

Christin Robisch, Steffen Tröbst & Kornelia Möller

## Hypothesenbezogene Schlussfolgerungen im Grundschulalter fördern

Hypothesenbezogene Schlussfolgerungen werden bereits in einem inquiry-orientierten (Sach-)Unterricht gefordert. Entwicklungspsychologisch stellen angemessene Schlüsse jedoch eine Herausforderung für Grundschul Kinder dar. In Experiment 1 wird gezeigt, dass eine Verbesserung adäquater Schlüsse durch ein Fördersetting im naturwissenschaftlichen Kontext bereits bei Drittklässlern möglich ist. Experiment 2 bekräftigt den Befund auf Basis zweier Fördersettings mit unterschiedlichem Maß an Unterstützung und zeigt, dass Fördereffekte auch bei einer Rücknahme der Unterstützung erhalten bleiben.

**Schlüsselwörter:** Scaffolding; schlussfolgerndes Denken; scientific inquiry; Sachunterricht

*Inquiry-based early science education affords students to reason about the truth of hypotheses. However, developmental research has highlighted substantial deficits in elementary school children's reasoning. In experiment 1 it is demonstrated that third-graders' conditional reasoning in scientific contexts can be successfully scaffolded. In experiment 2 this finding was replicated by scaffolds with varying intensity of support. Moreover, here effects of scaffolding survived fading of support.*

**Key words:** scaffolding; conditional reasoning; scientific inquiry; primary science education

### 1. Hypothesenbezogene Schlussfolgerungen

Eine Möglichkeit der Beantwortung (natur-)wissenschaftlicher Fragen ist die theoretisierte Ableitung und Prüfung von Hypothesen (vgl. z.B. Lederman 2006). In Bezug auf die exemplarische Projektfrage „Wie kommt es, dass Dinge springen?“ (in Anlehnung an Thiel 1987) wird von Kindern häufig die Hypothese aufgestellt, dass Dinge, die leicht sind, springen. Die aus einer vermutlich intuitiven Gewichtstheorie abgeleitete deterministisch-universelle Hypothese kann z.B. durch einen Wattlehausch, der leicht ist und nicht springt, widerlegt werden. Die formallogische Fähigkeit ein gegebenes Ereignis hinsichtlich seiner Bedeutung für die Prüfung der Gültigkeit einer Hypothese zu beurteilen wird kognitionspsychologisch als hypothesenbezogenes Schlussfolgen (reasoning about truth-values) bezeichnet (vgl. Barrouillet, Gauffroy & Lecas 2008). Insgesamt vier kombinatorische Ereignisse können aus dem Vorhandensein und Nichtvorhandensein ( $\neg$ ) des Antezedens p (leicht) und der Konsequenz q (springt) unterschieden werden: pq,  $\neg$ pq,  $\neg$ pq und  $\neg$ pq. In Bezug auf deterministisch-universelle Hypothesen gelten Ereignisevaluierungen dann als angemessen, wenn ein Ereignis der Form pq (leicht und

springt) als bestätigend, ein Ereignis der Form p $\neg$ q (leicht und springt nicht) als widerlegend und Ereignisse der Form  $\neg$ pq (schwer und springt) und  $\neg$ p $\neg$ q (schwer und springt nicht) als irrelevant für die Hypothesenprüfung beurteilt werden (vgl. Johnson-Laird & Byrne 2002; Lecas & Barrouillet 1999). Zahlreiche Studien belegen jedoch, dass insbesondere die adäquate Evaluation irrelevanter Ereignisse für Grundschul Kinder herausfordernd ist (vgl. z.B. Barrouillet u.a. 2008; Gauffroy & Barrouillet 2009, 2011; Tröbst, Hardy & Möller 2011). Einen Erklärungsansatz bietet die (*dual-process*) *mental models theory*, welche die Evaluation von Ereignissen auf Basis der Bildung entsprechender mentaler Modelle im Arbeitsgedächtnis erklärt (vgl. Gauffroy & Barrouillet 2009; Vergauwe, Gauffroy, Morsanyi, Dagrý & Barrouillet 2013).

**Tab. 1:** Ereignisbezogene Interpretationsmuster konditionaler Hypothesen der Form „Wenn p, dann q.“

Ereignis: Beispiel	Interpretationsmuster		
	Hypothese: Dinge, die leicht sind, springen.	konjunktiv bikonditional	defektiv konditional
p q	Die Styrporknagel ist leicht und springt.	+	+
$\neg$ p q	Der Wattlehausch ist leicht und springt nicht.	-	-
p $\neg$ q	Die Boocidangel ist schwer und springt.	-	0
$\neg$ p $\neg$ q	Der Luftballon mit Sand ist schwer und springt nicht.	0	0

$\neg$  = Negation, + = bestätigend, - = widerlegend, 0 = irrelevant.

Barrouillet u.a. konnten 2008 einen dreistufigen Entwicklungstrend – vgl. Tabelle 1 – konditionaler Interpretationen nachweisen. In Studien mit Drit-, Sechst-, Neunklässlern und Studierenden wurde für abstrakte konditionale Hypothesen nachgewiesen, dass die adäquate Evaluation irrelevanter Ereignisse mit steigendem Alter zunimmt (vgl. Barrouillet u.a. 2008; Gauffroy & Barrouillet 2011). Drit- und Sechsklässler nehmen die Hypothese vornehmlich als eine konjunktive Verknüpfung von Antezedens und Konsequenz wahr (vgl. Barrouillet u.a. 2008). Neunklässler interpretieren das Antezedens p primär als ausschließlichen Grund für die Konsequenz q. Diese Interpretationsform wird als defektiv bikonditional bezeichnet (vgl. ebd.). Erst Studierende evaluieren vornehmlich das Antezedens p als möglichen, aber nicht ausschließlichen Grund für die Konsequenz q. Diese adäquate Form der Interpretation wird als defektiv konditional bezeichnet (vgl. ebd.). Die Ergebnisse bekräftigen die (*dual-process*) *mental models theory*, da die altersabhängige Zunahme adäquater Evaluationen mit der zunehmenden Bildung entsprechender mentaler Modelle erklärt werden kann (vgl. Gauffroy & Barrouillet 2009; Vergauwe u.a. 2013). Grundschul Kinder sind folglich zu einer logisch adäquaten Beurteilung der Bedeutung von Ereignissen für die Prüfung von Hypothesen noch nicht in der Lage.

