

Schecker, Horst; Gerdes, Jörn

Messung von Konzeptualisierungsfähigkeit in der Mechanik - Zur Aussagekraft des Force Concept Inventory

Zeitschrift für Didaktik der Naturwissenschaften : ZfDN 5 (1999) 1, S. 75-89



Quellenangabe/ Reference:

Schecker, Horst; Gerdes, Jörn: Messung von Konzeptualisierungsfähigkeit in der Mechanik - Zur Aussagekraft des Force Concept Inventory - In: Zeitschrift für Didaktik der Naturwissenschaften : ZfDN 5 (1999) 1, S. 75-89 - URN: urn:nbn:de:0111-pedocs-315261 - DOI: 10.25656/01:31526

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-315261>

<https://doi.org/10.25656/01:31526>

in Kooperation mit / in cooperation with:



IPN

Leibniz-Institut für die Pädagogik der
Naturwissenschaften und Mathematik

<https://www.leibniz-ipn.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung

E-Mail: pedocs@dipf.de

Internet: www.pedocs.de

Digitalisiert

Mitglied der


Leibniz-Gemeinschaft

HORST SCHECKER UND JÖRN GERDES

Messung von Konzeptualisierungsfähigkeit in der Mechanik - Zur Aussagekraft des Force Concept Inventory

Zusammenfassung:

Der Force Concept Inventory (FCI) ist ein weltweit verwendeter Standardtest zur Untersuchung der Lernwirkungen von Mechanikunterricht. Inwieweit er geeignet ist, die Konzeptualisierung des Kraftbegriffs im Detail zu erfassen, ist jedoch umstritten. Auf Grundlage von FCI-Daten aus einer Vergleichsstudie in Mechanikkursen mit und ohne Nutzung grafikorientierter Modellbildungssoftware hinterfragen wir die Aussagekraft des Tests. Während globale Aussagen zu den Unterrichtswirkungen möglich sind, bereiten Profilaussagen zum Kern des Kraftverständnisses Schwierigkeiten. Insgesamt ergaben sich keine wesentlichen Unterschiede zwischen den beiden Probandengruppen. Die Untersuchung der Fehlantworten am Ende des Kurses deutet auf einen Mischzustand zwischen newtonschen Denkmustern und dominierenden Impetusvorstellungen hin.

Abstract:

The Force Concept Inventory (FCI) is used worldwide as a standard test for the effects of teaching strategies in mechanics. It is, however, debated whether the test can help to pinpoint facets of students' force concepts. We investigate the power of the FCI on the basis of data from a comparative study on the effects of computer based modelling in mechanics. While the global learning outcome can be assessed reliably, it is difficult to employ the FCI for analyzing the core of the students' alternative frameworks in detail. In our study there were no essential differences between the experimental and the control groups. At the end of the mechanics course the students were in a mixed learning state between Newtonian thinking and dominating impetus notions.

1. Der Force Concept Inventory

Die weltweite Forschung über Schülervorstellungen ist so weit vorangeschritten, daß Standarduntersuchungsinstrumente entwickelt werden, die eine schnelle Diagnose der Konzeptualisierungsfähigkeit in einem Teilgebiet der Physik ermöglichen sollen. Der bekannteste Test ist der „Force Concept Inventory“ (FCI; Hestenes et al. 1992). Man findet darin Denkaufgaben zur Mechanik, die in ähnlicher Form bereits in empirischen Studien zur Exploration von Schülervorstellungen verwendet wurden. Um physikalische Denkaufgaben zu lösen, benötigen die Schüler keine genauen Formelkenntnisse oder mathematische Fähigkeiten im Lösen von Gleichungssystemen. Voraussetzung ist vielmehr ein solides Verständnis physikalischer Grundbegriffe wie

Kraft oder Trägheit (s. dazu Schecker, 1988). Standardisierte Tests zu anderen physikalischen Themengebieten sind in Arbeit.¹

Obwohl die Auswahlantworten zu den Aufgaben des FCI zu einem großen Teil auf den Befunden über alternative Konzeptualisierungen der Mechanik beruhen, wird der Test vorwiegend in dichotomisierter Form ausgewertet, also nach richtiger oder falscher Antwort. Dabei spielen die Arten der falschen Antworten und die damit verbundenen alternativen Konzepte keine Rolle. Während der FCI in der Originalveröffentlichung als Hilfe für Lehrkräfte vorgestellt wurde, das Denken ihrer Schüler zu ergründen (vgl. Hestenes et al. 1992, S. 142), findet er seine Hauptanwendung inzwischen beim Vergleich der Lernwirkungen unterschiedlicher Unterrichtskonzepte. Die Frage, ob der FCI seinen dia-

¹ Ein solches Projekt läuft z.B. an der Swinburne University of Technology in Australien (<http://www.swin.edu.au/bsee/mazzo/tcup>). Weitere Hinweise auf „conceptual tests“ finden sich in Beiträgen zur „Physics Learning Research List“ im Oktober 1998 (PHYSLRNR@LISTSERV.IDBSU.EDU).

gnostischen Anspruch überhaupt erfüllen kann, ist durchaus umstritten. Die Debatte hierüber wird unter Punkt 1.2 aufgegriffen.

In unserem Beitrag analysieren wir die Qualität des FCI - insbesondere seiner Subskalen zu Teilbereichen der Mechanik - diskutieren die Konsistenz der Schülerantworten und berichten über FCI-Ergebnisse zu zwei unterschiedlichen Unterrichtsansätzen. Hier geht es um spezifische Lernwirkungen eines Mechanikunterrichts, in dem die Schüler mit computergestützten Modellbildungssystemen numerische Simulationen entwerfen. Wir haben über den Kontext der Untersuchung an anderer Stelle berichtet (s. Schecker & Gerdes, im Druck).

1.1 Konstruktion des FCI

Der Force Concept Inventory besteht aus 29 multiple-choice Fragen zu Bewegungsvorgängen. Darin wird das Verständnis folgender Bereiche getestet:

- 1) Kinematik: Differenzierung zwischen Ort, Geschwindigkeit und Beschleunigung (2 Aufgaben), Bahnkurven bei konstanter Beschleunigung, Fall und Wurf (4);
- 2) Erstes Newtonsches Axiom: Trägheitsprinzip (8);
- 3) Zweites Axiom: Kraft und Beschleunigung (4);
- 4) Drittes Axiom: Wechselwirkungskräfte (4);
- 5) Superposition: Überlagerung von Kräften (4);
- 6) Spezielle Kräfte: Reibung (6), Gravitation (10).

