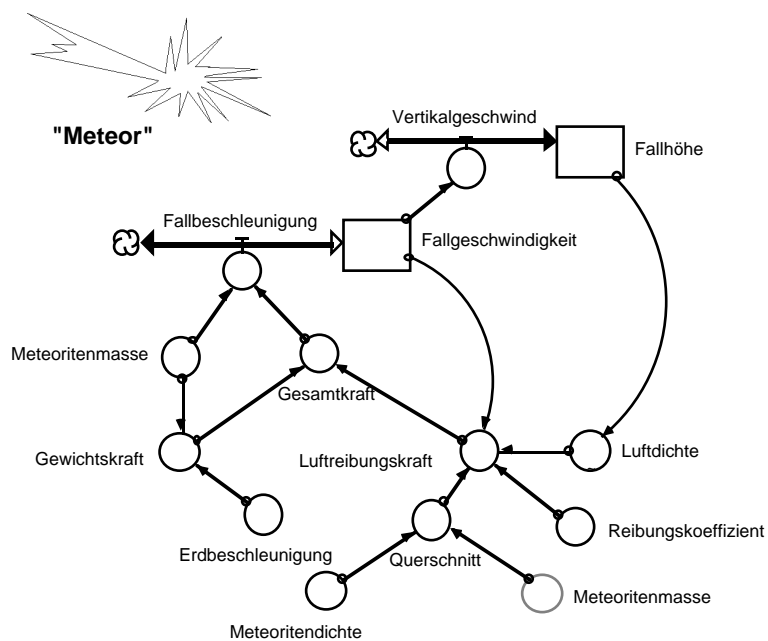


Abschlussbericht zum DFG-Projekt

Physiklernen mit Modellbildungssystemen

Förderung physikalischer Kompetenz und systemischen Denkens durch computergestützte Modellbildungssysteme



PD Dr. Horst Schecker (Universität Bremen)

Dr. Eckhard Klieme (Max-Planck-Institut für Bildungsforschung, Berlin)

Prof. Dr. Hans Niedderer (Universität Bremen)

Dipl.Psych. Judith Ebach (Institut für Bildungsforschung, Bonn)

Jörn Gerdes (Universität Bremen)

Bremen, Berlin und Bonn, im Juni 1999

Gliederung

1	Allgemeine Angaben	1
1.1	Formalia	1
1.2	Veröffentlichungen aus diesem Projekt.....	2
1.3	Tagungsbeiträge aus diesem Projekt.....	2
2	Arbeits- und Ergebnisbericht	3
2.1	Ausgangsfragen und Zielsetzung	3
2.2	Untersuchungsdesign und Treatment	3
2.3	Projektverlauf und -entwicklung.....	4
2.4	Testinstrumente	5
2.5	Skalenqualität.....	8
2.6	Ergebnisse im Bereich "Physikalische Kompetenz"	9
2.6.1	Physikalische Kompetenz.....	9
2.6.2	Zusammenhang zwischen Mechanikverständnis und Modellbildungsfähigkeit....	10
2.6.3	Konzeptuale physikalische Kompetenz	11
2.6.4	Fähigkeit zur Beschreibung von Bewegungsvorgängen.....	12
2.6.5	Fähigkeit zur Formalisierung systemdynamischer Modelle	13
2.7	Ergebnisse in den Bereichen "Systemisches Denken" und "Transfer"	13
2.7.1	Zum Konstrukt "Systemisches Denken"	13
2.7.2	Direkte Unterrichtseffekte.....	17
2.7.3	Transfereffekte.....	18
2.7.4	Profil der Lern- und Transfereffekte	21
2.8	Diskussion.....	21
2.9	Qualifikationsaspekte.....	24
3	Zusammenfassung.....	25
4	Anhang.....	26
4.1	Literatur	26
4.2	Anlagen	28

1 Allgemeine Angaben

1.1 Formalia

DFG-Geschäftszeichen:

NI 271/9-1 und NI 271/9-2 (Bremer Teilprojekt; Niedderer/Schecker)

KL 1057/1-1 und KL 1057/1-2 (Bonner Teilprojekt; Klieme)

Antragsteller

Hans NIEDDERER, Dr. rer. nat., Universitätsprofessor

geb. 23.12.1938, deutsch

dienstlich: Universität Bremen, Fachbereich Physik/Elektrotechnik, Institut für Didaktik der Physik, Postfach 33 04 40, 28334 Bremen

Tel: 0421 218 2484/4695, Fax: 0421 218 4015, e-mail: niedderer@physik.uni-bremen.de

privat: Schumannstr. 71b, 28790 Schwanewede; Tel: 04209 1008

Horst SCHECKER, Dr. rer. nat., Privatdozent

geb. 17.7.1955, deutsch

dienstlich: Universität Bremen, Fachbereich Physik/Elektrotechnik, Institut für Didaktik der Physik, Postfach 33 04 40, 28334 Bremen

Tel: 0421 218 2964/4695, Fax: 0421 218 4015, e-mail: schecker@physik.uni-bremen.de

privat: Fichtenweg 4, 27299 Langwedel; Tel.: 04232 1612

Eckhard KLIEME, Dr. phil., Wissenschaftlicher Mitarbeiter

geb. 8.10.1954, deutsch

dienstlich: Max-Planck-Institut für Bildungsforschung, Lentzeallee 94, 14195 Berlin

Tel. 030 82406 232 , Fax: 030 8249939 , email: klieme@mpib-berlin.mpg.de

bis 31.12.1997: Institut für Bildungsforschung e.V., (IBF), Heilsbachstr. 32, 53123 Bonn,

Tel: 0228 820 90 37, Fax: 0228 820 90 38

(Das psychologische Teilprojekt wurde am IBF verwaltet)

privat: Unter dem Klotenrech 6, 53347 Alfter; Tel: 02222 1361

Thema des Projekts

Förderung physikalischer Kompetenz und systemischen Denkens durch computergestützte Modellbildungssysteme

Berichtszeitraum

1.2.1996 - 28.2.1999

Förderungszeitraum insgesamt

1.2.1996 - 31.1.1999 (Bonner Teilprojekt)

1.3.1996 - 28.2.1999 (Bremer Teilprojekt)

1.2 Veröffentlichungen aus diesem Projekt

Die Punkte 2.6 und 2.7 des vorliegenden Berichts stellen die wesentlichen Ergebnisse des Projekts knapp dar. Zu einer Reihe von Ergebnissen liegen bereits Publikationen in Zeitschriften oder zumindest zur Einreichung vorgesehene Manuskripte vor, die wir diesem Bericht anlegen. Weitere Publikationen, die einen stärker zusammenfassenden Charakter tragen, sind in Vorbereitung.

Erschienen oder im Druck (Sonderdrucke beigelegt):

Schecker, H. & Gerdes, J.: Interviews über Experimente zu Bewegungsvorgängen. In: *Zeitschrift für Didaktik der Naturwissenschaften* **4** (1998), 3, 61-74.

Gerdes, J. & Schecker, H.: Der Force Concept Inventory — Ein diagnostischer Test zu Schülervorstellungen in der Mechanik. In: *Der Mathematische und Naturwissenschaftliche Unterricht* (im Druck).

Schecker, H. & Gerdes, J.: Messung von Konzeptualisierungsfähigkeit in der Mechanik — Zur Aussagekraft des Force Concept Inventory. In: *Zeitschrift für Didaktik der Naturwissenschaften* **5** (1999), H. 1, 75-89.

In Vorbereitung:

Ebach, J. & Gerdes, J., Klieme, E. & Schecker, H.: Modellbildungs- und Modellkonstruktionsprozesse (einzureichen bei: *Zeitschrift für Psychologie in Erziehung und Unterricht*).

Klieme, E., Schecker, H. & Niedderer, H.: Assessment of Physics Knowledge by Concept Mapping (to be submitted to: *Journal of Research in Science Teaching*).

1.3 Tagungsbeiträge aus diesem Projekt

Ebach, J. & Klieme, E.: Förderung systemischen Denkens durch den Einsatz eines Modellbildungssystems im Physikunterricht. Vortrag gehalten auf der 6. Tagung der Fachgruppe Pädagogische Psychologie in der Deutschen Gesellschaft für Psychologie in Frankfurt a.M. am 1.10.1997.

Ebach, J. & Klieme, E.: Förderung des systemischen Denkens durch Modellbildung und Simulation im Physikunterricht. Vortrag gehalten auf dem 41. Kongress der Deutschen Gesellschaft für Psychologie in Dresden am 29.9.1998.

Gerdes, J.: Physiklernen mit Modellbildungssystemen – Erste Ergebnisse. Vortrag gehalten auf der Jahrestagung der Gesellschaft für Didaktik der Chemie und Physik (GDCCP) in Potsdam am 23.9.1997. In: Behrendt, H. (Hrsg.): *Zur Didaktik der Chemie und Physik*. Alsbach: Leuchtturm 1998, 227-229.

Klieme, E.: Zur Nutzung von Analogien beim Bilden formaler Modelle. Poster-Beitrag auf der 54. Tagung der Arbeitsgemeinschaft für empirische pädagogische Forschung (AEPF) in Dresden, März 1997.

Klieme, E. & Ebach, J.: Physiklernen mit Modellbildungssystemen – Transfer im Bereich systemischen Denkens. Vortrag gehalten auf der 55. Tagung der Arbeitsgemeinschaft für empirische pädagogische Forschung (AEPF) in Berlin am 2.10.1997.

- Klieme E. & Ebach, J.: Modellbildung und Simulation im Physikunterricht: Transfer im Bereich des systemischen Denkens. Poster auf der Tagung der Arbeitsgemeinschaft für Empirische Pädagogische Forschung in Mannheim, 25.9.1998.
- Schecker, H.: Physiklernen mit Modellbildungssystemen – Forschungskonzeption. Vortrag gehalten auf der Jahrestagung der Gesellschaft für Didaktik der Chemie und Physik (GDPC) in Potsdam am 23.9.1997. In: Behrendt, H. (Hrsg.): Zur Didaktik der Chemie und Physik. Alsbach: Leuchtturm 1998, 230-232.
- Schecker, H. & Gerdes, J.: Physiklernen mit Modellbildungssystemen – Konzeptualisierungsfähigkeit in der Newtonschen Dynamik. Vortrag gehalten auf der 55. Tagung der Arbeitsgemeinschaft für Empirisch Pädagogische Forschung (AEPF) in Berlin am 2.10.1997.
- Schecker, H.: Entwicklung physikalischer Kompetenz bei unterrichtlicher Nutzung von Modellbildungssoftware. Vortrag gehalten auf der Jahrestagung der Gesellschaft für Didaktik der Chemie und Physik (GDPC) in Essen am 22.9.1998. In: Brechel, R. (Hrsg.): Zur Didaktik der Chemie und Physik. Alsbach: Leuchtturm 1999, 289-291.
- Schecker, H. & Gerdes, J.: Physiklernen mit Modellbildungssystemen. Poster auf der Tagung der Arbeitsgemeinschaft für Empirisch Pädagogische Forschung in Mannheim, 25.9.1998.

2 Arbeits- und Ergebnisbericht

2.1 Ausgangsfragen und Zielsetzung

Im Projekt werden die Effekte des Einsatzes grafikorientierter Modellbildungssysteme (MBS) im Physikunterricht untersucht. Die Leitfragen lauten:

- In welcher Weise fördert die Arbeit mit MBS bei Schülern der Sekundarstufe II die Fähigkeit zur Konzeptualisierung von Bewegungsvorgängen unter dem Einfluss von Kräften?
- Inwieweit fördert die längerfristige Arbeit mit einem MBS die Fähigkeit, physikalische Phänomene und komplexe lebensweltliche Phänomene in ihrem Wirkungsgefüge zu strukturieren sowie die Dynamik solcher Wirkungsgefüge zu erkennen und zu steuern?