## 2. Relevanz adäquater Schlussfolgerungen für den (Sach-)Unterricht

Wenngleich zahlreiche Evidenzen auf noch eingeschränkte Fähigkeiten von Grundschulkindern beim Schlussfolgern hinweisen (vgl. z.B. Barrouillet u.a. 2008; Gauffroy & Barrouillet 2011; Tröbst u.a. 2011), so wird dieses jedoch im Rahmen einer naturwissenschaftlichen Grundbildung bereits vorausgesetzt (Prenzel, Rost, Senkbeil, Häußler & Klopp 2001). Es wird angenommen, dass hypothesenbezogene Schlussfolgerungen eine zentrale Rolle in scientific-inquiry-Prozessen spielen (vgl. de Jong 2006; Lawson 2010). So ist das Schlussfolgervermögen beispielsweise bei der Planung eines Experimentes zur Hypothesenprüfung erforderlich, um zu entscheiden, welche Ergebnisse – im Falle des angeführten Beispiels: welche Dinge – das Potenzial haben, Auskunft über den Wahrheitsgehalt einer Hypothese zu liefern (Lawson 2010). Erste Ergebnisse weisen zudem auf Zusammenhänge zwischen dem Schlussfolgervermögen und der Falsifikation von Vermutungen sowie argumentativen Prozessen – beispielsweise das Wählen adäquater Belege zur Unterstützung einer Meinung – hin (Robisch, i. V.). Auf Basis der angeführten Gründe muss die Angemessenheit von Schlüssen als ein fachdidaktisch relevanter Fördergegenstand erscheinen.

## 3. Bedingungsfaktoren konditionalen Schlussfolgerns

Aufgrund der zuvor angeführten Relevanz hypothesenbezogener Schlüsse für einen inquiry-orientierten Sachunterricht bedarf es der Betrachtung empirisch abgesicherter Bedingungsfaktoren, um fördernd auf den Lernprozess einwirken zu können. Vor dem Hintergrund der (dual-process) mental models theory (vgl. Gauffroy & Barrouillet 2009; Vergawe u.a. 2013) gibt es zahlreiche Belege für einen Zusammenhang zwischen der *Kapazität des Arbeitsgedächtnisses* und adäquaten Schlüssen (vgl. z.B. Gilhooly, Logie, Wetherick & Wynn 1993; Klauer, Stegmaier & Meiser 1997). Die Bildung mentaler Modelle kann jedoch auch in Abhängigkeit vom wahrgenommenen *Inhalt und Kontext* der konditionalen Aussage variieren: Wird die Hypothese als hinreichende oder notwendige Bedingung für die Konsequenz wahrgenommen (vgl. Cummins, Lubart, Alkaisis & Rist 1991; Markovits 1984; Standenmeyer 1975; Thompson 1994)? Wie wird die Stärke einer kausalen Relation eingeschätzt (vgl. Gauffroy & Barrouillet 2009)? Des Weiteren gilt die *Inhibition*, d.h. die Hemmung oberflächlicher zugunsten analytischer Verarbeitungprozesse, als starker Prädiktor adäquater Schlussfolgerungen (vgl. Handley, Capon, Beveridge, Dennis & Evans 2004). Zudem gibt es Belege für einen positiven Zusammenhang zwischen dem Wissenschaftsverständnis (*Theorie-Evidenz-Koordination*) und dem Schlussfolgervermögen (Stephan-Gramberg, Robisch, Tröbst, Löhner, Hardy & Möller 2012).

## 4. Die Förderung angemessener Schlussfolgerungen

Die Unterstützung von Lernenden bei der Bewältigung von Aufgaben, die sie ohne Hilfe nicht lösen könnten, wird als *Scaffolding* bezeichnet (vgl. van de Pol, Volman & Beishuizen 2010; Wood, Bruner & Ross 1976). Unabhängig vom Fördergegenstand können zwei Wirkmechanismen des Scaffolding unterschieden werden: Während eine *Aufmerk-*

*samkeitlenkung* auf relevante Aspekte (channeling and focussing) der Komplexitätsreduktion einer Aufgabe dient, intendiert eine *Modellierung* (modeling) die Einführung des Lernenden in fortgeschrittene Lösungsansätze, z.B. durch das Freiliegen von Gedankenketten (vgl. Pea 2004). Diese können präventiv, also *statisch*, oder im konkreten Problemfall, also *adaptiv*, eingesetzt werden (vgl. Brush & Saye 2002). Scaffolds sollen bedarfsgerecht reduziert werden, um dem Lernenden sukzessive mehr Verantwortung für die Aufgabenbewältigung zu übertragen (vgl. van de Pol u.a. 2010).

Empirische Belege zeigen, dass eine Förderung adäquater Schlussfolgerungen prinzipiell möglich erscheint (English 1997; Klauer u.a. 1997; Klauer, Meiser & Namer 2000; Rumain, Connel & Braine 1983; Tröbst u.a. 2011). Mit Ausnahme der Studie von Tröbst u.a. (2011) berücksichtigen die Studien jedoch weder hypothesenbezogene Schlussfolgerungen – sondern die damit verbundenen regelbasierten Schlussfolgerungen (reasoning about possibilities) – noch naturwissenschaftliche Kontexte. Zudem basieren die Fördermaßnahmen auf unterschiedlichen theoretischen Annahmen und in keiner Studie werden alle vier Bedingungsfaktoren konditionaler Schlüsse (vgl. Kap. 3) beachtet. Dieses Desiderat motivierte eine Durchführung der folgenden beiden Experimente mit dem Ziel, die Wirksamkeit von Maßnahmen zur Förderung der Angemessenheit hypothesenbezogener Schlüsse bei Grundschulkindern im naturwissenschaftlichen Kontext zu prüfen. Diese Fördermaßnahmen sollen künftig in einen Klassenunterricht implementiert werden. Da in Schulklassen heterogene Problemkonstellationen hinsichtlich hypothesenbezogener Schlüsse zu erwarten sind (Punambeckar & Hübscher 2005) wurden bei der Konstruktion der Scaffolds alle zuvor angeführten Bedingungsfaktoren berücksichtigt.