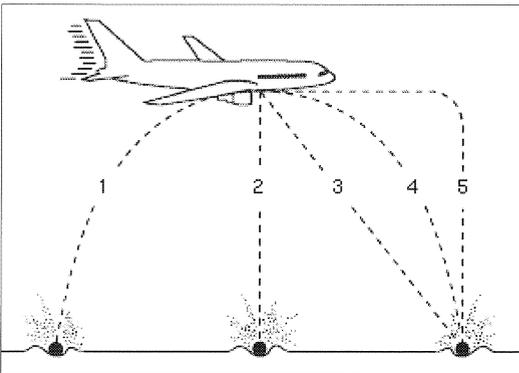


Abb. 1: Aufgabe Nr. 23 des FCI. Gefragt wird nach der Flugbahn der Bowlingkugel, wenn man den Vorgang von der Erde aus beobachtet.

Da bestimmte Aufgaben mehreren Subskalen zugeordnet sind, ist die Summe über die Bereiche größer als 29. Der Inhaltskanon der Kinematik und Dynamik in der gymnasialen Oberstufe wird weitgehend abgedeckt. Hestenes et al. (1992) postulieren, dass damit das „Kraftkonzept“ umfassend getestet werde.

In der Regel sind fünf Antwortmöglichkeiten vorgegeben, von denen eine im Sinne der Newtonschen Mechanik korrekt ist, während die anderen auf Schülervorstellungen beruhen. Hestenes et al. (1992) ordnen diese alternativen Sichtweisen nach fünf Hauptkategorien. Wir veranschaulichen die Kategorien an konstruierten idealtypischen Aussagen:

- Impetusvorstellung zum Verhältnis von Kraft und Bewegung: „Ein Körper bewegt sich aufgrund einer gespeicherten ‘Kraft’, die ihm übertragen wurde (z.B. durch einen Stoß)“;
- Aktivitätsvorstellung: „Nur aktive Körper können Kräfte ausüben“; „Bewegung impliziert Kraft“;
- Kraft und Gegenkraft: „Bei einer Wechselwirkung übt der schwerere bzw. aktivere Körper mehr Kraft aus“;
- Kräftewettstreit: „Die stärkere Kraft bestimmt die Bewegung“;
- inkonsequenter Kraftgebrauch: Nicht alles, was eine Bewegung beeinflusst, wird auf eine Kraftwirkung zurückgeführt: „Man kann zwischen ‘richtigen’ Kräften und Bewegungswiderständen unterscheiden“; „Schwere Körper fallen schneller“.

Die Kategorien überlappen in einigen Bereichen. Eine ausführliche Darstellung des FCI mit der Aufschlüsselung der Aufgaben nach den einzelnen Inhaltsbereichen und der genauen Zuordnung der Auswahlantworten zu Schülervorstellungen wird in Gerdes & Schecker (im Druck) gegeben. Der FCI wurde von uns ins Deutsche übersetzt und kann bei uns angefordert werden.²

² Möglichkeit zum Download im Internet unter <http://www.physik.uni-bremen.de/physics.education/niedderer/download/index.html>

In Mazur (1997) findet man eine von Halloun und anderen leicht überarbeitete Fassung des FCI. Eine Aufgabe wurde ergänzt. Dazu kommen zahlreiche redaktionelle Änderungen am Text und verbesserte Abbildungen. Einige Aufgaben sind umformuliert. Die Fassung ist klarer formuliert und für zukünftige Untersuchungen vorzuziehen.

1.2 Die Debatte um den FCI

In einer Veröffentlichung zum Vorläufer des FCI (Mechanics Diagnostic Test) diskutierten Halloun und Hestenes (1985a, S. 1044) die Validität und Reliabilität des Tests. In Interviews mit fortgeschrittenen Studenten und Befragung von Physikexperten wurde geklärt, daß ein Konsens über die richtigen Antworten besteht. An Anfängern wurde getestet, daß die vorgelegten Antwortalternativen inhaltlich korrekt verstanden wurden. Reliabilitätsstudien beruhten z.B. auf Nachinterviews zum Test, in denen die Probanden die gleichen Antworten gaben und begründeten, die sie vorher schriftlich gewählt hatten. Auch der Vergleich von multiple-choice Fassungen mit freien schriftlichen Antworten ergab eine große Übereinstimmung. Wiederholungseffekte von Vor- und Nachtest zeigten sich nicht. Die Ergebnisse nach dem Unterricht hingen nicht davon ab, ob die Studenten am Vortest teilgenommen hatten. Aufgrund der großen Ähnlichkeiten zu seinem Vorgänger verzichteten die Autoren auf weitergehende Untersuchungen der Testqualität des FCI (vgl. Hestenes et al. 1992, S. 148).

Was man in den bisherigen Veröffentlichungen von Hestenes und Mitarbeitern kaum findet, ist eine Überprüfung der Zuordnungen von Aufgaben zu Inhaltsbereichen - d.h. eine empirische Überprüfung der Reliabilität der Subskalen des FCI - und von Items zu „misconceptions“ - d.h. eine Überprüfung der Konsistenz der (Fehl-) Antworten. In Halloun und Hestenes (1985b) wird nur eine grobe Analyse der Auswahlantworten einiger Aufgaben des FCI-Vorläufers Mechanics Diagnostic Test nach den Kategorien aristotelisch,

newtonsch oder Impetus vorgenommen. In Tests mit Physik-Anfängerstudenten (Universität) vor dem Unterricht wurden etwa zwei Drittel der Antworten als Typ Impetus eingestuft. Der Rest verteilte sich gleichgewichtig auf „überwiegend aristotelisch“ und „überwiegend newtonsch“ (vgl. Halloun & Hestenes, 1985b, S. 1058). Betrachtete man die Antworten jeweils einzelner Studenten, dann zeigte sich ein wenig konsistentes Antwortverhalten. So blieben z.B. nur 2% (bzw. 1% im Nachtest) konsequent bei der Annahme, daß ein Körper sich bei einer konstanten Kraft mit gleichbleibender Geschwindigkeit bewege, obwohl 66% (bzw. 44%) mindestens einmal eine entsprechende Auswahlantwort gewählt hatten. Genaue Angaben über die Anzahl der für die Analyse herangezogenen Aufgaben finden sich in der Veröffentlichung jedoch nicht.