Projektziele sind:

- eine Einschätzung der Wirkungen computergestützter, systemdynamischer Modellierung im Physikunterricht auf den Erwerb physikalischer Kompetenz,
- die Untersuchung der Transfers von Modellbildungsfähigkeiten, die in der Mechanik erworben wurden, auf andere physikalische und auf lebensweltliche Domänen,
- die Präzisierung und Validierung des Konstrukts "systemisches Denken".

2.2 Untersuchungsdesign und Treatment

Die empirische Untersuchung folgte einem Kontrollgruppen-Design. Die Versuchsgruppe besteht aus Schülern, die in zwei Leistungskursen Physik (5 Unterrichtswochenstunden) im Schuljahr 1996/97 wiederholt mit einem MBS arbeiteten. Schüler aus zwei parallelen Leistungskursen ohne Computereinsatz bilden die Kontrollgruppe. Wesentliche Voraussetzungen für die Auswahl der Versuchskurse waren die Vertrautheit der Kurslehrer mit dem didaktischen Konzept grafikorien-

tierter Modellbildung im Physikunterricht und praktische Erfahrungen aus vorhergehenden Schuljahren sowie eine im Physikunterricht verfügbare Computerausstattung für Klassenunterricht und Gruppenarbeit. Durch inhaltliche Absprachen mit den vier beteiligten Lehrern wurde die curriculare Vergleichbarkeit der Unterrichtsinhalte und -ziele im Halbjahr Mechanik sichergestellt. Ausschlaggebend war die Ähnlichkeit des Zeitaufwands für "Kinematik" und "Newtonsche Axiome". Auch die zentralen Experimente stimmten in den Kursen weitgehend überein.

Differenzierungen ergaben sich aus der quantitativen Behandlung von Reibungskräften in den Versuchskursen. Dies ist im Sinne des speziellen Treatments schlüssig und gewollt. In den Versuchskursen entfiel etwa ein Viertel der Unterrichtszeit auf Modellbildungseinheiten, einschließlich der damit verbundenen Experimente. Dafür wurden lehrerzentrierte Unterrichtsanteile entsprechend gekürzt. Das Treatment entsprach somit in beiden Versuchskursen dem angestrebten Konzept (s. dazu Schecker 1998). Unterschiede in den Schüleraktivitäten ergaben sich aus der höheren Schülerfrequenz in Versuchskurs 2 (VK2). Dadurch musste bei Gruppenunterricht teilweise in Dreiergruppen gearbeitet werden. In VK1 war dagegen auch Einzelarbeit möglich.

2.3 Projektverlauf und -entwicklung

Der Projektverlauf kann in vier Phasen eingeteilt werden:

1. Vorbereitung der empirischen Untersuchungen (Februar bis Juli 1996)

In diese Phase fällt die Entwicklung der Testinstrumente (s. Tab. 1) und die Auswahl der Kurse. Eine wissenschaftliche Mitarbeiterin und ein wissenschaftlicher Mitarbeiter wurden eingestellt und eingearbeitet. Für Instrumente zur Messung der physikalischen Konzeptualisierungsfähigkeit im Bereich Kräfte und Bewegungen konnte zu einem Teil auf internationale Tests zurückgegriffen werden, die übersetzt und adaptiert wurden. Andere Tests wurden grundlegend überarbeitet oder neu entwickelt. Letzteres gilt vor allem für die Tests zur Abgestuften Modellentwicklung, die in beiden Domänen (Physik und Ökonomie) zum Einsatz kamen. In der Domäne Ökonomie war die Entwicklung von Aufgabenkontexten aufwändig, die inhaltlich stimmig, für die Schüler nachvollziehbar und strukturell zugleich systemdynamisch modellierbar sind.

2. Durchführung der empirischen Untersuchungen (August 1996 bis Juli 1997)

Die empirischen Arbeiten begannen mit der Eingangserhebung in der ersten Versuchsgruppe und den beiden Kontrollgruppen. Die Tests nahmen hier wie auch in der Zwischen- und der Abschlusserhebung jeweils fünf Unterrichtsstunden in Anspruch und waren nur Dank der großen Kooperationsbereitschaft von Schülern und Lehrern zu realisieren. Der Versuchsplan wird in Tab. 1 gezeigt.

Im Verlauf des Schulhalbjahres 1996/97 verließen aus persönlichen Gründen drei Schüler den Versuchskurs. Angesichts der — allerdings bei Physik-Leistungskursen nicht ungewöhnlichen — geringen Ausgangsgröße von acht Schülern erschienen damit statistische Untersuchungen gefährdet. Drop-outs sind bei Feldstudien über lange Zeiträume besonders dann ein Problem wenn, wie im vorliegenden Fall, wegen der Rechner-Ausstattung und der notwendigen Erfahrung der Lehrkräfte mit Modellbildungssystemen die Auswahl sehr begrenzt ist. Es ergab sich jedoch die Möglichkeit, einen weiteren Versuchskurs (VK2) in die Untersuchung einzubeziehen. Daten aus diesem Kurs stehen ab der Zwischenerhebung zur Verfügung. Ein weiteres Problem der Feldsituation sind Datenausfälle aufgrund persönlicher (Erkrankung) oder schulorganisatorischer Gründe (z.B. Exkursionen oder Klausuren in anderen Fächern).

3. Datenauswertung (August 1997 bis April 1998)

Mit der Kodierung von Daten aus Multiple-choice Tests wurde bereits während der Durchführungsphase begonnen. Die Differenzierung und Operationalisierung des Konstrukts "systemisches Denken" nach vier Repräsentations- und zwei Anforderungsebenen (s. dazu Punkt 2.7.1) führte zu entsprechend differenzierten diagnostischen Instrumenten mit hohem Auswertungsaufwand. Einen Schwerpunkt bildete die Entwicklung geeigneter Skalen zur Beschreibung zentraler Aspekte des systemischen Denkens und der physikalischen Kompetenz.

4. Zusammenführung der Ergebnisse (ab Mai 1998)

Die letzte Projektphase war bzw. ist der Zusammenführung der Einzelergebnisse und qualitativen Detailstudien gewidmet. In Einzelfallstudien wurde der interaktive Prozess der Modellerstellung am Computer analysiert. Erste Ergebnisse sind in Zeitschriften publiziert, bzw. liegen als Manuskripte vor. Die Arbeiten an Veröffentlichungen werden nach Ende der Förderungsdauer fortgeführt.

2.4 Testinstrumente

Im folgenden beschreiben wir die eingesetzte Testbatterie und die jeweils wesentlichen Parameter, die später ausgewertet und diskutiert werden.

FCI Force Concept Inventory Test (in eigener Übersetzung; s. Gerdes & Schecker, im Druck)

Der weltweit verwendete Test zur Konzeptualisierungsfähigkeit in der Mechanik (Hestenes, Wells & Swackhamer 1992) besteht aus 29 multiple-choice Fragen mit jeweils 5 Antwortmöglichkeiten. Davon entspricht genau eine der Newtonschen Formulierung der Mechanik. Die restlichen Antwortmöglichkeiten sind Alltagsvorstellungen zugeordnet. Als wesentliche Parameter werden der Gesamtscore ausgewertet sowie der Teilscore zu sieben Items, die sich speziell mit dem zweiten Axiom zum Zusammenhang von Kraft und Bewegung befassen.

EXI Experimentalinterviews

teilstrukturierte Einzelinterviews von etwa 15 Minuten über einen demonstrierten Bewegungsvorgang. Die Schüler sollen einen Geschwindigkeitsverlauf beschreiben, bzw. halb-quantitativ prognostizieren und seine Ursachen erläutern. Ihre Aussagen in drei Interviewphasen werden auf Basis eines Tonbandmitschnitts auf einer sechsstufigen Rangskala eingestuft, die den Grad der Annäherung an die Newtonsche Beschreibung von Bewegungen unter dem Einfluss von Kräften erfasst. Diese Skala kann auf drei Hauptstufen reduziert werden. Die Inter-Rater-Reliabilität liegt je nach Interviewphase zwischen 0,80 und 0,92 (Übereinstimmung zweier Bewerter; Kendalls τ_b).

CM Concept mapping

Die Schüler legen ein map aus neun zentralen Begriffen der Mechanik, die auf Kärtchen vorgegeben werden. Sie sollen auf dem Tisch so angeordnet werden, dass eng zusammengehörende Begriffe nah beieinander liegen. Die Schüler übertragen diese Struktur dann auf ein Blatt Papier. Anschließend zeichnen sie Verbindungslinien zwischen solchen Begriffen, die ihrer Meinung nach direkt miteinander zusammenhängen. Ein Anwendungskontext wird nicht vorgegeben. Der Auswertungsschwerpunkt liegt auf der inhaltlichen Güte der map, wobei drei unterschiedliche Auswertungsverfahren verglichen werden: eine holistische Beurteilung durch Experten, eine analytische Beurteilung der Übereinstimmung zwischen Probanden-map und Experten-map sowie das Vorkommen fachlich zentraler Komponenten im Probanden-map.

	Instrument	Kürzel (Dauer)	Eingangserhebung (1 VK, 2 KK)	Zwischenerhebung (2 VK, 2 KK)	Abschlusserhebung (2 VK, 2 KK)	ergänzende Erhebungen
Physikalische Kompetenz	concept mapping	CM (15)	CM	CM		
	Force Concept Inventory	FCI (30)	FCI	FCI		
	Graphenverständnis	GRA (15)	GRA	GRA		
	Experimentalinterview Mechanik	EXI (15)	EXI <i>Fall von Papier- kegeln</i>	EXI <i>Beschleun. mit Gliederkette</i>	EXI <i>Elektronenbewe- gung im Kond.</i>	
	Abgestufte Modellerstellung und -interpretation Physik	AME (30)	AME <i>Fall von Papier- kegeln</i>	AME <i>Beschleun. mit Gliederkette</i>	AME <i>Abkühlung Tasse Kaffee</i>	AME <i>Kondensator- entladung (4-5/97)</i>
	Interaktive Modellentwicklung Physik (nur Versuchsgruppe)	IME (45)		IME <i>Beschleun. mit Gliederkette</i>	IME <i>Abkühlung Tasse Kaffee</i>	IME <i>Kondensator- entladung (4-5/97)</i>
Systemisches Denken	Abgestufte Modellerstellung und -interpretation Ökonomie	AME (45)	AME <i>Entwicklung der Mitarbeiterzahl</i>	AME <i>Ausbildungs- platzabgabe</i>	AME <i>Preiskalkulation für Jeans</i>	
	Interaktive Modellentwicklung Ökonomie	IME (45)		IME <i>Ausbildungs- platzabgabe</i>	IME <i>Preiskalkulation für Jeans</i>	
	Freie Modellerstellung	FME (20)	FME <i>Trinkwasser</i>	FME <i>Trinkwasser</i>		
	Schriftliche Modellinterpretation	SMI (20)	SMI <i>Absatzprognose CD-Brenner</i>	SMI <i>Absatzprognose Pay-TV</i>	SMI <i>Absatzprognose Internetanschluss (nur VG)</i>	
	Exploration und Steuerung eines dynamischen Systems	ESS (90)	ESS <i>Hunger in Nord- afrika</i>		ESS <i>Hunger in Nord- afrika</i>	
Begleitende In- strumente	Fragebogen zu computerbezo- genen Einstellungen (nur VG)	FCE (15)	FCE (VK 1)	FCE (VK 2)		
	Fähigkeit zum Umgang mit dem Modellbildungssystem (nur VG)	FBS (15)		FBS	FBS	FBS mehrfach in 11/1
	Intelligenztest nach Wilde	INT (10)			INT	

Tab. 1: Messinstrumente und deren Einsatzzeiten im Verlauf der Datenerhebung.