## 5. Experiment 1

Zur Verbesserung des Schlussfolgervermögens bei Grundschulkindern im naturwissenschaftlichen Kontext wurde ein Fördersetting entwickelt. Dieses wurde durch eine Kombination statisch modellierender und statisch sowie adaptiv aufmerksamskeitslenkender Scaffolds realisiert. Das Fördersetting wurde am Beispiel der Frage „Wie kommt es, dass Dinge springen?“ in Einzelsituationen durchgeführt.

Es wird der Frage nachgegangen, ob das hypothesenbezogene Schlussfolgervermögen bereits bei Drittklasslern gefördert werden kann. Das Experiment umfasst eine Experimentialgruppe und eine Kontrollgruppe. Aufgrund der Hinweise, dass eine Förderung angemessener Schlüsse prinzipiell möglich ist (vgl. Kap. 4), wird erwartet, dass die geförderte Experimentialgruppe der nicht geförderten Kontrollgruppe hinsichtlich angemessener Ereignisbewertungen überlegen sein wird.

### 5.1 Stichprobe

Den folgenden Auswertungen liegen die Daten von 100 Drittklässlerinnen und Drittklässlern (46 Mädchen und 54 Jungen) mit einem Durchschnittsalter von 8,83 Jahren ( $SD = .53$ ) aus fünf Klassen vier nordrhein-westfälischer Grundschulen zugrunde, die freiwillig an der Studie teilnahmen.

## 5.2. Instrumente und Versuchsablauf

Das experimentelle Design umfasste drei Erhebungstage pro Klasse. Am ersten Tag wurden ein Wissens-Prätest sowie ein Inhibitionstest (Früchte-Gemüse-Stroop) durchgeführt. Auf Basis des Inhibitionstests wurden die Kinder den Untersuchungsbedingungen mit Scaffolds (MIT) respektive ohne Scaffolds (OHNE) parallelisiert zugewiesen. Am zweiten und dritten Tag wurden vier truth-testing-tasks, ein Wissens-Posttest und bei der Hälfte der Stichprobe eine Anforderung zur Gesamtevaluation der betrachteten Hypothesen instruiert. Alle Instrumente wurden im Rahmen von leitfadengestützten Einzelinterviews von sechs geschulten Versuchsleitern in separaten Räumen der Schulen eingesetzt und audiographiert.

Im Folgenden wird das für die hier verfolgte Fragestellung relevante Instrument der truth-testing-task dargestellt. Das Ziehen angemessener Schlussfolgerungen über den Wahrheitsgehalt von Hypothesen wurde durch vier *truth-testing-tasks* (vgl. z.B. Barrouillet u.a. 2008) erfasst. Bei jeder truth-testing-task wurden die Probanden mit einer Hypothese in Form eines Relativsatzes (z.B. „Dinge, die leicht sind, springen.“) konfrontiert. Bei der zuerst präsentierten truth-testing-task evaluierten alle Kinder ein Beispielereignis. Anschließend wurden nacheinander die vier kombinatorisch möglichen Ereignisse (vgl. Tab. 1) präsentiert. Bei jedem Ereignis wurden die Kinder aufgefordert zu entscheiden, ob das Ereignis zeigt, dass die Vermutung stimmt (+), nicht stimmt (-) oder ob es nichts mit der Vermutung zu tun hat (o). Die drei Antwortoptionen wurden den Kindern auf bedruckten Karten vorgegeben. Der Ablauf war bei allen truth-testing-tasks identisch. Die Vermutungen rotierten ab der zweiten truth-testing-task zur Kontrolle etwaiger Reihenfolge-Effekte.

## 5.3. Das Fördersetting

Das Setting zur Förderung adäquater Schlussfolgerungen umfasste eine Kombination statischer Modellierung mit einer statischen sowie adaptiven Aufmerksamkeitslenkung (vgl. Kap. 4). Bei der Konstruktion des Fördersettings wurden die vier Bedingungs-konditionalen Schlüsse – Arbeitsgedächtnis, Inhalt und Kontext, Inhibition sowie Theorie-Evidenz-Koordination (vgl. Kap. 3) – berücksichtigt.

Bei der Präsentation einer neuen Vermutung erhielten die Kinder der Bedingung MIT unter Rückgriff auf einen Setzkasten (vgl. Abb. 1) eine *statische Modellierung*. In Anlehnung an Rumann u.a. (1983) wurde im Rahmen der Modellierung die Bedeutung der Hypothese expliziert, um die inhaltsbasierte Wahrnehmung der Vermutung zu erleichtern (z.B. „Wenn ich eine Vermutung prüfen möchte, dann überlege ich zuerst, was mir die Vermutung genau sagt. Die Vermutung sagt ja, dass Dinge, die leicht sind [auf das Feld *leicht* im Setzkasten zeigend], springen [auf das Feld *springt* zeigend]. Die Vermutung sagt also nur etwas über Dinge, die leicht sind. Schwere Dinge [auf das Feld *schwer* zeigend] können also springen oder nicht springen [auf die Felder *springt/springt nicht* zeigend] – das ist egal, weil die Vermutung gar nichts über schwere Dinge sagt.“). Die intensive Auseinandersetzung mit der Hypothese intendierte zudem die Inhibition spontaner zugunsten analytischer Verarbeitungsprozesse.

Der *statisch aufmerksamkeitslenkende Scaffold* wurde mit der Intention entwickelt, die Aufmerksamkeit auf die Eigenschaft und das Sprungverhalten des zu evaluierenden Gegenstands zu lenken, um das Arbeitsgedächtnis zugunsten des Aufbaus der erforderlichen mentalen Modelle zu entlasten. Der Scaffold wurde, in Anlehnung an Tröbst u.a. (2011), in Form eines Setzkastens realisiert. Der Setzkasten unterschied in der oberen Hälfte die dichotomen Eigenschaften der zu prüfenden Vermutungen (z.B. mit Luft und ohne Luft) und in der unteren Hälfte jeweils das Sprungverhalten (springt, springt nicht). Vor jeder Evaluation sortierten die Kinder den jeweiligen Gegenstand in einem Zwischenschritt – erst Eigenschaft, dann Sprungverhalten – in den Setzkasten, sodass beide Merkmale während des Evaluationsprozesses salient waren. Zur Konstanthaltung der time on task wurden die Kinder der OHNE-Gruppe gebeten, sich den jeweiligen Gegenstand nochmals genau anzusehen, während die Kinder der MIT-Gruppe den Setzkasten nutzten. Die Nutzung des Setzkastens wurde nach der Evaluation des Beispielereignisses theoretisch gelbt (z.B. „In welches Feld würdest du ein Ding legen, das leicht ist? Und in welches Feld würdest du dieses Ding anschließend legen, wenn es springt?“).