Mit Bezug auf die mangelnde Überprüfung der Subskalen stellten Huffman und Heller (1995a) in einer kritischen Analyse die Frage „What does the force concept inventory actually measure?“. Sie unterzogen die FCI-Daten von 145 Schülern und 750 Studenten einer Faktorenanalyse. Es ergaben sich viele indifferente Faktoren und nur ein bzw. zwei interpretierbare Faktoren, die begrenzte Übereinstimmung mit dem der Aufgabengruppe „Drittes Axiom“ und der unscharfen Subskala „Spezielle Kräfte“ zeigten.

„The items on the inventory appear to be only loosely related to each other, and instructors should be cautious about concluding that the inventory actually measures students' understanding of a 'force concept'. It seems more likely that the inventory actually measures bits and pieces of students' knowledge that do not necessarily form a coherent force concept.“ (Huffman & Heller, 1995a, S. 141).

Huffman und Heller bezweifelten nicht die Reliabilität des Tests oder seinen Wert als Werkzeug zum Vergleich der Wirkungen unterschiedlicher Unterrichtskonzeptionen. Sie stellten allerdings in Frage, ob aus dem FCI-Ergebnis auf „das“ Kraftkonzept eines Schülers geschlossen werden kann.

In einer heftigen Erwiderung verteidigten Hestenes und Halloun (1995) die Validität des FCI, verschoben dabei aber die Gewichtung des Einsatzzwecks von der Diagnose des Kraftverständnisses hin zur Bewertung des Effektes bestimmter Instruktionsverfahren. Dafür sei der Gesamtscore heranzuziehen. Bei newtonschen Denkern werde man ein sehr konsistentes Antwortverhalten feststellen. Huffman und Heller hätten Daten vorwiegend von nicht-newtonschen Probanden verwendet. Daß die sich inkonsistent verhalten, sei bekannt (s. Halloun und Hestenes, 1985b). Der Wert der Auswahlantworten, die auf Schülervorstellungen beruhen, wurde nicht als diagnostischer dargestellt, sondern als Distraktoren von der newtonschen Lösung. Die sechs Ordnungskategorien für die 29 Aufgaben hätten einen rein logischen Wert und entsprächen nicht der Struktur von Schülerkonzepten zum Kraftbegriff.

Mit Recht wiesen Hestenes und Halloun darauf hin, daß eine Faktorenanalyse dichotomisierter Antworten keinen Aufschluß über die Struktur des Denkens von Nicht-Newtonianern geben kann. Sie selbst haben jedoch keine Anstrengungen dazu unternommen, die Fehlantworten auf mögliche Muster zu analysieren, wie man es bei einem „diagnostischen“ Test erwarten sollte.

„We would welcome a more rigorous statistical analysis, but that would be difficult, because the response pattern of each student must be analyzed separately and compared to identify possible groups of students with similar patterns.“ (Hestenes & Halloun, 1995, S. 504)

In einer erneuten Replik führten Heller und Huffman (1995b) an, daß ihre umfangreichen Daten selbst bei Probanden mit relativ hohem FCI-Score (60% bis 85% richtige Antworten) keine hohen Korrelationen zwischen den einzelnen Aufgaben aufweisen (nur 5 von 406 Korrelationen liegen über $\phi=0,3$). Sie bleiben dabei, daß auch ein hoher FCI-Wert nicht auf ein solides, konsistentes Grundverständnis des Kraftkonzepts schließen lasse. Darüberhinaus hätten unterschiedliche Kurskonzeptionen – „traditional“ versus „interactive-

engagement“ (s.u.) - an der University of Minnesota keinen Einfluß auf die Testscores gehabt.

Zu diesem Punkt liegt inzwischen eine Meta-Analyse von 6542 Testergebnissen aus 62 Kursen an amerikanischen Universitäten und Colleges vor (Hake, 1998). Hake vergleicht darin sogenannte „interactive-engagement“ (IE) Kurse mit „traditionellen“ Kursen. IE-Kurse werden durch „hands-on“ Aktivitäten sowie Unterrichtsgespräche der Lernenden untereinander und Diskussionen mit dem Lehrenden gekennzeichnet. Traditionelle Kurse entsprechen Vorlesungen ohne direkte Ansprache der Studierenden mit algorithmisch zu lösenden Aufgaben und rezeptartig abzuarbeitenden Praktika. Verglichen werden die Vor- und Nachtestergebnisse im FCI unter dem Aspekt des relativen Zuwachses. Als Maß wird der Faktor „g“ (normalisierter mittlerer Zuwachs) in einem Kurs definiert:

$$g = \frac{\text{erreichter Zuwachs}}{\text{maximal möglicher Zuwachs}} = \frac{\% \text{Nachtest} - \% \text{Vortest}}{100\% - \% \text{Vortest}}$$

Dieser Faktor ermöglicht es, Kurse mit unterschiedlichen Eingangsvoraussetzungen miteinander zu vergleichen. Weder der Absolutwert im Nachtest noch der absolute Zuwachs alleine sagen etwas über die Unterrichtsqualität aus.

Die g-Werte sind bei den 48 von Hake herangezogenen IE-Kursen ($0,48 \pm 0,14$) deutlich höher als in den 14 traditionellen Kursen ($0,23 \pm 0,04$). Nahezu alle traditionellen Kurse liegen damit unterhalb der Grenze von $g=0,3$, die Hake als „low g“-Region bezeichnet. Mit der Metastudie von Hake wird der Wandel des FCI vom diagnostischen Test zu einer Art Rating-Koeffizienten für Kurse evident.

Als Fazit der Debatte um den FCI kann man festhalten, daß die Inhaltsvalidität des Tests für das generelle Mechanikverständnis auf der Hand liegt, während sein innerer Aufbau aus Aufgabengruppen zu hinterfragen ist. Gleiches gilt für die wirkliche Diagnose alternativer Konzeptualisierungen der Mechanik. Eine

Analyse der Struktur von Fehlantworten ist bisher kaum vorgenommen worden. Weite Anwendung findet der FCI im Vergleich von Unterrichtskonzeptionen.

