f. und 173f.).

4.5 Zusammenfassung der Ergebnisse

Die Ergebnisse der Generalisierbarkeitsanalyse zeigen, dass für fast alle Items gute bis sehr gute Generalisierbarkeitskoeffizienten zwischen .71 und .97 erreicht werden. Nur für wenige Items kann keine ausreichende Interrater-Reliabilität erzielt werden. Der Grund hierfür ist jeweils die geringe Varianz in den Videos, so dass in diesen Fällen die beobachtete Gesamtvarianz vor allem durch die Unterschiede zwischen den Ratern bzw. die Interaktion zwischen Ratern und Videos erklärt wird.

Bei den übrigen Items steht die Varianzkomponente Video als die im Sinne der klassischen Testtheorie 'wahre Varianz' durchweg in einem günstigen Verhältnis zu den systematischen Raterunterschieden und der unsystematischen Fehlervarianz. Die Werte der Varianzkomponente Rater fallen stets niedrig aus. Die Verankerung der zu beurteilenden Items an konkret beobachtbarem Verhalten scheint bei diesen Skalen gut gelungen, so dass systematische Fehlervarianzen aufgrund von Niveau-Unterschieden in den Raterurteilen (also schlechter 'Adjustierung') sehr gering gehalten werden konnten.

Die Berechnung der internen Konsistenzen, gemessen durch Cronbachs Alpha, liefert durchgängig hohe Werte der Koeffizienten, so dass von einer hinreichenden Reliabilität der Messwerte ausgegangen werden kann.

Die zur Einschätzung der Kriteriumsvalidität des Instruments durchgeführten Analysen von Unterschieden im Unterrichtshandeln der Lehrkräfte in den Fortbildungsgruppen zeigen zum Zeitpunkt der Unterrichtsbeobachtung „Schwimmen und Sinken“ (nach der Intervention) signifikante Mittelwertsunterschiede zugunsten der Experimentalgruppe in den erwarteten Konstrukten. Im Unterricht zum Thema „Wasserkreislauf“ unterscheiden sich die Lehrkräfte in ihrem Unterrichtsverhalten wie erwartet nicht signifikant voneinander. Die Ergebnisse der Analysen sprechen für die Kriteriumsvalidität des Instruments.

Die Ergebnisse der Expertenvalidierung zeigen, dass sich Validitätsprobleme für die Konstrukte „Ideen – gemeinsam diskutieren“ und „Eigene Formulierungen“ ergeben. Die beiden Transmissions-Konstrukte werden als ein großes Konstrukt beurteilt, gleiches gilt für die Konstrukte „Conceptual Change – kognitive Konflikte anregen“ und „Scaffolding – Perturbieren“. Für die übrigen Konstrukte kann von einer zufriedenstellenden Expertenvalidierung ausgegangen werden.

5 Diskussion und Ausblick

Im Zentrum des Interesses der Arbeit stand das Ziel, kognitiv anregende Verhaltensweisen im naturwissenschaftsbezogenen Sachunterricht zu konzeptualisieren und zu erfassen. Dies geschah zum einen vor dem Hintergrund der Forderungen nach einem anspruchsvollen naturwissenschaftlichen Sachunterricht und zum anderen in Anbetracht der Bedeutung kognitiv anregender Unterrichtsqualitätsmerkmale für den Lernerfolg von Lernenden, worauf erstmals die Mathematikdidaktik hingewiesen hat.

Die Konzeptualisierung erfolgte auf der Grundlage von zwei Theoriebereichen: Auf der Basis von Hinweisen aus der Instruktionspsychologie und auf der Basis von Hinweisen aus sachunterrichtsdidaktischen Konzeptionen. Herausgearbeitet wurden Lehrerverhaltensweisen zur Anregung aktiven und konstruktiven Lernens, zur Umsetzung sozialen und kooperativen Lernens, zur Umsetzung unterstützten Lernens und zur Umsetzung situierten Lernens. Um derartige Verhaltensweisen in Folgestudien hinsichtlich ihrer Wirksamkeit untersuchen zu können und da bisher kein Videoanalyse-Instrument besteht, welches derartige Verhaltensweisen von Lehrkräften hoch-inferent misst, wurde der Versuch unternommen, ein Videoanalyse-Instrument zur Erfassung der konzeptualisierten Verhaltensweisen zu entwickeln und in Bezug auf testtheoretische Gütekriterien zu analysieren. Wie lassen sich die Ergebnisse der Analysen nun abschließend beurteilen? Welche Konsequenzen ergeben sich aus den Testanalysen für den Einsatz des Instruments zu Forschungszwecken? Welche Folgerungen können für (inhaltliche und methodische) Überarbeitungen der Konstrukte des Instruments gezogen werden? Welche weiteren Forschungsfragen wirft die Arbeit auf? Diesen Fragen soll in diesem, die Arbeit abschließenden Kapitel nachgegangen werden.

Im Folgenden werden zunächst die Befunde zu den Testanalysen zusammenfassend vor dem Hintergrund methodischer Aspekte der Instrumententwicklung diskutiert. Auf der Basis der Ergebnisse der Expertenvalidierung werden die Skalen noch einmal abschließend betrachtet und bewertet (5.1). Daran anschließend werden Probleme der vorliegenden Untersuchung diskutiert (5.2). In Abschnitt 5.3 wird der Frage nach der Fachspezifität der konstruierten und erfassten Lehrerhandlungen nachgegangen. In Abschnitt 5.4 wird Bezug zu schon vorliegenden und veröffentlichten Ergebnissen von Zusammenhängen der Videoskalen mit einem Instrument zu Erfassung von Lehrervorstellungen zum Lehren und Lernen, das im Rahmen des BIQUA-Projekts von T. Kleickmann (2008) entwickelt wurde, hergestellt. Im Ausblick (5.5) werden Fragen und 'Wünsche' für weiterführende Arbeiten, Folgerungen für (inhaltliche) Überarbeitungen des Instruments und weiterführende Fragestellungen diskutiert. Schließlich wird kurz auf den Einsatz von Videoanalyse-Instrumenten in der Aus- und Fortbildung von Lehrpersonen sowie in der Unterrichtsentwicklung eingegangen.

5.1 Zusammenfassende Diskussion der Ergebnisse der Testanalysen

Im Rahmen der Testung des Instruments hinsichtlich seiner testtheoretischen Güte waren vor allem die Frage nach der Reliabilität und Validität des Instruments zentral. Dabei wurde bezüglich der Interrater-Reliabilität aufgrund der Anfälligkeit hoch-inferenter Ratings für systematische und unsystematische Beurteilungsfehler erwartet, dass die Ergebnisse der Generalisierbarkeitsanalysen tendenziell eher niedrig ausfallen. Hingegen wurden in Bezug auf die Validität des Instruments aufgrund des starken Bezugs zur Theorie gute Ergebnisse bei der Expertenvalidierung erwartet.

Zu den Ergebnissen der Generalisierbarkeitsanalyse

Hinsichtlich der Interrater-Reliabilität zeigen die Ergebnisse, dass insgesamt gute bis sehr gute absolute Generalisierbarkeitskoeffizienten (von .71 bis .97) erreicht werden. Zudem zeigt sich ein durchweg günstiges Verhältnis von 'wahrer' Varianz zu systematischen Raterunterschieden und systematischer Fehlervarianz. Auch konnten systematische Raterunterschiede (aufgrund von Niveau-Unterschieden) gering gehalten werden. In Anbetracht dessen, dass hoch-inferente Ratings besonders anfällig für systematische und unsystematische Beurteilungsfehler sind, fallen die Ergebnisse höher und damit besser aus als erwartet. Zurückzuführen ist dies wahrscheinlich auf die umfangreichen Schulungsmaßnahmen⁴⁵. Die Schulung der Rater an Extrembeispielen zur Verankerung der höchsten und niedrigsten Werte (für 1 *trifft überhaupt nicht zu* und 4 *trifft voll und ganz zu*) der Beobachtungsdimensionen ist zweifellos ein unabdinglicher Bestandteil bei der Vorbereitung der Ratings. Zusätzlich hat sich auch die Spezifikation weiterer Abstufungen (in diesem Fall die Stufen 2 *trifft eher nicht zu* und 3 *trifft eher zu*) anhand von Beispielfideos als effektive Ergänzung zur Erzeugung zusätzlicher Urteilssicherheit bei den Ratern erwiesen. Insbesondere hat es sich auch als hilfreich herausgestellt, bei einzelnen Items die Vergabe der Abstufungen schriftlich festzuhalten, um Unsicherheiten bei den Beurteilungen entgegenzuwirken.

Auch der Einsatz der Generalisierbarkeitstheorie als Methode zur Prüfung der Güte der Raterurteile hat möglicherweise zu den positiven Ergebnissen der Interrater-Reliabilität beigetragen. Weshalb dem so ist, soll näher beleuchtet werden: Die Varianzkomponente 'Video' gibt Informationen über die Streuung in den Videos. Fehlende Varianz in den Videos führt zu einem ungünstigen Verhältnis der Varianzkomponenten zueinander. Entsprechend niedrig fallen die Generalisierbarkeitskoeffizienten aus. Erreicht ein Item aufgrund fehlender Varianz in den Videos einen niedrigen Wert beim Generalisierbarkeitskoeffizienten und somit keine ausreichende Reliabilität, ist klar, dass auch weitere Schulungsmaßnahmen erfolglos verlaufen würden. Das Nicht-Funktionieren eines Items liegt in diesem Fall nicht an den Ratern, sondern an der Stichprobe. Ein derartiges Item kann in einer anderen Stich-

⁴⁵Eine etwas weniger intensive Schulung dürfte wahrscheinlich ausreichen, um dennoch akzeptable Übereinstimmungskoeffizienten zu erreichen.

probe durchaus erfolgreich geratet werden. Die Varianzkomponente 'Rater' gibt systematische Niveau-Unterschiede in den Beurteilungen der Rater an. Nehmen die systematischen Raterunterschiede einen hohen Anteil an der Gesamtvarianz ein und führen zu einem unzureichenden Generalisierbarkeitskoeffizienten, so liegt dies möglicherweise an einer schlechten Adjustierung des Items, an der Itemformulierung o.ä. In weiteren Schulungen kann dem durch passende Maßnahmen Abhilfe geschaffen werden (z.B. durch Umformulierung des Items, durch stärkere Herausarbeitung der Bewertungsgrundlage des Items: Welche Aussagen, Handlungen, Situationen im Unterricht sind ausschlaggebend für die Beurteilung des Items?, ...). Die Varianzkomponenten liefern also wertvolle Informationen über den reinen Generalisierbarkeitskoeffizienten hinaus. Durch die genaue Betrachtung der Varianzkomponenten ist es demnach möglich, das Zustandekommen der Generalisierbarkeitskoeffizienten genau nachzuvollziehen. Bei zu geringen Werten können entsprechende Folgerungen für weitere Schulungsmaßnahmen gezogen werden⁴⁶. Es lässt sich somit festhalten, dass die Generalisierbarkeitsanalyse einen sehr geeigneten Ansatz zur Prüfung der Interrater-Reliabilität darstellt.

Zu den Ergebnissen der Item- und Skalenanalyse

Die Darstellung der Mittelwerte und Standardabweichungen der Items und Skalen hat gezeigt, dass für einige Items Decken- bzw. Bodeneffekte vorliegen. Die Deckeneffekte bei Items der Skala „Fehleroffenheit“ sowohl für das Thema „Wasserkreislauf“ als auch für das Thema „Schwimmen und Sinken“ sind insofern nicht verwunderlich, als Grundschullehrkräfte generell schülerorientiert denken und wahrscheinlich auch handeln.

Die Bodeneffekte bei Items der beiden Transmissiv-Skalen und der Skala „Laisser-faire“ beim Thema „Schwimmen und Sinken“ lassen sich inhaltlich wahrscheinlich auf den Einfluss der Fortbildung zurückführen. Methodisch wäre die Itemformulierung zu überdenken. Möglicherweise sind die Items zu 'extrem', d.h. in diesem Fall zu negativ formuliert. Dies könnte zu einer Art 'Vorbewertung' geführt haben. Denkbar ist, dass eine schwächere Itemformulierung zu einer größeren Streuung in den Videos geführt hätte, was aus statistischer Sicht wünschenswert gewesen wäre, auch wenn es natürlich ebenso wünschenswert ist, dass derartiges Lehrerverhalten selten beobachtet wird.

Die Ergebnisse zur Prüfung der Trennschärfen der Items zeigen, dass die korrigierten Item-Skalenwert-Korrelationen (r_{it}) für fast alle Items erstaunlich hoch ausfallen. Die Ergebnisse bezüglich der Reliabilitätsanalysen der Skalen mittels der internen Konsistenzen liefern durchweg gute bis sehr gute Werte für Cronbachs Alpha. Die hohen Trennschärfen und die daraus resultierenden hohen Cronbachs Alpha-Werte sind möglicherweise darauf zurückzuführen, dass die Rater – wie im Rahmen der methodischen Ausführungen beschrieben – neben den einzelnen Items der theoretischen Konstrukte

⁴⁶Zwar ist eine endgültige Testung der Reliabilität der Raterurteile erst möglich, nachdem das ganze Datenmaterial vollständig geratet worden ist, dennoch können als Orientierung auch während des Ratingprozesses die Interrater-Reliabilitäten anhand der zu den Zeitpunkten bereits beurteilten Videos bestimmt werden, um abschätzen zu können, ob reliable Werte zu erwarten sind (vgl. dazu die Ausführungen zur Entwicklung des Ratingverfahrens in Abschnitt 3.4.1.3 und zu Methoden der Videoanalyse in Unterkapitel 1.5.1).

auch eine Gesamteinschätzung eines jeden Konstrukts vornahmen. Dies hat wahrscheinlich dazu geführt, dass die Rater sich zunächst eine Gesamteinschätzung überlegten und daraufhin die Items ähnlich der Gesamteinschätzung bewerteten statt sie individuell zu beurteilen. Die Entscheidung für eine Gesamteinschätzung der Skalen hat sich demnach als nicht empfehlenswert erwiesen.

Zudem liegen (teilweise größere) Unterschiede in den Cronbachs Alpha-Werten zwischen den Themen vor; das Alpha ist beim Thema „Schwimmen und Sinken“ in fast allen Skalen höher als beim Thema „Wasserkreislauf“. Die Rater scheinen bei den Unterrichtsvideos zum Thema „Schwimmen und Sinken“ noch mehr Schwierigkeiten gehabt zu haben, die Items individuell zu beurteilen. Dies ist möglicherweise darauf zurückzuführen, dass die Unterrichtsaufzeichnungen zum „Schwimmen und Sinken“ ähnliche Oberflächenstrukturen aufweisen, was bei den Wasserkreislauf-Videos nicht der Fall ist⁴⁷. Denkbar ist auch, dass sich die für das Rating der Videos getroffene Auswahl der gesamten Unterrichtsstunde als Analyseeinheit ausgewirkt hat. Die ganzheitliche Betrachtung der Videos hat den Ratern zwar prinzipiell keine Probleme bereitet, sie könnte jedoch dazu geführt haben, dass die Rater sich bei ihren Urteilen an (wenigen) prägnanten Augenblicken im Video orientiert haben. Dies würde dafür sprechen, mehrere Sinnabschnitte als Analyseeinheit zu wählen (z.B. Hinführung Arbeitsphase, Reflexion). Zwar geht es bei hoch-inferenten Ratings um eine Einschätzung des Lehrenden im Sinne eines Gesamteindrucks, allerdings besteht dabei auch die Gefahr, die Lehrperson zu global zu erfassen und verschiedene Facetten des Handelns der Lehrperson in unterschiedlichen Phasen des Unterrichts zu vernachlässigen. Bei getrennter Betrachtung einzelner Unterrichtsabschnitte gelingt es Ratern ggf. besser, das Augenmerk auch auf Feinheiten des Unterrichts zu lenken und diese in die Bewertungen einzubeziehen.

Zu den Ergebnissen von Unterschieden im Unterrichtshandeln der Lehrkräfte in den Fortbildungsgruppen zur Analyse der 'concurrent validity'

Mit der Berechnung von Unterschieden im Unterrichtshandeln der Lehrkräfte in den Fortbildungsgruppen wurde im Rahmen dieser Studie das Ziel verfolgt, die 'concurrent validity' als einer Form der Kriteriumsvalidität des Videoanalyse-Instruments zu analysieren. Unter der Annahme des unterschiedlichen Einflusses der Fortbildungsarten auf das Unterrichtshandeln der Lehrkräfte, sollte das Instrument Unterschiede im unterrichtlichen Lehrerverhalten nach der Intervention in den Fortbildungsgruppen (Experimentalgruppen, EGs, N=19 vs. Kontrollgruppe, KG, N=10) messen können.

Die Analysen zeigen fast durchgängig die zum Zeitpunkt der Videoaufzeichnung „Schwimmen und Sinken“ erwarteten Mittelwertsunterschiede in den Gruppen: Die Skalen „Präkonzepte – Vorwissen aktivieren und aufgreifen“, „Conceptual-Change – kognitive Konflikte erzeugen“, „Ideen gemeinsam diskutieren“ und die drei Scaffolding-Skalen zeigen höhere Mittelwerte für die Lehrkräfte der Experi-

⁴⁷Die ähnlichen Oberflächenstrukturen in den Unterrichtsvideos zum „Schwimmen und Sinken“ sind sehr wahrscheinlich darauf zurückzuführen, dass die Lehrkräfte für die Durchführung des Unterrichts Lehrerhandreichungen mit Stundenverlaufsplänen erhalten haben. Sie dazu auch die Ausführungen in den Methoden in 3.2.

mentalgruppen; die Skala „Laisser-faire“ zeigt einen höheren Mittelwert für die Lehrkräfte der Kontrollgruppe. Die Unterschiede werden mit Ausnahme der Skala „Scaffolding-Hervorhebungen“ signifikant, wobei der mittlere Effekt bei der Scaffolding-Skala „Hervorhebungen“ zumindest auf eine tendenzielle Überlegenheit der Lehrkräfte aus den Experimentalgruppen weist. Ein Blick auf die Boxplots beider Gruppen, die in Anhang A.5 (Abbildung A.5.20) dargestellt sind, gibt einen Hinweis darauf, weshalb der Unterschied nicht signifikant wird. Ein Ausreißer in der Kontrollgruppe führt vermutlich dazu, dass der Mittelwert der Gruppe, der ohnehin schon auf einem sehr kleinen N von 10 basiert, deutlich höher liegt als der Median und damit recht nahe am Mittelwert der Experimentalgruppe.

Die nicht direkt erwarteten Unterschiede zwischen den Gruppen zugunsten der Experimentalgruppen in den Skalen „Motivation des Unterrichts“, „Eigene Ideen zulassen“ und „eigene Formulierungen“ sind wohl darauf zurückzuführen, dass die Fortbildungen auch auf diese Bereiche des unterrichtlichen Lehrerhandelns indirekt Auswirkungen haben. Denkbar ist z.B., dass Lehrpersonen, die besonderen Wert auf einen adäquaten Umgang mit Präkonzepten legen, auch vermehrt eigene Ideen und eigene Formulierungen der Kinder zulassen. Lassen Lehrpersonen eigene Ideen und Formulierungen der Kinder zu, so kann dies wiederum dazu führen, dass ein Unterricht motivierend auf die Lernenden wirkt.

Die im Vergleich zur Selbststudiumsgruppe geringere Heterogenität in den Experimentalgruppen, auf die bei der Ergebnisdarstellung hingewiesen wurde, ist höchstwahrscheinlich auf die tutorielle Unterstützung, welche die Lehrenden in den Experimentalgruppen erfahren haben, zurückzuführen⁴⁸.

Zum Zeitpunkt der Videoaufzeichnung des Unterrichts zum Thema „Wasserkreislauf“ zeigen die Skalen wie berichtet erwartungsgemäß ähnliche Werte für die Lehrkräfte der Experimentalgruppen einerseits und die Lehrkräfte der Kontrollgruppe andererseits. Zu diesem Messzeitpunkt liegen keine statistisch bedeutsamen (signifikanten) Unterschiede vor.

Insgesamt lässt sich also festhalten: Die Analysen deuten darauf hin, dass das Videoanalyse-Instrument die Unterschiede im Lehrerverhalten im Unterricht zum Thema „Schwimmen und Sinken“, die in den Gruppen vorzuliegen scheinen, messen kann. Insgesamt sprechen die Ergebnisse also für die 'concurrent validity' des Instruments. Die Validierungstechnik der 'bekannten Gruppen' konnte im Rahmen der vorliegenden Arbeit demnach sinnvoll eingesetzt werden.

⁴⁸Insgesamt lassen sich aus den berichteten Ergebnissen nicht nur Rückschlüsse auf die Kriteriumsvalidität des Instruments ziehen. Die Ergebnisse verdeutlichen auch die Relevanz von Lehrerfortbildungen, die an Conceptual Change-Ansätzen und Theorien des Scaffolding orientiert sind.

Zu den Ergebnissen der Expertenvalidierung zur Analyse der Konstruktvalidität

Hinsichtlich der Konstruktvalidität des Instruments zeigen die Ergebnisse der Expertenvalidierung entsprechend den Ausführungen in Kapitel 4.3 das folgende Bild:

Bewährt haben sich die folgenden sieben Skalen:

- „Motivation des Unterrichts“
- „Präkonzepte – Vorwissen aktivieren und aufgreifen“
- „Scaffolding – Hervorhebungen“
- „Scaffolding – Sequenzierungen“
- „Anwendungsbezogenes Lernen“
- „Eigene Ideen zulassen“
- „Fehleroffenheit“

Bewährt haben sich auch die folgenden drei Skalen:

- „Konfrontation“ (als Neukonstruktion aus den Skalen „Conceptual Change – kognitive Konflikte anregen“ und „Scaffolding – Perturbieren“)
- „Transmission von Wissen“ (als Neukonstruktion aus den Skalen „Transmission von Wissen – Erklärungen“ und „Transmission von Wissen – Denkwege ablehnen“)
- „Laisser-faire“ (als Kurzskala mit zwei Items)

Nicht bewährt haben sich zwei Skalen:

- „Ideen gemeinsam diskutieren“
- „Eigene Formulierungen“

Gemäß den Ergebnissen der Expertenvalidierung sind die Items, die den Konstrukten „Motivation des Unterrichts“, „Präkonzepte – Vorwissen aktivieren und aufgreifen“, „Scaffolding – Hervorhebungen“, „Scaffolding – Sequenzierungen“, „Anwendungsbezogenes Lernen“, „Eigene Ideen zulassen“ und „Fehleroffenheit“ zugeordnet sind, durch die fünf an der Expertenvalidierung teilnehmenden Personen korrekt bzw. einheitlich zugeordnet. Diese sieben Skalen lassen sich somit wie theoretisch angenommen bilden. Die Rückbindung der Skalen an die Theorie scheint gelungen zu sein.

Die Skala „Konfrontation“ ergibt sich als Neukonstruktion aus den Skalen „Conceptual Change – kognitive Konflikte anregen“ und „Scaffolding – Perturbieren“, da die Items, die diesen beiden Beobachtungsdimensionen zugeordnet sind, wechselseitig zugeordnet sind. Das Gesamtkonstrukt kann als ausreichend valide betrachtet werden. Wie lassen sich die wechselseitigen Zuordnungen erklären? Mit der Skala „Conceptual Change – kognitive Konflikte anregen“ soll erhoben werden, inwiefern die Lehr-

person die Schüler mittels objektiver Evidenz auf Widersprüche in ihrem Denken aufmerksam macht. Mit der Skala zum Perturbieren wird das Ziel verfolgt, Problematisierungshilfen und Provokationen zu erfassen. Kognitive Konflikte sind Problematisierungshilfen. Die Skala zum Anregen kognitiver Konflikte sollte durch die Fokussierung des Konfrontierens der Kinder mit *objektiver* Evidenz enger gefasst sein als die Scaffolding-Skala zum Perturbieren. Dies ist offensichtlich nicht deutlich genug hervorgehoben worden, so dass es absolut nachvollziehbar ist, dass beide Konstrukte für die Experten (außer für einen Experten) nicht trennscharf sind.

Die Skala „*Transmission von Wissen*“ ergibt sich als Neukonstruktion aus den beiden Skalen „Transmission von Wissen – Erklärungen“ und „Transmission von Wissen – Denkwege ablehnen“. Für beide Skalen liegen keine eindeutigen, da wechselseitige Zuordnungen der Items vor. Zudem sind der Skala „*Laisser-faire*“ nur zwei von ursprünglich fünf Items eindeutig zugeordnet. Ein Item wird von drei Experten in erster und von einem Experten in zweiter Priorität der Skala zusortiert. Die übrigen zwei Items, sind einheitlich den Transmissions-Skalen zugeordnet. Zurückzuführen sind diese Probleme sowohl bei den Transmissions-Skalen als auch bei der Skala „*Laisser-faire*“ wahrscheinlich auf unzureichende theoretische Ausarbeitungen und Operationalisierungen. Bei Betrachtung der den Transmissions-Konstrukten zugrunde liegenden theoretischen Ausführungen fällt auf, dass diese deutlich abstrakter und allgemeiner gehalten sind, als die theoretischen Darlegungen, aus denen sich die Konstrukte, die sich auf ein konstruktives bzw. kognitiv-konstruierendes Verständnis beziehen ableiten (z.B. Umgang mit Präkonzepten, Einsatz von Scaffolding-Maßnahmen). Entsprechend allgemein gehalten sind die Konstruktbeschreibungen der Transmissiv-Skalen. Zudem scheint die Abgrenzung beider Konstrukte zu unpräzise zu sein. Das Konstrukt „*Laisser-faire*“ ist wie auch die Transmissiv-Konstrukte negativ belegt. Es ist als Negation aus den Ausführungen zur Umsetzung unterstützten Lernens entstanden. Auch dieses Konstrukt ist recht allgemein gehalten und besteht aus Items, die ebenso wie die Transmissiv-Items negativ formuliert sind. Bei allen drei Skalen führen die beschriebenen Probleme dazu, dass bei der Zuordnung der Items zu den Konstrukten nicht ganz deutlich wird, welchem Konstrukt sie nun zugeordnet sind.

Sowohl für die beiden neu konstruierten Gesamtskalen „Konfrontation“ und „Transmission von Wissen“ als auch für die Kurzskala „*Laisser-faire*“ zeigen die nachträglich durchgeführten Reliabilitätsanalysen wie im Ergebnisteil bereits berichtet, dass bei allen drei Skalen von einer hohen Reliabilität ausgegangen werden kann (vgl. Tabelle 4.4-3 und die Tabelle A.3-1 in Anhang A.3).

Offen bleibt die Validität der Konstrukte „Ideen gemeinsam diskutieren“ und „Eigene Formulierungen“, da für diese Konstrukte uneinheitliche Zuordnungen vorliegen. Für beide Skalen sind Neukonstruktionen notwendig. Dabei stellen sich die Fragen, worin die Probleme der unzureichenden Validierung begründet liegen und wie eine Weiterentwicklung der genannten Beobachtungsdimension aussehen könnte. Im Folgenden werden mögliche Probleme der betroffenen Konstrukte thematisiert. Auf

die Frage, welche inhaltlichen Überarbeitungen als sinnvoll betrachtet werden, wird in einem gesonderten Abschnitt im Rahmen des Ausblicks in 5.5 ausgeführt.