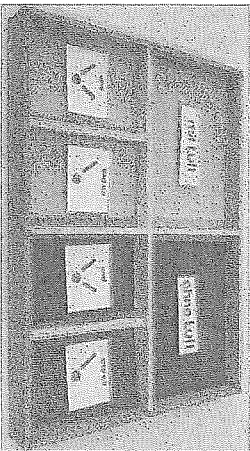


Abb. 1: Ereignisbezogene Interpretationsmuster-Konditionaler Hypothesen der Form „Wenn p, dann q.“

Im Falle einer inadäquaten Evaluation erhielten Kinder der Bedingung MIT unter Rückgriff auf den Setzkasten einen ereignisspezifischen *adaptiv aufmerksamkeitslenkenden Scaffold*, welcher die Aufmerksamkeit des Kindes auf relevante Aspekte von Hypothese und Gegenstand lenkte, um die inhaltsbasierte Wahrnehmung der Vermutung sowie die Koordination von Theorie und Evidenz zu erleichtern. Der Scaffold wurde in Form von Strukturbäumen realisiert, welche den Versuchsleitern die Impulse vorgeben (z.B. „In welchem Feld [des Setzkastens] lag der Wartebausch zuletzt? Lege ihn dort noch mal hinein. Jetzt lies die Vermutung noch mal laut vor. Über welche Dinge sagt die Vermutung etwas? Und wie ist der Wartebausch? Was sagt die Vermutung über Dinge, die leicht sind? Was ist denn mit dem Wartebausch passiert, als du ihn fallen gelassen hast? Der Wartebausch ist also auch leicht, aber nicht gesprungen. Was zeigt dir der Wartebausch dann über die Vermutung: Dinge, die leicht sind, springen?“). Anschließend wurden die Kinder aufgefordert, den Gegenstand erneut zu evaluieren und ihre Wahl zu begründen. Der anfängliche Impuls („Über welche Dinge sagt die Vermutung etwas aus?“) bezweckte eine Erinnerung an die bereits bei der Modellierung intendierte Erleichterung der inhaltlichen Wahrnehmung der Vermutung. Auf einer übergeordneten

Ebene diente auch der adaptiv aufmerksamkeitslenkende Scaffold, durch die intensive Auseinandersetzung mit der Aufgabe, der Inhibition spontaner Verarbeitungprozesse.

#### 5.4 Ergebnisse und Diskussion

Zur Prüfung der Wirksamkeit der Förderung in der Gruppe MIT gegenüber der Gruppe OHNE hinsichtlich angemessener Schlussfolgerungen wurden die über die vier Vermutungen gemittelten adäquaten Ereignisevaluationen (pq+, p-q-, -pq, 0, -p-q, 0) einer repeated measures ANOVA mit dem Innersubjektfaktor Ereignis (pq, p-q, -pq, -p-q) und dem Zwischensubjektfaktor Bedingung (MIT, OHNE) unterzogen. Die prozentuale Verteilung der über die vier Vermutungen gemittelten Ereignisevaluations ist Tabelle 2 zu entnehmen.

**Tab. 2:** Prozentuale Verteilung der über die vier truth-testing-tasks gemittelten Ereignisevaluations

Ereignis	OHNE Scaffolds (n = 48)			MIT Scaffolds (n = 52)		
	+	0	-	+	0	-
p q	89 <sup>a</sup>	8	3	98 <sup>a</sup>	2	0
p-q	4	17	79 <sup>a</sup>	0	6	94 <sup>a</sup>
-p q	12	52 <sup>a</sup>	36	1	83 <sup>a</sup>	16
-p-q	25	48 <sup>a</sup>	27	9	85 <sup>a</sup>	6

+ = bestätigend, 0 = irrelevant, - = widerlegend, <sup>a</sup> = Bestantwort

Die inferenzstatistische Auswertung ergab signifikante Haupteffekte der Faktoren Bedingung,  $F(1, 98) = 48.64$ ,  $p < .001$ ,  $\eta_p^2 = .33$ , und Ereignis,  $F(2.38, 234.04) = 34.91$ ,  $p < .001$ ,  $\eta_p^2 = .26$  sowie einen signifikanten Interaktionseffekt der Faktoren Bedingung und Ereignis,  $F(2.39, 234.04) = 7.72$ ,  $p < .001$ ,  $\eta_p^2 = .07$ . Der signifikante Interaktionseffekt der Faktoren Bedingung und Ereignis wurde durch einfaktorielle ANOVAs spezifiziert, die für alle Ereignisse signifikante Ergebnisse lieferten – pq:  $F(1, 99) = 8.90$ ,  $p = .004$ ,  $\eta_p^2 = .08$ ; p-q:  $F(1, 99) = 10.21$ ,  $p = .002$ ,  $\eta_p^2 = .09$ ; -pq:  $F(1, 99) = 23.99$ ,  $p < .001$ ,  $\eta_p^2 = .20$ ; -p-q:  $F(1, 99) = 37.81$ ,  $p < .001$ ,  $\eta_p^2 = .28$ . Die Effekte der relevanten Ereignisse pq und p-q sind nach Cohen (1988) als mittlere, die Effekte der irrelevanten Ereignisse -pq und -p-q als starke Effekte zu interpretieren.

Die offenen Begründungen der Neuevaluation nach Erhalt eines adaptiv aufmerksamkeitslenkenden Scaffolds wurden transkribiert und auf Grundlage eines Codierschemas analysiert. Insgesamt wurden 20 % der offenen Antworten doppelt codiert – die Interrater-Reliabilität betrug  $\kappa = .82$ . Eine Begründung wurde bei den Ereignissen pq und p-q als adäquat codiert, wenn die angemessene Neuevaluation mit der Erfüllung des Antezedens und der (Nicht-)Erfüllung der Konsequenz oder ausschließlich mit Letzterem erklärt wurde. Eine Begründung bei den Ereignissen -pq und -p-q wurde als adäquat codiert, wenn die Nichterfüllung des Antezedens (ggf. mit Anführung des Sprungverhaltens) als Erklärung angeführt wurde. Insgesamt wurden 78 % der Wechsel zur angemessenen Antwortoption adäquat erklärt.