2. Einbettung des FCI in die Studie „Physiklernen mit Modellbildungssystemen“

Der FCI ist Teil des Instrumentariums, mit dem wir die Wirkungen der Nutzung computergestützter grafikorientierter Modellbildungssysteme im Mechanikunterricht der Klasse 11 getestet haben. Das Projekt „Physiklernen mit Modellbildungssystemen“ wird in Schecker (1998a) näher beschrieben. Die Schülerinnen und Schüler der beiden Versuchskurse (Leistungskurse, 11. Jahrgangsstufe, fünf Wochenstunden Physik, Schuljahr 1996/97) erstellten an Computern mit Hilfe des Programms Stella (1997) im Verlaufe des Halbjahres Mechanik fünf numerische Modelle zu Bewegungen auf Experimentierbahnen, Wurfbewegungen, Bewegungen mit konstanter beschleunigender Kraft und Bewegungen mit Luftreibungskraft. Diese Unterrichtseinheiten hatten einen Anteil von 20% bis 25% an der Unterrichtszeit. Zwei weitere Leistungskurse bearbeiteten den gleichen Phänomenbereich ohne Nutzung grafikorientierter Modellbildungssysteme.

Die zentrale Wirkungshypothese lautet:

Die Schüler der Versuchsgruppe entwickeln in höherem Maße die Fähigkeit zur halbquantitativen Beschreibung von Bewegungsvorgängen unter dem Einfluß von Kräften.

Begründet ist diese Erwartung dadurch, daß das erfolgreiche Konstruieren der Modelle ein Durchlaufen der folgenden Überlegungskette erfordert: Man ermittle die auf einen Körper wirkenden Einzelkräfte und bilde daraus die Resultierende; diese bestimmt die Beschleunigung des Körpers, d.h. die Geschwindigkeitsänderung und damit deren zeitlichen Verlauf. Konzeption und Beispiele sind an anderer Stelle ausführlich beschrieben (Schecker, 1998b).

Neben dem FCI haben wir als Untersuchungsinstrumente strukturierte Einzelinterviews über Bewegungsvorgänge (s. dazu Schecker & Gerdes, im Druck) - sowie Concept Mapping und weitere schriftliche Tests verwendet.

Der FCI wurde unter folgenden Aspekten eingesetzt:

- Zum Vergleich der globalen Unterrichtswirkungen zwischen den beiden Versuchskursen mit Computereinsatz und den beiden Kontrollkursen ohne Nutzung von Modellbildungssystemen.
- Zum Vergleich von Unterrichtswirkungen im Zusammenhang mit dem zweiten Newtonschen Axiom. In diesem Bereich erwarten wir aufgrund der Unterrichtskonzeption in den Versuchskursen spezifische positive Effekte.
- Zur Untersuchung der Konsistenz der Antworten in diesem Bereich.

Eine wichtige Rolle im folgenden Ergebnisteil spielt die Frage, ob der FCI geeignet ist, über den Test eines globalen Verständnisses der Mechanik hinaus Profilaussagen über das Verständnis des Zusammenhangs zwischen Kraft und Bewegungsänderung zu gewinnen.

3. Globalergebnisse

In unserer Studie wurde der FCI den Schülern am Beginn des Physikunterrichts in Klasse 11 (Vortest) und nach Abschluß des Mechanikkurses am Ende des ersten Halbjahres vorgelegt (Nachttest). Da der Versuchskurs 2 (VK2) erst später in die Studie einbezogen wurde, liegen hier nur Ergebnisse des Nachttests vor (s. Tab. 1). Für einen Vergleich der Kurse wird in Abb. 2 das Ergebnis der Vortests gegen den prozentualen Zuwachs aufgetragen. Die übrigen Daten stammen aus der Meta-Studie von Hake (1998), der aus den Daten zahlreicher FCI-Studien zu einer Unterscheidung unterschiedlicher Typen von Mechanikkursen gelangt ist (vgl. Punkt 1.2):

- Kurse, die auf traditionelle Art und Weise unterrichtet wurden. Diese Kurse gruppieren sich um eine Gerade, die einen gewichteten Zuwachs von $g=0,23$ repräsentiert.

• Interaktive Kurse, die sich in einem Bereich wiederfinden lassen, der von zwei Geraden aufgespannt wird, die gewichtete Zuwächse von $g=0,34$ bzw. $g=0,69$ repräsentieren. Hake bezeichnet diesen Bereich als interaktive Zone.

Fügt man die Kurse unserer Studie in das Hake-Diagramm ein, dann liegen VK1 und KK1 in der interaktiven Zone, während KK2 im Bereich der Kurse mit traditionellen Unterrichtsmethoden liegt. „Interaktivität“

kann sich in Kursen in ganz unterschiedlicher Weise ausdrücken. Hake (1998) beschreibt dieses Merkmal nur sehr grob (s.o.). Von der Arbeit mit einem Modellbildungssystem sollte man ein hohes Maß an Interaktion erwarten - in Gruppenarbeitsphasen am Computer unter den Schülern und im Klassenunterricht mit der Lehrkraft im Gespräch über die Struktur der Modelle. Daran gemessen liegt der g -Wert von Versuchskurs 1 unterhalb unserer Erwartungen.

Kurs	Kursgrößen		Testteilnehmer			Lösungsanteile		
	Vortest	Nachtest	Vortest	Nachtest	Vor & Nach	Vortest	Nachtest	g
VK1	8	5	8	5	5	0,42	0,67	0,43
VK2	19	19	—	18	—	—	0,44	—
KK1	20	18	20	15	15	0,31	0,57	0,38
KK2	15	11	15	11	10	0,33	0,44	0,24

Tab. 1: Hauptuntersuchung: Kursgrößen zu den zwei Meßzeitpunkten (VK: Versuchskurs, KK: Kontrollkurs), Testteilnehmer und Ergebnisse. Bei den Ergebnissen sind nur die Schüler berücksichtigt, die sowohl am Vor- als auch am Nachtest teilnahmen.