Mit dem Konstrukt „*Ideen gemeinsam diskutieren*“ soll erfasst werden, inwiefern die Lehrperson Unterrichtsinhalte mit den Kindern gemeinsam diskutiert und inwiefern sie die gemeinsame Diskussion unterstützt. Die Beschreibung der Beobachtungsdimension liegt also auf einer sehr allgemeinen Ebene. Genau darin könnten die Probleme, welche die Experten bei der Zuordnung der Items zum Konstrukt hatten, begründet liegen. Das Konstrukt wirkt unklar. Für den außenstehenden Betrachter stellen sich Fragen, wie: Was wird konkret diskutiert? Wozu wird diskutiert? Wie fördert die Lehrperson die Diskussion? Das Konstrukt muss also eine Konkretisierung erfahren.

Bei der Skala „*Eigene Formulierungen*“, welche erfassen soll, ob die Lehrperson Schülerformulierungen im Unterricht Platz einräumt, haben zwei Items Schwierigkeiten bei der Zuordnung bereitet. Das Item ei_e3 „*Die Lehrperson greift die Formulierungen der Kinder auf und arbeitet damit weiter*“ wird mit Ausnahme eines Experten dem Konstrukt „Präkonzepte – Vorwissen aktivieren und aufgreifen“ zugeordnet. Bei diesem Konstrukt wird hervorgehoben, dass es wichtig ist, dass die Lehrperson die Konzepte der Kinder nicht nur erfragt, sondern in den weiteren Unterrichtsverlauf auch einbezieht, also auf die Präkonzepte der Kinder zurückkommt. Legt man die Betonung beim genannten Item nicht auf die Worte '*greift die Formulierungen der Kinder auf*', sondern auf die Worte '*und arbeitet damit weiter*', lässt sich die Zuordnung des Items bei der Skala „Präkonzepte – Vorwissen aktivieren und aufgreifen“ erklären. Darin zeigt sich, wie problematisch es sein kann, wenn Items mehr als eine Aussage enthalten.

Das Item ei_e5 „*Die Lehrperson sorgt nicht dafür, dass 'Fachtermini' mit eigenen Worten erklärt werden*“ ist mehrheitlich den Transmissiv-Skalen zugeordnet, was eventuell an der negativen Formulierung des Items liegt. Da die Beobachtungsdimensionen zur Transmission von Wissen durchweg mit negativen Attributen belegt sind (Autorität, ohne Begründungen, vorschnell, unterbindet, ignoriert, ...) ist eine Zuordnung des genannten Items zu diesen Konstrukten nachvollziehbar. Hinzu kommt, dass es – wie ein Experte im Rahmen seiner Ausführungen anmerkt – generell schwieriger ist, bei negativ formulierten Items einzuordnen, für welches Konstrukt sie angedacht sind.

Es lässt sich also festhalten, dass sich bei beiden Items Indizien finden lassen, die Bezüge zu anderen Konstrukten schaffen. Als weitere Schwierigkeit kommt beim Konstrukt „eigene Formulierungen“ laut der Aussage eines Experten hinzu, dass 'eigene Formulierungen' auch in vielen anderen Items unterschwellig enthalten sind.

Die abschließende Betrachtung und Bewertung der Ergebnisse der Expertenvalidierung zeigen, dass bei einigen wenigen Konstrukten des Instruments Mängel vorliegen, so dass diese Konstrukte eine Überarbeitung erfahren sollten. Größtenteils lassen sich die Skalen aber wie theoretisch angenommen bilden. Für sieben von 14 Konstrukten liegen eindeutige Zuordnungen der Items vor, vier Skalen lie-

ßen sich zu zwei Skalen zusammenfassen, die Skala „Laisser-faire“ funktioniert als Kurzskala, so dass für zehn Skalen von einer ausreichenden Validität ausgegangen werden kann und diese Skalen eingesetzt werden können. Nur zwei Skalen mussten aufgrund unzureichender Validität ausgeschlossen werden. Die Zuordnung der Items zu den theoretischen Konstrukten fällt also insgesamt zufriedenstellend aus. Die erwarteten guten Ergebnisse der Expertenvalidierung bestätigen sich also.

Fazit

Was lässt sich rückblickend auf die Fragestellung feststellen? Kann die Entwicklung des hoch-inferenten Videoanalyse-Instruments zur Erfassung kognitiv anregenden Verhaltensweisen von Lehrkräften im naturwissenschaftsbezogenen Sachunterricht als erfolgreich betrachtet werden?

Insgesamt zeigen die Ergebnisse der Güteprüfung, dass davon ausgegangen werden kann, dass das entwickelte Instrument mit Ausnahme der zwei Skalen „Ideen gemeinsam diskutieren“ und „Eigene Formulierungen“ hinreichend reliabel und valide misst. Die Befunde der Arbeit können als Hinweis verstanden werden, dass es möglich ist, Lehrerverhaltensweisen, wie sie in dieser Arbeit konzeptualisiert wurden, mittels eines hoch-inferenten Videoratings zu erfassen. Zudem scheint es möglich zu sein, kognitiv anregendes Lehrerverhalten in verschiedenen Facetten zu erheben. Es soll noch einmal hervorgehoben werden, dass es sich bei dem vorliegenden Instrument um den ersten Versuch handelt, Lehrerverhaltensweisen, wie sich hier aus der Theorie abgeleitet wurden, hoch-inferent zu messen. Das Instrument bietet sich als Ausgangspunkt für weitere Videoanalysen an. Es wurde zur Auswertung der BIQUA-Untersuchung eingesetzt (vgl. 5.4).

Zu berücksichtigen sind jedoch auch Einschränkungen der Studie, die sich zum einen aus der methodischen Anlage der Untersuchung und zum anderen durch die erläuterten Validitätsprobleme ergeben. Im folgenden Abschnitt wird auf diese Probleme der Studie näher eingegangen.

5.2 Probleme der Studie

Einschränkungen der Studie ergeben sich vor allem durch methodische Aspekte.

Zur Repräsentativität und zur Größe der Stichprobe

Repräsentativität der Stichprobe. Zur Einschätzung der Repräsentativität der dieser Untersuchung zugrunde liegenden Lehrerstichprobe wurde neben demographischen Daten (Geschlechter- und Altersverteilung; Berufserfahrung) auch ein Vergleich der Lehrerstichprobe mit einer für Nordrhein-Westfalen weitgehend repräsentativen Stichprobe von Grundschullehrkräften berücksichtigt. Der Vergleich erfolgte hinsichtlich der Variablen „Fachinteresse Physik“, „Sachinteresse Physik“, „Fähigkeitsselbstkonzept“, „Selbstwirksamkeitserwartungen bezüglich des Unterrichts physikbezogenen Sachunterrichts“ und „Interesse am Unterrichten physikbezogenen Sachunterrichts“ (Selbsteinschätzungen der

Lehrkräfte). Dabei zeigen sich wie berichtet zum Zeitpunkt der Unterrichtsbeobachtung „Wasserkreislauf“ (vor der Intervention) mit Ausnahme des Fähigkeitsselbstkonzepts keine oder nur als 'klein' zu bewertende Abweichungen der Untersuchungsstichprobe von der NRW-Stichprobe, so dass zu diesem Messzeitpunkt von einer repräsentativen Stichprobe und von repräsentativen Unterrichtsbeobachtungen ausgegangen werden kann. Dahingegen zeigen sich zum Zeitpunkt der Unterrichtsbeobachtung „Schwimmen und Sinken“ (nach der Intervention) mittlere bis stärkere Abweichungen der Videostichprobe von der repräsentativen Vergleichsstichprobe insbesondere hinsichtlich der interessenbezogenen Variablen und hinsichtlich der Selbstwirksamkeitserwartungen bezüglich des Unterrichts physikbezogenen Sachunterrichts. Nach der Intervention zeichnen sich die Lehrkräfte also durch ein höheres Interesse sowohl an der Sache Physik als auch am Unterrichten naturwissenschaftlichen Sachunterrichts aus, sowie durch positivere Selbsteinschätzungen der eigenen Fähigkeiten im Unterrichten solcher Themen. Diese Abweichungen sind vermutlich durch die Teilnahme der Lehrkräfte an den Fortbildungen zu erklären (vgl. die Abschnitte 3.1, 3.2 und 3.3). Außerdem zeigen Ergebnisse von Untersuchungen, die im Rahmen des BIQUA-Projekts durchgeführt wurden, dass Vorstellungen von Lehrkräften zum Lehren und Lernen (wie z.B. „Lernen als Conceptual Change“, „Schüler haben Präkonzepte“ oder in Form von Ablehnung auch praktizistische und sehr offene Vorstellungen zum Lehren und Lernen („Praktizismus und Laisser-faire“) durch die Fortbildungen verändert werden konnten (vgl. Möller et al., 2006) und dass diese modifizierten Lehrervorstellungen auch handlungsrelevant zu sein scheinen (vgl. Vehmeyer, Kleickmann & Möller, 2007; Kleickmann, Vehmeyer & Möller, i. D.; siehe dazu auch die Ausführungen in Abschnitt 5.4). Zum zweiten Messzeitpunkt kann deshalb zum einen nicht mehr von einer repräsentativen Lehrerstichprobe ausgegangen werden. Zum anderen ist es plausibel, dass das beobachtete Unterrichtshandeln der Lehrkräfte zum Thema „Schwimmen und Sinken“ über der durchschnittlichen Qualität des Unterrichts zu diesem Thema liegt.

Der Aspekt, dass es sich bei den Unterrichtsbeobachtungen zum Thema „Schwimmen und Sinken“ um Unterricht von fortgebildeten Lehrkräften handelt, muss allerdings nicht nur als Einschränkung betrachtet werden. Für die Instrumententwicklung kann die Teilnahme der Lehrkräfte an den Fortbildungen sogar von Vorteil gewesen sein. Möglicherweise hat die vermutlich überdurchschnittliche Qualität des Unterrichtshandelns der Lehrkräfte zu mehr Varianz in den 59 Videos geführt, was den Ratern die Beurteilung der Videos sicherlich erleichtert hat.

Größe der Stichprobe. Zu berücksichtigen ist bei der Interpretation der Ergebnisse ebenfalls, dass die Entwicklung des Instruments auf einer recht kleinen Stichprobe von 30 Lehrkräften basiert.

Zur Untersuchungsanlage

Neben der geringen Stichprobengröße müssen weitere für Videostudien typische Einschränkungen berücksichtigt werden. Das Instrument wurde auf der Basis von zwei Unterrichtsthemen – den Themen „Wasserkreislauf“ und „Schwimmen und Sinken“ – entwickelt und getestet, wobei ausschließlich je-

weils die erste Doppelstunde in den Einheiten zu beobachten war. Zudem standen für die Entwicklung des Instruments nur ein Video pro Lehrer und Thema zur Verfügung.

Zu berücksichtigen ist weiterhin, dass die Fortbildungsteilnehmer Materialpakete (Materialboxen und Lehrerhandreichungen) für die Durchführung des Unterrichts zum Thema „Schwimmen und Sinken“ erhielten (vgl. 3.2). Angemerkt werden soll an dieser Stelle, dass die in den Handbüchern enthaltenen Vorschläge für die Unterrichtsgestaltung und methodisch-didaktischen Hilfen für den Unterricht zwar dazu beigetragen haben, dass die Oberflächenstrukturen des videographierten Unterrichts ähnlich sind; sie haben jedoch nicht dazu beigetragen, dass die Lehrpersonen sich in Verhaltensweisen, die auf 'Verständnis' abzielen (z.B. Umgang mit Präkonzepten und Anregung von Konzeptwechseln, Anbieten von Scaffolding-Maßnahmen) ähneln. Hier gibt es zwischen den Lehrpersonen deutliche Unterschiede (siehe dazu die Ausführungen in 4.3 und 5.1 und die Boxplots in Anhang A.5).

Zur Validitätsprüfung des Instruments

'Concurrent validity'. Bezüglich der Analyse der Mittelwertsunterschiede im Unterrichtshandeln der Lehrkräfte in den Fortbildungsgruppen (EGs und KG) zur Testung der 'concurrent validity' des Instruments muss sicherlich beachtet werden, dass die angenommenen Unterschiede im Unterrichtshandeln in den Fortbildungsgruppen bisher nur durch das vorliegende Instrument nachgewiesen wurden. Wünschenswert wäre es sicherlich, die Unterschiede, die im Lehrerverhalten vorzuliegen scheinen, auch mittels eines anderen Instruments zu bestätigen. Des Weiteren sollte auch hier berücksichtigt werden dass die Analysen auf einer kleinen Stichprobe (Schwimmen und Sinken: N=29 und Wasserkreislauf: N=30) basieren. Zu berücksichtigen ist ebenfalls, dass mit der 'concurrent validity' nur eine Form der Kriteriumsvalidität getestet worden ist.

Konstruktvalidität. Die wohl bedeutsamste methodisch bedingte Beschränkung der Studie liegt darin, dass die Überprüfung der Konstruktvalidität des Instruments aufgrund des für Videostudien bekannten Problems der geringen Teststärke ausschließlich theoretisch erfolgte. Der Stichprobenumfang von 59 Videos⁴⁹ hat im Kombination mit der hohen Itemanzahl des Instruments dazu geführt, dass keine faktorenanalytische Validitätsprüfung möglich war (vgl. 3.5.3.2). Die Größe der Stichprobe ist hier somit vor allem vor dem Hintergrund der Validität (als dem wichtigsten Gütekriterium) zu diskutieren.

Konvergente und diskriminante Validität. Im Rahmen der Ausführungen zur Auswahl der Methoden zur Überprüfung der Qualität des Videoanalyse-Instruments in Abschnitt 3.5.3.2 wurde diskutiert, dass ein Einsatz des Multitrait-Multimethod-Verfahrens zur Einschätzung der konvergenten und diskriminanten Validität des Instruments aufgrund der vorliegenden Datenbasis nicht möglich war. Die Interkalenkorrelationen geben bereits Hinweise auf die diskriminante Validität der Skalen des Instruments.

⁴⁹Angemerkt sei an dieser Stelle, dass der Umfang von 59 Videos zum damaligen Zeitpunkt der Datenerhebung für Videoanalysen üblich hoch war.

Validitätsprobleme bei der Erfassung einiger Lehrerhandlungen. Die Expertenvalidierung hat, wie bereits erwähnt, deutlich gemacht, dass bei den Konstrukten „Ideen gemeinsam diskutieren“ und „Eigene Formulierungen“ Probleme mit der Konstruktvalidität vorliegen.

5.3 Zur Fachspezifität der konzeptualisierten und erfassten Lehrerhandlungen

In der Arbeit ging es darum, fachspezifische Unterrichtsqualitätsmerkmale, also Merkmale fachlich guten Lehrerhandelns für den naturwissenschaftlichen Sachunterricht zu konzeptualisieren und zu erfassen. Begründet wurde dieser Ansatz zum einen über aktuelle Modelle der Unterrichtsforschung zur Wirksamkeit von Unterricht (Angebots-Nutzungs-Modelle), die Merkmale der Lehrer- und Unterrichtsqualität als bedeutsame Faktoren für den Lernerfolg der Lernenden herausstellen (vgl. die Ausführungen in der Einleitung) und zum anderen über Forschungen aus der deutschsprachigen Mathematikdidaktik, die für die Annahme, dass fachspezifische Unterrichtsmerkmale mit positiven (kognitiven) Lernergebnissen einhergehen, mittlerweile deutliche empirische Evidenzen vorweisen kann.

Zur theoretischen Umsetzung des Vorhabens der Arbeit wurden aus Ansätzen der Instruktionspsychologie und aus sachunterrichtsdidaktischen Ansätzen Verhaltensmerkmale abgeleitet, die als relevant für die Umsetzung eines fachlich anspruchsvollen und auf Verstehen ausgerichteten Unterrichts anzusehen sind. Während bei Konstrukten, die das Anregen von Conceptual Change, den Umgang mit Präkonzepten oder den Einsatz von Scaffolding-Maßnahmen beschreiben, die Fachbezogenheit deutlich ist, stellt sich bei anderen Konstrukten, wie z.B. „Ideen gemeinsam diskutieren“, „Motivation des Unterrichts“ oder „Fehleroffenheit“ die Frage, ob diese Konstrukte tatsächlich spezifisch für den naturwissenschaftlichen (Sach-)Unterricht sind. Ein Vergleich der in dieser Arbeit konzeptualisierten Unterrichtsmerkmale mit den in Kapitel 1.1 berichteten Elementen kognitiv aktivierenden Mathematikunterrichts zeigt Parallelen auf. Auch hier gelten die Motivierungsfähigkeit der Lehrperson, anspruchsvolle Unterrichtsgespräche oder ein positiver Umgang mit Fehlern als wichtige Elemente fachlichen Lernens. Ebenso wurden in einer Videostudie zum Englischunterricht, die im Rahmen von DESI⁵⁰ durchgeführt wurde, Unterrichtselemente, wie ein konstruktiver Umgang mit Schülerfehlern oder die (thematische) Motivierung der Lernenden herausgearbeitet (vgl. Helmke & Schrader, 2006b). Der Vergleich zeigt also, dass auch in anderen Disziplinen von ähnlichen unterrichtlichen Handlungsweisen von Lehrern berichtet wird. Stellt der Vergleich die Fachspezifität von einigen der entwickelten Konstrukte nun in Frage? Es soll noch einmal ein Blick auf die lernpsychologischen Grundlagen der Arbeit geworfen werden.

⁵⁰DESI-Studie: „Deutsch-Englisch-Schülerleistungen-International“, eine Studie zur Erfassung der sprachlichen Leistungen in Deutsch und Englisch von Lernenden in der Sekundarstufe I in Deutschland (erfasst wurden alle Schulformen in allen Bundesländern).

Aktuelle instruktionspsychologische Sichtweisen zum Lernen und Lehren gelten natürlich nicht nur für den Sachunterricht, sondern sind für alle Fächer relevant. So sollte der Lernende in jedem Fach sein Wissen aktiv selbst konstruieren, zum Lernen motiviert werden oder auch Gelegenheiten zu sozialen Austauschprozessen haben. Entsprechend liegen Parallelen bei der theoretischen Konzeptualisierung fachlicher Unterrichtsmerkmale für verschiedene Fächer nahe. Die Umsetzung der einzelnen Merkmale im Unterricht sieht jedoch in jedem Fach anders aus. Ein Beispiel soll dies näher beleuchten: Zur Umsetzung sozial-kooperativen Lernens spielen gemeinsame Unterrichtsgespräche in jedem Unterricht eine Rolle, ob in der Mathematik, in den Naturwissenschaften oder im sprachbezogenen Unterricht. Der Fokus bei gemeinsamen Unterrichtsgesprächen ist jedoch in den einzelnen Fächern ein ganz anderer. Während es z.B. im Fremdsprachenunterricht in gemeinsamen Gesprächen primär um die Kommunikation in der Sprache, also um die Verwendung der Sprache geht (vgl. Helmke & Schrader, 2006b), steht im naturwissenschaftlichen (Sach-)Unterricht dahingegen das tiefere Verständnis der zu erlernenden Unterrichtsinhalte und der Austausch über Konzepte im Vordergrund. Entsprechend fällt die Operationalisierung eines solchen Konstrukts in jedem Fach anders aus. Der gewählte Bewertungshintergrund im Unterricht ist in jedem Fach ein ganz anderer. Es stellt sich also nicht die Frage, ob es sich bei bestimmten Konstrukten um fachspezifische Konstrukte handelt. Es stellt sich vielmehr die Frage, wann die Fachspezifität der Konstrukte beginnt – nämlich nicht mit der Konzeptualisierung, sondern mit der Operationalisierung der Konstrukte.

5.4 Einsatz des Instruments zur Erfassung von Zusammenhängen zwischen Lehrervorstellungen und -handlungen

Das im Rahmen dieser Arbeit vorgestellte hoch-inferente Videoanalyse-Instrument konnte zur Auswertung der in der Einleitung genannten und bei der Darstellung der methodischen Grundlagen dieser Arbeit (Kapitel 3) kurz vorgestellten BIQUA-Untersuchung zur Wirksamkeit von Fortbildungen im naturwissenschaftlichen Sachunterricht angewendet werden.

In der Studie richtete sich das Forschungsinteresse zunächst auf Vorstellungen von Lehrkräften zum Lehren und Lernen im naturwissenschaftlichen Sachunterricht und deren Modifikation durch Fortbildungen. Die Ergebnisse zeigen, dass die Vorstellungen der Lehrkräfte zum Lehren und Lernen durch Fortbildungsmaßnahmen verändert bzw. erweitert werden konnten (Möller et al., 2006) und dass die veränderten Lehrervorstellungen einen Einfluss auf Lernzuwächse der Lernenden haben (Kleickmann, 2008)⁵¹. Da davon auszugehen ist, dass Zusammenhänge zum Lehren und Lernen mit Lernzuwachsen

⁵¹Mittels Mehrebenenanalysen konnte T. Kleickmann (2008) nachweisen, dass höhere Leistungsfortschritte der Lernenden (konzeptuelles Verständnis zum Thema „Schwimmen und Sinken“) insbesondere durch Lehrervorstellungen erklärt werden, gemäß denen Lernen als Conceptual Change zu verstehen ist („Conceptual Change“) und Lernende über z.T. tief verwurzelte Vorstellungen über Naturphänomene verfügen, die Lernschwierigkeiten verursachen können („Schülervorstellungen“). Vorstellungen von Lehrkräften, denen zufolge Lernende am besten weitgehend selbstgesteuert ohne Hilfestellung und Strukturierungshilfen lernen („Laisser-faire“) oder denen zufolge Lernende am besten durch Erklärungen der Lehrkraft lernen („Transmission“), stehen in negativem Zusammenhang mit den Lernzuwachsen der Schüler.

der Schüler über unterrichtliche Prozesse vermittelt werden (Bromme, 1997), wurden auch Zusammenhänge der untersuchten Vorstellungen der Lehrkräfte mit deren unterrichtlichen Handlungsweisen vermutet. Somit wurde im Projekt der Frage nachgegangen, ob sich durch Fortbildung modifizierte Vorstellungen der Lehrkräfte zum Lehren und Lernen auch in entsprechenden unterrichtlichen Handlungsweisen widerspiegeln. Die Vorstellungen der Lehrkräfte wurden mittels eines Fragebogens⁵² und das Unterrichtshandeln der Lehrkräfte mit dem in dieser Arbeit vorgestellten hoch-inferenten Video-Rating nach der Intervention erhoben (vgl. Tabelle 3.1-1 zur Untersuchungsanlage der DFG-Studie).

Die Ergebnisse von bivariaten Korrelationen (n. Pearson 2-seitig, N=28) zeigen substantielle Zusammenhänge zwischen den Vorstellungen der Lehrkräfte zum Lehren und Lernen und deren unterrichtlichen Handlungsweisen in den erwarteten Richtungen. Besonders hoch korrelieren die Vorstellungen „Conceptual Change“ (naturwissenschaftliches Lernen ist das Verändern von Vorstellungen) und „Schülvorstellungen“ (Schüler haben Präkonzepte, die zu Lernschwierigkeiten führen können) mit den korrespondierenden Unterrichtshandlungen. Diese verständnisorientierten Vorstellungen der Lehrkräfte spiegeln sich also in Verhaltensweisen der Lehrkräfte wider (Vehmeier, Kleickmann & Möller, 2007). Zudem zeigen sich signifikante bis hoch-signifikante Korrelationen zwischen den Vorstellungen „Conceptual Change“ und „Schülvorstellungen“ mit der Lehrerhandlung „Laisser-faire“. Lehrkräfte, die Lernen als Conceptual Change verstehen und die Präkonzepte der Lernenden Bedeutung beimessen, zeigen anscheinend weniger Laisser-faire-Verhalten im Unterricht.

In einer weiteren Veröffentlichung (Kleickmann, Vehmeier & Möller, i. D.) konnten mittels Regressionsanalysen signifikante Effekte zwischen den Lehrervorstellungen „Conceptual Change“, „Schülvorstellungen“, „Praktizismus und Laisser-faire“ und den hoch-inferent eingeschätzten Unterrichtsmerkmalen zum Scaffolding (Hervorhebungen bedeutsamer Unterrichtsinhalte, Problematisierungshilfen und Sequenzierungsmaßnahmen) nachgewiesen werden. Je stärker die untersuchten Lehrkräfte davon ausgingen, dass naturwissenschaftliches Lernen als Konzeptwechsel zu verstehen ist und vorunterrichtliche Vorstellungen der Lernenden eine besondere Rolle dabei spielen, desto mehr setzten die Lehrpersonen Scaffolding-Maßnahmen ein. Diese Vorstellungen der Lehrkräfte zum Lehren und Lernen scheinen also den Einsatz von Strukturierungshilfen durch die Lehrkraft im Unterricht zu unterstützen. Bei der Lehrervorstellung, welche die Selbststeuerung des Lernens der Schüler überbetont und die Strukturierung und Unterstützung der Lernprozesse durch die Lehrkraft vernachlässigt bzw. ablehnt, fanden sich negative Effekte mit dem Einsatz von Scaffolding-Maßnahmen. Keine Zusammenhänge zeigen sich bei transmissiven Lehrervorstellung zum Lehren und Lernen mit den hier untersuchten Unterrichtsmerkmalen zum Scaffolding.

⁵²Der Fragebogen zur Erfassung der Vorstellungen von Lehrkräften zu Lehren und Lernen im naturwissenschaftlichen Sachunterricht wurde von T. Kleickmann entwickelt (Kleickmann, 2008). Wie bereits in 3.4.1.1 erläutert, erfolgte die Entwicklung der Beobachtungsdimensionen des Videoanalyse-Instruments weitgehend analog zu den Konstrukten des Fragebogens.

Die Ergebnisse deuten darauf hin, dass durch Fortbildungen aufgebaute Vorstellungen der Lehrkräfte zum Lehren und Lernen, die positive Effekte auf den kognitiven Lernerfolg der Schüler haben, für die Unterrichtsgestaltung relevant sind. Diese Ergebnisse sind vor allem vor dem Hintergrund von Bedeutung, als die analysierten Lehrerhandlungen unter der Annahme entwickelt worden sind, dass sie für die Umsetzung eines kognitiv anregenden Unterrichts von Bedeutung sind und für verständnisvolle Lernprozesse der Schüler als wichtig erachtet werden.

5.5 Ausblick

Zunächst ergeben sich Fragen und 'Wünsche' für weiterführende Arbeiten aus der methodischen Anlage der Arbeit. Auf diese wird in einem ersten Schritt eingegangen. Konsequenzen für weiterführende inhaltliche Überarbeitungen der Beobachtungsdimensionen des Instruments ergeben sich aus den berichteten Ergebnissen der Testanalysen, insbesondere aus den Ergebnissen der Expertenvalidierung. Welche Überarbeitungen für sinnvoll erachtet werden und wie diese konkret aussehen können, wird in einem zweiten Schritt näher beschrieben. Daran anschließend werden (offene) Forschungsfragen und Untersuchungsansätze skizziert, die interessante Ergebnisse vermuten lassen und denen mit einem Videoanalyse-Instrument, wie es in dieser Arbeit entwickelt wurde, nachgegangen werden könnte. Dieses Kapitel abschließend wird über Einsatzmöglichkeiten von Videoanalyse-Instrumenten über Forschungsfragen hinaus im Bereich der Lehreraus- und Lehrerfortbildung und in der Unterrichtsentwicklung nachgedacht.

Folgerungen aus der methodischen Anlage der Untersuchung

Die eingeschränkte Repräsentativität der Lehrerstichprobe zum zweiten Messzeitpunkt der Studie (Unterrichtsbeobachtung „Schwimmen und Sinken“) führt zu der Frage, ob ein Einsatz des Instruments bei einer Stichprobe nicht fortgebildeter Lehrer zum Thema „Schwimmen und Sinken“ erfolgreich wäre. Zum Zeitpunkt der Unterrichtsbeobachtung „Wasserkreislauf“ handelt es sich um nicht fortgebildete Lehrkräfte, da die Aufzeichnungen der Unterrichtsvideos vor der Intervention stattgefunden haben. Die Beobachtungsdimensionen bzw. die Items konnten bei den Videos zum Thema „Wasserkreislauf“ erfolgreich bewertet werden. Demnach kann generell davon ausgegangen werden, dass eine Anwendung der Konstrukte und Items auch beim Thema „Schwimmen und Sinken“ bei nicht fortgebildeten Lehrkräften funktionieren würde.