AME *Abgestufte Modellentwicklung*

Das Verfahren erfasst den Umgang mit System-Modellen auf unterschiedlichen Repräsentationsebenen, d.h. auf qualitativer, halb-quantitativer, quantitativer und systemdynamischer Ebene. Auf jeder Ebene wird sowohl die Erstellung eines Modells als auch die Interpretation eines vorliegenden Systemmodells verlangt. Um nicht mit wechselnden Inhalten arbeiten zu müssen, deren Effekte unkontrollierbar wären, wird ein Thema über alle Teilaufgaben hinweg durchgehalten. Um die Bearbeitungsschritte unabhängig voneinander zu halten, werden nach jedem Schritt Modelllösungen vorgegeben, von denen der Bearbeiter in den nachfolgenden Teilaufgaben ausgehen kann. Im ersten Schritt (qualitative Modellerstellung) werden ihm Begriffe vorgegeben, die er in einem concept map anordnen soll. In Schritt 2 wird ihm ein solches map vorgegeben, zu dem dann Schlussfolgerungen zu ziehen sind. Anhand des vorgegebenen concept map und eines ergänzenden Textes, der relevante Zusammenhänge aufzeigt, soll der Bearbeiter sodann ein Wirkungsdiagramm mit gerichteten Pfeilen und Vorzeichen zeichnen (halb-quantitative Modellerstellung). Schritt 3 beginnt mit der Vorlage eines fertigen Wirkungsdiagramms, zu dem wiederum interpretative Schlüsse gezogen werden müssen; anschließend sollen Gleichungen erstellt oder aus Alternativen ausgewählt werden, die dem Wirkungsdiagramm entsprechen (quantitative Modellerstellung). Schritt 4 schließlich beinhaltet interpretative Fragen zu einem vorgegebenen Gleichungssystem. Für Schüler der Versuchsgruppe schlossen sich — außer bei der Vorerhebung — Schritte der systemdynamischen Modellerstellung und -interpretation an, die wir unten als „Integrierte Modellerstellung“ (IME) beschreiben. Die Auswertung erfolgt in weiten Teilen nach einem eindeutigen Lösungsschlüssel, da viele Aufgaben mit gebundenen Antwortformaten vorgegeben waren; die erstellten Diagramme (concept maps, Wirkungsdiagramme und Stella-Modelle) wurden im Hinblick auf Strukturiertheit und (wie beim CM-Test) auf inhaltliche Güte ausgewertet.

Zu jedem Erhebungszeitraum wurden zwei AME-Verfahren entwickelt, eines in der Physik und eines aus der Transferdomäne, Ökonomie. Bei den Transferthemen wurde darauf geachtet, dass die Aufgaben ohne fachliches Vorwissen bearbeitet werden konnten. Themen waren in der Vorerhebung die Entwicklung der Mitarbeiterzahl eines Unternehmens, im Zwischentest die Veränderung des Ausbildungsplatzangebotes durch staatliche Fördermaßnahmen und im Abschlusstest die Entwicklung und Kalkulation des Preises für ein Massenprodukt.

Die Kontexte der Physik-AMEs stimmten in der Eingangs- und der Zwischenerhebung mit den Gegenständen der Experimentalinterviews überein. Für das Verständnis der Newtonschen Dynamik werden die inhaltliche Güte der auf der qualitativen und halb-quantitativen Repräsentationsebene angesiedelten Begriffsnetze bzw. Wirkungsnetze sowie die Fähigkeit zur quantitativen (gleichungsorientierten) Beschreibung von Bewegungsvorgängen herangezogen.

IME *Interaktive Modellentwicklung*

Für Schüler in den Versuchskursen schließt sich unmittelbar an Schritt 4 der AME unter Beibehaltung des Themas und der in der AME vorgegebenen Modellannahmen eine IME an. Die Schüler sollen hier ein lauffähiges Modell erstellen, welches eine vorgegebene Verlaufskurve der Zielgröße (z.B. Geschwindigkeit oder Betriebsvermögen) replizieren kann. Vor der interaktiven Arbeit am Computer sollen die vorgegebenen Modellgrößen hinsichtlich ihres systemdynamischen Typus (Zustand, Rate, Funktion) eingestuft werden. Als Bewertungsmaße werden die inhaltliche Güte der erstellten Modelle und ihre systemdynamische Differenziertheit betrachtet.

FME *Freie Modellerstellung*

Auch bei diesem Verfahren hat der Bearbeiter nach einer Textvorlage ein Wirkungsdiagramm mit gerichteten Pfeilen und Vorzeichen zu zeichnen. Ausgewertet werden wiederum Strukturiertheit

und inhaltliche Güte der Schülerzeichnung. Im Gegensatz zur AME wurde hier zu allen drei Erhebungszeitpunkten derselbe Text vorgegeben, um Lerneffekte ohne wechselnde Kontexteinflüsse abschätzen zu können. Das Referenzmodell zum Thema „Wasserverbrauch“ ist so komplex, dass keine Deckeneffekte auftreten.

SMI *Schriftliche Modellinterpretation*

Mit der SMI wird von den Schülern die Interpretation eines Modells auf halb-quantitativer, quantitativer und systemdynamischer Repräsentationsebene erwartet. Hier wurde zu den drei Erhebungszeitpunkten ein in seiner Struktur gleichbleibendes Modell zum Ansatz eines Produktes verwandt. So konnten auch in allen drei Versionen äquivalente Aufgabenstellungen vorgelegt werden. Die graphische Darstellung sowie die inhaltliche Einbettung des Modells wurden jedoch variiert.

ESS *Exploration und Steuerung eines dynamischen Systems*

Zur Eingangs- und Schlusserhebung wurden die Schüler gebeten, das dynamische System „Hunger in Nordafrika“ (vgl. Leutner 1992) zu explorieren und zu steuern. Diese Aufgabe stellt — für das Niveau von Sekundarschulen angepasst — eine gute Operationalisierung des „komplexen Problemlösens“ sensu Dörner (1983) dar. Der Bearbeiter ist mit einem intransparenten, dynamischen und stark vernetzten System konfrontiert, dessen strukturelles Gefüge und Gesetzmäßigkeiten er erkunden und in einer Steuerungsaufgabe zielorientiert beeinflussen soll. Im Fall des von uns verwendeten Computerprogramms beschreibt das System einen bäuerlichen Familienbetrieb in der Sahel-Zone mit seiner ökonomischen und ökologischen Umwelt. Erfasst wird, wie viele Jahre und mit welchem ökonomischen Erfolg der Bearbeiter seinen simulierten Betrieb führen kann, und (mit Hilfe eines Wissenstest) wie gut er die zugrundeliegenden Zusammenhänge und die quantitativen Relationen erkennt.

2.5 Skalenqualität

In den Tests wurde eine Vielzahl einzelner Indikatorvariablen erhoben. Jede Variable wurde einer Plausibilitätsprüfung unterzogen, d.h. es wurden ihre Häufigkeitsverteilung betrachtet und insbesondere die Minima und Maxima auf mögliche Ausreißer hin geprüft. Die Einzelitems der Verfahren sind häufig nominal skaliert (z.B. FCI und viele Items der AME). Der Zusammenfassung dieser Items zu Gesamtwerten (Anzahl der richtig gelösten Aufgaben) wird jedoch Intervallskalenniveau zugeschrieben, ebenso wie einer recht großen Anzahl der aus den Concept maps und Wirkungsnetzen gewonnenen Indikatorvariablen (z.B. Anzahl von Größen und Relationen). Für die aus diesen Variablen abgeleiteten Maße (z.B. Linkage, Dichte) kann demnach ebenfalls von Intervallskalenniveau ausgegangen werden. Einige Variablen sind eindeutig ordinal skaliert, so z.B. die Argumentationsniveaus in den physikalischen Experimentalinterviews oder auch der Typus des Modells in den AMEs und der FME.

Für die Auswahl geeigneter statistischer Verfahren galt es zu prüfen, ob für die intervallskalierten Variablen Normalverteiltheit angenommen werden kann. Es zeigte sich, dass das nicht durchgängig der Fall ist, und es daher für die statistischen Auswertungen vielfach sinnvoll ist, auf nicht-parametrische Verfahren zurückzugreifen.

Gruppenunterschiede zwischen Versuchs- und Kontrollgruppe werden mit dem Mann-Whitney U-Test für ordinale Daten aufgrund eines Vergleichs der Rangplätze untersucht. Angesichts der geringen und überdies ungleichen Stichprobengrößen setzen wir für einseitige Tests, an denen nur

der Versuchskurs 1 beteiligt ist (vgl. Tab. 2), eine konservative α -Grenze, ab der die Nullhypothese der Gruppengleichheit verworfen werden kann, von $\alpha=0,3$ an. Die zugelassene Irrtumswahrscheinlichkeit liegt damit deutlich höher als die übliche 1%- oder 5%-Niveaus. Bei unseren von kleinen Zahlen gekennzeichneten Statistiken ergibt sich aber nur so eine Gleichgewichtung statistischer Fehler erster und zweiter Art, d.h. der Gefahr, Ungleichheit der Gruppen zu behaupten, obwohl die Nullhypothese gilt bzw. Gleichheit zu behaupten, obwohl die Gruppen sich unterscheiden. Für den speziellen Fall eines zweiseitigen Tests, der nur in der Eingangserhebung zum Zuge kam, gilt die Grenze von $\alpha=0,4$. Für die Abschätzung eines sinnvollen α -Niveaus bedienen wir uns eines Programms (G-Power), das uns von E. Erdfelder (Universität Bonn) zur Verfügung gestellt wurde.

Für die Berechnung korrelativer Zusammenhänge verwenden wir Kendalls τ_b als Maß der Assoziation zwischen ordinalen Variablen. Metrische oder quasi-metrische Variablen (wie der FCI-Gesamtscore) werden damit im Skalenniveau abgesenkt.

2.6 Ergebnisse im Bereich "Physikalische Kompetenz"

Wir orientieren uns bei der Ergebnisdarstellung eng an den Hypothesen der Studie, die im Zwischenbericht (Schecker, Klieme & Niedderer 1997) präzisiert wurden. Zunächst soll jedoch das Konstrukt der physikalischen Kompetenz beleuchtet werden.

2.6.1 Physikalische Kompetenz

Im Projekt wird das Konstrukt der physikalischen Kompetenz auf einen bestimmten Inhaltsbereich bezogen: die begriffliche Strukturierung der Mechanik, speziell die Beschreibung des Zusammenhangs zwischen Kraft und Bewegung (besser: *Bewegungsänderung*). Ein weiterer, allerdings nachrangiger Aspekt ist die Fähigkeit zur Lösung quantitativer Aufgaben. Parameter für physikalische Kompetenz werden aus folgenden Tests abgeleitet:

- FCI: Gesamtscore (qualitatives Verständnis grundlegender mechanischer Begriffe und Prinzipien und Sub-Score zum zweiten Newtonschen Axiom ("Grundgesetz der Mechanik")).
- EXI: höchstes erreichtes Argumentationsniveau bei der Beschreibung bzw. Vorhersage eines Bewegungsvorgangs,
- CM: inhaltliche Güte der qualitativen Beziehungen zwischen Begriffen der Mechanik,
- AME: qualitative (concept map), halb-quantitative (Wirkungsnetz) und quantitative Aufgabenteile (algebraisch zu lösende Aufgaben).

Es wurde per Faktorenanalyse geprüft, inwieweit sich ein Faktor "mechanische Kompetenz" aus den verfügbaren Parametern konstruieren lässt. Damit lassen sich jedoch nur 20-30 % gemeinsame Varianz abdecken. Bei Mehrfaktoren-Lösungen werden tendenziell die einzelnen Tests reproduziert. Um die Zusammenhänge zwischen den Parametern für physikalische Kompetenz besser einschätzen zu können, wurden für den Zeitpunkt der Zwischenerhebung, d.h. nach dem Kurshalbjahr Mechanik, Rangkorrelationen zwischen den Testergebnissen berechnet. Vier Parameter weisen untereinander signifikante Zusammenhänge auf und können somit vorrangig für die weiteren Auswertungsschritte herangezogen werden (s. Tab. 2). Die inhaltliche Güte der Concept maps

steht in keinem Zusammenhang dazu. Bei den qualitativen und halbquantitativen Anteilen der AME sind die Ergebnisse uneinheitlich.