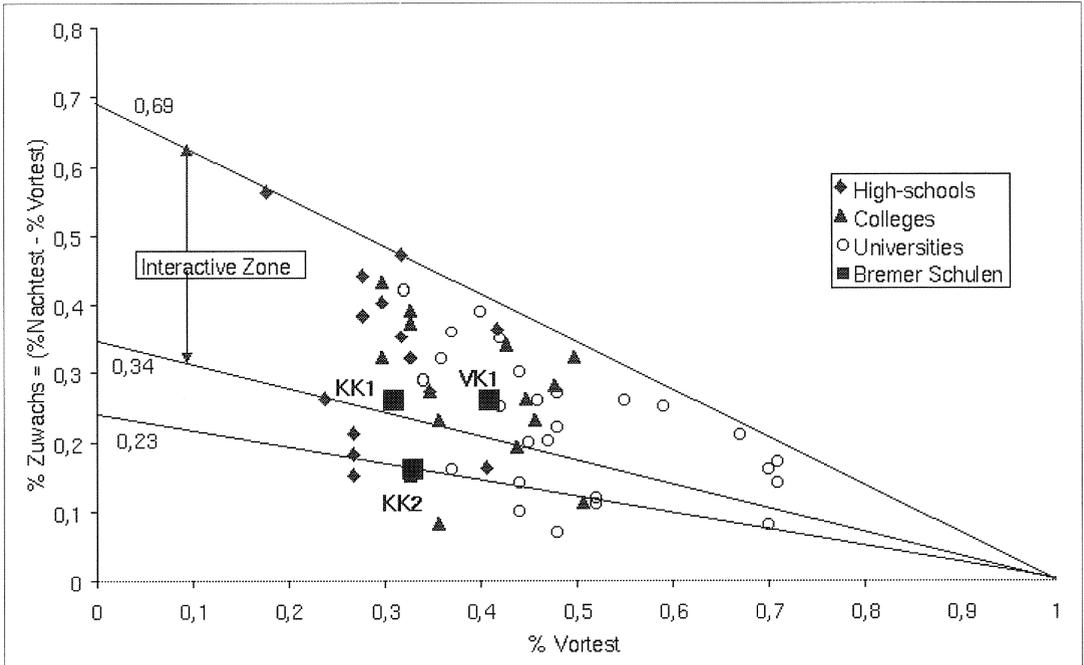


Abb. 2: Graphische Repräsentation von FCI Vor- und Nachtestergebnissen; nach Hake (1998).

Im Nachtest, an dem beide Versuchskurse teilnahmen, erreichten die Schüler in der Versuchsgruppe einen Lösungsanteil von 48,7% gegenüber 53,7% in der Kontrollgruppe. Der Unterschied ist signifikant³, beträgt aber nur etwa ein Viertel Standardabweichung. Auch wenn man über Lernzuwächse bei einem solchen Zustandsvergleich keine Aussagen treffen kann, lassen sich damit - entgegen unserer Hypothese - globale positive Wirkungen bezüglich des Verständnisses der Newtonschen Mechanik in der Versuchsgruppe nicht nachweisen.

Spezifische Wirkungen grafikorientierter Modellbildungssysteme haben wir beim halbquantitativen Verständnis des Zusammenhangs zwischen Kraft und Bewegungsänderung postuliert (s.o.), während z.B. das dritte Axiom bei der Entwicklung der Modelle keine Rolle gespielt hat. Solchen Detailanalysen (Profilaussagen) soll mit Subskalen der FCI-Items nachgegangen werden.

4. Eignung des FCI für Profilaussagen

Im Folgenden gehen wir der Frage nach, in wie weit der FCI geeignet ist, über den Gesamtscore hinaus Aussagen über das Verständnis bestimmter Teilbereiche der Mechanik zu machen (Profildaten).

4.1 Reliabilitäten der FCI-Subskalen

Ein Mittel zur Einschätzung der inneren Konsistenz einer Skala sind Reliabilitätsanalysen aufgrund von Cronbachs α (s. Tab. 3). Die Werte von fünf der sieben Subskalen liegen z.T. deutlich unterhalb des in der Psychometrie üblichen Grenzwerts von $\alpha = 0,7$. Nur die Itemgruppe zum 3. Axiom und der FCI-Gesamtscore überschreiten diese Grenze. Das Gesamtergebnis ist demnach ein konsistenter Indikator für ein Konstrukt „allgemeines Verständnis der Mechanik“. Profilaussagen über einzelne Elemente dieses Verständnisses, wie zu Aspekten des Kraftkonzepts, sind dagegen problematisch. Die recht geringen Zusammenhänge zwischen

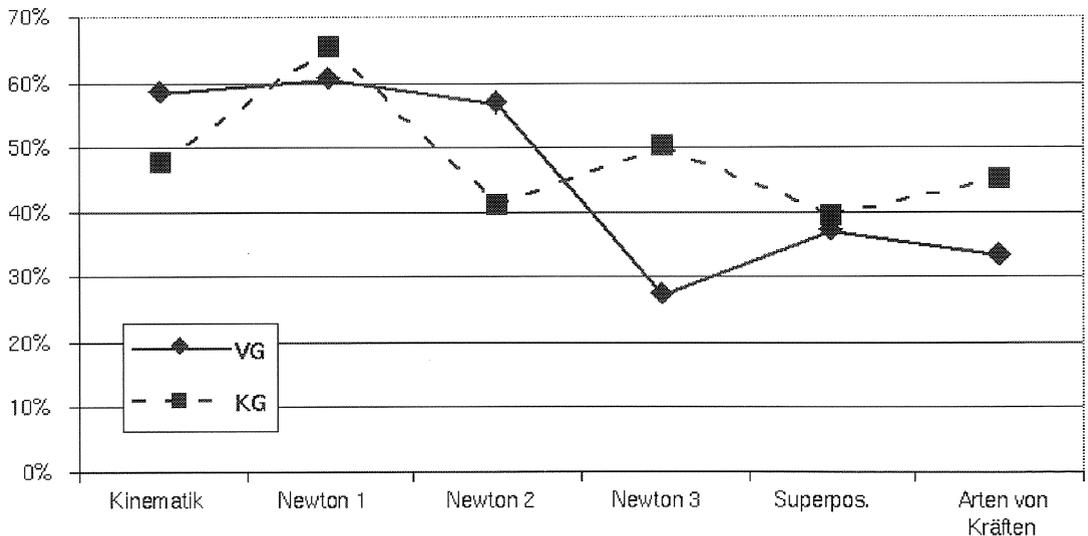


Abb. 3: Vergleich der Gruppen bezüglich der Ergebnisse in den FCI-Subskalen (Anteile richtiger Lösungen).