Es konnte gezeigt werden, dass eine Förderung des Schlussfolgerungsvermögens im naturwissenschaftlichen Kontext bereits bei Drittklasslern möglich ist. Die Scaffolds

wirkten dabei stärker auf die adäquate Evaluation der irrelevanten Ereignisse (-pq und -p-q) als auf die adäquate Evaluation der relevanten Ereignisse (pq und p-q). Dieser Effekt entsprach insofern den bisherigen Befunden zur Schlussfolgerungsleistung im Grundschulalter, als irrelevante Ereignisse ohne Unterstützung häufiger inadäquat evaluiert werden als relevante Ereignisse (vgl. z.B. Tröbst u.a. 2011). Hieraus kann ein erhöhter Förderbedarf bei der Evaluation irrelevanter Ereignisse abgeleitet werden, der ein entsprechend erhöhtes Förderpotenzial erwarten lässt. Aussagen über das Förderpotenzial der einzelnen Scaffolds waren auf Basis dieses Experiments nicht möglich und gaben Anlass zu weiteren Untersuchungen. Hinsichtlich der adaptiv aufmerksamkeitslenkenden Scaffolds ist anzumerken, dass diese den geförderten Kindern eine Neuevaluation der inadäquat evaluierten Ereignisse ermöglichten, während nicht geförderte Kinder im Falle einer Fehlévaluation keine Möglichkeit zur Neuevaluation hatten. Dass jedoch die bloße Möglichkeit zur Korrektur die Wechsel zur angemessenen Antwortoption verursachte, ist aufgrund der hohen Rate adäquater Begründungen sowie vor dem Hintergrund der Kenntnis der Bedingungsfaktoren (vgl. Kap. 3), welche durch die reine Möglichkeit der Neuevaluation nicht bedient werden, unwahrscheinlich.

## 6. Experiment 2

Nachdem im ersten Experiment gezeigt wurde, dass eine Förderung adäquater Schlüsse bei Drittklasslern in einem naturwissenschaftlichen Kontext möglich ist, wurde im Rahmen des zweiten Experiments erkundet, (1) inwiefern prozessbegleitende Scaffolds mit unterschiedlichem Maß an Unterstützung angemessene Schlussfolgerungen über den Wahrheitsgehalt von Hypothesen fördern können. Zu diesem Zweck wurden zwei Fördersettings (EG 1 und EG 2) gebildet: Das erste Fördersetting (EG 1) beinhaltete die bereits in Experiment 1 genutzten aufmerksamkeitslenkenden Scaffolds (statisch und adaptiv), während das zweite Fördersetting (EG 2) zusätzlich eine veränderte Form der Modellierung umfasste. Da Ereignisevaluations in der EG 2 bei der ersten truth-testing-task bis zur richtigen Antwort modelliert wurden, wurde eine Überlegenheit der EG 2 gegenüber der EG 1 erwartet, wobei beide Experimentalgruppen (EGn) der KG ohne Unterstützung überlegen sein sollten. Ebenso wurde der Frage nachgegangen, (2) inwiefern der etwaige Fördererfolg bei völliger Rücknahme der Unterstützung erhalten bleibt. Eine Überlegenheit der EG 2 gegenüber der EG 1, die beide wiederum der KG hinsichtlich angemessener Ereignisevaluations überlegen sein sollten, wurde erwartet.

### 6.1 Stichprobe

Für die Auswertungen des zweiten Experiments legen Daten von 101 Drittklässlerinnen und Drittklasslern (50 Mädchen und 51 Jungen) mit einem Durchschnittsalter von 8.14 Jahren ( $SD = .53$ ) aus fünf Klassen drei nordrhein-westfälischer Grundschulen vor, die freiwillig am Experiment teilnahmen.

## 6.2 Instrumente und Versuchsablauf

Das Experiment umfasste drei Erhebungstage pro Klasse. Beim ersten Termin wurden ein Wissens-Prüfetest sowie ein Inhibitionstest (Frühle-Gemüse-Stroop) durchgeführt. Auf Basis des Inhibitionsvermögens wurden die Kinder drei Untersuchungsgruppen – keine Scaffolds (KG), aufmerksamskeislenkende Scaffolds (EG 1), aufmerksamskeislenkende und modellierende Scaffolds (EG 2) – parallelisiert zugewiesen. Am zweiten Termin wurden vier truth-testing-tasks am Inhalt „Springen und Nichtspringen“ durchgeführt. Am dritten Tag wurde eine weitere truth-testing-task zum gleichen Inhalt eingesetzt, bei welcher keine der Untersuchungsgruppen Unterstützung erhielt. Außerdem wurden drei truth-testing-tasks zum Inhalt „Schwimmen und Sinken“ gestellt. Die inhaltlichen Vermutungen sowie die Reihenfolge der präsentierten Ereignisse rotierten. Abschließend wurden ein Wissens-Posttest sowie ein Argumentationstest instruiert. Alle Tests wurden von geschulten Versuchsleitern im Rahmen von leitfadengestützten Einzelinterviews in separaten Räumen durchgeführt und audiographiert. Das für die folgenden Auswertungen relevante Instrument ist die truth-testing-task (vgl. Experiment 1).