	FCI N2	EXI	AMEqn
FCI	,502***	,331**	,275*
FCI N2		,368*	,380**
EXI			,341*

Tab. 2: Korrelationen zwischen ausgewählten Tests zur physikalischen Kompetenz aus der Zwischenerhebung; N=33; Kendalls τ_b ; * $p < .05$, ** $p < .01$, *** $p < .001$. FCI: Gesamtscore im FCI, FCI N2: FCI-Subscore für Items zum 2. Newtonschen Axiom, EXI: höchstes erreichtes Argumentationsniveau im Experimentalinterview, AMEqn: Prozentsatz korrekt gelöster quantitativer Aufgaben.

2.6.2 Zusammenhang zwischen Mechanikverständnis und Modellbildungsfähigkeit

H1.1 Die Fähigkeit zur eigenständigen Konstruktion systemdynamischer Modelle zur Physik hängt mit der physikalischen Kompetenz der Schüler zusammen.

In Tab. 3 werden die oben ausgewählten Tests im Bereich der physikalischen Kompetenz folgenden Parametern der interaktiven Modellentwicklung gegenübergestellt:

- Qualität der Symbolisierung und Quantifizierung der Modellgrößen (IMEq)
- inhaltliche Güte (IMEi)
- Lauffähigkeit (IMEI)

	IMEq	IMEi	IMEI
FCI	,363*	,366*	,462**
FCI N2	,492*	,577***	,455**
EXI	,438*	,337	,330
AMEqn	,594*	,402*	,354

Tab. 3: Korrelationen zwischen ausgewählten Kompetenz-Parametern und Parametern über Leistungen bei der interaktiven Modellentwicklung (Zwischenerhebung); N=14.

Klammert man die Lauffähigkeit des Modells aus, welche bereits bei inhaltlich weniger entscheidenden Fehlern beeinträchtigt sein kann, so wird die Hypothese H1.1 bestätigt: Schüler, die erfolgreich Modelle zu Bewegungsvorgängen konstruieren, verfügen über eine hohe physikalische Kompetenz in diesem Bereich. Damit ist eine wichtige Voraussetzung gegeben, um die weiteren Hypothesen sinnvoll testen zu können.

2.6.3 Konzeptuale physikalische Kompetenz

H1.2: Im Vergleich zur Arbeit ohne MBS erreichen Schüler durch die Arbeit mit MBS in erhöhtem Maße solche Stufen konzeptueller physikalischer Kompetenz, die durch eine Fähigkeit zur begrifflich-qualitativen bzw. halb-quantitative Analyse des zu untersuchenden Vorgangs gekennzeichnet sind.

Hypothese 1.2 geht von einem besseren qualitativen und halbquantitativen Verständnis der Mechanik in der Versuchsgruppe aus. Spezielle Wirkungen im quantitativen Bereich werden dagegen nicht erwartet. Wir greifen für die Prüfung auf die FCI-Ergebnisse zurück.

Die Analysen führen zu uneinheitlichen Ergebnissen, welche die Hypothese nicht bestätigen. Bei absolut gesehen geringen Leistungsdifferenzen (VG: 48%, KG: 53% korrekte Lösungen) schneiden die Schüler der Versuchsgruppe beim FCI-Gesamtergebnis nach dem Mechanikunterricht (*Lernzustand*) signifikant schlechter ab als die Kontrollgruppe (Mann-Whitney U-Test; $p=0.09$ und damit deutlich unterhalb der Fehlergrenze von 0,3, so dass die Nullhypothese der Gruppengleichheit verworfen wird). Der gewichtete *Lernzuwachs* (realisierter Leistungszuwachs im Verhältnis zum maximal möglichen Zuwachs, s. Hake 1998) war in Versuchskurs 1 dagegen höher als in der Kontrollgruppe. (Für Versuchskurs 2 kann der Lernzuwachs aufgrund fehlender Vortestergebnisse nicht berechnet werden.)

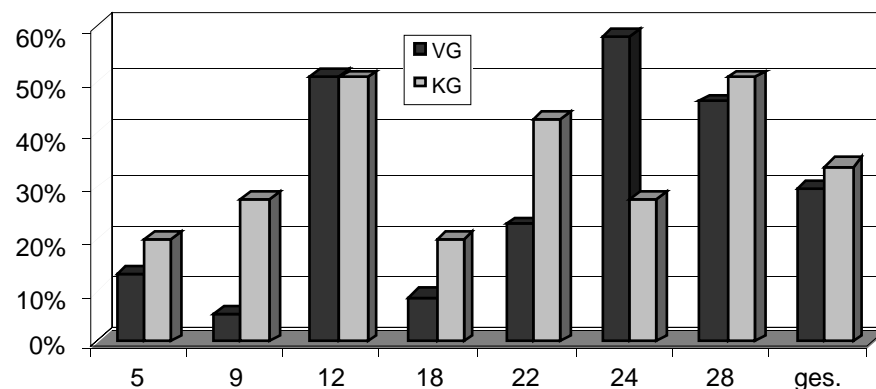


Abb. 1: Lösungshäufigkeiten der Aufgaben in der Skala „Kraftverständnis“ (Nachttest); Abzisse: Item-Nr. des FCI, Ordinate: Anteil richtiger Lösungen; VG: N=24, KG: N=27.

Eine Profilanalyse des uns besonders interessierenden Verständnisses des Zusammenhangs zwischen Kraft und Bewegungsänderung ist wegen der nicht befriedigenden inneren Konsistenz der Subskalen des FCI schwierig (unbefriedigende Skalenreliabilitäten gemessen mit Cronbachs α). Betrachtet man dennoch die für den Zusammenhang zwischen Kraft und Bewegung inhaltlich relevante Aufgabengruppe des FCI (s. Abb. 1), so erreichten Versuchs- und Kontrollgruppe im Mittel gleiche Ergebnisse. Sowohl in der Versuchs- als auch in der Kontrollgruppe erreichte der Unterricht einen Mischzustand zwischen Denkweisen nach dem Muster der Impetustheorie ("Kraft" als ein Körpern innenwohnender Antrieb) und newtonschen Denkweisen, wobei die "gespeicherte Kraft"-Vorstellung überwog. Konsequenterweise newtonsche Denker waren in beiden Gruppen rar.

Es bestehen — hier erwartungsgemäß — keine signifikanten Leistungsunterschiede (U-Test) bei den gleichungsorientierten (Rechen-) Aufgaben, die im quantitativen Teil der AME getestet wur-

den. Das gilt sowohl für den Lernzustand am Ende des Halbjahres Mechanik wie auch für den Lernzuwachs von der Eingangs- zur Zwischenerhebung.

(Eine ausführliche Darstellung der Ergebnisse zu Punkt 2.6.3 findet sich in Schecker & Gerdes 1999.)

2.6.4 Fähigkeit zur Beschreibung von Bewegungsvorgängen

Hypothese 1.3 postuliert in der Versuchsgruppe eine bessere Fähigkeit zur halbquantitativen Beschreibung bzw. Vorhersage von Bewegungsvorgängen, die sich in einer bestimmten Überlegungskette manifestiert, welche bei der erfolgreichen Konstruktion von Modellen angewendet und eingeübt wird:

H1.3: Schüler, die mehrfach mit MBS im Bereich Mechanik gearbeitet haben, wenden im Bereich „Bewegungen und Kräfte“ die folgende Problemlösungsstrategie auch in Situationen ohne Computereinsatz häufiger und konsequenter an als Schüler aus Kursen ohne MBS-Einsatz:

1. Bestimmung der auf einen bestimmten Körper einwirkenden Kräfte und der Einflüsse, denen diese Kräfte unterliegen (Komplexität).
2. Ermittlung der *Geschwindigkeitsänderung* aus der *Wirkung der Beschleunigung* und der Änderung der Position des Körpers aus der Wirkung der Geschwindigkeit (zeitliche Dynamik).
3. Berücksichtigung der Rückwirkungen einer veränderten Geschwindigkeit oder Position auf bestimmte Einzelkräfte (Rückkopplung).

Das Verfahren der Experimentalinterviews war spezifisch auf die Überprüfung dieser Annahme ausgerichtet. Sie wird durch die signifikant höheren Argumentationsniveaus in der Zwischenerhebung (Fahrbahnexperiment mit nicht konstanter Kraft) gestützt. 27% der Probanden in der Versuchsgruppe und 55% in der Kontrollgruppe fanden keine oder nur eine fehlerhafte kinematische Beschreibung des Fahrbahnexperiments. Ein korrekter Verlauf des Geschwindigkeits-Zeit-Graphen wurde von 27% (VG) bzw. 15% (KG) skizziert und erläutert. Weiterführend verbanden damit 47% bzw. 30% eine korrekte dynamische Erklärung. Geschlossene Argumentationsketten mit Rückkopplungsschleifen kamen ausschließlich in der Versuchsgruppe vor. Die Unterschiede zwischen den Qualitäten der maximal erreichten Argumentationsniveaus in Versuchs- und Kontrollgruppe sind statistisch signifikant (Mann-Whitney U-Test: $p=0,07$).

Für den Wissenszuwachs können nur die drei Kurse herangezogen werden, aus denen auch Daten aus der Eingangserhebung vorliegen (VK1, KK1 und KK2). Bei diesen Schülern war die physikalische Kompetenz — gemessen an Experimentalinterviews und Force Concept Inventory — am Beginn von 11/1 (Eingangserhebung) vergleichbar. Der Wissenszuwachs der Schüler in VK1 ist signifikant höher als der in der Vergleichsgruppe. Diese Ergebnisse der quantitativen Analyse auf Basis des Ratings der Argumentationsniveaus deckt sich mit früheren Ergebnissen qualitativer Studien (s. Schecker & Niedderer, 1991).

Beim Transfer auf Bewegungsvorgängen mit nicht-mechanischen Kräften (Elektronenablenkrohre) zeigen sich hingegen keine Vorteile der Versuchsgruppe. Dies beeinträchtigt die Hypothese nur bedingt. Das Experimentalinterview der Abschlusserhebung trägt zur Überprüfung des Beitrags von Modellbildungssystemen zum vertieften Verständnis des Zusammenhangs von Kraft und

Bewegung weniger bei als der Zwischenerhebung. Die Lösung erforderte keine Überlegungen über variable Kräfte oder Rückkopplungen.

(Eine ausführliche Darstellung der Ergebnisse zu Punkt 2.6.4 findet sich in Schecker & Gerdes 1998.)

2.6.5 Fähigkeit zur Formalisierung systemdynamischer Modelle

Hypothese 3.1 geht von einer habitualisierten Verwendung zentraler Zustandsgröße-Rate-Paare bei neuen Modellen innerhalb eines vertrauten Themenbereichs aus, während beim Transfer in andere Domänen die Identifizierung der dort zu verwendenden Paare deutlich schwieriger ist.

H3.1: Bei der systemdynamischen Formalisierung von Größen greifen die Schüler in mechanischen Kontexten auf bereits bekannte und wiederverwendbare Modell-Substrukturen zurück. In anderen physikalischen Kontexten benötigt die Identifizierung von Raten und Zustandsgrößen mehr Zeit und sie gelingt dort erst nach einer höheren Zahl von fehlerhaften Ansätzen als im Bereich der Mechanik.