³ Gruppenunterschiede werden mit dem Mann-Whitney U-Test für ordinale Daten aufgrund eines Vergleichs der Rangplätze untersucht. Der p-Wert 0,09 liegt deutlich unterhalb der Fehlergrenze von 0,3, so daß die Nullhypothese der Gruppengleichheit verworfen werden kann. Der vergleichsweise hohe Schwellenwert (üblich sind sonst z.B. 0,01 oder 0,05) ergibt sich aufgrund der geringen Probandenzahlen bei der Notwendigkeit, Fehler erster und zweiter Ordnung gleichgewichtig zu vermeiden.

Subskalen	Anzahl Items	Reliabilität
Kinematik	6	0,53
1. Axiom	8	0,62
2. Axiom	4	0,44
3. Axiom	4	0,78
Superpositionsprinzip	4	0,18
Arten von Kräften	16	0,52
FCI Gesamt	29	0,77

Tab. 2: Reliabilitäten (Cronbachs α) der Subskalen die von Hestenes und Halloun (1995) für die Konstruktion des FCI angegeben werden; berechnet aus den Ergebnissen des Nachtests für alle Schüler (N=49). Die Summe der Items der Subskalen liegt über 29, weil einzelne Aufgaben zwei Skalen zugeordnet sind.

der Beantwortung von FCI-Aufgaben drücken sich auch darin aus, daß nur 10 von insgesamt 406 möglichen Korrelationen über $\phi=0,5$ liegen. Diese Ergebnisse korrespondieren eng mit der in den Befunden von Huffman und Heller zur Frage „What does the force concept inventory actually measure?“ (s. dazu Punkt 1.2).

Vergleicht man trotz dieser Einschränkungen die Ergebnisse in den 6 Aufgabengruppen des FCI, dann haben die Versuchsgruppenschüler in der Kinematik und beim zweiten Axiom Vorteile (s. Abb. 4), die jedoch durch einen „Trade Off“ in anderen Bereichen, wie dem 3. Axiom, bezahlt werden.

4.2 Antwortverhalten im FCI bei Variationen der Aufgabenstellungen

In einer Voruntersuchung haben wir untersucht, inwieweit unterschiedliche Formulierungen des Aufgabentextes oder der Beispiele das Antwortverhalten der Schüler beeinflussen. Dabei wurden Schülern, die nicht an der Hauptuntersuchung beteiligt waren, Versionen vorgelegt, die gleichzeitig die originale Formulierung sowie alternative Formulierungen bestimmter Aufgaben enthielten. Der physikalische Inhalt blieb dabei erhalten, nur der Kontext, in den der physikalische Inhalt eingebunden war, wurde variiert.

		<i>FCI Aufgabe 22 alternative Formulierung</i>						Total
		Golfball - Fußball-Kontext						
		A	B	C	D	E	X	
<i>FCI</i>	A		2					2
<i>Aufgabe 22</i>	B	17	14	1	6	3	1	42
<i>Original</i>	C	1	1	26	2	1		31
	D				7		1	8
	E					2		2
	X		1		1			2
Total		18	18	27	16	6	2	87

		<i>FCI Aufgabe 5 alternative Formulierung</i>						Total	
		Stahlkugel versus Pistolenschuss-Kontext							
		?	A	B	C	D	E	X	
<i>FCI</i>	?	2			1	1		1	5
<i>Aufgabe 5</i>	A			1		3			4
<i>Original</i>	B								0
	C		2		9	10		2	23
	D		1		7	35		2	45
	E						10		10
Total		2	3	1	17	49	10	5	87

Tab. 3: Kreuztabellierung der Antwortmuster zu zwei variierten FCI-Fragen. Bei Aufgabe 22 (oben) steht Alternative B für eine Impetus-Vorstellung, während A und D newtonsche Lösungen kennzeichnen. Bei Aufgabe 5 (unten) unterstellt man Auswahlantwort C den Impetus-Ansatz; D ist die newtonsche Lösung.

Die größten Unterschiede in den Antworten waren bei solchen Aufgaben zu verzeichnen, die zur Erläuterung eine Abbildung enthielten. So wurde z.B. die Golfball-Aufgabe (s. Abb. 6 weiter unten) durch eine Aufgabe ersetzt, bei der ein Fußballspieler einen Ball auf ein Tor schießt. Von 23 Schülern, die bei der Golfballaufgabe noch der Meinung waren, es wirkten Gravitations- und Abschlagkraft, waren bei der alternativen Aufgabenformulierung 17 Schüler der Meinung, auf den Ball wirke nur die Gravitationskraft, und 6 Schüler gaben die vollständig richtige Antwort, es wirkten Gravitations- und Luftwiderstandskraft (s. Tab. 1). Dieses stark abweichende Antwortmuster läßt sich nicht einfach durch ein probabilistisches Modell des Antwortverhaltens erklären. Es ist eher zu vermuten, dass die unterschiedliche Erfahrung und Assoziationen der Schüler bei dieser Frage eine große Rolle spielen.

Unterschiedliche Antwortmuster zeigten sich aber auch bei Aufgaben, die keine Abbildungen enthielten. Bei Aufgabe 5 wird im FCI nach den Kräften auf eine Stahlkugel gefragt, wenn man sie in die Luft hochwirft. Der alternative Kontext bestand in einem vertikalen Pistolenschuß. Hier haben zahlreiche Wechsel in beiden Richtungen zwischen den Auswahlantworten C (Impetus-Ansatz) und D (newtonsch) stattgefunden (s. Tab. 3). Dies zeigt, wie sehr kontextabhängig die jeweils vom Schüler gegebene Antwort ist, bzw. wie stark das Antwortverhalten fluktuiert. Die Vorhersage eines bestimmten Antwortverhaltens ist nur in sehr begrenztem Maße möglich. Ergebnisse zu einzelnen Aufgaben lassen keine Schlüsse auf Vortellungskategorien zu. Um so wichtiger sind reliable Skalenbildungen.

5. Detailanalysen zum Zusammenhang von Kraft und Bewegung

Tab. 4 zeigt die Ergebnisse der einzelnen Kurse in der Subskala zum zweiten Newtonschen Axiom aus Vor- und Nachtest sowie den gewichteten Zuwachs g in dieser Skala. Diese Subskala besteht aus vier Aufgaben. Es handelt sich um je zwei Teilaufgaben mehrstufiger Fragenkomplexe. Bei den Vortestergebnissen zeigt sich, daß VK1

einen höheren Anfangsscore als die Vergleichskurse hatte, aber dennoch im Verlauf des Mechanikunterrichtes ein höherer gewichteter Zuwachs erreicht wurde. Für Versuchskurs 2 lässt sich aufgrund der fehlenden Daten aus der Eingangserhebung kein Zuwachs berechnen. Vergleicht man den die Nachtestscores, so schneidet die Versuchsgruppe insgesamt mit einem Lösungsanteil von 57% signifikant ($p=0,04$) besser ab als die Kontrollgruppe (41%).