## 6.3 Die Fördersettings

Das erste Fördersetting (EG 1) umfasste die statisch und adaptiv aufmerksamskeislenkenden Scaffolds aus Experiment 1. Das zweite Fördersetting (EG 2) beinhaltete darüber hinaus eine im Vergleich zu Experiment 1 veränderte Modellierung, die zunächst statisch und später adaptiv eingesetzt wurde. Bei der ersten truth-testing-task wurde, nach der Explikation der Bedeutung der Hypothese (vgl. Experiment 1), die angemessene Evaluation der vier Ereignisse demonstriert (z.B. „nach der Einsortierung des Gegenstandes in den Setzkasten, vgl. Experiment 1) In der Vermutung wird ja etwas über leichte Dinge gesagt. Ist der Wartebausch leicht? Genau, er ist leicht. Der Wartebausch hat also etwas mit der Vermutung zu tun. Und was sollen leichte Dinge machen – springen oder nicht springen? Richtig, die sollen springen. Also müsste der Wartebausch springen. Und ist der Wartebausch gesprungen? Genau, er ist nicht gesprungen. Wir wissen also jetzt: Der Wartebausch zeigt, dass die Vermutung nicht stimmt, weil der Wartebausch zwar leicht ist, aber gar nicht springt. Und die Vermutung hat ja gesagt, leichte Dinge springen.“) Während die Explikation der Bedeutung der Hypothese eine Erleichterung der inhaltlichen Wahrnehmung der Vermutung intendierte, bot die Demonstration eine Möglichkeit der Koordination von Theorie und Evidenz an. Die intensive Auseinandersetzung mit der Aufgabe bezweckte die Inhibition spontaner Reaktionen bei weiteren truth-testing-tasks. Der Rückgriff auf den Setzkasten diente der Entlastung des Arbeitsgedächtnisses, sodass erneut alle vier Bedingungsfaktoren berücksichtigt wurden. Bei der zweiten truth-testing-task wurde lediglich die Bedeutung der Hypothese expliziert (vgl. Experiment 1). Die Explikation wurde ab der dritten truth-testing-task nur dann eingesetzt, wenn ein Ereignis in der vorausgegangen truth-testing-task inadäquat evaluiert worden ist. In beiden EGN wurden alle vier Bedingungsfaktoren (vgl. Kap. 3) bei der Konstruktion der Fördersettings berücksichtigt, sodass sich die EGN nur im Ausmaß an Unterstützung unterscheiden.

## 6.4 Ergebnisse und Diskussion

Zur Prüfung des (1) Einflusses der prozessbegleitenden Scaffolds auf das angemessene Schlussfolgern wurde erneut eine repeated measures ANOVA berechnet. Die erste truth-testing-task mit ihren vier zugehörigen Ereignissen wurde aus der Analyse ausgeschlossen, da die adäquaten Antworten in der EG 2 modelliert wurden. Die über die drei übrigen truth-testing-tasks gemittelten angemessenen Ereignisbewertungen (pq<sub>+</sub>, p<sub>+</sub>q<sub>-</sub>, <sub>-</sub>pq<sub>+</sub>, <sub>-</sub>p<sub>+</sub>q<sub>-</sub>) wurden mit dem Innersubjektfaktor Ereignis (pq, p<sub>+</sub>q<sub>-</sub>, <sub>-</sub>pq<sub>+</sub>, <sub>-</sub>p<sub>+</sub>q<sub>-</sub>) und dem Zwischensubjektfaktor Bedingung (KG, EG 1, EG 2) analysiert. Die gruppenspezifische prozentuale Verteilung der gemittelten Ereignisbewertungen ist Tabelle 3 zu entnehmen.

**Tab. 3: Prozentuale Verteilung der über die drei truth-testing-tasks gemittelten Ereignisbewertungen**

Ereignis	KG (n = 32)				EG 1 (n = 33)				EG 2 (n = 36)			
	+	0	-	+	0	-	+	0	-	+	0	-
p q	86 <sup>a</sup>	8	6	94 <sup>a</sup>	6	0	96 <sup>a</sup>	2	2			
p <sub>+</sub> q	8	24	68 <sup>a</sup>	3	10	87 <sup>a</sup>	6	12	82 <sup>a</sup>			
<sub>-</sub> p q	20	27 <sup>a</sup>	53	6	80 <sup>a</sup>	14	6	89 <sup>a</sup>	5			
<sub>-</sub> p <sub>+</sub> q	49	26 <sup>a</sup>	25	14	80 <sup>a</sup>	6	5	89 <sup>a</sup>	6			

+ = bestätigend, 0 = irrelevant, - = widerlegend, <sup>a</sup> = Bestantwort  
 aS = aufmerksamskeislenkende Scaffolds, aS+mS = aufmerksamskeislenkende und modellierende Scaffolds.

Die inferenzstatistische Auswertung ergab neben dem signifikanten Haupteffekt des Faktors Bedingung,  $F(2, 98) = 49.27$ ,  $p < .001$ ,  $\eta_p^2 = .50$ , einen signifikanten Haupteffekt des Faktors Ereignis,  $F(2, 15, 210.49) = 24.65$ ,  $p < .001$ ,  $\eta_p^2 = .20$  sowie eine signifikante Interaktion der Faktoren Bedingung und Ereignis,  $F(4, 30, 210.49) = 11.62$ ,  $p < .001$ ,  $\eta_p^2 = .19$ . Zur Spezifizierung des Haupteffekts des Faktors Bedingung wurden paarweise Vergleiche durchgeführt. Diese zeigen, dass sich die KG jeweils signifikant von der EG 1,  $p < .001$ ,  $r = .73$  und von der EG 2,  $p < .001$ ,  $r = .77$  unterschied. Der geringe Unterschied in der angemessenen Ereignisbewertung zwischen den EGN wurde nicht statistisch signifikant,  $p = .352$ ,  $r = .10$ . Zur Präzisierung des Interaktionseffektes der Faktoren Bedingung und Ereignis wurden einfache ANOVAs mit der jeweils angemessenen Ereignisbewertung pro Ereignis als abhängiger Variable und der Bedingung (KG, EG 1, EG 2) als unabhängiger Variable berechnet. Bei allen vier Ereignissen wurde der Faktor Bedingung statistisch signifikant – pq:  $F(2, 98) = 3.38$ ,  $p = .038$ ,  $\eta_p^2 = .07$ ; p<sub>+</sub>q:  $F(2, 98) = 3.42$ ,  $p = .037$ ,  $\eta_p^2 = .07$ ; <sub>-</sub>pq:  $F(2, 98) = 35.76$ ,  $p < .001$ ,  $\eta_p^2 = .42$ ; <sub>-</sub>p<sub>+</sub>q:  $F(2, 98) = 43.47$ ,  $p < .001$ ,  $\eta_p^2 = .47$ . Die Effektstärken sind nach Cohen (1988) bei den Ereignissen pq und p<sub>+</sub>q als mittel und bei den Ereignissen <sub>-</sub>pq und <sub>-</sub>p<sub>+</sub>q als groß zu interpretieren.

Die Begründungen einer Neu-evaluation wurden mittels eines Codierschemas (vgl. Experiment 1) analysiert. Die Interrater-Reliabilität der Doppelcodierung (20%) betrug  $\kappa = .91$ . Insgesamt wurden 88 % der Wechsel zur angemessenen Antwort adäquat begründet.