Darüber hinaus wäre es wünschenswert, das Instrument auch anhand einer größeren Lehrerstichprobe und anhand weiterer naturwissenschaftlicher Unterrichtsthemen zu testen. Interessant wäre es auch, wenn mehrere Videos pro Lehrer und Thema zur Verfügung stehen würden. Die 90-minütigen Unterrichtsvideos stellen nur sehr kleine Ausschnitte des Lehrerhandelns dar. Je mehr Videos zu einer Lehr-

person und einer Unterrichtseinheit vorliegen, desto aussagekräftigere Beurteilungen lassen sich über das Lehrerverhalten im Unterricht machen.

Thematisiert wurde die recht schmale Validierung des Instruments und die damit verbundene Notwendigkeit, umfangreichere Validierungsmaßnahmen in Folgeuntersuchungen durchzuführen. Angedeutet wurde in diesem Zusammenhang auch, welche Validierungsmaßnahmen als lohnend erscheinen. Sie sollen an dieser Stelle noch einmal skizziert werden: Insbesondere fehlt eine empirische Überprüfung der Konstruktvalidität, welche die Ergebnisse der Expertenvalidierung unterstützt (und nicht ersetzt). Faktorenanalysen – im Idealfall konfirmatorische, zumindest aber explorative – ließen sich schon mit einer größeren Teststärke realisieren.

Über die Dimensionalität der Skalen hinaus könnten Faktorenanalysen auch Aufschluss darüber geben, ob die Entscheidung, alle 14 Konstrukte von allen vier Ratern in einem Ratingdurchgang beurteilen zu lassen, sinnvoll war. Das aus verschiedenen Gründen gewählte Vorgehen⁵³ stellte nach Aussagen der Rater trotz der Vielzahl von zu bewertenden Merkmalen keine Überlastung dar und erwies sich als praktikabel. Sollte sich aber zeigen, dass sich die theoretisch angenommenen Konstrukt-Dimensionen nicht empirisch mittels Faktorenanalysen rekonstruieren lassen, so könnte dies u.a. ein Hinweis darauf sein, dass die Rater Schwierigkeiten hatten, die Konstrukte zu trennen und die Bewertung der Ratingdimensionen besser mit mehreren Durchgängen durchführen zu lassen oder aber die Konstrukte von verschiedenen Ratern beurteilen zu lassen, um die Rater entlasten.

Lohnend wären auch MTMM-Analysen. Bisher fehlte dazu eine weitere Methode, mittels der das Unterrichtshandeln der Lehrkräfte beurteilt werden konnte. Im Rahmen eines DFG-Kooperationsprojekts der Didaktik des Sachunterrichts (Möller, Kleickmann) mit der Forschergruppe 'Naturwissenschaftlicher Unterricht' der Universität Essen-Duisburg (Fischer)⁵⁴, ist auf der Basis des in dieser Arbeit vorgestellten Videoanalyse-Instruments ein Fragebogen zur Erfassung der von Schülern wahrgenommenen Verständnisorientierung des Unterrichts entwickelt worden. Die Lernenden geben eine Einschätzung, wie kognitiv anregend sie das Unterrichtshandeln ihrer Lehrkräfte wahrnehmen. Unter der Voraussetzung der Replikation der Ergebnisse der Testanalysen in Folgestudien ließe sich die konvergente Validität des Rating-Instruments mit den mittels Fragebogen erfassten Schülerwahrnehmungen zur Verständnisorientierung des Unterrichts beurteilen. Würden die Werte der hoch-inferent erfassten Unterrichtsbeobachtungen mit den mittels Fragebogen erfassten Schülerwahrnehmungen in den korrespondierenden Konstrukten korrelieren, so könnte dies als Hinweis auf konvergente Validität der jeweiligen Videoanalyse-Skalen interpretiert werden. Zur Messung der diskriminanten Validität der

⁵³Die Entscheidung, alle Konstrukte von allen Ratern in einem Durchgang beurteilen zu lassen, wurde zum einen aus ökonomischen (zeitlichen und finanziellen) Gründen getroffen, zum anderen in Anbetracht der Tatsache, dass es bei hoch-inferenten Ratings um ein 'ganzheitliche Betrachtung des Unterrichts geht (vgl. 3.4.1.1).

⁵⁴PLUS-Studie: „Professionswissen von Lehrkräften, naturwissenschaftlicher Unterricht und Zielerreichung im Übergang der Primar- zur Sekundarstufe“.

Skalen würden die Interkorrelationen der Skalen berechnet. Die Korrelationen sollten aus methodischer Sicht möglichst gering sein.

Folgerungen für weiterführende Überarbeitungen des Instruments

Im Rahmen der Diskussion über die Ergebnisse der Expertvalidierung wurde bereits herausgearbeitet, dass die Skalen „Ideen gemeinsam diskutieren“ und „Eigene Formulierungen“ eine Überarbeitung oder Neukonstruktion erfahren sollten.

Bei der Skala „Ideen gemeinsam diskutieren“ geht es bisher darum, zu erfassen, ob und in welchem Ausmaß die Lehrperson den Lernenden Gelegenheiten zum sozialen Austausch bietet, indem sie mit den Kindern gemeinsam über die Unterrichtsinhalte diskutiert und gemeinsame Diskussionen fördert. In der Theorie (Unterkapitel 1.4.2) wurde beschrieben, dass sprachliche Interaktionen die Lernenden in ihrer konzeptuellen Entwicklung unterstützen. Durch gemeinsame Diskussionen über eine Sache, führen sich die Lernenden ihr Verständnis dieser Sache vor Augen, überdenken ihre Konzepte und durchdringen sie besser (Palincsar, 1998, Duschl & Hamilton, 1998). Dieser Aspekt, die Bedeutung gemeinsamer Aushandlungsprozesse für das Verstehen der Kinder, wurde bisher vernachlässigt und sollte bei Überarbeitungen des Instruments berücksichtigt werden. Ebenso wäre es denkbar, zu fokussieren, wie die Lehrperson die Diskussion fördert bzw. was sie tut, um eine fruchtbare Diskussion zu gewährleisten. Dies geschieht im Rahmen des Dissertationsvorhabens von Anne Ewerhardy, einer Doktorandin am Seminar für Didaktik des Sachunterrichts. Sie entwickelt auf der Grundlage des vorliegenden Videoanalyse-Instruments ein Videorating zur Erfassung von Verständnisorientierung im naturwissenschaftlichen Sachunterricht. Mit dem Konstrukt „Kommunizieren und Aushandeln von Bedeutungen“ fokussiert sie verschiedene Aspekte gemeinsamer inhaltlicher Diskussionen: klare Formulierung von Lehreräußerungen, die Verwendung einer angemessenen Fachsprache, das Einfordern von klaren Schüleräußerungen in inhaltlichen Gesprächen durch die Lehrperson, das Sicherstellen des einander Zuhörens und des aufeinander Reagierens in gemeinsamen inhaltlichen Gesprächen (Ewerhardy, i. V.).

Eine Überarbeitung der Skala „Eigene Formulierungen“ könnte folgendermaßen aussehen: Möglich wäre es, das Konstrukt aufzulösen und das Konstrukt „Eigene Ideen zulassen“ um den Aspekt der eigenen Formulierungen zu ergänzen. Entsprechend würde das Items ei_e2 *„Die Lehrperson legt Wert darauf, dass die Kinder Sachverhalte in ihren eigenen Worten erklären“* der Skala „Eigene Ideen zulassen“ zugeordnet. Das Item ei_e3 *„Die Lehrperson greift die Formulierungen der Kinder auf und arbeitet damit weiter“* ließe sich gemäß den Ausführungen zu den Ergebnissen der Expertvalidierung dem Konstrukt zum Umgang mit Präkonzepten zuordnen. Das Items ei_e5 *„Die Lehrperson sorgt nicht dafür, dass 'Fachtermini' mit eigenen Worten erklärt werden“* könnte den Dimensionen zur Transmission von Wissensinhalten zugeordnet werden. Dieses Vorgehen hat in dieser Untersuchung nicht zu einer Verbesserung der internen Konsistenzen der Skalen geführt und wurde deshalb abgelehnt; in einem anderen Datensatz könnten die Ergebnisse ggf. anders ausfallen.

Denkbar wäre es aber auch, eine Ratingdimension zu entwickeln, mit der nicht nur erhoben wird, ob die Lehrperson den Kindern im Unterricht Raum gibt, Sachverhalte mit ihren eigenen Formulierungen auszudrücken, sondern mit der im Sinne Wagenscheins erfasst wird, wie im Unterricht mit Umgangssprache (Sprache der Kinder) und mit Fachsprache umgegangen wird. Da der Aspekt 'Umgang mit Umgangssprache und Fachsprache' eng mit gemeinsamen inhaltlichen Gesprächen verbunden ist, könnte dieser Aspekt in den Bereich „Ideen gemeinsam diskutieren“ integriert werden. Ein ähnliches Vorgehen wählt A. Ewerhardy bei der Entwicklung ihres Rating-Instruments. Sie fokussiert allerdings nur den Umgang mit Fachbegriffen (d.h. sie erfasst inwiefern die Lehrperson eine für die Lerngruppe angemessene Fachsprache einführt) und ordnet diesen Aspekt der Dimension „Kommunizieren und Aushandeln von Bedeutungen“ zu.

Um die im Bereich der Transmission von Wissen ursprünglich angedachten theoretischen Facetten „Erklärungen“ und „Denkwege ablehnen“ trennscharf abzubilden, müssten beide Transmissiv-Skalen eine Überarbeitung erfahren. Hier wären bessere theoretische Anbindungen und 'fachlichere', d.h. differenziertere Operationalisierungen erforderlich. Gleiches gilt für die Skala „Laisser-faire“. Bei letzterer sollte zudem die Itemanzahl erweitert werden. Diese Überarbeitungen werden aber aus den folgenden Gründen als nicht lohnend erachtet: Die in Abschnitt 5.4 dargestellten Ergebnisse der BIQUA-Studie zeigen auf, dass transmissive und Laisser-faire-Vorstellungen zum Lehren und Lernen von Lehrkräften wie erwartet negative Effekte auf die Lernzuwächse der Lernenden haben. Zudem zeigen sich weder Zusammenhänge zwischen transmissiven Lehrervorstellungen und -handlungen, noch haben transmissive Lehrervorstellungen einen positiven Effekt auf den Einsatz von Scaffolding-Maßnahmen durch die Lehrkräfte. Im Bereich Laisser-faire zeigen sich zwar Zusammenhänge zwischen den korrespondierenden Konstrukten in der erwarteten Richtung (Vehmeier, Kleickmann & Möller, 2007), zu beachten ist jedoch, dass es sich beim Bereich Laisser-faire im Grunde um eine Negation des Scaffolding handelt. Somit wird es als sinnvoll angesehen, auf eine Überarbeitung der genannten Konstrukte zu verzichten und den Fokus lieber auf die für einen kognitiv anregenden, auf Verstehen ausgerichteten Unterricht wesentlichen Bereiche zu konzentrieren. So sollten die Bereiche Conceptual Change und Scaffoldingmaßnahmen eine inhaltliche Ausweitung erfahren. In beiden Bereichen wurden bisher nur Teilaspekte berücksichtigt.

Conceptual Change. Aus dem Bereich Conceptual Change wurden bisher nur das Anregen kognitiver Konflikte (als Strategie zur Förderung von Wissensumstrukturierungen) aufgegriffen, was auch in der Beschreibung der Grundidee der Ratingdimension begründet wurde. Keine Berücksichtigung fanden Strategien zur Förderung der Wissensausdifferenzierung (Anknüpfungs- und Umdeutungsstrategien). Wünschenswert wäre es also, dass in Folgestudien auch solche Konzeptwechselstrategien erfasst werden. Um den Bereich Konzeptwechsel komplett abzudecken, sollte ebenso die Konzeptwechselbedingung 'Plausibilität' stärker herausgearbeitet werden. Gemäß den theoretischen Ausführungen in 1.4 ist es möglich, den Kindern neue Konzepte plausibel zu machen, indem Experimente ausgewählt werden, die das Potential haben, die Kinder von neuen Konzepten zu überzeugen. Im Instrument gibt es ein

Item dazu: „Die Lehrperson hat Versuche ausgewählt, die für die Kinder überzeugend sind“. Es wurde jedoch versäumt, den Bezug zum Konstrukt explizit hervorzuheben; in der Konstruktbeschreibung wurde nur das Anregen kognitiver Konflikte mittels objektiver Evidenz deutlich hervorgehoben. So war für die Teilnehmer der Expertenvaldierung nicht ganz klar, welchem Konstrukt das Item zugeordnet war.

Scaffolding. Im Sinne des Scaffolding wurden eine Maßnahme zur äußeren Strukturierung und zwei Maßnahmen zur diskursiven Unterstützung herausgearbeitet. Die äußere Strukturierung bezieht sich auf die Sequenzierung von Inhalten (hier: Gliederung der fachlichen Inhalte). Die diskursiven Unterstützungsformen beziehen sich auf das Hervorheben bedeutsamer Inhalte und Schüleräußerungen im Unterricht („Scaffolding – Hervorhebungen“) und auf Problematisierungshilfen und Provokationen (ursprünglich: „Scaffolding – Perturbieren“; nach der Expertenvaldierung: Gesamtkonstrukt „Konfrontation“). Nicht erfasst werden konnte der Einsatz materieller Denkhilfen (z.B. Experimente, Repräsentationsformen, Modelle). Auch das mit dem Scaffolding eng verbundene so genannte 'fading', das Zurücknehmen der Lernunterstützung bei zunehmender Befähigung der Lernenden konnte nicht erfasst werden (vgl. Kleickmann, Vehmeyer & Möller, i. D.), da die zur Verfügung stehenden Unterrichtsvideos lediglich Einstiegsstunden in die Unterrichtsthemen zeigten. Um das Zurücknehmen von Lernhilfen beobachten zu können, bräuchte man mehrere Stunden einer Unterrichtseinheit von einer Lehrperson.

Zuletzt soll als weiterführende Überarbeitung über eine 'Abspeckung' des Instruments nachgedacht werden. Dabei handelt es sich um eine inhaltliche Überarbeitung zu methodischen Zwecken. Ursprünglich beinhaltete das Instrument 14 Beobachtungsdimensionen mit Items, die als verhaltensnahe Indikatoren der Konstrukte direkt geratet wurden. Denkbar wäre es auch, die Konstrukte unter Einbezug der Items direkt zu raten. Die direkt beurteilten Konstrukte würden übergeordneten Dimensionen zugeordnet, im Idealfall den aus der Theorie gewonnenen Indikatoren für kognitiv anregendes Lehrerverhalten, die in Kapitel 1.4. beschrieben wurden. Ein Beispiel soll dieses Vorgehen näher beleuchten: Aus den sozial-konstruktivistischen Theorien wurde der Indikator '*Anleitung und Unterstützung der Lernenden – Scaffolding*' abgeleitet. Daraus wurden ursprünglich die Scaffolding-Skalen „Hervorhebungen“, „Perturbieren“⁵⁵, und „Sequenzierungen“ operationalisiert. Die Skalen würden gemäß des vorgeschlagenen Vorgehens unter Berücksichtigung der ursprünglichen Items direkt geratet. Die theoretisch angenommene übergeordnete Dimension '*Anleitung und Unterstützung der Lernenden - Scaffolding*' sollte sich in einer empirisch gefundenen Komponente wiederfinden lassen, d.h. in Faktorenanalysen sollte sich zeigen, dass die drei Scaffolding-Skalen auf einem gemeinsamen Faktor laden.

Dieses Vorgehen würde den Vorteil bieten, dass man ein besseres Verhältnis von Items zur Anzahl der Videos erhalten würde, so dass Faktorenanalysen auch bei kleineren Teststärken möglich wären. Demgegenüber steht der Nachteil, dass der bisher vorhandene Facettenreichtum des Instruments (das Erfas-

⁵⁵Diese Skala geht nach der Expertenvaldierung in das Gesamtkonstrukt „Konfrontation“ ein.

sen des Lehrerhandelns in so vielen verschiedenen Konstrukten) verloren gehen würde. Entsprechend dem Beispiel würde also nur noch die Anleitung und Unterstützung der Lernenden durch die Lehrkraft erfasst, nicht aber einzelne Aspekte der Unterstützung (z.B. das Fokussieren wichtiger Inhalte und Schüleräußerungen, das Provozieren, das Sequenzieren der Unterrichtsinhalte, ...). Die berichteten Interskalenkorrelationen sprechen für eine Reduktion der bisherigen Beobachtungsdimensionen (vgl. Tabelle 4.2-3 auf Seite 138).

Weiterführende Forschungsfragen

Die in dieser Arbeit konzeptualisierten und erfassten Lehrerhandlungen sind unter der Annahme entwickelt worden, dass sie das Potential haben, Lernende im naturwissenschaftlichen Sachunterricht kognitiv anzuregen. Um nun sagen zu können, ob besagte Lehrerhandlungen die Lernenden tatsächlich kognitiv anregen, müssen die entwickelten Verhaltensweisen der Lehrkräfte auf ihre Lernwirksamkeit hin untersucht werden, d.h. das Instrument müsste also zur Untersuchung der Relevanz der Lehrerhandlungen für Zielkriterien von Unterricht eingesetzt werden. Dabei sind sowohl leistungsbezogene als auch motivationale und persönlichkeitsbezogene Zielkriterien von Unterricht von gleicher Bedeutung. Erste (bisher noch nicht veröffentlichte) Analysen, die im Münsteraner BIQUA-Projekt durchgeführt worden sind, deuten darauf hin, dass verständnisorientierte Lehrerhandlungen, zu denen der Umgang mit Präkonzepten und der Einsatz von Scaffolding-Maßnahmen gehören, positiven Einfluss auf den kognitiven Lernerfolg der Grundschul Kinder haben. Anne EweAnne Ewerhardy konnte für das Gesamtscore ihrer hoch-inferent erfassten auf Verständnisorientierung ausgerichteten Unterrichtsmerkmale signifikante Zusammenhänge mit den Lernzuwächsen der Schüler beim Thema „Wasserkreislauf“ finden (i. V.). Somit verlaufen diese ersten Analysen zu den Wirkungen eines kognitiv anregenden, auf Verständnis ausgerichteten Sachunterrichts analog zu den in Kapitel 1.1 berichteten Studien aus der Mathematikdidaktik⁵⁶.

Wie bereits näher ausgeführt wurde, konnte das Videoanalyse-Instruments erfolgreich zur Analyse von Zusammenhängen zwischen Vorstellungen von Lehrkräften zum Lehren und Lernen und ihren Handlungen im Unterricht eingesetzt werden. Bei den untersuchten Lehrervorstellungen handelt es sich um Vorstellungen, die dem fachspezifisch-pädagogischen Wissen ('pedagogical content knowledge') der Lehrkräfte zuzuordnen sind. Ebenfalls von Interesse wäre es, neben dem fachspezifisch-pädagogischem Wissen auch andere Vorstellungen, wie z.B. das fachliche Wissen ('content knowledge') der Lehrkräfte zu berücksichtigen, um dann Zusammenhängen zwischen den unterrichtlichen Handlungen der Lehrkräfte und ihrem fachlichen Wissen nachgehen zu können. Einer derartigen Fragestellung wird m. E. in der genannten PLUS-Studie nachgegangen.

⁵⁶Die Ergebnisse von Untersuchungen in der Mathematikdidaktik zeigen, dass Leistungsunterschiede der Lernenden durch ein unterschiedliches Maß an kognitiv aktivierenden Unterrichtsmerkmalen erklärt werden kann. Ebenso zeigen die Ergebnisse, dass derartige Unterrichtsmerkmale auch einen positiven Einfluss auf das Interesse der Lernenden haben können (vgl. die Ausführungen in Kapitel 1.1).

In der BIQUA-Untersuchung wurden neben den hoch-inferent erfassten Lehrerhandlungen auch die Oberflächenstrukturen der Unterrichtsstunden kategorienbasiert erfasst⁵⁷. Somit bietet es sich an, Zusammenhängen zwischen der sichtbaren Struktur des Unterrichtsablaufs und den kognitiv anregenden unterrichtlichen Handlungsweisen der Lehrkräfte nachzugehen. Zeigen sich keine bedeutsamen Zusammenhänge – wovon ausgegangen wird – so würde dies die These unterstützen, dass fachliche Merkmale von Unterricht, die auf Verständnisorientierung ausgerichtet sind, losgelöst von oberflächlichen Gestaltungsmerkmalen von Unterricht zu sehen sind (siehe dazu die Ausführungen in 1.1.2).

In den obigen Ausführungen wurde erwähnt, dass auf der Basis des in dieser Arbeit vorgestellten Videoanalyse-Instruments im Rahmen der PLUS-Studie ein Fragebogen zur Erfassung der von den Schülern wahrgenommenen Verständnisorientierung im Unterricht entwickelt worden ist. Somit wäre es interessant zu untersuchen, ob sich Zusammenhänge zwischen dem durch externe Beobachter eingeschätzten Grad kognitiv anregender Lehrerhandlungen und dem durch Schüler wahrgenommenen Grad der Verständnisorientierung im Unterricht finden lassen.

Bei den bisher skizzierten Untersuchungsansätzen handelt es sich um variablenbezogene Ansätze. Interessant wäre es auch, mit Hilfe der vorliegenden Daten einen personenzentrierten Ansatz zu verfolgen, um die Lehrpersonen mit ihren spezifischen Merkmalskonfigurationen nicht aus dem Blick zu verlieren (vgl. Helmke, 2009). In einem ersten Schritt würden z.B. Lehrpersonen als 'Experten' ausgewählt, die in Klassen unterrichten, in denen die Lernenden besonders hohe Lernfortschritte erzielen. In einem zweiten Schritt würde im Sinne eines *best practice*-Ansatzes untersucht, was den Unterricht in diesen leistungsstarken Klassen auszeichnet bzw. welche Handlungsweisen der Lehrkräfte den Unterricht ausmachen. Ergebnisse von *best practice*-Studien eignen sich besonders in Verbindung mit videobasierten Unterrichtsbeispielen als Material für die Aus- und Fortbildung von Lehrkräften und für die Unterrichtsentwicklung, da sie in anschaulicher Weise Profile erfolgreicher Lehrpersonen verdeutlichen können (vgl. Helmke, 2009).

Schließlich ist es möglich, der Frage nachzugehen, wie kognitiv anregend sich die in dieser Arbeit untersuchten Lehrkräfte in ihrem Unterricht zum „Wasserkreislauf“ verhalten. Da es allerdings keine Vergleichswerte aus anderen Studien gibt, ist eine Bewertung der Ergebnisse nur innerhalb der Stichprobe möglich.

Zum Einsatz von Videoverfahren über Fragen der Unterrichtsforschung hinaus

Analysen von Unterricht bereichern nicht nur die Unterrichtsforschung, sie stellen auch für die Lehreraus- und Lehrerfortbildung sowie für die Unterrichtsentwicklung eine interessante und fruchtbare Perspektive dar – darüber herrscht in der Literatur Konsens (vgl. z.B. Helmke, 2009; Helmke & Schrader, 2006b, Seidel & Prenzel, 2007; Pauli & Reusser, 2006). Somit ist es eine Überlegung wert, Vi-

⁵⁷Das Kategoriensystem erfasst u.a. Unterrichtsphasen und öffentliche und individualisierte Interaktionsarten und wurde von B. Gais (2009) entwickelt.

deanalyse-Instrumente neben dem Einsatz zur Hypothesenprüfung auch für die Aus- und Fortbildung von Lehrkräften und für die Entwicklung von Unterricht zu nutzen.

In der Lehreraus- und Lehrerfortbildung können Studierenden und Lehrenden mittels (computerbasierter) Analysen von Unterrichtsvideos theoretische Konzepte konkretisiert und empirisch fundiertes Wissen anhand von praktischen Beispielen vermitteln werden⁵⁸. Was die Studierenden in ihrer Ausbildung und Lehrer in Fortbildungen praktisch erfahren, hat größere Chancen in der Praxis eingesetzt zu werden (vgl. Helmke, 2009).

Für eine anspruchsvolle Unterrichtsentwicklung bietet gemäß Helmke und Schrader (2006b) insbesondere die videografierte Unterrichtsreflexion hervorragendes Potential, welches nach Ansicht der Autoren bisher nur unzureichend genutzt wird. Ergebnisse von Studien zu den Wirkbedingungen von Unterrichtsvideos bei praktizierenden Lehrkräften haben gezeigt, dass der Lerngewinn für die Lehrkräfte u.a. besonders hoch ist, wenn die Betrachtung der Unterrichtsvideos mit einem strukturierten (und nicht offenen) Beobachtungsauftrag verknüpft war (Seidel & Prenzel, 2007; Helmke, 2009). Instrumente zur Beschreibung und Analyse von Unterricht spielen somit eine wichtige Rolle für eine sinnvolle, auf Verbesserung von Unterricht angelegte Unterrichtsreflexion. Solche Instrumente dienen dabei nicht nur dazu, die Erkenntnisse aus der Unterrichtsforschung der Praxis zugänglich zu machen, sie bieten sich auch als besonders hilfreiches Mittel an, den eigenen Unterricht zu evaluieren, sich über den Unterricht auszutauschen, das (eigene) unterrichtliche Handeln in Bezug auf Unterrichtsqualitätsstandards zu reflektieren und sich selbst und sich gegenseitig besser beobachten zu lernen (vgl. Helmke & Schrader, 2006b; Pauli & Reusser, 2006). Letzteres ist insofern von Bedeutung, als auch die Fähigkeit, den eigenen Unterricht angemessen diagnostizieren zu können, dazu beiträgt, die Qualität des eigenen Unterrichts zu verbessern (Helmke & Schrader, 2006b).