Kurse	Subskala „2. Newtonsches Axiom“		
	Vortest	Nachtest	g
VK1	0,55	0,75	0,44
VK2	/	0,51	/
KK1	0,32	0,50	0,27
KK2	0,25	0,30	0,06

Tab. 4: Ergebnisse zur Subskala „2. Newtonsches Axiom“

Die Aussagekraft dieses Befundes sollte aus den bereits genannten Gründen nicht überbewertet werden: Die Skala hat eine geringe innere Konsistenz, und sie besteht aus nur vier Aufgaben, die zudem paarweise zusammenhängen.

Aus fachdidaktischer Sicht kommen wesentliche inhaltliche Kritikpunkte hinzu: Zieht man die Bedeutung des zweiten Newtonschen Axioms für das Newtonsche Kraftkonzept in Betracht, dann muß man feststellen, daß seine Gewichtung im Vergleich zu anderen Kategorien, wie z.B. der Kategorie „Gravitation“ mit neun Items, zu niedrig ist. In den beiden betroffenen Aufgabenpaaren (6/7 und 24/25) geht es um die Bahnkurven und Tempoänderungen eines Eishockey-Pucks nach einem Schlag und die einer Rakete während der Brenndauer des Triebwerks. Vorgänge mit mehreren wirkenden Kräften kommen nicht vor. Andere Aufgaben des FCI erscheinen uns zumindest ebenso relevant für ein grundlegendes Verständnis des Zusammenhangs zwischen Kraft und Bewegungsänderung zu sein. Wir haben daher aus den Aufgaben des FCI eine eigene Subskala gebildet, um die Frage nach den Wirkungen des Unterrichtsansatzes mit Modellbildungssystemen zuverlässiger überprüfen zu können.

5.1 Konstruktion einer Subskala „Kraftverständnis“

Die von uns zusammengestellte Subskala soll zentrale Elemente des Zusammenhangs zwischen Kräften und Bewegungen testen, die sich in Aufgaben des FCI wiederfinden lassen. Für eine begründete Auswahl aus den 29 Aufgaben wurden folgende Kriterien festgelegt:

- Zunächst wurden alle Aufgaben einbezogen, die mit Kräften und Bewegungsänderungen zu tun haben. Ausgeschlossen wurden rein kinematische Aufgabenstellungen ohne dynamische Zusammenhänge und Aufgaben, die sich mit Randproblemen des Kraftbegriffes wie Auftriebskräften beschäftigen.
- Ausgeschlossen wurden Aufgaben zum Wechselwirkungsprinzip (drittes Newtonsches Axiom), da hier keine spezifischen Wirkungen eines Unterrichts mit Modellbildungssystemen erwartet werden können. Vorteile der Versuchsgruppe in diesem Bereich wurden bereits angemerkt (s. Punkt 4.1).
- Internationale Standardaufgaben aus der Schülervorstellungsforschung sollen möglichst einbezogen sein. Bei Hestenes et al. (1992) werden manche davon in der recht diffusen Kategorie „Arten von Kräften“ subsumiert, z.B. die „book on the table“-Aufgabe nach Arons (1981, S. 169).
- Mindestens drei Auswahlantworten zu einer Aufgabe sollen den Vorstellungskategorien „newtonsch“, „aristotelisch“ oder „Impetus“ zuzuordnen sein, um auch Aussagen über die Art der Fehler machen zu können.

Die Reliabilität der so erhaltenen Skala mit acht Aufgaben wurde berechnet. Durch Streichung einer weiteren Aufgabe konnte ein Cronbachs α von 0,53 erreicht werden. Es blieben damit sieben Items übrig: FCI-Aufgaben 5, 9, 12, 18, 22, 24, 28. Das zentrale Item in dieser Aufgabengruppe, d.h. das Item, bei dessen Streichung der Reliabilitätswert am stärksten sinkt, ist die Golfballaufgabe (s. Abb. 4). Diese Aufgabe ist in der äquivalenten „bouncing ball“-Fassung aus vielen internationalen Studien bekannt - nicht zuletzt aus der TIMSS-Studie (Population III). Wie neue analytische

Studien zur TIMSS-Studie ergaben (Klieme 1998), charakterisiert die Aufgabe ein besonders hohes Fähigkeitsniveau.



Abb. 4: Aufgabe 22 des FCI; gefragt ist nach den Kräften, die am Golfball während der Flugphase angreifen.

5.2 Lösungshäufigkeiten in der Subskala „Kraftverständnis“

Obwohl die Konsistenz der neu gebildeten Skala „Kraftverständnis“ mit $\alpha=0,53$ weiterhin zu wünschen übrig lässt - sie liegt allerdings deutlich über der Hesteneschen „Newton 2“-Gruppierung von $\alpha=0,44$ - soll auf ihrer Grundlage ein erneuter Gruppenvergleich angestellt werden. Die Begründung liegt in der inhaltlichen Bedeutung der betroffenen Aufgaben. Abb. 5 zeigt die Lösungshäufigkeiten der sieben Aufgaben.

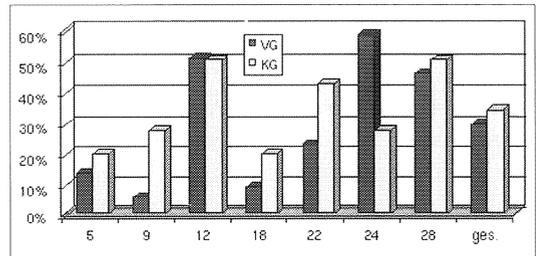


Abb. 5: Lösungshäufigkeiten der Aufgaben in der Skala „Kraftverständnis“ (Nachttest); einbezogen sind alle Schüler, die am Nachttest teilgenommen haben.