Grundlage der Überprüfung dieser Hypothese ist die qualitative Einzelfallanalyse kodierter Verlaufsprotokolle von Videomitschnitten der interaktiven Modellbildung am Computer. Die Annahme, dass Schüler bei mechanischen Kontexten auf bekannte Substrukturen zurückgreifen, kann bestätigt werden. Besonders bei der IME Physik in der Zwischenerhebung zeigt sich, dass Schüler zu Beginn der Modellbildungsphase stets die allgemeingültige Substruktur „Newtonsches Kraftmodell“ modellieren und dann die phänomenspezifische Peripherie um diesen Kern herum aufbauen. Dieser Trend setzt sich in der Modellieraufgabe „Zweidimensionale Wurfbewegung mit Luftreibung“ fort.

Die Annahme, dass Schülern die Identifizierung von Raten und Zustandsgrößen bei nicht-mechanischen Kontexten schwerer fällt, kann nicht eindeutig bestätigt werden. Schülern mit sehr guten Modellbildungsfähigkeiten gelingt auch in anderen Themenbereichen schnell die Identifizierung der zentralen Raten und Zustandsgrößen, während leistungsschwächeren Schülern dieses erst nach einer längeren trial-and-error Phase gelingt, in der sie die Qualität ihres Modells immer wieder mit Testläufen überprüft haben.

(Eine ausführliche Darstellung der Ergebnisse zu Punkt 2.6.5 findet sich in Ebach, Gerdes, Klieme & Schecker (in Vorbereitung), s. Anlage).

2.7 Ergebnisse in den Bereichen "Systemisches Denken" und "Transfer"

Auch in diesem Abschnitt folgen wir den Hypothesen, die im Zwischenbericht (Schecker, Klieme & Niedderer 1997) in überarbeiteter Form vorgelegt wurden.

2.7.1 Zum Konstrukt "Systemisches Denken"

H 2.1 Zu jeder Kombination einer Anforderungsebene (Modellentwicklung oder –interpretation) mit einer Repräsentationsebene (qualitativ, halb-quantitativ, quantitativ, systemdynamisch) lässt sich ein konsistentes Maß des systemischen Denkens definieren.

Bei der Definition dieser Maße haben wir über verschiedene Untersuchungsinstrumente, Kriterien des systemischen Denkens und Maßzahlen (Indikatoren) hinweg aggregiert. Dieses Vorgehen soll hier am Beispiel der halb-quantitativen Modellerstellung erläutert werden. Operationalisiert wurde dieser Teilaspekt des systemischen Denkens durch die Aufgabe, auf der Basis einer verbalen Beschreibung eines Wirkungsgefüges ein Wirkungsdiagramm zu zeichnen. Diese Aufgabe ist in der AME und der FME enthalten. Auswertungskriterien waren Strukturiertheit, inhaltliche Güte und systemdynamische Differenziertheit. Die Schülerdiagramme wurden zunächst mittels mehrerer Indikatoren bewertet:

- Typus des Modells (von linearen Ketten bis zu Netzdiagrammen), maximale Länge einer Verbindung zwischen zwei Größen, *linkage* (Anzahl der Relationen/Anzahl der Größen) sowie Anzahl der unverbundenen Teile (umgepolt) als Indikatoren der Strukturiertheit des Diagramms,
- ein Index für die Übereinstimmung mit dem Expertendiagramm als Indikator der inhaltlichen Güte und schließlich
- die Anzahl der Elemente des Diagramms, die als Zustandsgrößen erkennbar (d.h., mit einer Änderungsrate verknüpft) sind, als Indikator für die systemdynamische Differenziertheit.

Diese zwölf Indikatoren (je sechs aus AME und FME) wurden in einem Skalenwert „halb-quantitative Modellerstellung“ integriert.

Skala	Physik			Ökonomie		
	Vor- erhebung	Zwischen- erhebung	Abschluss- erhebung	Vor- erhebung	Zwischen- erhebung	Abschluss- erhebung
qualitative ME	.68 (5)	.71 (5)	.76 (5)	.74 (5)	.84 (5)	.70 (5)
halb-quant. ME	.81 (6)	.81 (6)	.81 (6)	.86 (12)	.86 (12)	.87 (6)
halb-quant. MI	.53 (9)	.68 (11)	.46 (8)	.35 (15)	.52 (16)	.62 (15)
quantitative ME	.35 (7)	.63 (6)	.38 (5)	.00 (3)	.67 (6)	.65 (6)
quantitative MI	.59 (3)	.34 (3)	.44 (2)	.57 (6)	.85 (9)	.71 (5)
systemdyn. ME	-	.80 (9)	.71 (9)	-	.92 (9)	.89 (9)

Tab. 4: Konsistenzwerte (Cronbachs α) für Skalen des systemischen Denkens. In Klammern: Zahl der Einzelindikatoren. Basis: Auswertungen der Arbeitsschritte der AME sowie (für einzelne Skalen im Bereich der Ökonomie) FME und SMI. Fallzahlen: $17 < n < 50$, Median=43. Fettgedruckt = keine signifikante Abweichung von der Normalverteilung. ME = Modellerstellung, MI = Modellinterpretation.

Zur Überprüfung der Konsistenz von Indikatoren und zu ihrer Integration in einer Skala stehen grundsätzlich unterschiedliche Verfahren zur Verfügung wie z.B. die Analyse latenter Klassen, die Faktorenanalyse oder verschiedene testtheoretische Modelle. Sie sind jedoch hier angesichts der

Fallzahlen von maximal 50, der logischen Abhängigkeiten zwischen den Indikatoren und der unterschiedlichen Verteilungsformen nicht sinnvoll anwendbar. Wir entschieden uns, die Aggregation der Basisindikatoren auf dem Modell der Klassischen Testtheorie aufzubauen, das als universelle und relativ robuste Theorie von Messfehlern angesehen werden kann (Krauth 1995). Als Konsistenzmaß verwenden wir Cronbachs α . Der Skalenwert errechnet sich als Summe der Einzelindikatoren, die zuvor standardisiert wurden und somit gleichgewichtig eingehen. Die Skalenwerte wurden sodann mittels eines nonparametrischen Tests auf Normalverteiltheit geprüft.

Tab. 4 gibt die Konsistenzwerte der je 5 bis 6 Skalen für Physik- und Ökonomie-Aufgaben zu den drei Messzeitpunkten wieder. Die Konsistenzwerte für die Erstellung von concept maps, Wirkungsdiagrammen und systemdynamischen Modellen sind ausnahmslos gut. Problematisch sind zu einen die Messungen im quantitativen Bereich der Physik-Aufgaben; hier interferiert vermutlich das systemische Denken mit spezifischen Komponenten des physikalischen Wissens (vgl. Abschnitt 2.6.1). Zum anderen fallen einige Kennwerte der Ökonomie-Aufgaben aus der Vorerhebung relativ schlecht aus; hier gab es bei dem neuartigen Verfahren AME z.T. Verständnisprobleme. Insgesamt erlauben die ermittelten Kennwerte jedoch, die Skalenwerte als Basis von Gruppenvergleichen zu nutzen. Dabei lassen sich auch statistische Verfahren nutzen, die Normalverteilungen annehmen.

H2.2 Innerhalb einer Domäne sind die in H2.1 angesprochenen Maße des systemischen Denkens korreliert und bilden eine Gesamtskala „Qualität der Modellerstellung und –interpretation in der Domäne xy“.

Zur Prüfung dieser Hypothese ermittelten wir für jede Spalte in Tab. 5 die Interkorrelationen der Skalen und berechneten, wieviel Varianz der erste gemeinsame Faktor einer Hauptkomponentenanalyse zu erklären vermag. Mittels des Bartlett-Tests auf Sphärizität wurde überprüft, ob die Matrix der Interkorrelationen von einer Einheitsmatrix unterschieden werden kann. Aus Gründen der Vergleichbarkeit wurde bei diesen Analysen zunächst die systemdynamische Ebene ausgelassen.

	Physik			Ökonomie		
	Vor- erhebung	Zwischen- erhebung	Abschluss- erhebung	Vor- erhebung	Zwischen- erhebung	Abschluss- erhebung
sign. Korrelationen ¹	1 von 10	3 von 10	5 von 10	2 von 10	5 von 10	7 von 10
1 Faktor	33 %	40 %	39 %	43 %	45 %	52 %
2 Faktoren	63 %	60 %	63 %	65 %	66 %	72 %
Zweiter Faktor	+ qualit. - QUANT. MI	- quant. ME +quant. MI	quantitat. Ebene	quantitat. Ebene	quantitat. ME	qualitative Ebene
Bartlett α	.19	.00	.00	.37	.01	.00

Tab. 5: Ergebnisse von Korrelations- und Faktorenanalysen zum systemischen Denken.

¹ Kriterium: $p < .038$. Bei $\alpha = .1528$ und $n = 43$ ist sichergestellt, dass der beta-Fehler für die Entdeckung einer Korrelation von mittlerer Effekstärke nicht größer ist als der alpha-Fehler. Da jede Skala in 4 Korrelationen eingeht, führt alpha-Adjustierung auf das angegebene Kriterium.

Zum Zeitpunkt der Vorerhebung kann demnach weder in der Physik noch in der Transferdomäne von einer einheitlichen Kompetenzdimension ausgegangen werden. Im Verlauf der Untersuchung wird jedoch der Zusammenhang der jeweils fünf Skalen immer enger. Den Kern des ersten gemeinsamen Faktors bilden durchgängig die halb-quantitative Erstellung und Interpretation von Modellen, also das Denken in Wirkungsdiagrammen. Explorative Faktorenanalysen identifizieren jeweils einen zweiten Faktor, der zumeist spezifische Kompetenzen im quantitativen Bereich erfasst. Bezieht man die systemdynamische Modellerstellung mit ein, so bleiben die Varianzaufklärung und die Zahl der signifikanten Korrelationen im wesentlichen unverändert.

Die Homogenität des Konstrukts „systemisches Denken“ ist also vor allem dadurch in Frage gestellt, dass wir eine quantitative Ebene berücksichtigt haben. Ursache könnte – vor allem in der Physik – die Rolle spezifischer Fähigkeiten im Umgang mit mathematischen Formalismen und Größengleichungen sein, aber auch die mangelnde Zuverlässigkeit unserer Aufgaben zum Erstellen und Auswerten von Gleichungen (s.o.). Jedenfalls ist das Ein-Faktor-Modell durchweg gut angepasst (erklärter Varianzanteil über 50 %), wenn man die quantitative Ebene herausnimmt. Im nicht-quantitativen Bereich - im Umgang mit Begriffsnetzen, Wirkungsdiagrammen und graphisch repräsentierten STELLA-Modellen - lässt sich systemisches Denken identifizieren.

Aus diesen Strukturanalysen ziehen wir die Konsequenz, das systemische Denken parallel auf zwei Arten zu messen: Zum einen durch die erste Hauptkomponente der fünf Teilskalen als bester Approximation der Gesamtleistung auf qualitativer, halb-quantitativer und quantitativer Ebene; zum anderen durch die Summe aller nicht-quantitativen Indikatoren als Maß für den Kernbereich des systemischen Denkens, das Denken in Begriffsnetzen und Wirkungsdiagrammen.

H2.3 Die Güte der Zielannäherung bei der Steuerung eines komplexen Systems (Aufgabentyp ESS) ist kein Indikator für die unter H2.2. definierte Gesamtskala.