Hypothesenkonform schlussfolgerten die beiden geförderten Gruppen angemessener als die nicht geförderte Kontrollgruppe. Die erwartete Überlegenheit der EG 2 gegenüber der EG 1 konnte nicht nachgewiesen werden. Sowohl die aufmerksamkeitlenkenden als auch die modellierenden Scaffolds berücksichtigten alle vier Bedingungsfaktoren konditionaler Schlüsse – die Förderstetigkeit unterschieden sich lediglich im Ausmaß an Unterstützung. Die zusätzliche Unterstützung der EG 2, insbesondere durch die stärkere Lenkung einer Modellierung im Vergleich zur Aufmerksamkeitslenkung, hatte aber keinen zusätzlichen Einfluss auf die Angemessenheit der Schlüsse. Unter Rückgriff auf die deskriptiven Daten ist jedoch zu bedenken, welcher zusätzliche Effekt durch die Modellierung noch hätte erwartet werden können, da bereits über 80 % der Kinder mit aufmerksamkeitlenkenden Scaffolds (EG 1) die Ereignisse angemessen evaluierten. Wie im Experiment 1 zeigten die Scaffolds zudem einen stärkeren Einfluss auf die angemessene Evaluation der Ereignisse  $-pq$  und  $-p-q$  als auf die Ereignisse  $pq$  und  $p-q$ , die bereits ohne Unterstützung vornehmlich adäquat evaluiert werden (vgl. z.B. Tröbst u.a. 2011). Weiterhin wurde untersucht, (2) ob die Überlegenheit der Experimentalgruppen hinsichtlich angemessener Ereignisevaluierungen auch bei Rücknahme der Unterstützung erhalten blieb. Die über vier Ereignisse gemittelte, angemessene Ereignisevaluation ( $pq$ ,  $+$ ,  $p-q$ ,  $-pq$ ,  $0$ ,  $-p-q$ ) wurde einer einfaktoriellen ANOVA mit dem Zwischenblckfaktor Bedingung (KG, EG 1, EG 2) unterzogen. Die gruppenspezifische prozentuale Verteilung der Ereignisevaluierungen ist Tabelle 4 zu entnehmen.

Tab. 4: Prozentuale Verteilung der Evaluierungen der vier Ereignisse der Vermutung ohne Scaffolds

	KG (n = 32)		EG 1 (n = 33)		EG 2 (n = 36)	
Ereignis	+	0	-	0	-	0
p q	78 <sup>a</sup>	13	9	88 <sup>a</sup>	6	94 <sup>a</sup>
p-q	3	19	78 <sup>a</sup>	6	12	80 <sup>a</sup>
-p q	16	25 <sup>a</sup>	59	3	61 <sup>a</sup>	36
-p-q	53	31 <sup>a</sup>	16	18	73 <sup>a</sup>	9
						17
						72 <sup>a</sup>
						11

+ = bestätigend, 0 = irrelevant, - = widerlegend, <sup>a</sup> = Bosantwort.

Die inferenzstatistische Auswertung ergab über die vier Ereignisse hinweg einen statistisch signifikanten Haupteffekt des Faktors Bedingung,  $F(2, 98) = 8.33$ ,  $p < .001$ ,  $\eta^2 = .15$ . Der Effekt ist nach Cohen (1988) als groß zu interpretieren. Paarweise Vergleiche zeigten, dass sowohl die EG 1,  $p < .001$ ,  $r = .46$ , als auch die EG 2,  $p = .002$ ,  $r = .38$ , der KG signifikant in Bezug auf angemessene Schlussfolgerungen überlegen waren, während sich die Experimentalgruppen nicht überzufällig unterscheiden,  $p = .463$ ,  $r = .09$ .

Die Überlegenheit der EGn gegenüber der KG hinsichtlich angemessener Schlüsse blieb also auch bei Rücknahme der Unterstützung bestehen. Die zusätzliche Modellierung der

EG 2 erzielte aber auch hier keinen zusätzlichen Effekt gegenüber der EG 1. Es ist anzumerken, dass die Effekte auf einer truth-testing-task basierten; allerdings wurde diese truth-testing-task einen Tag nach der Förderung instruiert und der Effekt der vorangegangenen Unterstützung blieb dennoch groß (vgl. Cohen 1988). Dieser Befund spricht neben der hohen Rate angemessener Begründungen für adäquate Neu-evaluierungen – dafür, dass die prozessbegleitend adäquaten Neu-evaluierungen nicht auf ein Raten, sondern auf ein verbessertes Verständnis zurückzuführen sind.

## 7. Zusammenfassung und Ausblick

In beiden Experimenten wurden durch Förderstetigkeit angemessene Schlüsse bei Drittklässlern in einem naturwissenschaftlichen Kontext gefördert. Im ersten Experiment umfasste das Förderstetigkeit eine statisch modellierende Explikation der Bedeutung der jeweils präsentierten Vermutung sowie statisch und adaptiv aufmerksamkeitlenkende Scaffolds. Die Ergebnisse zeigten, dass die geförderte Gruppe der nicht geförderten Gruppe überzufällig hinsichtlich adäquater Schlüsse überlegen war. Die Ergebnisse ergänzen den Befund der Studie von Tröbst u.a. (2011), in welcher bereits nachgewiesen werden konnte, dass ein adaptiv aufmerksamkeitlenkendes Nachfragen eine Verbesserung des Schlussfolgerungsvermögens von Viertklässlern bewirkt.

Im zweiten Experiment wurden zwei im Ausmaß der Unterstützung variierte Förderstetigkeiten untersucht. Die Ergebnisse replizierten den Befund, dass eine Förderung des Schlussfolgerungsvermögens bereits bei Drittklässlern im naturwissenschaftlichen Kontext möglich ist. Die Modellierung erzielte jedoch keinen zusätzlichen Effekt, was der hohen Effektivität der Aufmerksamkeitslenkung geschuldet sein könnte. Der Förderstetigkeit beider Förderstetigkeiten blieb auch bei Rücknahme der Unterstützung erhalten, sodass ein flexibler Einsatz beider Förderstetigkeiten möglich scheint. Vor dem Hintergrund des Prinzips des Scaffolding war es möglich, die Verantwortung für die Aufgabenbewältigung erfolgreich an die Lernenden zu übertragen (vgl. van de Pol u.a. 2010).