Schlussgedanken

Ausgehend von der Diskussion über den Einfluss von Lehrkräften und ihrer Unterrichtsgestaltung auf Zielkriterien von Unterricht wurden in der Einleitung fachspezifische Lehrerverhaltensweisen fokussiert. Die weiteren Ausführungen zeigen, dass es (fachliche) Merkmale guten Unterrichts gibt, die durch die Lehrkraft beeinflusst werden können. Auch wenn der erfolgreiche Verlauf von Unterricht selbstverständlich nicht nur vom Lehrer abhängt – vor allem auch die Schüler tragen zum Gelingen des Unterrichts maßgeblich bei (!) – so bleibt die Planung und Durchführung guten Unterrichts, wie Helmke und Schrader (2006b) sagen, doch das 'Kerngeschäft' des Lehrers. Aus diesem Grund wird der Lehrer mit seinem Professionswissen und Können auch weiterhin im Zentrum des Interesses von Öff-

⁵⁸An der Universität Landau wurde dies im Rahmen eines interdisziplinären Seminars in der Lehrerbildung von den Professoren A. Helmke und D. Gohrbrandt umgesetzt (Helmke, 2009). Für Studierende des Lehramts an der WW Münster werden computerbasierte Videoanalysen z.B. im Seminar 'Klassenführung und Unterrichtserfolg in der Primarstufe' (Leitung: Prof. Dr. M. Holodyski) am Psychologischen Institut V vollzogen, an dem die Autorin dieser Arbeit im Sommersemester 2007 im Rahmen ihres Promotionsstudium teilnahm. Die Art dieses Seminars wurde von den Studierenden als sehr fruchtbar empfunden und stellt eine gute Vorbereitung für die Referandarzeit dar.

fentlichkeit, Bildungspolitik und Unterrichtsforschung stehen. Die Verbesserung von Unterricht wird auch in Zukunft eine zentrale Forderung bleiben. Diesbezüglich lässt sich aus der Sicht des Lehrers sagen: „Fordern Bildungspolitiker und -forscher, dass Lehrer ihren Unterricht verbessern, so müssen sie auch dafür sorgen, dass eine Verbesserung von Unterricht gelingen kann!“. Dies verlangt – das darf die Autorin der Arbeit als Lehrerin im Schuldienst sagen – sinnvolle, praktikable und 'praxisfreundliche' Unterstützungsformen. Müssen die Lehrkräfte in mühevoller Arbeit z.B. selbst Instrumente zur Beobachtung von Unterricht oder zur Diagnose von Schülerverhalten und -leistungen erarbeiten, so kann eine vernünftige Unterrichtsentwicklung nicht gelingen. Umgekehrt verlangt es von Lehrern, dass sie sich aktiv um Hilfen bemühen⁵⁹ und diese Hilfen auch annehmen. Dass dieses Unterfangen ein äußerst anspruchsvolles und schwieriges ist, das viel Anstrengungsbereitschaft erfordert, sei nicht weiter ausgeführt. Nicht zuletzt verlangt die Verbesserung von Unterricht die Zusammenarbeit von den sich oft 'skeptisch' gegenüberstehenden Parteien. Lehrer müssen sich aufgrund ihres relativen Einflusses um einen qualitativ hochwertigen Unterricht bemühen. Ebenso verdienen sie es, dass man sich um sie 'kümmert'. Auch aus diesem zweiten Blickwinkel lassen sich die eingangs zitierten Worte „Auf die Lehrer kommt es an!“ verstehen. Mit diesen Worten begann die Arbeit und mit diesen Worten schließt die Arbeit.

⁵⁹Das Buch „Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts“ von A. Helmke (2009) bietet viele Anregungen für die Hand des Lehrers.

Literaturverzeichnis

- Adhami, M. (2002). Cognitive acceleration in mathematics education in years 5 and 6: Problems and challenges. In M. Shayer, P. S. Adey (Eds.), *Learning Intelligence: Cognitive Acceleration across the Curriculum from 5 to 15 years* (pp. 98-117). Buckingham: Open University Press.
- Aebli, H. (1969). *Grundformen des Lehrens* (6. Aufl.). Stuttgart: Klett.
- Aebli, H. (1981). *Denken: das Ordnen des Tuns*. Bd. 2: Denkprozesse. Stuttgart: Klett-Cotta.
- Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher*, 25, 5-11.
- Attelander, P. (2000). *Methoden der empirischen Sozialforschung* (9. überarb. Aufl.). Berlin, New York: de Gruyter.
- Ausubel, D. P., Novak, J. D., & Hanesian, H. (1981). Psychologische und pädagogische Grenzen des entdeckenden Lernens. In H. Neber (Hrsg.), *Entdeckendes Lernen* (3. völl. überarb. Aufl.). Weinheim: Beltz.
- Babu, S., & Mendro, R. (2003). *Teacher accountability: HLM-Based Teacher Effectiveness Indices in the Investigation of Teacher Effects on Student Achievement in a State Assessment Program*. Paper presented at the annual meeting of the American Educational Research Association, Chicago. URL:http://www.dallasisd.org/inside_disd/depts/evalacct/research/articles.htm.
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., Hosenfeld, I., Köller, O., & Neubrand, J. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske & Budrich.
- Baumert, J., & Köller, O. (2000). Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III. Dritte internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band II: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe* (S. 271-315). Opladen: Leske & Budrich.
- Baumert, J., Kunter, M., Brunner, M., Krauss, S. Blum, W., & Neubrand, M. (2004). Mathematikunterricht aus Sicht der PISA-Schülerinnen und -Schüler und ihrer Lehrkräfte. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand u.a. (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 314-354). Münster: Waxmann.
- Beinbrech, C. (2003). *Problemlösen im Sachunterricht der Grundschule. Eine empirische Studie zur Gestaltung von Lehr-Lernumgebungen im Hinblick auf die Förderung des Problemlöseverhaltens im Sachunterricht*. Dissertationsschrift. Westfälische Wilhelms-Universität Münster. <http://miami.uni-muenster.de/servlets/DocumentServlet?id=1326>
- Beinbrech, C. (2007). Problemorientierter Sachunterricht. In J. Kahlert, M. Fölling-Albers, M. Götz, A. Hartinger, D. von Reeken & S. Wittkowske (Hrsg.), *Handbuch Didaktik des Sachunterrichts* (S. 406-410). Bad Heilbrunn: Klinkhardt.
- Berg, H. C. (1995). Genetische Methode, In H- C. Berg & T. Schulze (Hrsg.), *Lehrkunstwerkstatt II. Lehrkunst und Schulvielfalt* (S. 349-360). Neuwied: Luchterhand.
- Biester, W., & Soostmeyer, M. (1997). Das Unterrichtsgespräch im Problemlösungsprozess. In E. Glumpler & S. Luchtenberg (Hrsg.), *Jahrbuch Grundschulforschung, Bd. 1* (S. 212-220). Weinheim: Deutscher Studien Verlag.

- Bliss, J. (1996). Piaget und Vygotsky: Ihre Bedeutung für das Lehren und Lernen der Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften*, 2(3), 3-16.
- Bliss, J., Askew, M., & Macrae, S. (1996). Effective teaching and learning: scaffolding revisited. *Oxford Review of Education*, 22(1), 37-61.
- Blum, W. (1998): Ursachen der TIMSS-Ergebnisse und Ansätze für Veränderungen des Mathematikunterrichts. In W. Blum & M. Neubrand (Hrsg.), *TIMSS und der Mathematikunterricht* (S. 11-16). Hannover: Schroedel.
- Blum, W., Drüke-Noe, C., Köller, O., & Hartung, R. (2006). *Praxisbuch Bildungsstandards Mathematik: konkret*. Berlin: Cornelsen Scriptor.
- Blumberg, E. (2008). *Multikriteriale Zielerreichung im naturwissenschaftsbezogenen Sachunterricht der Grundschule – Eine Studie zum Einfluss von Strukturierung in schülerorientierten Lehr-Lernumgebungen auf das Erreichen kognitiver, motivationaler und selbstbezogener Zielsetzungen*. Dissertationsschrift. Westfälische Wilhelms-Universität Münster. <http://miami.uni-muenster.de/servlets/DocumentServlet?id=4618>
- Bönsch, M. (1985). Problemorientierter Unterricht. *Neue Deutsche Schule*, 5, 22-25.
- Bos, W., & Tarnai, C. (1999). Content analysis in empirical social research. *International Journal of Educational Research*, 31, 659-671.
- Bortz, J. (2005). *Statistik für Sozialwissenschaftler* (6. vollst. überarb. u. aktual. Aufl.). Berlin: Springer.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. überarb. Aufl.). Berlin: Springer.
- Bransford, J. D., Sherwood, R. D., Hasselbring, T. S. Kinzer, CH. K., & Williams, S. M. (1990). Anchored Instruction: Why we need it and how technology can help. In D. Nix & R. J. Spiro (Eds.), *Cognition, education, and multimedia: Exploring ideas in high technology* (pp. 115-141). Hillsdale, NJ: Erlbaum.
- Brennan, R. L. (2001). *Generalizability Theory*. New York: Springer.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Psychologie des Unterrichts und der Schule* (Bd. 3, S. 177-212). Göttingen: Hogrefe.
- Brophy, J. E. (2004). *Motivating students to learn* (2nd ed.). Mahwah, NJ: Erlbaum..
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32-41.
- Brown, A. L., & Palincsar, A. S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 293-451). Hillsdale, NJ: Erlbaum.
- Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, 31, 21-32.
- Bruner, J. S. (1981). Der Akt der Entdeckung. In H. Neber (Hrsg.), *Entdeckendes Lernen* (3. völl. überarb. Aufl., S. 15-29). Weinheim: Beltz.
- Brunner, M., Kunter, M., Krauss, S., Klusmann, U., Baumert, J., Blum, W., et al. (2006). Die professionelle Kompetenz von Mathematiklehrkräften: Konzeptualisierung, Erfassung und Bedeutung für den Unterricht. Eine Zwischenbilanz des COAKTIV-Projekts. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 54-82). Münster: Waxmann.

- Bühl, A., & Zöfel, P. (2005). *SPSS 12. Einführung in die moderne Datenanalyse unter Windows* (9. überarb. u. erweit. Aufl.). München: Pearson Studium.
- Bühner, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2. Aufl.). München: Pearson Studium.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 103, 276-279.
- Caramaza, A., McCloskey, M., & Green, B. (1981). Naive beliefs in 'sophisticated' subjects: Misconceptions about the trajectories of objects. *Cognition*, 9, 117-123.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: The MIT Press.
- 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: Erlbaum.
- Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science education. *Review of Educational Research*, 63, 1-19.
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*. Münster: Waxmann.
- Clausen, M., Reusser, K., & Klieme, E. (2003). Unterrichtsqualität auf der Basis hoch-inferenter Unterrichtsbeurteilungen. Ein Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, 31(2), 122-141.
- Clement, J. (2000). Model based learning as a key research area for science education. *International Journal of Science Education*, 22(9), 1041-1053.
- Cobb, P., Wood, T., & Yackel, E. (1993). Discourse mathematical thinking, and classroom practice. In E. A. Forman, N. Minick & C. Addison Stone (Eds.), *Contexts for learning. Sociocultural dynamics in children's development* (pp. 91-119). New York: Oxford University Press.
- Cognition and Technology Group at Vanderbilt (1993). Designing learning environments that support thinking: The Jasper series as a case study. In T. M. Duffy, L. Lowyck, D. H. Jonassen & T. M. Welsh (Eds.), *Designing environments for constructive learning* (pp. 9-36). Berlin: Springer.
- Cognition and Technology Group at Vanderbilt (1994). Multimedia environments for developing literacy in at-risk students. In B. Means (Ed.), *Technology and education reform: The reality behind the promise* (pp. 23-56). San Francisco: Jossey-Bass.
- Cognition and Technology Group at Vanderbilt (1997). *The Jasper project: Lessons in curriculum, instruction, assessment, and professional development*. London: Erlbaum.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction. Essays in honour of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Erlbaum.
- Cronbach, L. J. (1951) Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: the theory of generalizability for scores and profiles*. New York: Wiley.
- Davis, E. A., & Linn, M. C. (2000). Scaffolding students' knowledge integration: prompts for reflection in KIE. *International Journal of Science Education*, 22(8), 819-837.

- Davis, E., & Miyake, N. (2004). Explorations of scaffolding in complex classroom systems. *The Journal of the Learning Sciences*, 13(3), 265-272.
- Deci, E., & Ryan, R. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39(2), 223-238.
- Deutscher Bildungsrat (1970). *Strukturplan für das Bildungswesen. Empfehlungen der Bildungskommission*. Stuttgart: Klett.
- diSessa, A. (1988). Knowledge in pieces. In G. Forman & P. Pulfall (Eds.), *Constructivism in the computer age* (pp. 49-70). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Doll, J., & Prenzel, M. (2004). Das DFG-Schwerpunktprogramm „Bildungsqualität von Schule (BIQUA): Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen“. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule. Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategie der Qualitätsverbesserung* (S. 9-26). Münster: Waxmann.
- Bortz, J. (2005) *Statistik für Human- und Sozialwissenschaftler* (6. Aufl.). Heidelberg: Springer.
- Drollinger-Vetter, B., & Lipowsky, F. (2006). Fachdidaktische Qualität der Theoriephasen. In I. Hugener, C. Pauli & K. Reusser. Videoanalysen. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Materialien zur Bildungsforschung: Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“* (Bd. 15, S. 189-205). Frankfurt am Main: GFPP.
- Dubs, R. (1995). Konstruktivismus: Einige Überlegungen aus der Unterrichtsgestaltung. *Zeitschrift für Pädagogik*, 41(6), 889-903.
- Duit, R. (1993). Alltagsvorstellungen berücksichtigen! *PdN-Ph.*, 42(6), 7-11.
- Duit, R. (1995). Zur Rolle der konstruktivistischen Sichtweise in der naturwissenschaftsdidaktischen Lehr- und Lernforschung. *Zeitschrift für Pädagogik*, 41, 905-923.
- Duit, R. (1996). Lernen als Konzeptwechsel im naturwissenschaftlichen Unterricht. In R. Duit & C. Rhöneck (Hrsg.), *Lernen in den Naturwissenschaften* (S. 145-162). Kiel: IPN.
- Duit, R. (1997). Alltagsvorstellungen und Konzeptwechsel im naturwissenschaftlichen Unterricht – Forschungsstand und Perspektiven für den Sachunterricht in der Primarstufe. In W. Köhnlein, B. Marquard-Mau & H. Schreier (Hrsg.), *Kinder auf dem Weg zum Verstehen der Welt* (S. 233-245). Bad Heilbrunn: Klinkhardt.
- Duit, R., & Häußler, P. (1997). Physik und andere naturwissenschaftliche Lernbereiche. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Psychologie des Unterrichts und der Schule* (Bd. 3, S. 427-460). Göttingen: Hogrefe.
- Duit, R. (1999). Conceptual change approaches in science education. In W. Schnotz, S. Vosniadou & M. Carretero (Eds.), *New perspectives on conceptual change* (pp. 263-314). Oxford: Elsevier.
- Duit, R., & Teagust, D.F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education*, 25(6), 671-688.
- Duschl, R. A., & Hamilton, R. J. (1998). Conceptual change in science and in the learning of science. In B. J. Fraser & K. G. Tobin (Eds.), *International handbook of science education* (pp. 1047-1065). Dordrecht: Kluwer.

- Edelmann, W. (2000). *Lernpsychologie* (6. vollst. überarb. Aufl.). Weinheim: Psychologie Verlags Union.
- Edwards, A., & Mercer, N. (1987). *Common knowledge: the development of understanding in the classroom*. London: Routledge.
- Einsiedler, W. (1976). *Lehrstrategie und Lernerfolg. Eine Untersuchung zur lehrziel- und schülerorientierten Unterrichtsforschung*. Weinheim: Beltz.
- Einsiedler, W. (1985). Problemlösen als Ziel und Methode des Sachunterrichts in der Grundschule. In W. Einsiedler & R. Rabenstein (Hrsg.), *Grundlegendes Lernen im Sachunterricht* (S. 126-146). Bad-Heilbrunn: Klinkhardt.
- Einsiedler, W. (1994). Aufgreifen von Problemen – Gespräche über Probleme – Problemorientierter Unterricht in der Grundschule. In L. Duncker & W. Popp (Hrsg.), *Kind und Sache: Zur pädagogischen Grundlegung des Sachunterrichts* (S. 199-212). Weinheim: Juventa Verlag.
- Einsiedler, W., & Treinies, G. (1997). Effects of teaching methods, class effects, and patterns of cognitive teacher-pupil interaction in an experimental study in primary school classes. *School Effectiveness and Improvement*, 8(3), 327-353.
- Einsiedler, W. (2000). Der Sachunterricht der Grundschule als Voraussetzung für Allgemeinbildung. In W. Hinrichs & H. F. Bauer (Hrsg.), *Zur Konzeption des Sachunterrichts* (S. 68-80). Donauwörth: Auer Verlag.
- Einsiedler, W. (2005). Lehr- Lern-Konzepte für die Grundschule. In W. Einsiedler, M. Götz, H. Hacker, J. Kahlert, R. W. Keck, E. Sandfuchs (Hrsg.), *Handbuch Grundschulpädagogik und Grundschuldidaktik* (2. überarb. Aufl., S. 373-385). Bad Heilbrunn: Klinkhardt.
- Einsiedler, W. (2007). Methoden und Prinzipien des Sachunterrichts. In J. Kahlert, M. Fölling-Albers, M. Götz, A. Hartinger, D. von Reeken & S. Wittkowske (Hrsg.), *Handbuch Didaktik des Sachunterrichts* (S. 389-400). Bad Heilbrunn: Klinkhardt.
- Ewerhardy, A. (in Vorbereitung). *Zusammenhänge zwischen Verständnisorientierung von naturwissenschaftsbezogenem Sachunterricht und Fortschritten im Verständnis naturwissenschaftlicher Konzepte bei Lernenden der Grundschule*. Dissertationsschrift. Westfälische Wilhelms-Universität Münster.
- Fend, H. (1981). *Theorie der Schule*. München: Urban.
- Foster, J. (1993). *Entdeckendes Lernen in der Grundschule*. München: Ehrenwirth.
- Fraser, B., & Walberg, H. J. (1981). Psychosocial learning environment in science classrooms: A review of research. *Studies in Science Evaluation*, 8, 67-92.
- Fraser, B., Walberg, H., Welch, W., & Hattie, J. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, 11, 147-252.
- Früh, W. (1991). *Inhaltsanalyse. Theorie und Praxis* (3. überarb. Aufl.). Leipzig: Ölschläger.
- Gais, B. (2009). *On the relationship between a university-based teacher preparation program for primary science and teachers' instructional practice*. Dissertationsschrift. Westfälische Wilhelms-Universität Münster. <http://miami.uni-muenster.de/servlets/DocumentServlet?id=5037>
- Gais, B., Kleickmann, T., Möller, K., Oerke, B., & Wedemeier, K. (2004). *Abschlussbericht zum Projekt „Zum Einfluss von Lehrerausbildungskonzepten auf Lehrerkognition, Lehrerhandeln und Unterrichtserfolg im naturwissenschaftlichen Lernfeld der Grundschule“ für die Landesarbeitsgemeinschaft Wirksamkeit unserer Bildungssysteme*. Unveröffentlichtes Manuskript.
- Gerstenmaier, J., & Mandl, H. (1995). Wissenserwerb und konstruktivistischer Perspektive. *Zeitschrift für Pädagogik*, 41(6), 867-887.

- Gesellschaft für Didaktik des Sachunterrichts (2002). *Perspektivrahmen Sachunterricht*. Bad Heilbrunn: Klinkhardt.
- Gilbert, J. K., Watts, D. M., & Osborne, R. J. (1985). Eliciting students' views using an interview-about-instances technique. In L. H. T. West & A. L. Pines (Eds.), *Pines, cognitive structure and conceptual change* (pp. 1-7). Orlando: Academic Press.
- Goldhaber, D. D., & Anthony, E. (2004). *Can teacher quality be effectively assessed?* URL:<http://www.urban.org/url.cfm?ID=410958>.
- Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. C. Berliner, R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 15-46). New York.
- Gräber, W., Nentwig, P., Koballa, T., & Evans, R. (2002). *Scientific Literacy: Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung*. Opladen: Leske & Budrich.
- Greeno, J. G., Smith, D. R., & Moore, J. C. (1993). Transfer of situated learning. In D. K. Dettermann & T. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 99-167). Norwood: Ablex.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen. Schüler als Quellen der Unterrichtsbeschreibung*. Münster: Waxmann.
- Hameyer, U., Dudek, H., Friis, H., & Strenge, B. (1987). *AKTIV. Ein Lernangebot für den naturwissenschaftlichen Sachunterricht und naturwissenschaftliche Fächer (3. bis 7. Klasse)*. Kiel: Schmidt & Klaunig.
- Hagstedt, H., & Spreckelsen, K. (1986). Wie Kinder physikalischen Phänomenen begegnen. *Sachunterricht und Mathematik in der Primarstufe*, 14(9), 318-323.
- Hardy, I., Jonen, A., Möller, K., & Stern, E. (2004). Die Integration von Repräsentationsformen in den Sachunterricht der Grundschule. In J. Doll, & M. Prenzel (Hrsg.), *Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategie der Unterrichtsverbesserung* (S. 267-283). Münster: Waxmann.
- Hardy, I. (2007). *Die Bedeutung der kognitiven Strukturierung von Lernumgebungen für die Förderung mathematisch-naturwissenschaftlicher Verständnisprozesse*. Publikationsbasierte Habilitationsschrift, FU Berlin.
- Harrison A., & Tregust, D. (2000). A typology of school science models. *International Journal of Science Education*, 22, (9), 1011-1026.
- Hartinger, A., & Mörtl-Hafizović, D. (2004). Situiertes Lernen – ein aktuelles Thema der Lehr-Lernforschung und seine Relevanz für den Sachunterricht. In M. Hempel (Hrsg.), *Sich bilden im Sachunterricht* (S. 61-78). Bad Heilbrunn: Klinkhardt.
- Hartinger, A. (2005). Entdeckendes Lernen. In W. Einsiedler, M. Götz, H. Hacker, J. Kahlert, R. W. Keck & U. Sandfuchs (Hrsg.), *Handbuch Grundschulpädagogik und Grundschuldidaktik* (2. überarb. Aufl., S. 386-395). Bad Heilbrunn: Klinkhardt.
- Hashweh, M. Z. (1986). Towards an explanation of conceptual change. *European Journal of Science Education*, 8(3), 229-249.
- Hattie, J. (2003). *Teachers make difference. What is the reseach evidence?* Auckland: Australian Council for Educational Research.
- Häußler, P., Bündler, W., Duit, R., Gräber, W., & Mayer, J. (1998). *Naturwissenschaftsdidaktische Forschung – Perspektiven für die Unterrichtspraxis*. Kiel: IPN.

- Heinze, A. (2004). Zum Umgang mit Fehlern im Unterrichtsgespräch der Sekundarstufe I. *Journal für Mathematik-Didaktik*, 4(3), 221-244.
- Helmke, A., & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Psychologie des Unterrichts und der Schule* (Bd. 3, S. 71-176). Göttingen: Hogrefe.
- Helmke, A., Hosenfeld, I., Schrader, F.-W. (2002). Unterricht, Mathematikleistung und Lernmotivation. In A. Helmke & R. S. Jäger (Hrsg.), *Das Studie MARKUS – Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext* (S. 413-480). Landau: Verlag Empirische Pädagogik.
- Helmke, A. (2004). *Unterrichtsqualität – erfassen, bewerten, verbessern*. 3. Auflage. Seelze: Kallmeyersche.
- Helmke, A. (2006). Was wissen wir über guten Unterricht? *Pädagogik*, 2, 42-45.
- Helmke, A., & Schrader, F.-W. (2006a). Determinanten der Schulleistung. In D.H. Rost. (Hrsg.), *Handwörterbuch der pädagogischen Psychologie* (3. überarb. u. erweit. Aufl., S. 83-94). Weinheim: Beltz.
- Helmke, A., & Schrader, F.-W. (2006b). Lehrerprofessionalität und Unterrichtsqualität. Den eigenen Unterricht reflektieren und beurteilen. *Schulmagazin 5 bis 10*, 9, 5-12.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Kallmeyersche.
- Hewson, P. W. (1982). A case study of conceptual change in special relativity: The influence of prior knowledge in learning. *European Journal of Science Education*, 4, 61-78.
- Hewson, P. W., & Hewson, M. G. A'B. (1992). The status of students' conceptions. In R. Duit, F. Goldberg & H. Niedderer (Eds.), *Research in physics learning: theoretical issues and empirical studies; proceedings of an international workshop* (pp. 59-73). Kiel: IPN.
- Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research*, 66, 549-571.
- Hidi, S., & Anderson, V. (1992). Situational interest and its impact on reading and expository writing. In K. A. Renninger, S. Hidi & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 215-238). Hillsdale, NJ: Erlbaum.
- Hogan, K., & Pressley, M. (1997). Scaffolding scientific competencies within classroom communities of inquiry. In K. Hogan & M. Pressley (Eds.), *Scaffolding student learning: Instructional approaches and issues* (pp. 74-107). Louiseville, Quebec: Brookline Books.
- Holodynski, M. (1998). Klassenführung – eine notwendige Aufgabe erfolgreichen Unterrichtens. *Grundschule*, 9, 22-24.
- Hoppe-Graff, S. (2000). Spielen und Fernsehen: Phantasietätigkeiten des Kindes. In S. Hoppe-Graff & R. Oerter (Hrsg.), *Spielen und Fernsehen. Über Zusammenhänge von Spiel und Medien in der Welt des Kindes* (S. 179-189). Weinheim: Juventa.
- Hugener, I. (2006). Überblick über die Beobachtungsinstrumente. In I. Hugener, C. Pauli & K. Reusser. Videoanalysen. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Materialien zur Bildungsforschung: Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“* (Bd. 15, S. 45-54). Frankfurt am Main: GFPPF.

- Hugener, I., Rakoczy, K., Pauli, C., & Reusser, K. (2006). Videobasierte Unterrichtsforschung: Integration verschiedener Methoden der Videoanalyse für eine differenzierte Sicht auf Lehr-Lernprozesse. In S. Rahm, I. Mammes & M. Schratz (Hrsg.), *Schulpädagogische Forschung: Unterrichtsforschung - Perspektiven innovativer Ansätze* (Bd. 1, S. 41-53). Innsbruck: Studienverlag
- Hugener, I., Pauli, C., & Reusser, K. (2006). Videoanalysen. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Materialien zur Bildungsforschung: Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“* (Bd. 15). Frankfurt am Main: GfPF.
- Hugener, I., Pauli, C., & Reusser, K. (2007). Inszenierungsmuster, kognitive Aktivierung und Leistung im Mathematikunterricht. In D. Lemmermöhle, M. Rothgangel, S. Bögeholz, M. Hasselhorn & R. Watermann (Hrsg.), *Professionell Lehren. Erfolgreich Lernen* (S. 109-122). Münster: Waxmann
- Hugener, I. (2008). *Inszenierungsmuster im Unterricht und Lernqualität. Sichtstrukturen schweizerischen und deutschen Mathematikunterrichts in Beziehung zu Schülerwahrnehmung und Lernleistung – eine Videoanalyse*. Münster: Waxmann.
- Hugener, I., Pauli, C., Reusser, K., Lipowsky, F., Rakoczy, K., & Klieme, E. (2009). Teaching patterns and learning quality in Swiss and German mathematics lessons. *Learning and Instruction*, 19, 66-78.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence; An essay on the construction of formal operational structures*. New York: Basic Books.
- Jacobs, J. K., Kawanaka, T., & Stigler, J. W. (1999). Integrating qualitative and quantitative approaches to the analysis of video data on classroom teaching. *International Journal of Educational Research*, 31, 717-724.
- Jacobs, J., Garnier, K., Gallimore, R., Hollingsworth, J., Givving, K. B., Rust, K., Kawanaka, T., u.a. (2003), *Third International Mathematics and Science Study 1999 Video Study Technical Report. Vol. 1: Mathematics*. Washington D.C.: National Center for Educational Statistics, US. Department of Education.
- Jonen, A., & Möller, K. (2005). *Die KiNT-Boxen – Kinder lernen Naturwissenschaft und Technik. Klasse(n)kisten für den Sachunterricht. Paket 1: Schwimmen und Sinken*. Essen: Spectra.
- Jung, W. (1986). Alltagsvorstellungen und das Lernen von Physik und Chemie. *Naturwissenschaften im Unterricht. Physik/Chemie*, 34(13), 2-6.
- Klauer, K. J. (2006). Situiertes Lernen. In D. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. überarb. u. erweit. Aufl., S. 699-705). Weinheim: Beltz.
- Klauer, K. J., & Leutner, D. (2007). *Lehren und Lernen. Einführung in die Instruktionspsychologie*. Weinheim: Beltz, PVU.
- Kleickmann, T., Möller, K., & Jonen, A. (2006). Zur Wirksamkeit von Fortbildungen und die Bedeutung von tutorieller Unterstützung. In R. Hinz & T. Pütz (Hrsg.), *Professionelles Handeln in der Grundschule. Entwicklungslinien und Forschungsbefunde* (S. 121-128). Hohengehren: Schneider.
- Kleickmann, T. (2008). *Zusammenhänge fachspezifischer Vorstellungen von Grundschullehrkräften zum Lehren und Lernen mit Fortschritten von Schülerinnen und Schülern im konzeptuellen naturwissenschaftlichen Verständnis*. Dissertationsschrift. Westfälische Wilhelms-Universität Münster. <http://miami.uni-muenster.de/servlets/DocumentServlet?id=4597>
- Kleickmann, T., Vehmeyer, J., & Möller, K. (in Druck). Die Bedeutung von Lehrervorstellungen zum Lehren und Lernen für die Strukturierung von Unterricht am Beispiel von Scaffolding-Maßnahmen. *Unterrichtswissenschaft*.