In der Tat korreliert die Güte der Zielannäherung sowohl bei der Vorerhebung als auch bei der Nacherhebung nicht signifikant mit der jeweiligen Gesamtskala für Modellerstellung und –interpretation im Bereich der Ökonomie ($\tau = .05$, $p = .40$ bzw. $\tau = .16$, $p = .17$). Allerdings gibt es spezifische Zusammenhänge mit einzelnen Skalen: Je genauer zu Beginn des Erhebungszeitraums das System „Hunger in Nordafrika“ gesteuert werden konnte, desto besser waren die Leistungen im qualitativen und halb-quantitativen Kernbereich des systemischen Denkens ($\tau = .27$, $p = .016$). Bei der Wiederholung am Ende des Schuljahres war dieser Zusammenhang völlig verschwunden ($\tau = -.03$, $p = .42$); jetzt korrelierte die Steuerungsleistung signifikant mit der systemdynamischen Modellerstellung ($\tau = .45$, $p = .014$). Vermutlich stellt das Computerspiel bei der Wiederholung andere Anforderungen: Jetzt unterscheiden sich die Probanden nicht mehr danach, wie gut sie strukturelle Zusammenhänge im simulierten Szenario erschließen, sondern nach ihrer Fähigkeit, das Gefüge der Eingabeparameter zu optimieren.

H2.4 Zwischen den in H2.2 definierten Gesamtskalen für unterschiedliche Domänen bestehen positive Zusammenhänge; diese sind enger als der Zusammenhang mit externen Variablen wie z.B. Intelligenz und physikalisches Wissen.

Die folgende Tabelle demonstriert, dass der Erwerb systemischen Denkens im Bereich der Physik mit dem Erwerb systemischen Denken in der Transferdomäne zusammenhängt. Allerdings ist die Fähigkeit zum systemischen Denken auch vom fachlichen Wissen abhängig und somit domänenspezifisch geprägt: Von der allgemeinen Intelligenz lassen sich die hier erfassten Fähigkeiten zwar

klar abgrenzen (auch die nicht tabellierten Korrelationen zwischen Ökonomie-Skalen und Intelligenz sind niedrig), nicht jedoch von fachspezifischem Wissen (FCI-Wert). Systemisches Denken muss daher – anders als in Hypothese 2.4 vermutet – als einer Kombination zwischen domänenübergreifenden Kompetenzen und domänenspezifischem Verständnis konzipiert werden.²

Messzeitpunkt	Physik-Skala	Skala in der Transferdomäne Ökonomie		FCI (Vor- u. Zwischenerhebung) bzw. Intelligenz (Abschluss erh.)
		Gesamt	systemdynamisch.	
Vorerhebung	Gesamt	.273 (.121)	-	-.075 (.323)
Zwischenerh.	Gesamt	.438 (.003)	.477 (.021)	.525 (.000)
	Systemdynamische Modellerstellung	.515 (.010)	.209 (.086)	.529 (.002)
Abschluss erh.	Gesamt	.439 (.004)	.414 (.051)	.143 (.110)
	Systemdynamische Modellerstellung	.412 (.032)	.449 (.029)	-.008 (.483)

Tab. 6: Korrelationen (Kendalls τ) und Signifikanzgrenzen (in Klammern) für die Korrelation von Skalen des systemischen Denkens in der Physik mit systemischem Denken in der Transferdomäne sowie mit physikalischem Wissen und Intelligenz (Gesamt= Erste Hauptkomponente aus qualitativer, halb-quantitativer und quantitativer Ebene).

2.7.2 Direkte Unterrichtseffekte

H2.5 Schüler, die im Mechanikunterricht mit einem Modellbildungssystem gearbeitet haben, zeigen im Vergleich zu Schülern mit herkömmlichem Unterricht höhere Fähigkeiten beim systemischen Denken in dieser Domäne.

Wir prüfen diese Hypothese anhand der Kennwerte für das systemische Denken bei der Physik-AME in der Zwischenerhebung. Die Schüler der Versuchsgruppe sind im Gesamtwert des systemischen Denkens ($p=.116$ eins.³) und vor allem im Kernbereich ($p=.031$) signifikant besser als die Schüler der Kontrollgruppe. Allerdings galt dies auch schon im Vortest ($p=.029$ bzw. $p=.021$). Die Vortestwerte müssen also „auspartialisieren“ werden. Angesichts der geringen Fallzahlen – ganze 5 Schüler in der Versuchsgruppe und 15 in der Kontrollgruppe absolvierten bei beiden Testterminen das AME-Verfahren – können hierzu multivariate Verfahren nicht angewandt werden. Ersatzweise ermitteln wir für jeden Schüler die Differenz der Skalenwerte in Vor- und

² Völlig analoge Befunde ergeben sich, wenn man anstelle der „Gesamt“-Skalen die Skalen für den nicht-quantitativen Kernbereich des systemischen Denkens untersucht.

³ Die Signifikanzgrenze wurde hier und bei folgenden Tests mittels des Programms GPOWER so festgelegt, dass bei einer mittleren Effekstärke von $d=.5$ alpha- und beta-Fehler gleich unwahrscheinlich sind.

Zwischenerhebung und prüfen sodann mit einem non-parametrischen Verfahren (Mann-Whitney-Test), ob die Versuchsschüler – wie vermutet – höhere Zuwächse haben als die Kontrollschüler.

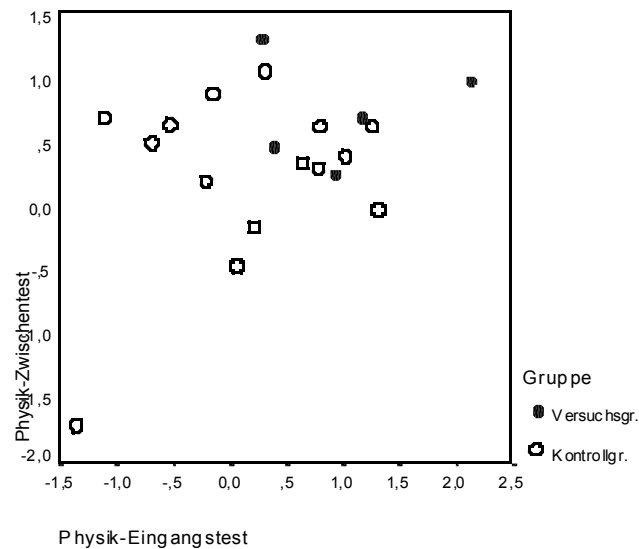


Abb. 2: Entwicklung der Gesamtleistung beim systemischen Denken im Bereich der Physik von der Eingangs- zur Zwischenerhebung.

Tatsächlich finden wir in der Versuchsgruppe ($n=5$) *niedrigere* Zuwächse als in der Kontrollgruppe. Die Hypothese muss daher zurückgewiesen werden⁴. Die obige Abbildung hilft, diesen Befund zu verstehen. Sie zeigt, dass Versuchs- und Kontrollgruppe beim zweiten Testtermin Werte erreichen, die jeweils unabhängig sind von den Ergebnissen der ersten Erhebung ($\tau = .01$ in der KG und $.00$ in der VG; $p = .48$ bzw. $.50$). Gerade die eingangs schlechteren KG-Schüler können also beim Zwischentest ihre Leistung deutlich verbessern. Dies könnte ein bloßer Testwiederholungseffekt sein. Entscheidend ist, daß die Versuchsgruppe trotz des systemdynamischen Trainings nicht in der Lage war, den Vorsprung, der schon im Vortest bestand, beizubehalten.

2.7.3 Transfereffekte

H2.6 Schüler, die im Mechanikunterricht mit einem Modellbildungssystem gearbeitet haben, zeigen im Vergleich zu Schülern mit herkömmlichem Unterricht höhere Fähigkeiten beim systemischen Denken zu Phänomenen aus anderen physikalischen Domänen (Elektrostatik, Thermodynamik).

Diese Hypothese bezieht sich auf den Übergang vom Zwischentest, bei dem systemisches Denken mit einer Mechanik-Aufgabe erfasst wurde, zum Abschlusstest mit einer Aufgabe aus der Thermodynamik. Hier hat ein Transfer stattgefunden, d.h. die thermodynamische Aufgabe wurde aufbauend auf den Mechanik-bezogenen Kompetenzen gelöst, wie die Korrelation über die Messzeitpunkte hinweg für Kontroll- und Versuchsgruppe zeigt ($\tau = .244$ ($p=.05$) bzw. $.455$ ($p=.02$)).

⁴ Der Befund bleibt gültig, wenn man nur den Kernbereich des systemischen Denkens betrachtet.

Und in der Versuchsgruppe gelang der Transfer tatsächlich besser als in der Kontrollgruppe (Mann-Whitney U-Test für Unterschiede der Zuwächse, $p=.174$)⁵.

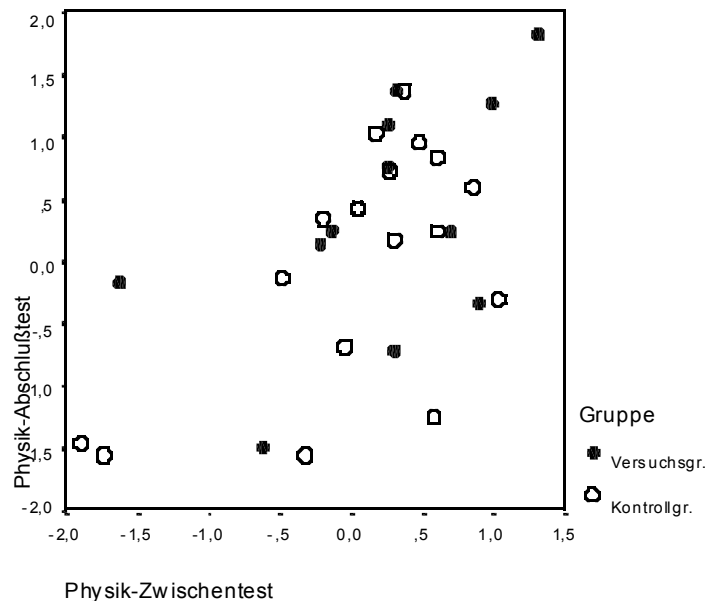


Abb. 3: Entwicklung der Gesamtleistung beim systemischen Denken von der Zwischenerhebung (Thema aus der Mechanik) zur Abschlusserhebung (Thema aus der Thermodynamik).

H2.7 Schüler, die im Mechanikunterricht mit einem Modellbildungssystem gearbeitet haben, zeigen im Vergleich zu Schülern mit herkömmlichem Unterricht höhere Fähigkeiten beim systemischen Denken im Kontext alltagsbezogener ökonomischer Fragestellungen.

Diese Hypothese vergleicht in ähnlicher Weise wie die beiden vorausgegangenen die Zuwächse im systemischen Denken bei Kontroll- und Versuchsschülern, allerdings jetzt in der Domäne der Ökonomie. Die statistischen Tests ergeben folgende Befunde:

- Im Laufe des ersten Halbjahres errang die Versuchsgruppe keinen Vorsprung beim systemischen Denken in der Transferdomäne – im Gegenteil: unter den wenigen Versuchspersonen, die mit einem vollständigen Datensatz einbezogen werden können, ergab sich ein höherer Zuwachs für die Kontrollschüler ($p=.334$). Beschränkt man sich jedoch auf den Kernbereich des systemischen Denkens, insbesondere die halb-quantitative Ebene, so ist die Versuchsgruppe eindeutig im Vorteil ($p=.170$ für halb-quant. Modellerstellung, $.176$ für Modellinterpretation, $p=.341$ für den qualitativen und halb-quantitativen Kernbereich).
- Im Lauf des zweiten Halbjahres fällt die Versuchsgruppe im Gesamtbereich des systemischen Denkens ($p=.139$) und insbesondere im nicht-quantitativen Bereich ($p=.039$) deutlich zurück.

Fazit: Es gibt zu Beginn des Beobachtungszeitraumes einen positiven Effekt in die Transferdomäne hinein, der aber auf die nicht-quantitativen Komponenten des systemischen Denkens beschränkt ist und im weiteren Lernprozess verlorenght.

Die Ergebnisse zu den Hypothesen 2.5 bis 2.7 scheinen inkonsistent. Kann man hier von Transfer sprechen, wenn in der eigentlichen Lerndomäne (Physik) erst im zweiten Halbjahr positive Effekte

⁵ Der Befund gilt auch, wenn man sich auf den Kernbereich des systemischen Denkens beschränkt.