In einer künftigen Studie soll das Förderstetigkeit in einem Unterricht für ganze Schulklassen überführt werden, um die Effekte unter natürlichen Bedingungen abzusichern.

## Literatur

- Barronillet, P., Gauthroy, C. & Lucas, J. F. (2008): Mental models and the suppositional account of conditionals. In: *Psychological Review*, 115, 760-772.
- Brunsh, T. A. & Sage, J. W. (2002): A summary of research exploring hard and soft scaffolding for teachers and students using a multimedia supported learning environment. In: *The Journal of Interactive Online Learning*, 1, 1-13.
- Cohen, J. (1988): *Statistical power analysis for the behavioral sciences*. 2. Aufl. Hillsdale, Cummings, D. D., Lubart, T., Alksnis, O. & Rist, R. (1991): Conditional reasoning and causation. In: *Memory & Cognition*, 19, 274-282.
- de Jong, T. (2006): Scaffolds for scientific discovery learning. In: Elen, J. & Clark, R. E. (Eds.), *Handling complexity in learning environments: Theory and research*. Amsterdam, 107-128.
- English, L. D. (1997): Interventions in children's deductive reasoning with indeterminate problems. In: *Contemporary Educational Psychology*, 22, 338-362.
- Gibhooly, K. J., Logie, R. H., Wehrick, N. E. & Wynn, V. (1993): Working memory and strategies in syllogistic reasoning tasks. In: *Memory and Cognition*, 21, 115-124.

- Gauffroy, C. & Barrouillet, P. (2009). Heuristic and analytic processes in mental models for conditionals: An integrative developmental theory. In: *Developmental Review*, 29, 249-282.
- Gauffroy, C. & Barrouillet, P. (2011). The primacy of thinking about possibilities in the development of reasoning. In: *Developmental Psychology*, 47, 1000-1011.
- Hardley, S. J., Capon, A., Beveridge, M., Dennis, I. & Evans, J. S. (2004). Working memory, inhibitory control and the development of children's reasoning. In: *Thinking and Reasoning*, 10, 175-195.
- Johnson-Laird, P. N. & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. In: *Psychological Review*, 109, 646-678.
- Klauer, K. C., Meiser, T. & Naumer, B. (2000). Training propositional reasoning. In: *Quarterly Journal of Experimental Psychology*, 53A, 868-895.
- Klauer, K. C., Stegmaier, R. & Meiser, T. (1997). Working memory involvement in propositional and spatial reasoning. In: *Thinking and Reasoning*, 3, H.1, 9-47.
- Lawson, A. E. (2010). Basic inferences of scientific reasoning, argumentation and discovery. In: *Science Education*, 94, H.2, 336-364.
- Lecca, J. F. & Barrouillet, P. (1999). Understanding of conditional rules in childhood and adolescence: A mental models approach. In: *Current Psychology of Cognition*, 18, 363-396.
- Liederman, N. G. (2006). Syntax of nature of science within inquiry and science instruction. In: Flick, L. B. & Liederman, N. G. (Eds.). *Scientific inquiry and nature of science. Implications for teaching, learning and teacher education*. Dordrecht, 301-317.
- Markovits, H. (1984). Awareness of the 'possible' as a mediator of formal thinking in conditional reasoning problems. In: *British Journal of Psychology*, 75, 367-376.
- Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. In: *Journal of the Learning Sciences*, 13, 423-451.
- Prenzel, M., Rost, J., Senfbeck, M., Häußler, P. & Klopp, A. (2001). *Naturwissenschaftliche Grundbildung: Testkonzeption und Ergebnisse*. In: Baumert, J., Klemme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W. u.a. (Hrsg.). *PISA 2000 - Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen, 192-250.
- Punamäke, S. & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? In: *Educational Psychologist*, 40, 1-12.
- Robisch, C. (Y.Y.). *Förderung von Schlussfolgerungen über den Wahrheitsgehalt von Hypothesen im naturwissenschaftlichen Kontext bei Grundschulkindern*. Dissertation.
- Rumann, B., Connell, J. & Braine, M. D. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: If is not the biconditional. In: *Developmental Psychology*, 19, 471-481.
- Staudemeyer, H. (1975). Understanding conditional reasoning with meaningful propositions. In: *Falmagne, R. J. (Ed.), Reasoning: Representation and process*. New York, 55-79.
- Stephan-Granberg, S., Robisch, C., Tröbst, S., Löhrer, K., Hardy, I. & Möller, K. (2012). Zusammenhang zwischen Wissenschaftsverständnis und Umgang mit Evidenz. Vortrag auf der 21. Jahrestagung der Gesellschaft für Didaktik des Sachunterrichts (GDSU) in Berlin.
- Thiel, S. (1987). *Wie springt ein Ball?* In: *Grundschule*, 19, H.1, 18-23.
- Thompson, V. A. (1994). Interpretational factors in conditional reasoning. In: *Memory & Cognition*, 22, H.6, 742-758.
- Tröbst, S., Hardy, I. & Möller, K. (2011). Die Förderung deduktiver Schlussfolgerungen bei Grundschulkindern in naturwissenschaftlichen Kontexten. In: *Unterrichtswissenschaft*, 39, 7-20.
- van de Pol, J., Volman, M. & Beishuizen, J. (2010). Scaffolding in teacher-student interaction: A decade of research. In: *Educational Psychology Review*, 22, 271-296.
- Vergaue, E., Gauffroy, C., Morsanyi, K., Dargy, I. & Barrouillet, P. (2013). Chronometric evidence for the dual-process mental models theory of conditionals. In: *Journal of Cognitive Psychology*, 25, 174-182.
- Wood, D., Bruner, J. S. & Ross, G. (1976). The role of tutoring in problem solving. In: *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 17, 89-100.

Christin Robisch, Westfälische Wilhelms-Universität Münster, Seminar für Didaktik des Sachunterrichts;  
E-Mail: christin.robisch@uni-muenster.de

Steffen Tröbst, Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik (IPN) Kiel;  
E-Mail: troebst@ipn.uni-kiel.de

Prof. Dr. Kornelia Möller, Westfälische Wilhelms-Universität Münster, Seminar für Didaktik des Sachunterrichts;  
E-Mail: molleko@uni-muenster.de