- Klewitz, E., & Mitzkat, H. (1973). Nuffield Junior Science Projekt. *Die Grundschule*, 5(3), 184-192.
- Klewitz, E., & Mitzkat, H. (1977). *Entdeckendes Lernen und offener Unterricht*. Braunschweig: Westermann.
- Klieme, E., & Clausen, M. (1999). *Identifying Facets of Problem Solving in Mathematics Instruction. Paper presented at the annual meeting of the American Educational Research Association (AERA)*. Montreal, CA.
- Klieme, E., & Bos, W. (2000). Mathematikleistungen und mathematischer Unterricht in Deutschland und Japan. Triangulation qualitativer und quantitativer Analysen am Beispiel der TIMS-Studie. *Zeitschrift für Erziehungswissenschaft*, 3(3), 359-380.
- Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I: „Aufgabenkultur“ und Unterrichtsgestaltung. In E. Klieme, J. Baumert u.a. (Hrsg.), *TIMSS – Impulse für Schule und Unterricht* (S. 43-57). Bonn: Bundesministerium für Bildung und Forschung (BMBF).
- Klieme, E., & Thußbas, C. (2001). Kontextbedingungen und Verständigungsprozesse im Geometrieunterricht: Eine Fallstudie. In S. v. Aufschnaiter, M. Welzel (Hrsg.), *Nutzung von Videodaten zur Untersuchung von Lehr-Lern-Prozessen. Aktuelle Methoden empirischer pädagogischer Forschung* (S. 41-59). Münster: Waxmann.
- Klieme, E., & Reusser, K. (2003). Qualität und mathematisches Verständnis im internationalen Vergleich – Ein Forschungsprojekt und erste Schritte zur Realisierung. *Unterrichtswissenschaft*, 3(13), 194-205.
- Klieme, E., Lipowsky, F., Rakoczy, K., & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts „Pythagoras“. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 127-146). Münster: Waxmann.
- Knierim, B. (2008). *Lerngelegenheiten anbieten - Lernangebote nutzen: Eine Videostudie im Schweizer Physikunterricht*. Hamburg: Verlag Dr. Kovač.
- Knoll, S., & Stigler, J.W. (1999). Management and analysis of large-scale video surveys using the software vPrism. *International Journal of Educational Research*, 31, 725-734.
- Knoll, S. (2003). *Verwendung von Aufgaben in Einführungsphasen des Mathematikunterrichts*. Dissertationsschrift, Berlin.
- Kobarg, M., & Seidel, T. (2003). Prozessorientierte Lernbegleitung im Physikunterricht. In T. Seidel, M. Prenzel, R. Duit & M. Lehrke (Hrsg.), *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“* (S.151-200). Kiel: IPN.
- Kohler, B. (2000). Problemorientierter Unterricht: Aktives Wissen vermitteln. *Schulmagazin 5 bis 10*, 3, 8-11.
- Köhnlein, W. (1973). *Die Pädagogik Martin Wagenscheins*. Dissertationsschrift. Erlangen, Nürnberg.
- Köhnlein, W. (1984). Zur Konzipierung eines genetischen, naturwissenschaftlich bezogenen Sachunterrichts. In F. Bauer & W. Köhnlein (Hrsg.), *Problemfeld Natur und Technik* (S. 193-215). Bad Heilbrunn: Klinkhardt.
- Köhnlein, W. (1991). Grundlegende Bildung und Curriculum des Sachunterrichts. In W. Biester (Hrsg.), *Denken über Natur und Technik: Zum Sachunterricht in der Grundschule* (S. 9-23). Bad Heilbrunn: Klinkhardt.
- Köhnlein, W. (1996). Leitende Prinzipien und Curriculum des Sachunterrichts. In W. Wittenbruch & P. Sorger (Hrsg.), *Allgemeinbildung und Grundschule* (S. 107-125). Münster.

- Köhnlein, W. (1996). Leitende Prinzipien und Curriculum des Sachunterrichts. In E. Glumpler, S. Wittkowske (Hrsg.), *Sachunterricht heute. Zwischen interdisziplinärem Anspruch und traditionellem Fachbezug* (S. 54-76). Bad Heilbrunn: Klinkhardt.
- Köhnlein, W. (1998a). Einführende Bemerkungen zum Leben und Werk Martin Wagenscheins sowie zu den Beiträgen. In W. Köhnlein (Hrsg.), *Der Vorrang des Verstehens: Beiträge zur Pädagogik Martin Wagenscheins* (S. 9-20). Bad Heilbrunn: Klinkhardt .
- Köhnlein, W. (1998b). Martin Wagenschein, die Kinder und naturwissenschaftliches Denken. In W. Köhnlein (Hrsg.), *Der Vorrang des Verstehens. Beiträge zur Pädagogik Martin Wagenscheins* (S. 66-87). Bad Heilbrunn: Klinkhardt.
- Kounin, J. S. (2006). *Techniken der Klassenführung*. Münster: Waxmann.
- Krapp, A. (1998). Entwicklung und Förderung von Interessen im Unterricht. *Psychologie in Erziehung und Unterricht*, 44, 185-201.
- Krapp, A. (2001). Interessenentwicklung in der beruflichen Ausbildung. In K. Beck (Hrsg.), *Lehren und Lernen in der beruflichen Erstausbildung* (S. 11-35). Opladen: Leske & Budrich.
- Krapp, A. (2002). An educational-psychological theory of interest and its relation to SDT. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research*. Rochester: University Press.
- Krauss, S., Kunter, M., Brunner, M., Baumert, J., Blum, W., Neubrand, M., et al. (2004). COACTIV: Professionswissen von Lehrkräften, kognitiv aktivierender Mathematikunterricht und die Entwicklung von mathematischer Kompetenz. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule. Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (S. 31-53). Münster: Waxmann.
- Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht*. Münster: Waxmann.
- Kunter, M., Brunner, M., Baumert, J., Klusmann, U., Krauss, S., Blum, W., Jordan, A., & Neubrand, M. (2005). Der Mathematikunterricht der PISA-Schülerinnen und Schüler. Schulformunterschiede in der Unterrichtsqualität. *Zeitschrift für Erziehungswissenschaft*, 8(4), 502-520.
- Lamnek, S. (1995). *Methodologie: Qualitative Sozialforschung* (Bd. 1, 3. Aufl.). Weinheim: Beltz.
- Landwehr, N. (1995). *Neue Wege der Wissensvermittlung*. Aarau: Sauerländer.
- Lankes, E. M. (2005). Problemorientiertes Lernen. In W. Einsiedler, M. Götz, H. Hacker, J. Kahlert, R.W. Keck, E. Sandfuchs (Hrsg.), *Handbuch Grundschulpädagogik und Grundschuldidaktik* (2. überarb. Aufl., S. 391-395). Bad Heilbrunn: Klinkhardt.
- Lauterbach, R. (1992). Naturwissenschaftlich orientierte Grundbildung im Sachunterricht. In K. Riquarts, u.a. (Hrsg.), *Didaktik: Naturwissenschaftliche Bildung in der Bundesrepublik Deutschland* (Bd. 3, S. 191-256). Kiel: IPN.
- Lauterbach, R. (2005). Naturwissenschaftlich-technischer Lernbereich im Sachunterricht. Naturwissenschaftliches und technisches Lernen im Sachunterricht. In W. Einsiedler, M. Götz, H. Hacker, J. Kahlert, R.W. Keck, & E. Sandfuchs (Hrsg.), *Handbuch Grundschulpädagogik und Grundschuldidaktik* (2. überarb. Aufl., S. 572-587). Bad Heilbrunn: Klinkhardt.
- Lave, J. (1991). Situated learning in communities of practice. In L. B. Resnick (Ed.), *Perspectives on socially shared cognition* (pp. 63-82). Washington, DC: American Psychological Association.
- Leutner, D. (2006). Instruktionspsychologie. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. überarb. Aufl., S. 261-270). Weinheim: PVU.

- Lienert, A., & Raatz, U. (1994). *Testaufbau und Testanalyse*. (5. völl. neu bearb. u. erweit. Aufl., S.283-285). Weinheim: Beltz
- Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: a critical appraisal. *Learning and Instruction*, 11, 357-380.
- Linn, M. C., & Eylon, B.-S. (2006). Science education: Integrating views of learning and instruction. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of Educational Psychology* (pp. 511-544). Mahwah: Lawrence Erlbaum Associates.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. *Zeitschrift für Pädagogik*. 51. Beiheft, 52, 47-65.
- Lipowsky, F. (2007a). Was wissen wir über guten Unterricht? In G. Becker, A. Feindt, H. Meyer, M. Rothland, L. Stäudel, E. Terhart (Hrsg.), *Guter Unterricht. Maßstäbe & Merkmale – Wege & Werkzeuge* (S. 26-30). Friedrich Jahresheft 2007.
- Lipowsky, F. (2007b). Unterrichtsqualität in der Grundschule. In K. Möller, P. Hanke, C. Beimbrech, A. K. Hein, T. Kleickmann & R. Schages (Hrsg.), *Qualität von Grundschulunterricht entwickeln, erfassen und bewerten* (S. 35-49). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lipowsky, F., Rakoczy, K., Pauli, C., Reusser, K., & Klieme, E. (2007). Gleicher Unterricht – gleiche Chance für alle? Die Verteilung von Schülerbeiträgen im Klassenunterricht. *Unterrichtswissenschaft*, 35, 125-147.
- Lipowsky, F., Pauli, C., & Rakoczy, K. (2008). Schülerbeteiligung und Unterrichtsqualität. In M. Gläser-Zikuda & J. Seifried (Hrsg.), *Lehrerexpertise – Analyse und Bedeutung unterrichtlichen Handelns* (S. 67-90). Münster: Waxmann.
- Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction*, 19(6), 527-537.
- Loska, R. (1995). *Lehren ohne Belehrung: Leonard Nelsons neosokratische Methode der Gesprächsführung*. Bad Heilbrunn: Klinkhardt.
- Maehr, M. L. & Anderman, E. M. (1993). Reinventing schools for early adolescents: Emphasizing task goals. *The Elementary School Journal*, 93, 593-610.
- Mandl, H., Gruber, H., & Renkl, A. (1995). Situiertes Lernen in multimedialen Lernumgebungen. In L. J. Issing & P. Klimsa (Hrsg.), *Informationen und Lernen mit Multimedia* (S. 167-178). Weinheim: Beltz.
- Mandl, H., Kopp, B., & Dvorak, S. (2004). *Aktuelle theoretische Ansätze und empirische Befunde im Bereich der Lehr-Lern-Forschung – Schwerpunkt Erwachsenenbildung – (Deutsches Institut für Erwachsenenbildung; Ludwig-Maximilian-Universität München: Institut für Pädagogische Psychologie*. URL:http://www.die-bonn.de/espid/dokumente/doc-2004/mandl04_01.pdf (02–05–09).
- Maturana, H. R., & Varela, F. J. (1987). *Der Baum der Erkenntnis*. Bern: Goldmann.
- Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, 59(1), 14-19.
- Mayring, P. (2003). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (8. Aufl.). Weinheim: Beltz.
- Mayring, P., & Brunner, E. (2007) Qualitative Inhaltsanalyse. In R. Buber & H. Holzmüller (Hrsg.), *Qualitative Marktforschung. Konzepte - Methoden – Analysen*. Wiesbaden: Gabler.

- McCombs, B. L. (1998): Integrating metacognition, affect, and motivation in improving teacher education. In N. Lambert & B. L. McCombs (Eds.), *How students learn: Reforming schools through learner-centered education* (pp. 379-408). Washington: American Psychological Association.
- Mc Kenzie, J. (1999): Scaffolding For Success. In: *FNO - The Educational Technology Journal*. URL: <http://fno.org/dec99/scaffold.html> (Stand: 12.5.2009)
- Mietzel, G. (2007). *Pädagogische Psychologie des Lernens und Lehrens*. (8. überarb. u. erweit. Aufl.). Göttingen: Hogrefe.
- Miller, G. A. (2003). The cognitive revolution: a historical perspective. *TRENDS in Cognitive Sciences*, 7(3), 141-144.
- Ministrell, J. (1992). Facets of students' knowledge and relevant instruction. In R. Duit, F. Goldberg & H. Niedderer (Eds.), *Research in physics learning: theoretical issues and empirical studies; proceedings of an international workshop* (pp. 110-128). Kiel: IPN.
- Möller, K. (1999). Konstruktivistisch orientierte Lehr-Lernprozessforschung im naturwissenschaftlich-technischen Bereich des Sachunterrichts. In W. Köhnlein (Hrsg.), *Vielperspektives Denken im Sachunterricht* (S. 125-191). Bad Heilbrunn: Klinkhardt.
- Möller, K. (2000). Lehr- und Lernprozesse im naturwissenschaftlich-technischen Bereich des Sachunterrichts. In R. Duit & C. Rhöneck (Hrsg.), *Ergebnisse fachdidaktischer und psychologischer Lehr-Lern-Forschung* (S. 131-155). Kiel: IPN.
- Möller, K. (2001a). Genetisches Lehren und Lernen – Facetten eines Begriffs. In D. Cech, B. Feige, J. Kahlert, G. Löffler, H. Schreier, H.-J. Schwier & U. Stoltenberg (Hrsg.). *Die Aktualität der Pädagogik Martin Wagenscheins für den Sachunterricht. Walter Köhnlein zum 65. Geburtstag* (S. 15-30). Bad Heilbrunn: Klinkhardt.
- Möller, K. (2001b). Konstruktivistische Sichtweisen für das Lernen in der Grundschule? In K. Czerwenka, K. Nölle & H.-G. Roßbach (Hrsg.), *Forschungen zu Lehr- und Lernkonzepten für die Grundschule* (S. 16-31). Opladen: Leske & Budrich.
- Möller, K. (2002). Anspruchsvolles Lernen in der Grundschule – am Beispiel naturwissenschaftlich-technischer Inhalte. *Pädagogische Rundschau*, 56(4), 411-435.
- Möller, K., Jonen, A., Hardy, I., & Stern, E. (2002). Die Förderung von naturwissenschaftlichem Verständnis bei Grundschulkindern durch Strukturierung der Lernumgebung. In M. Prenzel, J. Doll (Hrsg.), *Zeitschrift für Pädagogik. 45. Beiheft. Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (S. 176-191). Weinheim: Beltz.
- Möller, K. (2004a). Naturwissenschaftliches Lernen in der Grundschule – Welche Kompetenzen brauchen Grundschullehrkräfte? In H. Merkens (Hrsg.), *Lehrerbildung: IGLU und die Folgen* (S. 65-84). Opladen: Leske & Budrich.
- Möller, K. (2004b). Verstehen durch Handeln beim Lernen naturwissenschaftlicher und technikbezogener Sachverhalte. In R. Lauterbach, W. Köhnlein (Hrsg.), *Verstehen und begründetes Handeln* (S. 147-165). Bad Heilbrunn: Klinkhardt.
- Möller, K. (2005). Förderung von Problemlösekompetenzen durch Sachunterricht. Kommentar zum Beitrag von Christina Beinbrech: „Problemlösen im Sachunterricht“. In A. Hartinger & J. Kahlert (Hrsg.), *Förderung des wissenschaftlichen Nachwuchses im Sachunterricht. Perspektiven fachdidaktischer Forschung*. Bad Heilbrunn: Klinkhardt.

- Möller, K. (2006). Naturwissenschaftliches Lernen – eine neue Herausforderung für den Sachunterricht. In P. Hanke (Hrsg.), *Grundschule in Entwicklung. Herausforderungen und Perspektiven für die Grundschule heute* (S. 107-127). Weinheim: Waxmann.
- Möller, K., Hardy, I., Jonen, A., Kleickmann, T., & Blumberg, E. (2006). Naturwissenschaften in der Primarstufe. Zur Förderung konzeptuellen Verständnisses durch Unterricht und zur Wirksamkeit von Lehrerfortbildungen. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 162-193). Münster: Waxmann.
- Möller, K. (2007). Genetisches Lernen und Conceptual Change. In J. Kahlert, M. Fölling-Albers, M. Götz, A. Hartinger, D. von Reeken & S. Wittkowske (Hrsg.), *Handbuch Didaktik des Sachunterrichts* (S. 411-416). Bad Heilbrunn: Klinkhardt.
- Moser, U., Ramseier, E., Keller, C., & Huber, M. (1997). *Schule auf dem Prüfstand. Eine Evaluation der Sekundarstufe I auf der Grundlage der „Third International Mathematics and Science Study“*. Chur: Rüegger.
- Nachtigall, D. (1986). Vorstellungen im Bereich der Mechanik, Naturwissenschaften im Unterricht. *Physik/Chemie*, 34(13), 114-118.
- Neber, H. (1981). *Entdeckendes Lernen* (3. voll. überarb. Aufl.). Weinheim: Beltz.
- Neber, H. (2002). Entdeckendes Lernen. In U. Hameyer & F. Schlichting (Hrsg.), *Entdeckendes Lernen* (S. 10-26). Kronshagen: Körner Verlag GmbH.
- Neber, H. (2006). Entdeckendes Lernen. In D. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie*. (3. überab. u. erweit. Aufl.). Weinheim: Beltz.
- Neubrand, J. (2002). *Eine Klassifikation mathematischer Aufgaben zur Analyse Unterrichtssituationen: Selbstständiges Arbeiten in Schülerarbeitsphasen in den Stunden der TIMSS-Video-Studie*. Hildesheim, Berlin: div Verlag Franzenberger.
- Newton, L. D. (2001). Teaching for understanding in primary science. *Evaluation Research in Education*, 15(3), 143-153.
- Niedderer, H., & Schecker, H. (1992). Towards an explicit description of cognitive systems for research in physics learning. In R. Duit, F. Goldberg & H. Niedderer (Eds.), *Research in physics learning: theoretical issues and empirical studies; proceedings of an international workshop* (pp. 74-98). Kiel: IPN.
- Oser, F., & Patry, J.-L. (1990). Choreographien unterrichtlichen Lernens. Basismodelle des Unterrichts Freiburg. Pädagogisches Institut. *Berichte zur Erziehungswissenschaft* Nr. 89.
- Oser, F., Hascher, T., & Spychiger, M. (1999). Lernen aus Fehlern. Zur Psychologie des „negativen“ Wissens. In W. Althof (Hrsg.), *Fehlerwelten. Vom Fehlermachen und Lernen aus Fehlern* (S. 11-41). Opladen: Leske & Budrich.
- Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology*, 49, 345-375.
- Palincsar, A. S. (1998) Keeping the metaphor of scaffolding fresh – response to C. Addison Stone's "The metaphor of scaffolding: its utility for the field of learning disabilities." *Journal of Learning Disabilities*, 31, 370-373.
- Pauli, C., & Reusser, K. (2006). Von international vergleichenden Video Surveys zur videobasierten Unterrichtsforschung und -entwicklung. *Zeitschrift für Pädagogik*, 52(6), 774-798.
- Pauli, C., Drollinger-Vetter, B., Hugener, I., & Lipowsky, F. (2008). Kognitive Aktivierung im Mathematikunterricht. *Zeitschrift für Pädagogische Psychologie*, 22(2), 127-133.

- Pea, R. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences*, 13(3), 423-451.
- Petko, D., Waldis, M., Pauli, C., & Reusser, K. (2003). Methodologische Überlegungen zur videogestützten Forschung in der Mathematikdidaktik. *Zentralblatt für Didaktik der Mathematik (ZDM)*, 35(6), 265-280.
- Petko, D. (2006). Kameraskript. In I. Hugener, C. Pauli & K. Reusser. *Videoanalysen*. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Materialien zur Bildungsforschung: Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“* (Bd. 15, S. 15-37). Frankfurt am Main: GFPP.
- Pintrich, P. R. (1999). Motivational beliefs as resources for and constraints on conceptual change. In W. Schnotz, S. Vosniadou & M. Carretero (Eds.), *New perspectives on conceptual change* (pp. 33-50). Amsterdam: Pergamon.
- Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching. *Journal of Educational Psychology*, 95, 667-686.
- Pintrich, P. R., Marx, R.W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in process of conceptual change. *Review of Educational Research*, 63(2), 167-199.
- Posner, G., Strike, K., Hewson, P., & Gertzog, W. (1982). Accommodation of scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211-227.
- Preece, P. F. W. (1984). Intuitive science: learned or triggered? *European Journal of Science Education*, 6(1), 7-10.
- Prenzel, M., Geiser, H., Langeheine, R., & Lobemeier, K. (2003). Das naturwissenschaftliche Verständnis am Ende der Grundschule. In W. Bos, E.-M. Langes, M. Prenzel, K. Schwippert, G. Walther & R. Valtin (Hrsg.), *Erste Ergebnisse aus IGLU: Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich* (S. 143-187). Münster: Waxmann.
- Rakoczy, K., & Pauli, C. (2006). Hoch inferentes Rating: Beurteilung der Qualität unterrichtlicher Prozesse. In I. Hugener, C. Pauli & K. Reusser. *Videoanalysen*. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Materialien zur Bildungsforschung: Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie „Unterrichtsqualität, Lernverhalten und mathematisches Verständnis“* (Bd. 15, S. 206-233). Frankfurt am Main: GFPP.
- Reinmann-Rothmeier, G., & Mandl, H. (1998). Wissensvermittlung. Ansätze zur Förderung des Wissenserwerbs. In F. Klix & H. Spada (Hrsg.), *Wissen. Enzyklopädie der Psychologie. Bd. 2* (S. 457-500). Göttingen: Hogrefe.
- Reinmann, G., & Mandl, H. (2006). Unterrichten und Lernumgebungen gestalten. In A. Krapp & B. Weidemann (Hrsg.), *Pädagogische Psychologie* (5. vollst. überarb. Aufl., S. 613-658). Weinheim: Beltz.
- Reiser, B. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences*, 13(3), 273-304.
- Renkl, A., & Helmke, A. (1993). Prinzip, Nutzen und Grenzen der Generalisierungstheorie. *Zeitschrift zu Theorie und Praxis erziehungswissenschaftlicher Forschung*, 7(1), 63-85.
- Renkl, A., & Stern, E. (1994). Die Bedeutung der kognitiven Eingangsvoraussetzungen und schulischen Lerngelegenheiten für das Lösen von einfachen und komplexen Textaufgaben. *Zeitschrift für Pädagogische Psychologie*, 8(1), 27-39.

- Renkl, A. (1996). Träges Wissen: Wenn Erlerntes nicht genutzt wird. *Psychologische Rundschau*, 47, 78-92.
- Renkl, A., Gruber, H., & Mandl, H. (1999). Situated learning in instructional settings: From euphoria to feasibility. In J. Bliss, R. Säljö & P. Light (Eds.), *Learning sites: Social and technological resources for learning* (pp. 101-109). Amsterdam: Elsevier Pergamon.
- Renkl, A. (2006). Träges Wissen. In D. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 778-782) (3. überarb. u. erweit. Aufl.). Weinheim: Beltz.
- Resnick, L. B. (1987). Learning in school and out. *Educational Researcher*, 16(9), 13-20.
- Resnick, L. B. (1991). Shared cognition: thinking as a social practice. In L. B. Resnick, J. M. Levine & S. D. Teasdale (Eds.), *Perspectives on socially shared cognition* (pp. 1-20). Washington, DC: American Psychological Association.
- Reusser, K., & Reusser-Weyeneth, M. (1994). Verstehen als psychologischer Prozess und als didaktische Aufgabe: Einführung und Überblick. In K. Reusser & M. Reusser-Weyeneth (Hrsg.), *Verstehen. Psychologischer Prozess und didaktische Aufgabe* (S. 9-35). Bern, Göttingen, Toronto, Seattle: Huber.
- Reusser, K. (1999). Schülerfehler – die Rückseite des Spiegels. In W. Althof (Hrsg.), *Fehlerwelten. Vom Fehlermachen und Lernen aus Fehlern* (S. 203-231). Opladen: Leske & Budrich.
- Reusser, K. (2001). Bridging Instruction to Learning – Where we come from and where we need to go. A research strategy and its implementation in a national and crosscultural study in Switzerland. *Keynote address at the 9th European Conference for Research on Learning and Instruction (EARLI)*, Fribourg, Switzerland, August 2001, URL: http://www.ife.uzh.ch/index.php?treenode_id=117&staff_id=22.
- Reusser, K. (2006). Konstruktivismus – vom epistemologischen Leitbegriff zur Erneuerung der didaktischen Kultur. In M. Baer, M. Fuchs, P. Füglistner, K. Reusser & H. Wyss (Hrsg.), *Didaktik auf psychologischer Grundlage: Von Hans Aebli's kognitionspsychologischer Didaktik zur modernen Lehr- und Lernforschung* (S. 151-168). Bern: h.e.p.
- Reyer, T. (2004). *Oberflächenmerkmale und Tiefenstrukturen im Unterricht. Exemplarische Analysen im Physikunterricht der gymnasialen Sekundarstufe*. Berlin: Logos Verlag.
- Rieber, R. W., & Carton, A. S. (1997). *The collected works of L. S. Vygotsky*. Vol. 3: Problems of the theory and history of psychology. New York: Plenum.
- Rimmele, R. (2003). *Videograph. Multimedia Player zur Kodierung von Videos*. Kiel: IPN.
- Rivkin, S., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.
- Rogoff, B. (1990). *Apprenticeship in thinking. Cognitive development in social context*. New York, NY: Oxford University Press.
- Rogoff, B., & Wertsch, J. (1984). Editors' note. In B. Rogoff & J. Wertsch (Eds.), *New directions for child development: Children's learning in the zone of proximal development* (Vol. 23, pp. 1-6). San Francisco: Jossey Bass.
- Rosenshine, B. (1970). Evaluation of instruction. In P. L. Peterson & H. J. Walberg (Eds.), *Research in teaching* (pp. 28-56). Berkeley, CA: McCutchan.
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: review of the research. *Review of Educational Research*, 64(4), 479-530.
- Rost, D. (2005). *Interpretation und Bewertung pädagogisch-psychologischer Studien. Eine Einführung*. Weinheim: Beltz.