2.7.4 Profil der Lern- und Transfereffekte

H2.8 Der Transfer auf andere physikalische Domänen (nach H2.6) ist stärker als der Transfer auf alltagsbezogene ökonomische Fragestellungen (nach H2.7).

Diese Hypothese kann nach den im vorigen Abschnitt formulierten Befunden und Interpretationen als widerlegt gelten. Ein statistischer Vergleich der Zuwächse vom Zwischen- zum Abschlußtest ergab zudem, daß die Zuwächse in beiden Domänen gleich waren (Friedman Test, $p=.763$); dasselbe Ergebnis ist festzustellen, wenn die Zuwächse gegenüber dem Vortest betrachtet werden. Die Annahme, die Stärke des Transfers sei unmittelbar von der Nähe der Zieldomäne zur Quelldomäne abhängig, entspricht nicht der Wirklichkeit. Vielmehr ist mit Stern (1998) davon auszugehen, dass die Stärke des Transfers durch die Ähnlichkeit der jeweils eingesetzten Denkmodelle, insbesondere der Repräsentationsebenen, bedingt ist. Diese Vermutung wird in unserem Datensatz eindrucksvoll dadurch bestätigt, dass zeitgleiche und zeitverschobene Korrelationen der Einzelskalen (qualitative Ebene, halb-quantitative Modellerstellung usw.) durchweg maximal sind, wenn es sich um gleiche Anforderungen und Repräsentationsformate handelt.

H2.9 Die Lerneffekte (nach H2.5) und Transfereffekte (nach H2.6 und H2.7) sind bei der Modellerstellung auf quantitativer Ebene sowie der Modellinterpretation auf halb-quantitativer Ebene stärker als bei anderen Anforderungs- und Repräsentationsebenen.

Diese Hypothese konnte nicht bestätigt werden. Mit Hilfe der non-parametrischen Varianzanalyse nach Friedman wurde überprüft, ob die jeweils fünf skalenbezogenen Veränderungswerte beim Übergang vom Vor- zum Zwischentest und vom Zwischen- zum Abschlusstest in der Physik bzw. der Ökonomie zu unterschiedlichen Rangreihen der Versuchspersonen führen. Dies war nicht der Fall (χ^2 -Werte mit $p=.743$, $P=.923$ und $p=.884$) mit einer Ausnahme: Die Zuwächse des ersten Halbjahres in der Ökonomie-Domäne zeigen, wie oben erläutert, ein klares Profil ($\chi^2=5,3$; $df=4$; $p=.255$). Die Versuchsgruppe verbessert sich – verglichen mit der Kontrollgruppe – hauptsächlich im Umgang mit Wirkungsdiagrammen. Der vermutete Trainingseffekt auf quantitativer Ebene war aber in keinem Fall zu beobachten.

2.8 Diskussion

In der vorliegenden Feldstudie werden auf Basis eines langfristig angelegten Unterrichtskonzepts für den Fachunterricht die Wirkungen systemdynamisch orientierter Modellbildung auf domänen-spezifisches Wissen im Bereich der Mechanik und domänenübergreifendes systemisches Denken empirisch untersucht. Bisher vorliegende Untersuchungen zielen weniger auf die Vermittlung fachcurricularer Inhalte mit Hilfe systemdynamischer Werkzeuge als auf die generelle Förderung systemischen Denkens (Hassell 1987, Ossimitz 1996). Zumindest waren die Inhalte des Fachunterrichts spezifisch unter systemdynamischen Aspekten ausgewählt (Zuman & Weaver 1988). Andere Untersuchungen erfolgten in Laborsituationen (z.B. Bliss et al. 1993). So gewonnene Ergebnisse ließen vermuten, dass vernetztes Denken mit Hilfe systemdynamisch orientierter Softwarewerkzeuge im Unterricht gezielt vermittelbar ist. Eine Felderprobung von Modellbildungssystemen in unterschiedlichen Fächern (Klieme & Maichle 1994) konnte jedoch keine Effekte im Bereich des systemischen Denkens nachweisen. Unserer Studie liegt ein Unterricht zu Grunde, in dem systemdynamische Modellbildung primär als Instrument zur Vermittlung von Fachwissen

eingesetzt wird. Das Testinstrumentarium erlaubt es, fachliche und systemische Wirkungen gemeinsam zu erfassen und Querverbindungen herzustellen.

Der Aufbau newtonscher Denkmuster im Bereich der klassischen Mechanik wird nach unseren Befunden aus realem Physikunterricht durch computergestützte Modellbildungseinheiten nicht in dem Maße gefördert, wie dies nach der didaktischen Konzeption dieser Softwarewerkzeuge und den darauf aufbauenden Unterrichtsmodellen erwartet wird (z.B. Goldkuhle 1998, Schecker 1998). Gefördert werden halb-quantitative Fähigkeiten zur Beschreibung, bzw. Vorhersage von Bewegungsverläufen, während sich bezüglich des Grundverständnisses der zentralen Begriffe der Mechanik keine spezifischen Vorteile gegenüber konventionellem Unterricht (mehr fragend-entwickelnde Anteile anstelle interaktiver Arbeit mit dem Modellbildungssystem) zeigen. Der Befund zeigt, dass mit *Entwicklungsprojekten* oft starke Erwartungen an die Wirksamkeit von Curricula oder Lernumwelten verbunden sind, die dem kritischen Test im quasi-experimentellen Design nicht unbedingt standhalten.

Der Entwurf eines allgemeinen Konstrukts der "physikalischen Kompetenz" gestaltet sich sehr komplex. Bereits im Bereich der Mechanik/Dynamik zeigen sich Probleme ein homogenes Konstrukt zu spezifizieren und reliabel zu messen. Die Zusammenhänge zwischen den Leistungen in unterschiedlichen Tests (Vorstellungsinventare, Abgestufte Aufgaben zur Modellentwicklung, Interviews) zur gleichen Domäne sind begrenzt. Dieses Ergebnis sollte allzu simplen Konzeptualisierungen von "Kompetenz" sowohl in der Fachdidaktik als auch in der Erziehungswissenschaft vorbeugen. Die Beschreibung von Wissensstufen oder Kompetenzniveaus ist im Zusammenhang mit der Analyse der TIMSS-Ergebnisse ein aktuelles Thema.

Methodenkritik ist am "Force Concept Inventory" zu üben, dem weltweit verbreitetsten Test zur Ermittlung von Schülervorstellungen in der Mechanik. Wegen mangelnder Reliabilität der Subskalen (drei newtonsche Axiome, Kinematik usw.) erscheint er nicht geeignet, um Misconceptions oder Entwicklungsstufen der physikalischen Kompetenz feingradig zu bestimmen. Hier sind verbesserte Instrumente zu entwickeln, die psychometrischen Qualitätskriterien besser entsprechen. Ein guter Ansatz im Bereich teilstrukturierter Interviews sind die in der Studie entwickelten Experimentalinterviews. Concept mapping ohne Vorgabe eines bestimmten Bezugsphänomens erwies sich als wenig valides Verfahren zur Messung physikalischer Kompetenz. Als Schritt 1 der Abgestuften Modellentwicklung — und damit eingebunden in die Diskussion eines *bestimmten* Bewegungsvorgangs — sind Concept maps dagegen aussagekräftiger (Grundlage der Skala zur qualitativen Modellerstellung, s. Punkt 2.7.1)

Ein wesentlicher Teil der domänenübergreifenden Untersuchungen galt dem Konstrukt „Systemisches Denken“. Es werden vier Repräsentationsebenen des systemischen Denkens unterschieden (qualitativ, halb-quantitativ, quantitativ und systemdynamisch) und auf den Anforderungsebenen der Modellerstellung und Modellinterpretation operationalisiert. Als Kontexte dienen in den Tests physikalische Systeme und ökonomische Systeme mit Alltagsbezug. Es ließen sich — bei gewissen Problemen im Bereich der quantitativen Ebene in der Domäne Physik — aus den Indikatoren der Tests, mit denen die Repräsentations- und Anforderungsebenen erfasst wurden, jeweils konsistente Skalen formulieren.

Die zentrale Frage zur Validierung des Konstrukts „Systemisches Denken“ ist nun, ob sich über die Repräsentations- und Anforderungsebenen und möglicherweise auch über die Domänen (Physik, Ökonomie) hinweg eine entsprechende Kompetenz identifizieren lässt. Vor Beginn des Unterrichts, so zeigen unsere Daten aus der Vorerhebung, liegt eine solche homogene Kompetenzdimension sicherlich nicht vor. Danach werden jedoch die Zusammenhänge zwischen den Skalen

innerhalb des Physik- und des Ökonomie-Bereiches und auch der Zusammenhang zwischen den beiden Bereichen immer enger. Der Erwerb systemischen Denkens in der Basisdomäne Physik und der Zieldomäne Ökonomie hängen wechselseitig zusammen. Dies gilt vor allem für den Kernbereich des systemischen Denkens, das Erstellen und Interpretieren von Systemmodellen auf konzeptueller Ebene und in Gestalt von Wirkungsdiagrammen (also auf qualitativer und halb-quantitativer Ebene). Die Ebene der quantitativen Modellierung wie auch das zielorientierte Steuern eines komplexen dynamischen Systems (hier des Simulationsspiels „Hunger in Nordafrika“) sind hingegen nicht integrale Bestandteile des systemischen Denkens nach unserem Verständnis. Allerdings spielt bei der Entwicklung des systemischen Denkens auch fachbezogenes Wissen und Verständnis eine Rolle. Es erwächst demnach aus dem Zusammenwirken von fachübergreifenden Kompetenzen und domänenspezifischem Wissen.

Nachdem das Konstrukt „Systemisches Denken“ auf diese Weise präzisiert wurde, galt das Interesse den diesbezüglichen Effekten des Umgangs mit einem systemdynamisch orientierten Modellbildungsprogramm im Rahmen des Physikunterrichts. Die Befunde des Vergleichs von Versuchs- und Kontrollgruppenschülern bestätigen die skeptische Einschätzung der Möglichkeit, systemisches Denken mittels Modellbildungssystemen zu fördern, die Klieme und Maichle (1994) formulierten. In der Basisdomäne Mechanik ist der *Zuwachs* der Fähigkeit, systemisch zu denken, sogar in der Versuchsgruppe geringer. Der Transfer auf einen anderen physikalischen Bereich (Thermodynamik: Energiefluss) gelingt der Versuchsgruppe besser als der Kontrollgruppe. Beim Transfer auf alltagsbezogene ökonomische Fragestellungen aber sind Vorteile der Modellbildungsgruppen auf die halb-quantitative Repräsentationsebene beschränkt, und sie sind nicht nachhaltig, d.h. der Vorsprung der Versuchsgruppe geht im zweiten Unterrichtshalbjahr wieder verloren.

Dieser Befund korrespondiert mit dem obigen Ergebnis zur physikalischen Kompetenz, wonach der Wissenszuwachs in der Mechanik durch den Umgang mit Modellbildungssystemen vornehmlich im halb-quantitativen Bereich gefördert wird, was sich durch eine größere Fähigkeit zur Beschreibung und Vorhersage von Bewegungsverläufen ausdrückt. Bei gleichungsorientierten, quantitativen Anforderungen in der Mechanik gibt es hingegen keine signifikanten Gruppenunterschiede. Auch wenn die empirische Kontrolle nur spezifische und begrenzte Wirkungen zeigt, wertet das die Entwicklungsarbeit als Anregung für innovative Unterrichtsgestaltung nicht ab.

Die Rahmenbedingungen einer Langzeitstudie unter Realbedingungen hatten eine starke Begrenzung der empirischen Basis zur Folge (s. dazu Punkt 2.3). Die Ergebnisse der Studie wären auf Basis einer größeren Fallzahl besser abzusichern. Unter diesen Einschränkungen kann man als Fazit festhalten, dass ein Fachunterricht, der Modellbildungssysteme als Instrument zur Vermittlung von fachcurricularen Inhalten einsetzt, in der eigenen Domäne begrenzte Vorteile gegenüber konventionellem Unterricht aufweist, die im halb-quantitativen Bereich liegen (Abschätzung und Prognose von Systemverhalten). Vorteile hinsichtlich eines besseren Begriffsverständnisses stellen sich nicht ein. Ebenso wenig zeigen sich Effekte im quantitativen Bereich — dies war auch nicht erwartet worden. Für den Aufbau systemischen Denkens sind die Effekte ebenfalls begrenzt und liegen gleichfalls im halb-quantitativen Bereich (Konstruktion und Analyse von Wirkungsgefügen).

Wesentliche Ergebnisse der Studie liegen in der Konstruktion und Absicherung von Instrumenten zur Diagnostik des systemischen Denkens und der Validierung dieses Konstrukts. Dies Instrumentarium steht ausgearbeitet für eine breiter angelegte Studie mit höheren Probandenzahlen bereit. Um größere Fallzahlen handhaben zu können und Datenverluste zu begrenzen, ist z.B. an eine

Langzeit-Laborstudie im Hochschulbereich zu denken, die begleitend zu einem geeigneten Abschnitt des Fachstudiums (z.B. in Physik oder Ökonomie) durchgeführt werden könnte.

2.9 Qualifikationsaspekte

Im Projekt waren Frau Dipl.-Psych. Judith Ebach als wissenschaftliche Mitarbeiterin und Herr Jörn Gerdes (Lehrer für Physik und Deutsch) als wissenschaftlicher Mitarbeiter tätig. Beide haben wesentlich an der Erarbeitung der Testinstrumente und der Datenauswertung mitgewirkt. Herr Gerdes hat die umfangreiche Datenerhebung betreut. Für beide Mitarbeiter ist aus dem Projekt ein Promotionsvorhaben erwachsen. Herr Gerdes validiert theoretisch und empirisch das Konstrukt der physikalischen Kompetenz (Fachgebiet Didaktik der Physik, Universität Bremen). Das Promotionsthema von Frau Ebach lautet „Modellbildung im Unterricht - Lern-Wirkungen einer computergestützten Lernumgebung“. Das Vorhaben wird an der Universität Koblenz-Landau betreut (Fachgebiet Erziehungswissenschaften).

Da beide Mitarbeiter wegen parallel laufender Unterrichts- bzw. Dienstleistungsverpflichtungen in anderen Arbeitszusammenhängen tatsächlich nur mit halber Stelle im DFG-Projekt tätig sein konnten, sind die Promotionsvorhaben zwar weit fortgeschritten, sie werden aber erst gegen Ende dieses Jahres abgeschlossen sein.

3 Zusammenfassung

Im Vorhaben wurden die Effekte des Einsatzes grafikorientierter Modellbildungssysteme (MBS) im Physikunterricht auf die Entwicklung physikalischer Kompetenz im Themenbereich „Kräfte und Bewegungen“ und die Fähigkeit zur Strukturierung komplexer Wirkungszusammenhänge („Systemisches Denken“) in ökonomischen Kontexten mit Alltagsbezüge untersucht. Die Experimentalgruppe bestand aus Schülern, die in Physik-Leistungskursen (Jahrgang 11) wiederholt mit einem MBS gearbeitet haben. Schüler aus parallelen Leistungskursen ohne Computereinsatz bildeten die Kontrollgruppe.

Zu drei Messzeitpunkten — am Beginn, in der Mitte und am Ende des Schuljahres — wurde jeweils eine mehrstündige Testbatterie zur physikalischen Kompetenz und zum systemischen Denken bearbeitet. Es werden vier Repräsentationsebenen des systemischen Denkens unterschieden und auf den Anforderungsebenen der Modellerstellung und Modellinterpretation operationalisiert. Als Kontexte dienen in den Tests neben physikalischen Systemen ökonomische Systeme mit Alltagsbezug.

Der Aufbau newtonscher Denkmuster im Bereich der klassischen Mechanik wird nach den Befunden durch computergestützte Modellbildungseinheiten nicht in dem Maße gefördert, wie dies nach der didaktischen Konzeption dieser Softwarewerkzeuge und den darauf aufbauenden Unterrichtsmodellen erwartet wurde. Gefördert werden halb-quantitative Fähigkeiten zur Beschreibung, bzw. Vorhersage von Bewegungsverläufen, während sich bezüglich des Grundverständnisses der zentralen Begriffe der Mechanik und im quantitativen, gleichungsorientierten Bereich keine spezifischen Vorteile gegenüber konventionellem Unterricht zeigen.

Wesentliche Ergebnisse der Studie liegen in der Konstruktion und Absicherung von Instrumenten zur Diagnostik des systemischen Denkens und der Validierung dieses Konstrukts. Systemisches Denken konnte auf vier Repräsentations- und zwei Anforderungsebenen erfolgreich operationalisiert und durch Skalen beschrieben werden. Sein besonderes Kennzeichen ist die halb-quantitative Fähigkeit zur Erstellung und Interpretation von Wirkungsdiagrammen. Der Erwerb systemischen Denkens in der Basisdomäne Physik und der Zieldomäne Ökonomie hängen wechselseitig zusammen. Systemisches Denken erwächst aus dem Zusammenwirken von fachübergreifenden Kompetenzen und domänenspezifischem Wissen. Dagegen besteht kein globaler Zusammenhang zwischen systemischem Denken und der Fähigkeit zur Steuerung eines dynamischen Systems.

Der Vergleich von Versuchs- und Kontrollgruppenschülern ergibt kein einheitliches Bild hinsichtlich der Effekte des Unterrichts-Treatments. In der Basisdomäne Mechanik ist der *Zuwachs* der Fähigkeit, systemisch zu denken, in der Versuchsgruppe geringer als in der Vergleichsgruppe. Der Transfer auf einen anderen physikalischen Themenbereich (Thermodynamik) gelingt der Versuchsgruppe besser. Ein Transfer auf alltagsbezogene ökonomische Fragestellungen ist zwar im Kernbereich des systemischen Denkens, d.h. im halb-quantitativen Bereich, erkennbar, aber auch dort nicht nachhaltig. Die Befunde bestätigen somit den in früheren Felduntersuchungen der Autoren gewonnenen Eindruck, dass der weitreichende Anspruch, mit dem Einsatz von Modellbildungssystemen im Unterricht das systemische Denken fördern zu wollen, nicht eingelöst werden kann. Modellbildungssysteme bewähren sich im gewählten Unterrichtskonzept als Methoden im engeren Bereich der Förderung physikalischen Verständnisses, nicht jedoch als Mittel zur Förderung übergreifender Kompetenzen.

4 Anhang

4.1 Literatur

- Bliss, J., Mellar, H., Ogborn, J. & Nash, C. (1993). Tools for Exploratory Learning Programme - Technical Report 2 „Semi-Quantitative-Reasoning-Expressive“. London: University of London, King's College.
- Dörner, D., Kreuzig, H., Reither, F. & Stäudel, T. (Hrsg.) (1981): Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität. Bern: Huber.
- Gerdes, J. & Schecker, H.: Der Force Concept Inventory — Ein diagnostischer Test zu Schülervorstellungen in der Mechanik. In: *Der Mathematische und Naturwissenschaftliche Unterricht* (im Druck).
- Goldkuhle, P.: Modellbildung und Simulation mit dem Computer im Physikunterricht. Köln: Aulis 1997.
- Jäger, A.O. & Althoff, K. (1994). Wilde-Intelligenz-Test (WIT). Göttingen: Hogrefe.
- Hake, R. (1998). Interactive-engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. In: *American Journal of Physics* 66 (1998), 64-74.
- Hassell, D.J.: The role of modelling activities in the humanities curriculum, with special reference to geography: an investigate study. Dissertation for Associateship for Education, London: King's College.
- Hestenes, D., Wells, M. & Swackhammer, G. (1992). Force Concept Inventory. *The Physics Teacher*, 30 (4), 141-151.
- Klieme, E. & Maichle, U.: Modellbildung und Simulation im Unterricht der Sekundarstufe I — Auswertung von Unterrichtsversuchen mit dem Modellbildungssystem MODUS. Bonn: Institut für Bildungsforschung (im Auftrag des nordrhein-westfälischen Landesinstituts für Schule und Weiterbildung).
- Krauth, J.: Testkonstruktion und Testtheorie. Weinheim: Beltz, 1995.
- Leutner, D. (1992). Adaptive Lehrsysteme. Instruktionspsychologische Grundlagen und experimentelle Analysen. Weinheim: Psychologie - Verlags - Union.
- Ossimitz G. (1996): Können Schüler im Mathematikunterricht systemisch denken lernen? In: K.P. Müller (Hrsg.): Beiträge zum Mathematikunterricht 1996. Hildesheim: Franzbecker, 337-340.
- Schecker, H. (1998). Physik modellieren. Stuttgart: Klett.
- Schecker, H. & Gerdes, J.: Interviews über Experimente zu Bewegungsvorgängen. In: *Zeitschrift für Didaktik der Naturwissenschaften* 4 (1998), 3, 61-74.
- Schecker, H. & Gerdes, J.: Interviews über Experimente zu Bewegungsvorgängen. In: *Zeitschrift für Didaktik der Naturwissenschaften* 4 (1998), 3, 61-74.
- Schecker, H. & Gerdes, J.: Messung von Konzeptualisierungsfähigkeit in der Mechanik — Zur Aussagekraft des Force Concept Inventory. In: *Zeitschrift für Didaktik der Naturwissenschaften* 5 (1999), H. 1, 75-89.
- Schecker, H., Klieme, E. & Niedderer, H.: Physiklernen mit Modellbildungssystemen. Zwischenbericht zum DFG-Projekt. Bremen/Bonn 1997.

Schecker, H. & Niedderer, H.: Über den Beitrag des Computereinsatzes zur Entwicklung physikalischer Kompetenz. In: Wiebel, K.H. (Hrsg.): Zur Didaktik der Physik und Chemie. Alsbach: Leuchtturm, 147-149.

STELLA: Dartmouth, NH: High Performance Systems.

Stern, E.: Die Entwicklung des mathematischen Verständnisses im Kindesalter. Lengerich: Pabst, 1998.

Zuman, J.P. & Weaver, S.L.: Tools for teaching problem solving: An evaluation of a modeling and systems thinking approach. Paper presented at the 1988 Annual NARST Conference, Lake Ozark, MO, April 1988.

4.2 Anlagen

Schecker, H. & Gerdes, J.: Interviews über Experimente zu Bewegungsvorgängen. In: *Zeitschrift für Didaktik der Naturwissenschaften* **4** (1998), 3, 61-74.

Gerdes, J. & Schecker, H.: Der Force Concept Inventory — Ein diagnostischer Test zu Schüler- vorstellungen in der Mechanik. In: *Der Mathematische und Naturwissenschaftliche Unterricht* (im Druck).

Schecker, H. & Gerdes, J.: Messung von Konzeptualisierungsfähigkeit in der Mechanik — Zur Aussagekraft des Force Concept Inventory. In: *Zeitschrift für Didaktik der Naturwissenschaften* **5** (1999), H. 1, 75-89.

Zwei weitere in Vorbereitung befindliche Manuskripte werden in Kürze nachgereicht:

Ebach, J. & Gerdes, J., Klieme, E. & Schecker, H.: Modellbildungs- und Modellkonstruktions- prozesse (einzureichen bei: *Zeitschrift für Psychologie in Erziehung und Unterricht*).

Klieme, E., Schecker, H. & Niedderer, H.: Assessment of Physics Knowledge by Concept Map- ping (to be submitted to: *Journal of Research in Science Teaching*).