- Rowan, B., Chiang, F.-S., & Miller, R.J. (1997). Using research on employees' performance to study the effects of teachers on students' achievement. *Sociology of Education*, 70(4), 256-284.
- Schaub, H., & Zenke, K. G. (1995). *Wörterbuch zur Pädagogik*. München: DTV.
- Schecker, H. (1985). *Das Schülerverständnis zur Mechanik*. Dissertationsschrift. Universität Bremen.
- Schnell, R., Hill, P. B., & Esser, E. (2005). *Methoden der empirischen Sozialforschung*. München: Oldenbourg.
- Schreier, H. (1990). Der Begriff der Ganzheit und der Bildungsanspruch des Sachunterrichts. Zum Bildungswert des Sachunterrichts. In H. Schreier (Hrsg.), *Zum Bildungswert des Sachunterrichts. Wegweiser für die Lehrerfortbildung* (S. 7-31). Kiel: Verlag Schmidt & Klaunig.
- Scott, P. H., Asoko, H. M., & Driver, R. H. (1992). Teaching for conceptual change: A review of strategies. In R. Duit, F. Goldberg & H. Niedderer (Eds.), *Research in physics learning: theoretical issues and empirical studies; proceedings of an international workshop* (pp. 310-329). Kiel: IPN.
- Sedlmeier, P., & Renkewitz, F. (2008). *Forschungsmethoden und Statistik der Psychologie*. München: Pearson Studium.
- Seidel, T. (2003). Videobasierte Kodierverfahren in der IPN-Videostudie Physik – ein methodischer Überblick. In T. Seidel, M. Prenzel, R. Duit & M. Lehrke (Hrsg.), *Technischer Band zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“* (S. 99-111). Kiel: IPN.
- Seidel, T., Korbarg, M., & Rimmele, R. (2003). Aufbereitung der Videodaten. In T. Seidel, M. Prenzel, R. Duit & M. Lehrke (Hrsg.), *Technischer Band zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“* (S. 77-98). Kiel: IPN.
- Seidel, T., Prenzel, M., Duit, R., & Lehrke, M. (2003). *Technischer Band zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“*. Kiel: IPN.
- Seidel, T., Prenzel, M., Rimmele, R., Schwindt, K., Korbarg, M., Herweg, C., & Dalhefte, I. M. (2006a). Unterrichtsmuster und ihre Wirkungen. Eine Videostudie im Physikunterricht. In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 99-123). Münster: Waxmann.
- Seidel, T., Prenzel, M., Rimmele, R., Korbarg, M., Herweg, C., & Dalhefte, I. M. (2006b). Blicke auf Physikunterricht. Ergebnisse der IPN-Videostudie. *Zeitschrift für Pädagogik*, 52(6), 798-821.
- Seidel, T., & Prenzel, M. (2007). Wie Lehrpersonen Unterricht wahrnehmen und einschätzen – Erfassung pädagogisch-psychologischer Kompetenzen bei Lehrpersonen mit Hilfe von Videosequenzen. *Zeitschrift für Erziehungswissenschaft, Sonderheft 8*, 201-218.
- Shavelson, R. J., Webb, N. M., & Rowley, G. L. (1989). Generalizability Theory. *American Psychologist*, 44(6), 922-932.
- Shayer, M., & Adhami, M. (2007). Fostering cognitive development through the context of Mathematics: Results of the CAME project. *Educational studies in Mathematics*, 64(3), 265-291.
- Shuell, T. J. (1986). Cognitive conceptions of learning. *Review of Educational Research*, 56, 411-436.
- Siebert, H. (1999). *Pädagogischer Konstruktivismus*. Neuwied: Luchterhand.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.
- Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge transition. *The Journal of the Learning Sciences*, 3, 115-163.

- Sodian, B. (1995). Entwicklung bereichsspezifischen Wissens. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (S. 622-653). Weinheim: Psychologie Verlags Union.
- Sodian, B., Thoermer, C., Kircher, E., Grygier, P., & Günther, J. (2002): Vermittlung von Wissenschaftsverständnis in der Grundschule. In M. Prenzel, J. Doll (Hrsg.), *Zeitschrift für Pädagogik. 45. Beiheft. Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (S. 192-206). Weinheim: Beltz.
- Soostmeyer, M. (1977). *Das entdeckende und forschende Lernen als Ansatzpunkt einer Didaktik der Denkerziehung im naturwissenschaftlich-technischen Sachunterricht der Grundschule / Fachperspektive Physik – Versuch einer Grundlegung*. Dissertationsschrift. Essen.
- Soostmeyer, M. (1978). *Problemorientiertes Lernen im Sachunterricht*. Paderborn: Schöningh.
- Soostmeyer, M. (1986). Einwurzelung und Wissenschaftsorientierung. Zur Bedeutung des Exemplarischen, Sokratischen und Genetischen im Sachunterricht der Grundschule. *Sachunterricht und Mathematik in der Primarstufe*, 14(12), 436-445.
- Soostmeyer, M. (1988). *Zur Sache Sachunterricht*. Frankfurt am Main: Peter Lang.
- Soostmeyer, M. (1998). *Zur Sache Sachunterricht. Begründung eines situations-, handlungs- und sachorientierten Unterrichts in der Grundschule* (3. überarb. Aufl.). Frankfurt am Main: Peter Lang.
- Soostmeyer, M. (2002). *Genetischer Sachunterricht. Unterrichtsbeispiele und Unterrichtsanalysen zum naturwissenschaftlichen Denken bei Kindern in konstruktivistischer Sicht*. Baltmannsweiler: Schneider.
- Spiro, R. J., Feltovich, P. J., Coulson, R. L., & Anderson, D. K. (1989). Multiple analogies for complex concepts: Antidotes for analogy-induced misconception in advanced knowledge acquisition. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 498-531). Cambridge: Cambridge University Press.
- Spiro, R. J., & Jehng, J. C. (1990). Cognitive flexibility and hypertext: Theory and technology for the nonlinear and multidimensional traversal of complex subject matter. In D. Nix & R. J. Spiro (Eds.), *Cognition, education, and multimedia: Exploring ideas in high technology* (pp. 163-205). Hillsdale, NJ: Erlbaum.
- Spreckelsen, K. (1991). Bedeutung von Systematischem im physikalischen Elementarunterricht. *Physik in der Schule*, 29(1991)10, S. 389-391.
- Spreckelsen, K. (1993). Verstehen physikalischer Phänomene im frühen Schulalter. *Physik in der Schule*, 31(10), 326-328.
- Spreckelsen, K. (1997). Phänomenkreise als Verstehenshilfe. In W. Köhnlein, B. Marquart-Mau & H. Schreier (Hrsg.), *Kinder auf dem Weg zum Verstehen der Welt* (S. 111-127). Bad Heilbrunn: Klinkhardt.
- Spychiger, M., Oser, F., Hascher, T., & Mahler, F. (1999). Entwicklung einer Fehlerkultur in der Schule. In W. Althof (Hrsg.), *Fehlerwelten. Vom Fehlermachen und Lernen aus Fehlern* (S. 43-70). Opladen: Leske & Budrich.
- Stark, R., Gruber, R. & Mandl, H. (1998). Motivationale und kognitive Passungsprobleme bei komplexen situationsierten Lernen. *Psychologie in Erziehung und Unterricht*, 45(3), 202-215.
- Staub, A., & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology*, 94(2), 344-355.

- Stebler, R., & Reusser, K. (2000). Progressive, classical or balanced – a look at mathematical learning environments in Swiss-German lower-secondary schools. *Zentralblatt für die Didaktik der Mathematik (ZDM)*, 32(1), 1-10.
- Stern, E. (2002). Wie abstrakt lernt das Grundschulkind? Neuere Ergebnisse der entwicklungspsychologischen Forschung. In H. Petillon (Hrsg.), *Individuelles und soziales Lernen – Kindperspektive und pädagogische Konzepte: Handbuch der Grundschulforschung* (Bd. 5, S. 22-28). Leverkusen: Leske & Budrich.
- Stern, E., & Möller, K. (2004). Der Erwerb anschlussfähigen Wissens als Ziel des Grundschulunterrichts. In D. Lenzen, J. Baumert, R. Watermann, & U. Trautwein (Hrsg.), *Zeitschrift für Erziehungswissenschaft. 3. Beiheft: PISA und die Konsequenzen für die erziehungswissenschaftliche Forschung* (S. 25-36). Wiesbaden: VS.
- Stigler, J. W., & Hiebert, J. (1998). The TIMSS Videotape Study. *American Educator*, 22(4), 43-45.
- Stigler, J.W., & Hiebert, J. (1999). *The teaching gap*. New York: Free Press.
- Stigler, J., Gonzales, P., Kawanaka, T., Knoll, S., & Serrano, A. (1999). *The TIMSS Videotape classroom study. Methods and findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan and the United States*. Washington DC: U.S. Department of Education.
- Stigler, J. W., Gallimore, R., & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS video studies. *Educational Psychologist*, 35(2), 87-100.
- Stork, H. (1995). Was bedeuten die aktuellen Forderungen „Schülervorstellungen berücksichtigen, 'konstruktivistisch' lehren“ für den Chemieunterricht in der Sekundarstufe I? *Zeitschrift für Didaktik der Naturwissenschaften*, 1, 15-28.
- Strike, K. A., & Posner, G. J. (1992) A revisionist theory of conceptual change. In R. Duschl & R. J. Hamilton (eds.), *Philosophy of science, cognitive psychology, and educational theory and practice* (pp. 147-176). Albany, NY: SUNY.
- Terhart, E. (1989). *Lehr-Lern-Methoden. Eine Einführung in die Probleme der methodischen Organisation von Lehren und Lernen*. Weinheim: München.
- Terhart, E. (1999). Konstruktivismus und Unterricht. Gibt es einen neuen Ansatz in der Allgemeinen Didaktik? *Zeitschrift für Pädagogik*, 45(5), 629-647.
- Thiel, S. (1984). Kinder sprechen über Naturphänomene. „Kann Wasser auch den Berg hinauffließen“ In F. Bauer & W. Köhnlein (Hrsg.), *Problemfeld Natur und Technik* (S. 88-99). Bad Heilbrunn: Klinkhardt.
- Tiberghien, A. (1980). Modes and conditions of learning – an example: the learning of some aspects of the concept of heat. In F. Archenhold, R. Driver, A. Orton & C. Wood-Robinson (Eds.), *Cognitive development. Research in Science and Mathematics* (pp. 288-309). Leeds, UK: University of Leeds.
- Tytler, R. (2000a). Teaching for understanding in science: Student conceptions research, and changing views of learning. *Australian Science Teachers Journal*, 48(3), 14-21.
- Tytler, R. (2000b). Teaching for understanding in science: Constructivist / conceptual change teaching approaches. *Australian Science Teachers Journal*, 48(3), 30-35.
- Valsinier, J. (1987). *Culture and the development of children's action*. Cambridge, MA: Harvard Univ. Press.

- Vehmeier, J., Kleickmann T., & Möller, K. (2007). Zusammenhänge von Vorstellungen zum Lehren und Lernen mit unterrichtlichen Handlungen von Lehrkräften. In K. Möller, P. Hanke, C. Beinbrech, A. K. Hein, T. Kleickmann & R. Schages (Hrsg.), *Qualität von Grundschulunterricht entwickeln, erfassen und bewerten* (S. 317-320). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Vermunt, J. D., & Verloop, N. (1999). Congruence and friction between learning and teaching. *Learning and instruction*, 9, 257-280.
- Viennot, L. (1979). Spontaneous reasoning in elementary dynamics. *European Journal of Science Education*, 1, 205-221.
- Vosniadou, S. (1994). Capturing and modelling the process of conceptual change. *Learning and Instruction*, 4, 45-69.
- Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction*, 15, 317-419.
- Vygotsky, L. (1978). *Mind and society: The development of higher psychological processes*. Cambridge.
- Wagenschein, M. (1968). *Verstehen lehren*. Weinheim: Beltz.
- Wagenschein, M. (1976). *Die Pädagogische Dimension der Physik* (4. Aufl.). Braunschweig: Westermann.
- Wagenschein, M. (1984). Kinder auf dem Wege zur Physik. In F. Bauer & W. Köhnlein (Hrsg.), *Problemfeld Natur und Technik* (S. 65-77). Bad Heilbrunn: Klinkhardt.
- Wagenschein, M. (1986). Sprache zwischen Natur und Naturwissenschaft. In U. Knoop & H.-G. Schmitz (Hrsg.), *Das Jahrbuch der Henning-Kaufmann-Stiftung zur Pflege der Reinheit der deutschen Sprache* (S. 53-90). Marburg: Jonas Verlag.
- Wagenschein, M. (1992). *Verstehen lehren. Genetisch-sokratisch-exemplarisch*. (10. Aufl.). Basel, Weinheim: Beltz.
- Wagenschein, M. (1995). *Naturphänomene sehen und verstehen. Genetische Lehrgänge* (3. Aufl.). Dresden, Stuttgart: Klett.
- 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.
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63, 249-294.
- Weinert, F. E. (1996). Für und Wider die „neuen Lerntheorien“ als Grundlagen pädagogisch-psychologischer Forschung. *Zeitschrift für Pädagogische Psychologie*, 10(1), 1-12.
- Wertsch, J. V. (1991). *Voices of the mind: A sociocultural approach to mediated action*. Cambridge, MA: Harvard University Press.
- Whitehead, A. N. (1929). The aims of education. In A.N. Whitehead (Ed.), *The aims of education and other essays* (S. 1-28). New York: Macmillan.
- Wild, K. P. (2003). Videoanalysen als neue Impulsgeber für eine praxisnahe prozessorientierte empirische Unterrichtsforschung. *Unterrichtswissenschaft*, 31(2) 98-102.
- Wild, K. P. (1999). Catmovie3. Eine Software zur Unterstützung der Kodierung digitalen Videomaterials. In: *Arbeiten zur Empirischen Pädagogik und Pädagogischen Psychologie*, Nr. 37. Gelbe Reihe. Neubiberg: Universität der Bundeswehr.

- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.
- Wodzinski, R. (1996). *Untersuchungen von Lernprozessen beim Lernen Newtonscher Dynamik im Anfangsunterricht*. Münster: LIT.
- Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 17, 89-100.
- Woolfolk, A. (2008). *Pädagogische Psychologie* (10. Aufl.). München, Boston: Pearson Studium.

Abbildungsverzeichnis

Abb. 1.5-1	Überblick über die verschiedenen Arten von Reliabilitätsdefinitionen	87
Abb. 3.3-1	Alter Lehrkräfte NRW-Population	99
Abb. 3.3-2	Alter der Lehrkräfte Untersuchungsstichprobe	99
Abb. 3.4-1	Generalisierbarkeitskoeffizient in Abhängigkeit von der Anzahl der Rater	107
Abb. 4.1-1	Motivation des Unterrichts	125
Abb. 4.1-2	Eigene Ideen zulassen	126
Abb. 4.1-3	Eigene Formulierungen	126
Abb. 4.1-4	Fehleroffenheit	127
Abb. 4.1-5	Transmission von Wissen – Erklärungen	128
Abb. 4.1-6	Transmission von Wissen – Denkwege ablehnen	128
Abb. 4.1-7	Präkonzepte – Vorwissen aktivieren und aufgreifen	129
Abb. 4.1-8	Conceptual Change – kognitive Konflikte erzeugen	129
Abb. 4.1-9	Ideen gemeinsam diskutieren	130
Abb. 4.1-10	Scaffolding – Hervorhebungen	130
Abb. 4.1-11	Scaffolding – Perturbieren	131
Abb. 4.1-12	Scaffolding – Sequenzierungen	131
Abb. 4.1-13	Laisser-faire	131
Abb. 4.1-14	Anwendungsbezogenes Lernen	132

Tabellenverzeichnis

Tab. 1.5-1	Überblick über Verfahren der Videoanalyse zur Erfassung von Unterrichtsmerkmalen ...	79
Tab.1.5-2	Verfahren zur Entwicklung von Videoverfahren	82
Tab. 3.1-1	Vereinfachte Untersuchungsanlage der DFG-Interventionsstudie	97
Tab. 3.3-1	Vergleich der Untersuchungsstichprobe (N=30) mit einer für NRW weitgehend repräsentativen Stichprobe von Grundschullehrkräften (N=271)	100
Tab. 3.4-1	Zuwachs des Generalisierbarkeitskoeffizienten (absolut) bei steigender Anzahl der Rater	107
Tab. 3.5-1	Varianzanalytische Zerlegung der beobachteten Werte und Varianzkomponenten, die auf die verschiedenen Effekte zurückgehen	118
Tab. 3.5-2	Überblick über die Struktur der vorliegenden Daten	119
Tab. 3.5-3	Formeln für die Schätzer der Varianzkomponenten	119
Tab. 4.2-1	Statistische Kennwerte der Skalen und Items zum Thema „Wasserkreislauf“ (N=30)	134
Tab. 4.2-2	Statistische Kennwerte der Skalen und Items zum Thema „Schwimmen u. Sinken“ (N=29)	135
Tab. 4.2-3	Interskalenkorrelationen zum Messzeitpunkt „Wasserkreislauf“ (N=30)	138
Tab. 4.3-1	Unterschiede im Unterrichtshandeln der Lehrkräfte in den Fortbildungsgruppen zum MZP 2 (nach der Intervention)	140
Tab. 4.3-2	Unterschiede im Unterrichtshandeln der Lehrkräfte in den Fortbildungsgruppen zum MZP 1 (vor der Intervention)	142
Tab. 4.4-1	Expertenvalidierung (von Person 1 bis Person 5) durch Zuordnung der Items zu den theoretischen Konstrukten	143
Tab. 4.4-2	Bewertung der Zuordnung der Items zu den theoretischen Konstrukten	149
Tab. 4.4-3	Statistische Kennwerte der Skalen und Items der Konstrukte „Konfrontation“, „Transmission von Wissen“ und „Laisser-faire“ (Kurzsкала) zu den Themen „Wasserkreislauf“ (N=30) und „Schwimmen und Sinken“ (N=29)	155

Anhang A

A Analyseergebnisse

A.1 Ergebnisse der Generalisierbarkeitsanalyse

Tab. A.1-1 Ergebnisse der Generalisierbarkeitsanalyse

		Varianzkomponenten			Generalisierbarkeitskoeffizient (absolut)
Motivation des Unterrichts		$\text{Var}_{\text{Video}}$	$\text{Var}_{\text{Rater}}$	$\text{Var}_{\text{Video}*\text{Rater}}$	Φ
mt1	Die Lehrperson gestaltet den Unterricht spannend. (übernommen von Clausen, Reusser & Klieme, 2003)	.17	.03	.24	.71
mt2	Die Lehrperson kann auch trockenen Stoff interessant machen. (übernommen von Clausen, Reusser & Klieme, 2003)	.16	.07	.25	.67
mt3	Die Lehrperson kann die Schüler auch mal richtig begeistern. (übernommen von Clausen, Reusser & Klieme, 2003)	.24	.00	.23	.80
mt6	Die Lehrperson macht die Schüler neugierig.	.22	.00	.30	.74
mt7	Die Lehrperson ruft bei den Schülern Staunen hervor.	.25	.02	.21	.81
mt9	Die Lehrperson kann das Interesse der Kinder am Lerninhalt aufrecht erhalten.	.17	.01	.19	.78
Eigene Ideen zulassen		$\text{Var}_{\text{Video}}$	$\text{Var}_{\text{Rater}}$	$\text{Var}_{\text{Video}*\text{Rater}}$	Φ
ei_i1	Die Lehrperson ermuntert die Kinder, weiter nachzudenken/Ideen zu äußern.	.35	.03	.27	.83
ei_i2	Die Lehrperson räumt den Kindern die Möglichkeit ein, selbst Deutungen/Erklärungen für ein Phänomen zu finden.	.35	.01	.27	.83
ei_i3	Die Lehrperson verfolgt die Ideen der Kinder weiter.	.40	.01	.36	.81
ei_i4	Die Lehrperson greift Schülerbeiträge im weiteren Unterrichtsverlauf auf.	.34	.05	.36	.77
ei_i5	Die Lehrperson gibt den Kindern Zeit, Ideen zu entwickeln und Antworten zu finden.	.20	.01	.34	.69

ei_i6	Die Lehrperson fragt – ohne zu bewerten – nach, wenn ein Schüler eine unvollständige oder unklare Idee formuliert. (übernommen von Clausen, Reusser & Klieme, 2003)	.33	.09	.28	.78
Eigene Formulierungen		Var_{Video}	Var_{Rater}	$Var_{Video*Rater}$	Φ
ei_e1	Die Lehrperson lässt die Sprache der Kinder zu, auch wenn diese noch ungenau ist.	.05	.00	.16	.52
ei_e2	Die Lehrperson legt Wert darauf, dass die Kinder Sachverhalte in ihren eigenen Worten erklären.	.43	.03	.40	.80
ei_e3	Die Lehrperson greift die Formulierungen der Kinder auf und arbeitet damit weiter.	.40	.14	.49	.72
ei_e5	Die Lehrperson sorgt nicht dafür, dass "Fachtermini" mit eigenen Worten erklärt werden. (umgepoltes Items)	.78	.00	.44	.88
Fehleroffenheit		Var_{Video}	Var_{Rater}	$Var_{Video*Rater}$	Φ
ei_f1	Die Lehrperson ist geduldig, wenn ein Schüler im Unterricht ein Fehlkonzept nennt/hat. (in Anlehnung an Clausen, Reusser & Kieme, 2003)	.19	.00	.15	.83
ei_f2	Bei dieser Lehrperson ist das Nennen von falschen Vermutungen/Erklärungen nichts Schlimmes. (in Anlehnung an Clausen, Reusser & Klieme, 2003)	.05	.00	.09	.71
ei_f3	Die Lehrperson achtet darauf, dass keiner ausgelacht wird, der etwas Falsches sagt.	.02	.01	.23	.26
ei_f4	Die Lehrperson betont, dass auch falsche Ideen wichtig sind, um zum Ziel zu kommen.	.06	.00	.15	.94
ei_f7	Die Kinder dürfen auch falsche Vermutungen/Erklärungen äußern.	.02	.00	.15	.40
ei_f8	Die Lehrperson korrigiert nicht sofort jedes Fehlkonzept (in Anlehnung an Clausen, Reusser & Klieme, 2003)	.22	.01	.22	.79
Transmission von Wissen – Erklärungen		Var_{Video}	Var_{Rater}	$Var_{Video*Rater}$	Φ
il1	Die Lehrperson beantwortet aufkommende Fragen vorschnell selbst.	.38	.01	.30	.83
il2	Die Lehrperson gibt die Erklärungen für die zuvor durchgeführten Experimente vor.	.43	.03	.29	.84
il3	Die Lehrperson lässt die Kinder zuvor aufgeworfene Problemen nicht selbst lösen.	.39	.02	.31	.82
il5	Den Kindern werden 'fertige Erklärungen' durch Abbildungen vermittelt.	.96	.01	.13	.97

il6	Die Lehrperson gibt den Schülern Merksätze vor, die nicht mit den Kindern entwickelt worden sind.	.18	.00	.22	.76
il8	Die Lehrperson fragt nach den Vermutungen der Kinder, arbeitet damit aber nicht weiter.	.23	.03	.44	.67
Transmission von Wissen – eigene Denk- und Lernwege ablehnen		$\text{Var}_{\text{Video}}$	$\text{Var}_{\text{Rater}}$	$\text{Var}_{\text{Video}*\text{Rater}}$	Φ
il9	Die Lehrperson unterbindet eigene Denkwege der Kinder.	.61	.03	.29	.88
il13	Die Lehrperson berichtigt Fehlkonzepte ohne Begründung. (in Anlehnung an Heinze, 2004)	.18	.01	.18	.79
il4	Die Lehrperson korrigiert falsche Vermutungen qua Autorität.	.21	.00	.16	.84
il7	Die Lehrperson unterdrückt eigene Formulierungen der Kinder.	.07	.00	.07	.79
il12	Die Lehrperson ignoriert Fehlkonzepte.	.41	.13	.45	.74
il10	Die Lehrperson geht nicht auf Vorschläge der Kinder zur Überprüfung	.11	.00	.18	.71
il11	Die Lehrperson bricht Lösungsvorschläge der Kinder ab.	.02	.00	.11	.41
Präkonzepte – Vorwissen aktivieren und aufgreifen		$\text{Var}_{\text{Video}}$	$\text{Var}_{\text{Rater}}$	$\text{Var}_{\text{Video}*\text{Rater}}$	Φ
pk1	Die Lehrperson kommt im Verlauf der Stunde auf Präkonzepte der Kinder zurück.	.57	.00	.28	.89
pk3	Die Lehrperson knüpft im Unterricht an Präkonzepte der Schüler an. (in Anlehnung an Kobarg & Seidel, 2003)	.56	.02	.29	.88
pk4	Der Lehrer regt die Schüler an, über ihre Präkonzepte nachzudenken.	.52	.01	.31	.87
pk5	Die Lehrperson nimmt sich viel Zeit, Präkonzepte begründen zu lassen.	.44	.05	.30	.85
Conceptual Change – kognitive Konflikte erzeugen		$\text{Var}_{\text{Video}}$	$\text{Var}_{\text{Rater}}$	$\text{Var}_{\text{Video}*\text{Rater}}$	Φ
cc1	Die Lehrperson hat Versuche ausgesucht, die für die Kinder überzeugend sind.	.27	.00	.20	.85
cc4	Die Lehrperson konfrontiert die Kinder mit Beobachtungen oder Phänomenen, die den Erwartungen der Kinder widersprechen (objektive Evidenz).	.45	.03	.26	.86
cc6	Die Lehrperson widerlegt Fehlkonzepte der Kinder mit Hilfe ausgewählter Materialien (objektive Evidenz).	.40	.02	.23	.87
cc8	Der Lehrperson ist es besonders wichtig, dass objektive Evidenzen intensiv mit Präkonzepten in Beziehung gesetzt werden.	.53	.00	.30	.87

Ideen gemeinsam diskutieren		Var_{Video}	Var_{Rater}	$Var_{Video*Rater}$	Φ
dk4	Missverständliche, unvollständige oder unklare Äußerungen werden nicht ignoriert oder bloß kommentiert, sondern es wird nachgefragt. (übernommen von Clausen, Reusser & Klieme, 2003)	.36	.09	.27	.80
dk5	Die Lehrperson fordert die Kinder auf, ihre eigenen Vermutungen/Erklärungen untereinander zu diskutieren.	.51	.07	.29	.85
dk6	Die Lehrperson fordert die Kinder im Unterricht auf, ihre Deutungen zu einem Phänomen gegenüber Mitschülern zu begründen.	.64	.00	.33	.88
dk7	Die Lehrperson regt die Kinder auch dann an, ihre Vorstellungen untereinander zu diskutieren, wenn diese falsch sind.	.46	.08	.31	.83
dk8	Die Lehrperson diskutiert gemeinsam mit den Kindern Probleme, Idee, Lösungswege, Fehlkonzepte.	.43	.00	.30	.85
Scaffolding – Hervorhebungen		Var_{Video}	Var_{Rater}	$Var_{Video*Rater}$	Φ
sc_h1	Die Lehrperson weist darauf hin, was sich die Schüler merken sollen. (übernommen von Clausen, Reusser & Klieme, 2003)	.32	.01	.27	.82
sc_h2	Die Lehrperson hebt wichtige Fakten hervor. (übernommen von Clausen, Reusser & Klieme, 2003)	.35	.03	.37	.78
sc_h4	Die Lehrperson fasst Ergebnisse oder bereits Gesagtes häufig wiederholend zusammen. (in Anlehnung an Hartinger, Jonen)	.30	.02	.35	.76
sc_h6	Die Lehrperson fasst das Wichtigste nochmals zusammen. (übernommen von Clausen, Reusser & Klieme, 2003)	.42	.01	.27	.86
sc_h7	Die Lehrperson fokussiert die Aufmerksamkeit der Kinder auf wichtige inhaltliche Aspekte.	.38	.02	.34	.81
sc_h8	Die Lehrperson fordert die Kinder auf, verständlich zu wiederholen (auch Aufforderung zur Demonstration).	.24	.15	.36	.65
Scaffolding – Pertubieren		Var_{Video}	Var_{Rater}	$Var_{Video*Rater}$	Φ
sc_p1	Die Lehrperson fordert Begründungen ein.	.63	.00	.27	.90
sc_p2	Die Lehrperson fordert Beweise ein.	.41	.04	.36	.80
sc_p3	Die Lehrperson macht auf Widersprüche in Aussagen/Vermutungen aufmerksam.	.35	.01	.28	.83
sc_p5	Die Lehrperson hinterfragt Aussagen.	.41	.01	.37	.81
sc_p4	Die Lehrperson macht auf ungelöste Probleme aufmerksam.	.35	.01	.28	.83

Scaffolding – Sequenzierungen		Var_{Video}	Var_{Rater}	$Var_{Video*Rater}$	Φ
sc_s1	Der Unterricht ist so sequenziert, dass die Schüler die angestrebten Konzepte/Ziele verstehen können.	.90	.00	.28	.93
sc_s2	Die Lehrperson achtet darauf, dass das Gespräch nicht zwischen unterschiedlichen Aspekten hin und her springt.	.41	.03	.27	.84
Laisser-faire		Var_{Video}	Var_{Rater}	$Var_{Video*Rater}$	Φ
ol3	Die Lehrperson gibt den Kindern kaum Hilfe beim Entwickeln und Auswerten von Experimenten.	.54	.00	.29	.88
ol4	Die Lehrperson begnügt sich vorschnell damit, wenn ein Kind einen erwünschten Begriff/Zusammenhang genannt hat.	.64	.01	.37	.87
ol5	Die Lehrperson beschränkt den Umgang mit Präkonzepten auf ein oberflächliches Abfragen.	.72	.02	.48	.85
ol1	In Gesprächen über Deutungen von Naturphänomenen oder der Suche nach geeigneten Lösungswegen hält sich die Lehrperson zurück.	.49	.00	.21	.90
ol2	Die Lehrperson lässt die Kinder einfach nur 'laufen'.	.34	.00	.23	.85
Anwendungsbezogenes Lernen		Var_{Video}	Var_{Rater}	$Var_{Video*Rater}$	Φ
cc3	Die Lehrperson thematisiert die Anwendung neuer Konzepte.	.93	.00	.22	.94
aw2	Fragestellungen (Bezüge, Beobachtungen) aus der außerschulischen Welt werden angesprochen. (übernommen von Clausen, Reusser & Klieme, 2003)	.35	.04	.33	.79
aw4	Es werden Alltagsmaterialien benutzt.	.30	.01	.12	.90
aw5	Die Lehrperson stellt Bezüge zum außerschulischen Alltag der Kinder her.	.48	.09	.40	.80
aw6	Die Lehrperson stellt im Unterricht Bezüge zu Naturphänomenen her.	.66	.01	.37	.87
aw7	Die Lehrperson fragt die Kinder nach Alltagsbeispielen/Anwendungsbezügen	.64	.02	.28	.90

A.2 Ergebnisse der Item- und Skalenanalyse

In der folgenden Tabelle finden sich statistische Kennwerte (Mittelwerte, Standardabweichungen und Trennschärfen) pro Item und geordnet nach Skalen, jeweils zum Thema „Wasserkreislauf“ (prä) und zum Thema „Schwimmen und Sinken“ (post). Zudem ist pro Skala Cronbachs Alpha jeweils für prä und post angegeben. Die Stichprobengröße beträgt 30 (prä) bzw. 29 (post).

Tab. A.2-1 Ergebnisse der Items und Skalenanalyse

Conceptual Change – kognitive Konflikte erzeugen		M	SD	r_{it}
Alpha prä / post = .83 / .93		prä	prä	prä
		post	post	post
cc1	Die Lehrperson hat Versuche ausgesucht, die für die Kinder überzeugend sind.	2.38	.56	.56
		3.12	.45	.85
cc4	Die Lehrperson konfrontiert die Kinder mit Beobachtungen oder Phänomenen, die den Erwartungen der Kinder widersprechen.	2.55	.89	.63
		2.63	.63	.86
cc6	Die Lehrperson widerlegt Fehlkonzepte der Kinder mit Hilfe ausgewählter Materialien.	1.93	.65	.78
		2.75	.60	.90
cc8	Der Lehrperson ist es besonders wichtig, dass objektive Evidenzen intensiv mit Präkonzepten in Beziehung gesetzt werden.	1.75	.61	.77
		2.72	.81	.82
	Skala gesamt	2.15	.56	
		2.81	.57	
Präkonzepte – Vorwissen aktivieren und aufgreifen		M	SD	r_{it}
Alpha prä/post = .92/.95		prä	prä	prä
		post	post	post
pk1	Die Lehrperson kommt im Verlauf der Stunde auf Präkonzepte der Kinder zurück.	1.88	.75	.86
		2.85	.71	.89
pk3	Die Lehrperson knüpft im Unterricht an Präkonzepte der Schüler an. (in Anlehnung an Kobarg & Seidel, 2003)	1.79	.69	.93
		2.81	.73	.92
pk4	Der Lehrer regt die Schüler an, über ihre Präkonzepte nachzudenken.	1.63	.58	.85
		2.67	.80	.90

pk5	Die Lehrperson nimmt sich viel Zeit, Präkonzepte begründen zu lassen.	1.33 2.38	.39 .81	.78 .85
	Skala gesamt	1.66 2.68	.56 .71	
Laisser-faire		M	SD	r_{it}
Alpha prä/post = .89/.94		prä	prä	prä
		post	post	post
ol1	In Gesprächen über Deutungen von Naturphänomenen oder der Suche nach geeigneten Lösungswegen hält sich die Lehrperson zurück.	1.56 1.53	.82 .69	.81 .85
ol2	Die Lehrperson lässt die Kinder einfach nur 'laufen'.	1.44 1.39	.74 .61	.84 .85
ol3	Die Lehrperson gibt den Kindern kaum Hilfe beim Entwickeln und Auswerten von Experimenten.	1.92 1.58	.88 .73	.79 .88
ol4	Die Lehrperson begnügt sich vorschnell damit, wenn ein Kind einen erwünschten Begriff/Zusammenhang genannt hat.	2.60 1.62	.73 .75	.67 .74
ol5	Die Lehrperson beschränkt den Umgang mit Präkonzepten auf ein oberflächliches Abfragen.	2.55 1.76	.88 .89	.59 .89
	Skala gesamt	2.01 1.57	.68 .66	
Ideen gemeinsam diskutieren		M	SD	r_{it}
Alpha prä/post = .83/.92		prä	prä	prä
		post	post	post
dk4	Missverständliche, unvollständige oder unklare Äußerungen werden nicht ignoriert oder bloß kommentiert, sondern es wird nachgefragt. (dk4) (Übernommen von Clausen, Reusser & Klieme, 2003)	2.45 3.04	.72 .67	.59 .69
dk5	Die Lehrperson fordert die Kinder auf, ihre eigenen Vermutungen/Erklärungen untereinander zu diskutieren.	1.40 2.16	.50 1.00	.70 .84
dk6	Die Lehrperson fordert die Kinder im Unterricht auf, ihre Deutungen zu einem Phänomen gegenüber Mitschülern zu begründen.	1.69 2.69	.70 .89	.60 .79
dk7	Die Lehrperson regt die Kinder auch dann an, ihre Vorstellungen untereinander zu diskutieren, wenn diese falsch sind.	1.41 2.11	.55 .96	.66 .84

dk8	Die Lehrperson diskutiert gemeinsam mit den Kindern Probleme, Idee, Lösungswege, Fehlkonzepte.	1.83 2.47	.58 .86	.62 .89
	Skala gesamt	1.76 2.49	.47 .77	
Scaffolding – Hervorhebungen Alpha prä/post = .97/.97		M	SD	r_{it}
		prä	prä	prä
		post	post	post
sc_h1	Die Lehrperson weist darauf hin, was sich die Schüler merken sollen. (Übernommen von Clausen, Reusser & Klieme, 2003)	2.37 2.48	.66 .72	.91 .88
sc_h2	Die Lehrperson hebt wichtige Fakten hervor. (Übernommen von Clausen, Reusser & Klieme, 2003)	2.13 2.53	.64 .77	.92 .95
sc_h4	Die Lehrperson fasst Ergebnisse oder bereits Gesagtes häufig wiederholend zusammen. (In Anlehnung an Hartinger, Jonen)	1.65 1.89	.57 .87	.87 .91
sc_h6	Die Lehrperson fasst das Wichtigste nochmals zusammen. (Übernommen von Clausen, Reusser & Klieme, 2003)	2.30 2.70	.77 .69	.93 .90
sc_h7	Die Lehrperson fokussiert die Aufmerksamkeit der Kinder auf wichtige inhaltliche Aspekte. (In Anlehnung an Hartinger, Jonen)	2.15 2.47	.63 .81	.89 .93
	Skala gesamt	2.12 2.41	.62 .73	
Scaffolding – Pertubieren Alpha prä/post = .71/.91		M	SD	r_{it}
		prä	prä	prä
		post	post	post
sc_p1	Die Lehrperson fordert Begründungen ein.	1.70 2.90	.66 .66	.35 .81
sc_p2	Die Lehrperson fordert Beweise ein.	1.60 2.04	.66 .88	.35 .64
sc_p3	Die Lehrperson macht auf Widersprüche in Aussagen/Vermutungen aufmerksam.	2.22 2.67	.70 .72	.59 .86
sc_p5	Die Lehrperson hinterfragt Aussagen.	1.93 2.68	.65 .76	.75 .90

	Skala gesamt	1.86	.49	
		2.57	.67	
Scaffolding – Sequenzierungen		M	SD	r_{it}
Alpha prä/post = .77/.86		prä	prä	prä
		post	post	post
sc_s1	Der Unterricht ist so sequenziert, dass die Schüler die angestrebten Konzepte/Ziele verstehen können.	1.88	.94	
		3.18	.58	
sc_s2	Die Lehrperson achtet darauf, dass das Gespräch nicht zwischen unterschiedlichen Aspekten hin und her springt.	2.29	.71	
		2.83	.74	
	Skala gesamt	2.09	.75	
		3.00	.62	
Eigene Formulierungen		M	SD	r_{it}
Alpha prä/post = .73/.84		prä	prä	prä
		post	post	post
ei_e2	Die Lehrperson legt Wert darauf, dass die Kinder Sachverhalte in ihren eigenen Worten erklären.	2.22	.68	.72
		3.13	.64	.82
ei_e3	Die Lehrperson greift die Formulierungen der Kinder auf und arbeitet damit weiter.	2.23	.68	.61
		2.97	.76	.69
ei_e5	Die Lehrperson sorgt nicht dafür, dass "Fachtermini" mit eigenen Worten erklärt werden. (umgepoltes Item)	2.61	.99	.42
		3.58	.59	.62
	Skala gesamt	2.35	.64	
		3.22	.58	
Fehleroffenheit		M	SD	r_{it}
Alpha prä/post = .69/.91		prä	prä	prä
		post	post	post
ei_f1	Die Lehrperson ist geduldig, wenn ein Schüler im Unterricht ein Fehlkonzept nennt/hat. (angepasst an Clausen, Reusser & Kieme, 2003)	3.49	.69	.49
		3.84	.41	.83
ei_f2	Bei dieser Lehrperson ist das Nennen von falschen Vermutungen/Erklärungen nichts Schlimmes. (angepasst an Clausen, Reusser & Klieme, 2003)	3.80	.29	.71
		3.93	.28	.92
ei_f8	Die Lehrperson korrigiert nicht sofort jedes Fehlkonzept. (angepasst an Clausen, Reusser & Klieme, 2003)	3.46	.62	.57
		3.82	.46	.80

	Skala gesamt	3.58	.44	
		3.86	.36	
Eigene Ideen zulassen		M	SD	r_{it}
Alpha prä/post = .91/.93		prä	prä	prä
		post	post	post
ei_i1	Die Lehrperson ermuntert die Kinder, weiter nachzudenken/ Ideen zu äußern.	2.38	.63	.56
		3.10	.62	.87
ei_i2	Die Lehrperson räumt den Kindern die Möglichkeit ein, selbst Deutungen/Erklärungen für ein Phänomen zu finden.	2.55	.73	.77
		3.39	.52	.63
ei_i3	Die Lehrperson verfolgt die Ideen der Kinder weiter.	2.29	.70	.87
		3.00	.65	.91
ei_i4	Die Lehrperson greift Schülerbeiträge im weiteren Unter- richtsverlauf auf.	2.31	.64	.82
		2.97	.66	.82
ei_i6	Die Lehrperson fragt – ohne zu bewerten – nach, wenn ein Schüler eine unvollständige oder unklare Idee formuliert. (übernommen von Clausen, Reusser & Klieme, 2003)	2.61	.67	.84
		3.33	.55	.80
	Skala gesamt	2.43	.58	
		3.16	.53	
Motivation des Unterrichts		M	SD	r_{it}
Alpha prä/post = .70/.87		prä	prä	prä
		post	post	post
mt1	Die Lehrperson gestaltet den Unterricht spannend. (übernommen von Clausen, Reusser & Klieme, 2003)	2.55	.53	.61
		2.83	.58	.79
mt3	Die Lehrperson kann die Schüler auch mal richtig begeistern. (übernommen von Clausen, Reusser & Klieme, 2003)	2.29	.60	.65
		2.59	.63	.80
mt6	Die Lehrperson macht die Schüler neugierig.	1.86	.38	.26
		2.54	.60	.58
mt7	Die Lehrperson ruft bei den Schülern Staunen hervor.	1.80	.35	.24
		2.57	.64	.77
mt9	Die Lehrperson kann das Interesse der Kinder am Lerninhalt aufrecht erhalten.	2.58	.52	.53
		2.75	.55	.52

	Skala gesamt	2.22	.33	
		2.66	.48	
Anwendungsbezogenes Lernen		M	SD	r_{it}
Alpha prä/post = .84/.69		prä	prä	prä
		post	post	post
cc3	Die Lehrperson thematisiert die Anwendung neuer Konzepte.	2.48	.86	.77
		1.06	.14	.23
aw2	Fragestellungen (Bezüge, Beobachtungen) aus der außerschulischen Welt werden angesprochen. (übernommen von Clausen, Reusser & Klieme, 2003)	2.94	.65	.72
		2.57	.69	.29
aw5	Die Lehrperson stellt Bezüge zum außerschulischen Alltag der Kinder her.	2.60	.73	.79
		1.78	.60	.66
aw6	Die Lehrperson stellt im Unterricht Bezüge zu Naturphänomenen her.	2.56	.86	.55
		1.66	.65	.70
aw7	Die Lehrperson fragt die Kinder nach Alltagsbeispielen/Anwendungsbezügen.	2.13	.90	.50
		1.28	.43	.46
	Skala gesamt	2.54	.63	
		1.67	.36	
Transmission von Wissen – Erklärungen		M	SD	r_{it}
Alpha prä/post = .88/.86		prä	prä	prä
		post	post	post
il1	Die Lehrperson beantwortet aufkommende Fragen vorschnell selbst.	2.06	.77	.84
		1.35	.54	.87
il2	Die Lehrperson gibt die Erklärungen für die zuvor durchgeführten Experimente vor.	2.17	.78	.82
		1.29	.49	.88
il3	Die Lehrperson lässt die Kinder zuvor aufgeworfene Problemen nicht selbst lösen.	2.09	.72	.86
		1.22	.50	.87
il6	Die Lehrperson gibt den Schülern Merksätze vor, die nicht mit den Kindern entwickelt worden sind.	1.52	.74	.49
		1.12	.27	.23
	Skala gesamt	2.00	.65	
		1.25	.39	

Transmission von Wissen – eigene Denk- und Lernwege ablehnen		M	SD	r_{it}
Alpha prä/post = .82/.84		prä	prä	prä
		post	post	post
il9	Die Lehrperson unterbindet eigene Denkwege der Kinder.	2.30	.84	.67
		1.18	.46	.83
il13	Die Lehrperson berichtigt Fehlkonzepte ohne Begründung. (in Anlehnung an Heinze, 2004)	1.54	.74	.80
		1.14	.39	.82
II4	Die Lehrperson korrigiert falsche Vermutungen qua Autorität.	1.58	.59	.64
		1.17	.42	.85
II12	Die Lehrperson ignoriert Fehlkonzepte.	2.31	.74	.54
		1.74	.72	.45
	Skala gesamt	1.93	.59	
		1.31	.42	

Antwortformat von 1 bis 4 ((1) *trifft überhaupt nicht zu*, (2) *trifft eher nicht zu*, (3) *trifft eher zu* bis (4) *trifft voll und ganz zu*).

A.3 Nachtrag: Ergebnisse der Item- und Skalenanalyse

In der folgenden Tabelle finden sich für die auf der Basis der Ergebnisse der Expertenvalidierung gebildeten Gesamtskalen „Konfrontation“ (Conceptual Change – kognitive Konflikte + Scaffolding – Perturbieren) und „Transmission von Wissen“ (Erklärungen + Denkwege ablehnen“) sowie für die Kurzsкала „Laisser-faire“ statistische Kennwerte (Mittelwerte, Standardabweichungen und Trennschärfen) pro Item und geordnet nach Skalen, jeweils zum Thema „Wasserkreislauf“ (prä) und zum Thema „Schwimmen und Sinken“ (post). Zudem ist pro Skala Cronbachs Alpha jeweils für prä und post angegeben. Die Stichprobengröße beträgt 30 (prä) bzw. 29 (post).

Tab. A.3-1 Nachtrag: Ergebnisse der Item- und Skalenanalyse

Konfrontation		M	SD	r_{it}
Alpha prä/post = .87/.95		prä	prä	prä
		post	post	post
cc4	Die Lehrperson konfrontiert die Kinder mit Beobachtungen oder Phänomenen, die den Erwartungen der Kinder widersprechen.	2.55 2.63	.89 .63	.70 .85
cc6	Die Lehrperson widerlegt Fehlkonzepte der Kinder mit Hilfe ausgewählter Materialien.	1.93 2.75	.65 .60	.78 .89
cc8	Der Lehrperson ist es besonders wichtig, dass objektive Evidenzen intensiv mit Präkonzepten in Beziehung gesetzt werden.	1.75 2.72	.61 .81	.78 .88
sc_p1	Die Lehrperson fordert Begründungen ein.	1.70 2.89	.66 .66	.32 .84
sc_p2	Die Lehrperson fordert Beweise ein.	1.60 2.04	.66 .88	.40 .66
sc_p3	Die Lehrperson macht auf Widersprüche in Aussagen/Vermutungen aufmerksam.	2.22 2.67	.70 .72	.80 .93
sc_p5	Die Lehrperson hinterfragt Aussagen.	1.93 2.68	.65 .76	.85 .92
	Skala gesamt	1.95 2.63	.52 .64	
Transmission von Wissen		M	SD	r_{it}
Alpha prä/post = .92/.92		prä	prä	prä
		post	post	post
il1	Die Lehrperson beantwortet aufkommende Fragen vorschnell selbst.	2.06 1.35	.77 .54	.84 .88
il2	Die Lehrperson gibt die Erklärungen für die zuvor durchgeführten Experimente vor.	2.17 1.29	.78 .49	.78 .93
il3	Die Lehrperson lässt die Kinder zuvor aufgeworfene Problemen nicht selbst lösen.	2.09 1.22	.72 .50	.89 .92

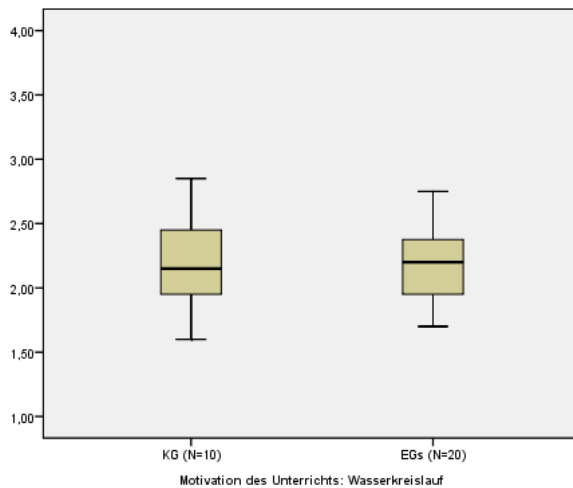
il6	Die Lehrperson gibt den Schülern Merksätze vor, die nicht mit den Kindern entwickelt worden sind.	1.52	.74	.54
		1.12	.27	.17
il9	Die Lehrperson unterbindet eigene Denkwege der Kinder.	2.30	.84	.79
		1.18	.46	.89
il13	Die Lehrperson berichtigt Fehlkonzepte ohne Begründung. (in Anlehnung an Heinze, 2004)	1.54	.74	.82
		1.14	.39	.86
il4	Die Lehrperson korrigiert falsche Vermutungen qua Autorität.	1.58	.59	.68
		1.17	.42	.88
il12	Die Lehrperson ignoriert Fehlkonzepte.	2.31	.74	.51
		1.74	.72	.54
Skala gesamt		1.94	.59	
		1.28	.40	
Laisser-faire (Kurzsкала)		M	SD	r_{it}
Alpha prä/post = .89/.94		prä	prä	prä
		post	post	post
ol2	Die Lehrperson lässt die Kinder einfach nur 'laufen'.	1.44	.74	
		1.39	.61	
ol3	Die Lehrperson gibt den Kindern kaum Hilfe beim Entwickeln und Auswerten von Experimenten.	1.92	.88	
		1.58	.73	
Skala gesamt		1.68	.77	
		1.48	.65	

Antwortformat von 1 bis 4 ((1) *trifft überhaupt nicht zu*, (2) *trifft eher nicht zu*, (3) *trifft eher zu* bis (4) *trifft voll und ganz zu*).

A.4 Verteilung der Stichprobe in den Beobachtungsdimensionen: Boxplot

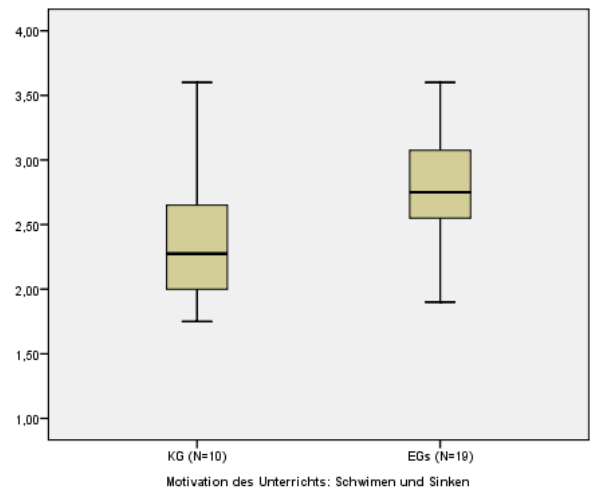
Motivation des Unterrichts

Abb. A.5.1 „Motivation des Unterrichts“ Boxplot
Wasserkreislauf (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

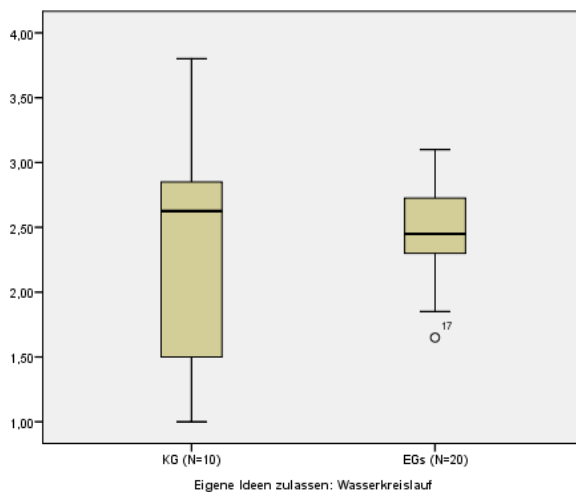
Abb. A.5.2 „Motivation des Unterrichts“ Boxplot
Schwimmen und Sinken (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

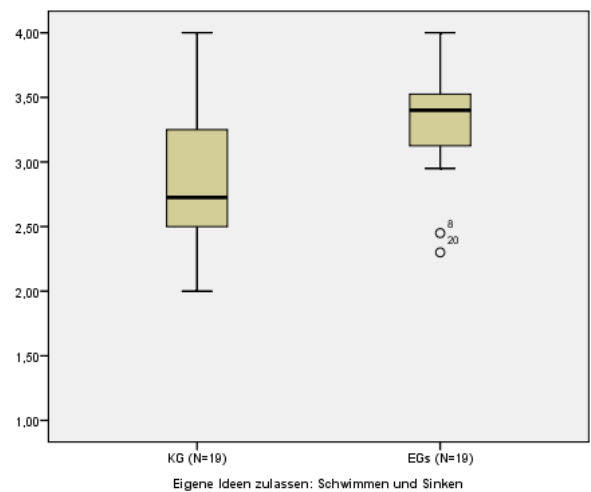
Eigene Ideen zulassen

Abb. A.5.3 „Eigene Ideen zulassen“ Boxplot
Wasserkreislauf (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

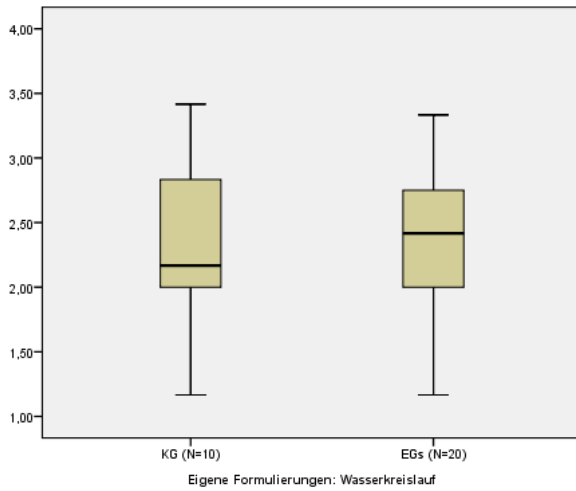
Abb. A.5.4 „Eigene Ideen zulassen“ Boxplot
Schwimmen und Sinken (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

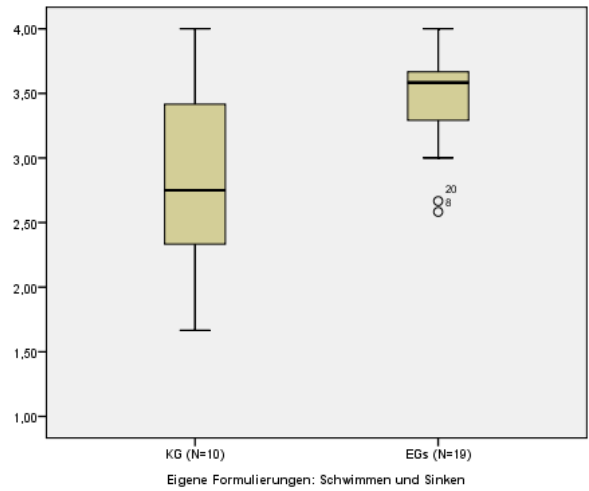
Eigene Formulierungen

Abb. A.5.5 „Eigene Formulierungen“ Boxplot
Wasserkreislauf (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

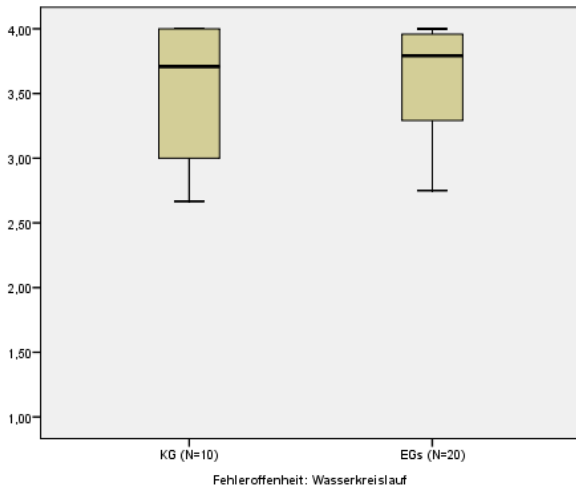
Abb. A.5.6 „Eigene Formulierungen“ Boxplot
Schwimmen und Sinken (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

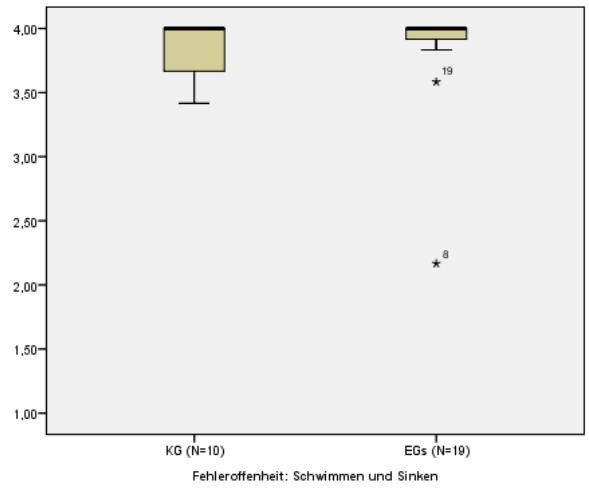
Fehleroffenheit

Abb. A.5.7 „Fehleroffenheit“ Boxplot
Wasserkreislauf (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

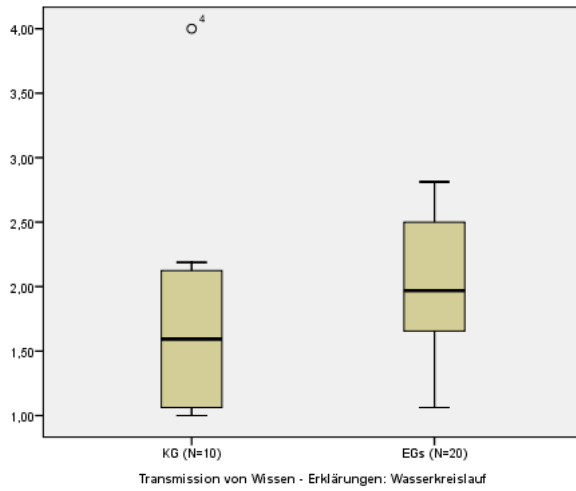
Abb. A.5.8 „Fehleroffenheit“ Boxplot
Schwimmen und Sinken (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

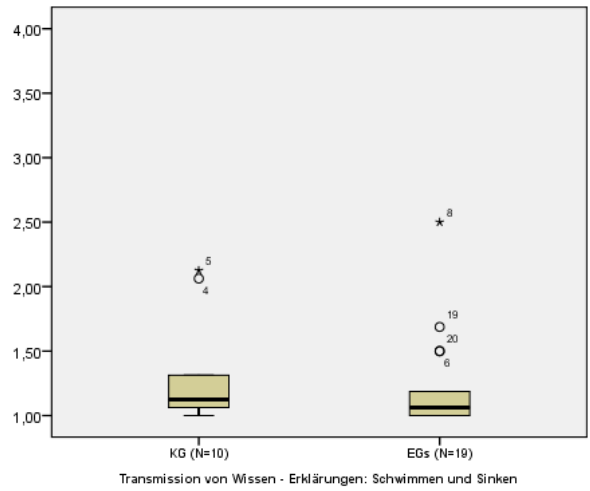
Transmission von Wissen - Erklärungen

Abb. A.5.9 „Transmission – Erklärungen“ Boxplot
Wasserkreislauf (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

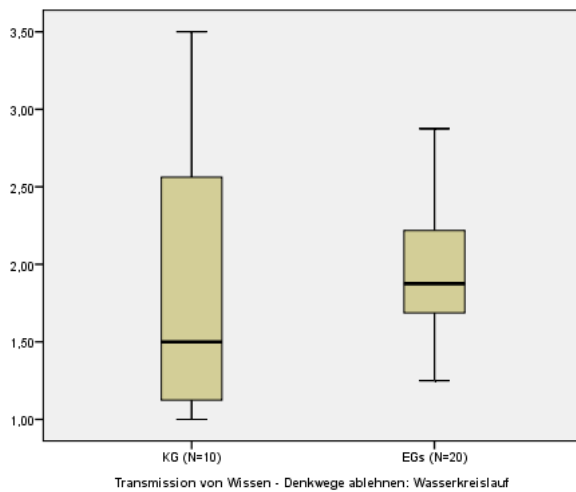
Abb. A.5.10 „Transmission – Erklärungen“ Boxplot
Schwimmen und Sinken (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

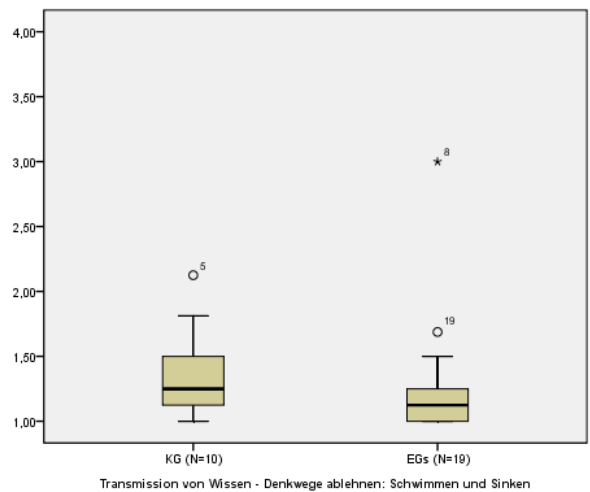
Transmission von Wissen – Denkwege ablehnen

Abb. A.5.11 „Transmission – Denkwege ablehnen“
Boxplot
Wasserkreislauf (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

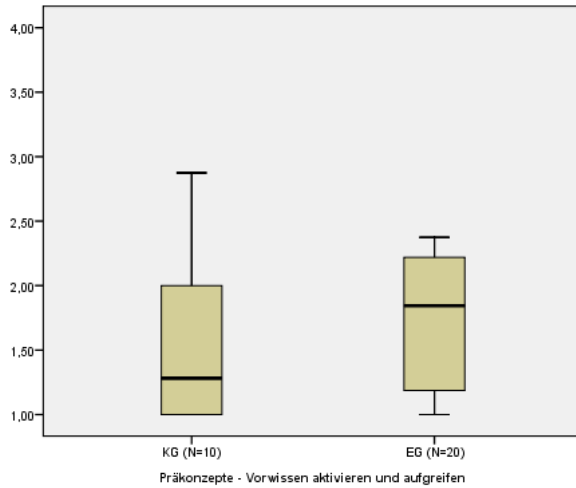
Abb. A.5.12 „Transmission – Denkwege ablehnen“
Boxplot
Schwimmen und Sinken (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

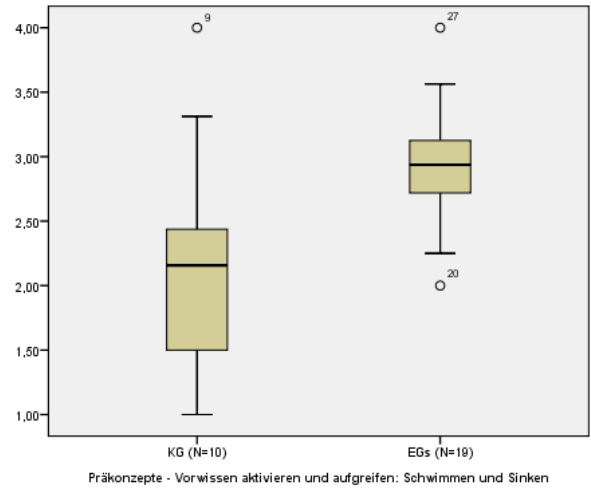
Präkonzepte – Vorwissen aktivieren und aufgreifen

Abb. A.5.13 „Präkonzepte – Vorwissen aktivieren und aufgreifen“ Boxplot
Wasserkreislauf (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

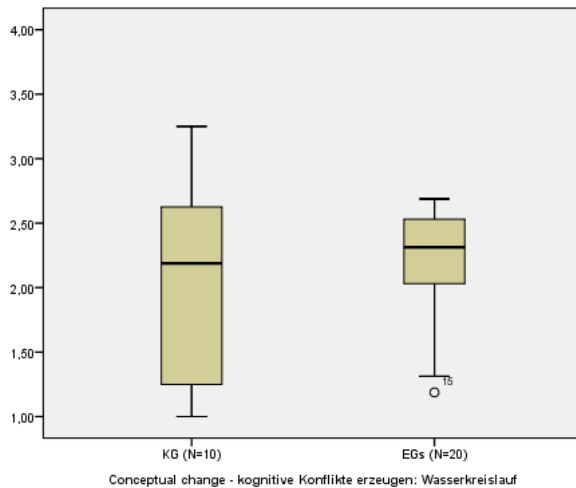
Abb. A.5.14 „Präkonzepte – Vorwissen aktivieren und aufgreifen“ Boxplot
Schwimmen und Sinken (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

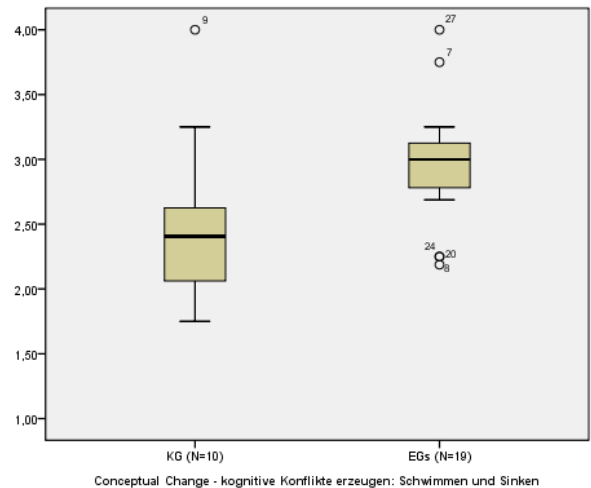
Conceptual Change – kognitive Konflikte erzeugen

Abb. A.5.15 „Conceptual Change – kognitive Konflikte erzeugen“ Boxplot
Wasserkreislauf (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

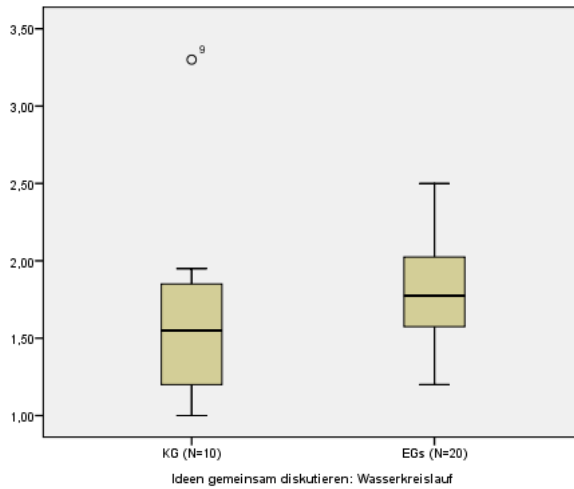
Abb. A.5.16 „Conceptual Change – kognitive Konflikte erzeugen“ Boxplot
Schwimmen und Sinken (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

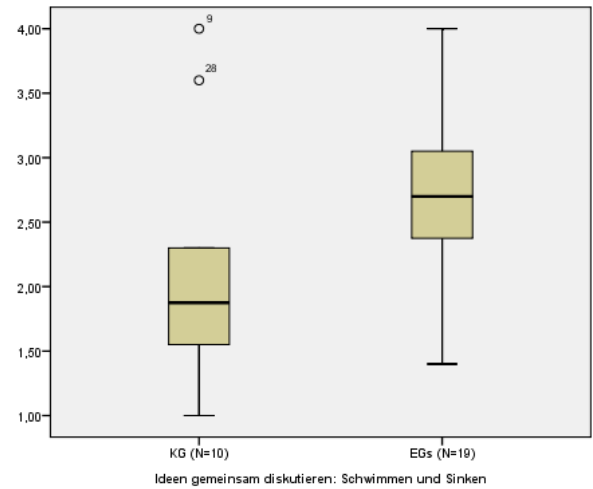
Ideen gemeinsam diskutieren

Abb. A.5.17 „Ideen gemeinsam diskutieren“ Boxplot
Wasserkreislauf (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

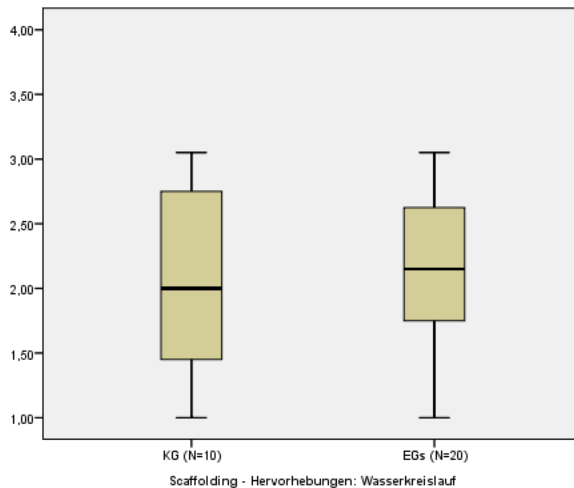
Abb. A.5.18 „Ideen gemeinsam diskutieren“ Boxplot
Schwimmen und Sinken (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

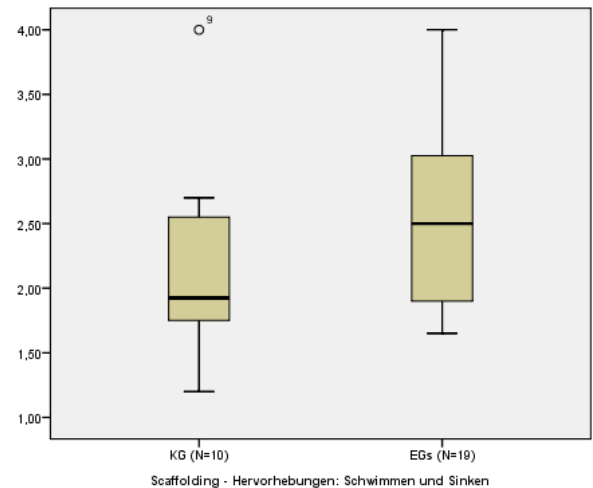
Scaffolding – Hervorhebungen

Abb. A.5.19 „Scaffolding - Hervorhebungen“ Boxplot
Wasserkreislauf (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

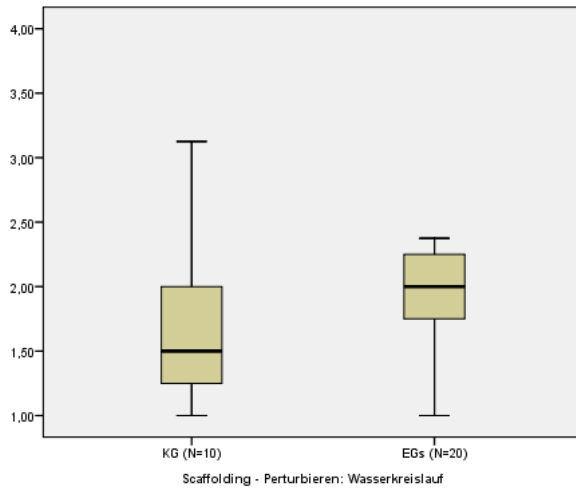
Abb. A.5.20 „Scaffolding - Hervorhebungen“ Boxplot
Schwimmen und Sinken (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

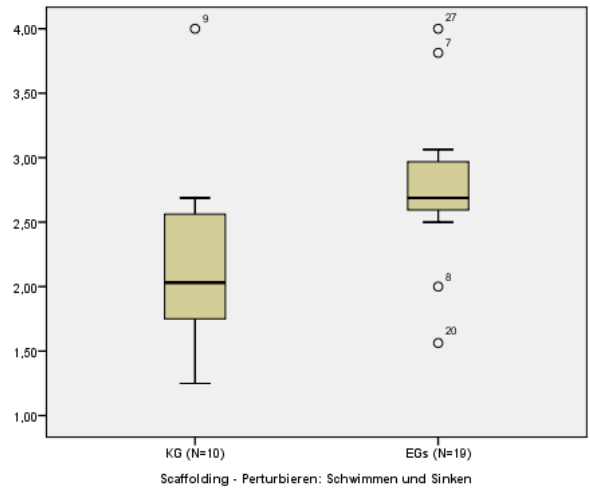
Scaffolding - Perturbieren

Abb. A.5.21 „Scaffolding – Perturbieren“ Boxplot
Wasserkreislauf (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

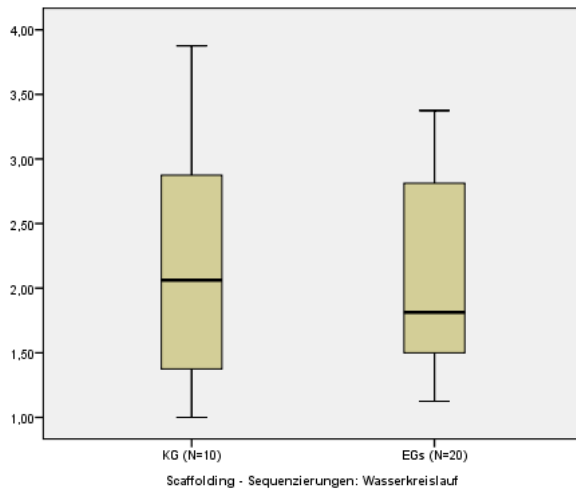
Abb. A.5.22 „Scaffolding – Perturbieren“ Boxplot
Schwimmen und Sinken (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

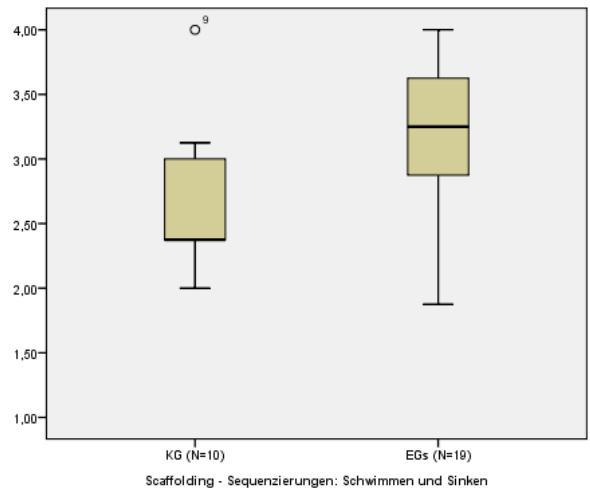
Scaffolding - Sequenzierungen

Abb. A.5.23 „Scaffolding - Sequenzierungen“ Boxplot
Wasserkreislauf (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

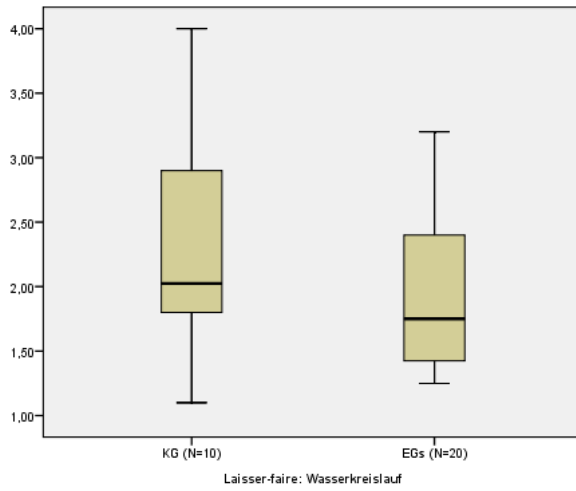
Abb. A.5.24 „Scaffolding - Sequenzierungen s“ Boxplot
Schwimmen und Sinken (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

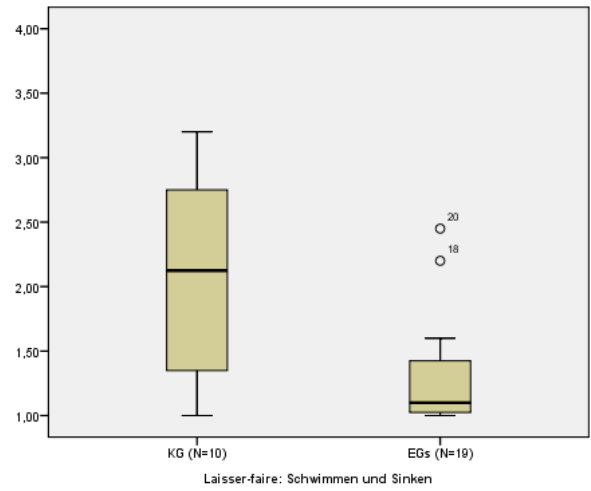
Laisser-faire

Abb. A.5.25 „Laisser - faire“ Boxplot
Wasserkreislauf (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

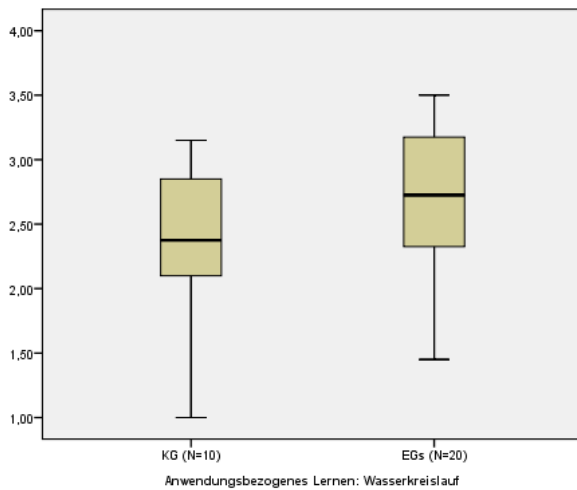
Abb. A.5.26 „Laisser-faire“ Boxplot
Schwimmen und Sinken (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

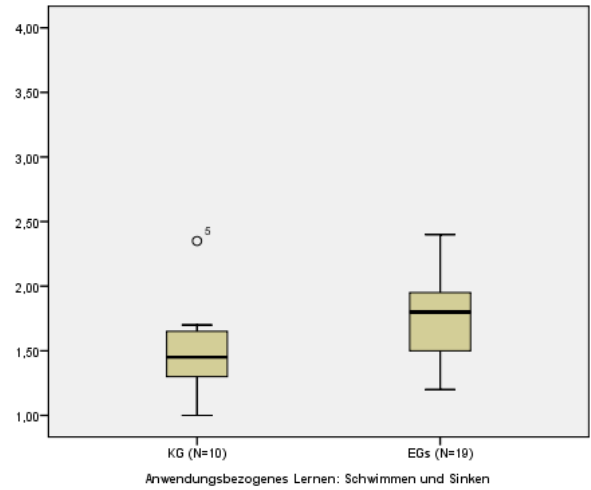
Anwendungsbezogenes Lernen

Abb. A.5.27 „Anwendungsbezogenes Lernen“ Boxplot
Wasserkreislauf (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

Abb. A.5.28 „Anwendungsbezogenes Lernen“ Boxplot
Schwimmen und Sinken (Vergleich KG – EGs)



Antwortformat von (1) *trifft überhaupt nicht zu* bis (4) *trifft voll und ganz zu*.

Anhang B

B Berechnungen

B.1 Der Zusammenhang des absoluten Generalisierbarkeitskoeffizienten mit der Anzahl der Rater

Wir bezeichnen mit $\Phi(n_r)$ den absoluten Generalisierbarkeitskoeffizienten bei n_r Ratern. Im Fall $n_r = 1$ gilt:

$$\Phi(1) = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_{vr}^2 + \sigma_r^2} = \frac{1}{\frac{\sigma_v^2}{\sigma_v^2} + \frac{1}{\sigma_v^2}(\sigma_{vr}^2 + \sigma_r^2)} = \frac{1}{1 + \frac{\sigma_{vr}^2 + \sigma_r^2}{\sigma_v^2}} .$$

Daraus ergibt sich

$$\frac{\sigma_{vr}^2 + \sigma_r^2}{\sigma_v^2} = \frac{1}{\Phi(1)} - 1 .$$

Für allgemeine Anzahl von Ratern folgt also

$$\begin{aligned} \Phi(n_r) &= \frac{\sigma_v^2}{\sigma_v^2 + \frac{\sigma_{vr}^2 + \sigma_r^2}{n_r}} = \frac{1}{\frac{\sigma_v^2}{\sigma_v^2} + \frac{1}{\sigma_v^2} \cdot \frac{\sigma_{vr}^2 + \sigma_r^2}{n_r}} \\ &= \frac{1}{1 + \frac{1}{n_r} \cdot \frac{\sigma_{vr}^2 + \sigma_r^2}{\sigma_v^2}} = \frac{1}{1 + \frac{1}{n_r} \cdot \left(\frac{1}{\Phi(1)} - 1 \right)} . \end{aligned}$$

Im Folgenden sollen noch einige Beispielrechnungen angeführt werden, welche die Basis für Abb. 3.4-1 bilden.

Tab. B.1-1 Absoluter Generalisierbarkeitskoeffizient bei unterschiedlicher Anzahl von Ratern

$\Phi(1)$	$\frac{1}{\Phi(1)} - 1$	$\Phi(2)$	$\Phi(3)$	$\Phi(4)$	$\Phi(5)$
.5	1.0000	.6667	.7500	.8000	.8333
.6	.6667	.7500	.8182	.8571	.8824
.7	.4286	.8235	.8750	.9032	.9211

B.2 Abschätzung der Varianzkomponenten

Ziel dieses Abschnittes ist es, kurz zu erläutern, wie sich das Verhältnis der Varianzkomponente σ_v^2 zur Summe der Varianzkomponenten σ_{vr}^2 und σ_r^2 auf den absoluten Generalisierbarkeitskoeffizienten auswirkt. Bezeichnen wir dieses Verhältnis mit α , so folgt für $\Phi = \Phi(n_r)$

$$\Phi(n_r) = \frac{\sigma_v^2}{\sigma_v^2 + \frac{\sigma_{vr}^2 + \sigma_r^2}{n_r}} = \frac{\sigma_v^2}{\sigma_v^2 + \frac{1}{n_r} \cdot \frac{\sigma_v^2}{\alpha}} = \frac{\sigma_v^2}{\sigma_v^2 \cdot \left(1 + \frac{1}{n_r \cdot \alpha}\right)} = \frac{1}{1 + \frac{1}{n_r \cdot \alpha}}.$$

Der absolute Generalisierbarkeitskoeffizient steigt also, wenn das Verhältnis α größer wird. Bei 4 Ratern gilt insbesondere, dass der Generalisierbarkeitskoeffizient genau dann größer als .70 ist, wenn α größer als .5833 ist, wenn also σ_v^2 etwa mindestens 60% der Summe von σ_{vr}^2 und σ_r^2 beträgt.

Anhang C

C NRW-Population

Tab. C.1-1 Altersstruktur NRW-Grundschullehrkräfte

Altersgruppe	Gesamt	Davon Frauen
Bis 24	43	39
25 – 29	2859	2757
30 – 34	5042	4640
35 – 39	5488	5077
40 – 44	3846	3421
45 – 49	4438	4094
50 – 54	6239	5753
55 – 59	9626	8368
60 – 64	3196	2534
65 und mehr	18	13
Insgesamt	40752	36657

Quelle: URL: <http://www.it.nrw.de/statistik/d/daten/eckdaten/r313schul2.html>, letzter Aufruf am 04.05.09.