Zunächst fällt auf, daß die Lösungsanteile bei diesen von uns als inhaltlich wichtige Aufgaben zum Kraftverständnis zusammengefaßten Subskala mit durchschnittlich 32 % deutlich niedriger liegt als der FCI-Gesamtscore, bzw. die mittlere Lösungswahrscheinlichkeit der übrigen 22 Aufgaben (58 %). Punkte wurden demnach vorwiegend bei den übrigen Aufgabengruppen gesammelt, während der Unterrichtserfolg im zentralen Bereich des Zusammenhangs von Kraft und Bewegungsänderung

unbefriedigend blieb. Der Zusammenhang zwischen dem Ergebnis in der Subskala und dem Lösungsanteil in den restlichen 22 Aufgaben ist allerdings hochsignifikant ($\tau_c=0,83$, 0,1% Niveau). Als schwierigstes Item erwies sich Aufgabe 18, bei der erkannt werden muß, daß für eine Bewegung mit konstanter Geschwindigkeit die Summe der angreifenden Kräfte Null ist (s. Abb. 6).

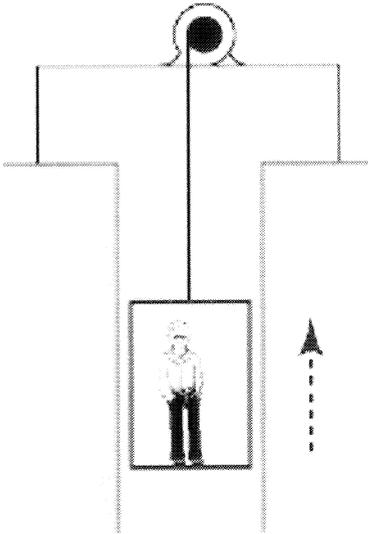


Abb. 6: Aufgabe 18 des FCI; gefragt ist nach dem Verhältnis von Seilkraft und Gravitationskraft auf die Fahrstuhlkabine bei konstanter Aufzugsgeschwindigkeit.

Es ergibt sich im Unterschied zur Hesteneschen „Newton 2“-Skala kein Vorteil für die Schüler, die mit dem Modellbildungssystem gearbeitet haben. Die Vorteile der Kontrollgruppe gegenüber der Versuchsgruppe (s. Abb. 5) sind statistisch nicht signifikant.

5.3 Analyse der Fehlantworten

In einem früheren Artikel über die Alltagsvorstellungen zum Bewegungsbegriff (Hallon & Hestenes, 1985a) weisen die FCI-Autoren auf eine Möglichkeit hin, Vorstellungen nach aristotelisch, impetustheoretisch und newtonsch zu kategorisieren. Wir greifen diesen Ansatz im folgenden auf. Die Möglichkeit zur Einteilung in die drei Kategorien war eine

Voraussetzung für Items, um in die Subskala „Kraftverständnis“ aufgenommen zu werden. Die Einteilung der Antwortmöglichkeiten in die Vorstellungskategorien erfolgte nach folgenden Kriterien:

- Newtonsch: Dies sind die im physikalischen Sinne korrekten Antworten.
- Aristotelisch: Die Gravitationskraft spielt keine Rolle; Körper fallen herunter, weil das ihrem natürlichen Verhalten entspricht; der natürliche Zustand eines Körpers ist die Ruhe; jede Bewegung impliziert eine wirkende Kraft. Die „aristotelischen“ Antwortmöglichkeiten des FCI weichen am weitesten von wissenschaftlichen Erklärungen ab. Sie entsprechen eher naiven Erklärungen von Alltagserfahrungen.
- Impetustheoretisch: Kern dieser Vorstellungskategorie ist die im Mittelalter entwickelte Vorstellung, daß jedem sich bewegenden Körper eine Kraft - ein Impetus - innewohnt, der sich im Verlauf der Bewegung verbraucht. So wirkt nach dieser Vorstellung auch immer eine Kraft in Richtung der Bewegung.

28 von insgesamt 35 Auswahlantworten der sieben Aufgaben der Subskala „Kraftverständnis“ ließen sich einer der drei beschriebenen Vorstellungen zuordnen (11 mal aristotelisch, 8 mal Impetus, 9 mal newtonsch). Nur zwei Aufgaben boten alle drei Möglichkeiten. Vier Aufgaben stellen aristotelische und newtonsche Antworten zur Wahl und eine Aufgabe Newton und Impetus. Diese Defizite der Aufgabenkonstruktionen sind bei der Interpretation der Antwortverteilungen zu berücksichtigen.

Am Nachtest haben 49 Schüler teilgenommen, insgesamt wurden also bei sieben Aufgaben insgesamt 343 Fragen beantwortet. Abb. 11 zeigt die prozentuale Verteilung der gegebenen Antworten auf die einzelnen Vorstellungskategorien, wobei unter der „sonstige“ die Wahlen all jener Antworten verzeichnet sind, die sich nicht den drei genannten Kategorien zuordnen lassen (wie z.B. nicht beantwortete Fragen, „inkonsistenter Kraftbegriff“ (s. Punkt 1.1) etc.).

Die Auswertung ergibt eine in etwa gleichgewichtige Verteilung auf die drei Hauptkategorien. Aristotelische Denkmuster kommen in gleichem Maße zum Zuge wie newtonsche. Die Virulenz der Disposition zum Impetus-Denken wird noch deutlicher, wenn man bedenkt, daß nur bei drei der sieben Aufgaben die Möglichkeit dazu bestand. Dann ist sie auch konsequent gewählt worden.

Das in Abb. 7 gezeigte Bild verändert sich, wenn man die Konsistenz der Antwortmuster einzelner Schüler betrachtet und vor dem Hintergrund der Wahlmöglichkeiten gewichtet. In Tab. 5 sind diese Antwortmuster nach starken und schwachen Tendenzen zu den jeweiligen Kategorien verzeichnet. Eine Tendenz wird dann als stark gesetzt, wenn sich die gewichtete Anzahl der Antworten (also inklusive der Fragen, bei denen es in den Antwortmöglichkeiten keine Lösungsalternative der entsprechenden Kategorie gab) zu mindestens 50% den einzelnen Kategorien zuordnen lässt. Eine schwache Tendenz wurde gesetzt, wenn diese Zahl zwar unter der 50%-Marke lag, der Schüler aber trotzdem mehr als 50% der ihm angebotenen Alternativen zu dieser Kategorie gewählt hat.

Tab. 5 zeigt, daß impetustheoretische Antwortmuster im Vordergrund stehen. Dagegen sind konsequent newtonsche Denker stark unterrepräsentiert. Bei keinem Schüler läßt sich ein Muster erkennen, das aristotelisch geprägt wäre.

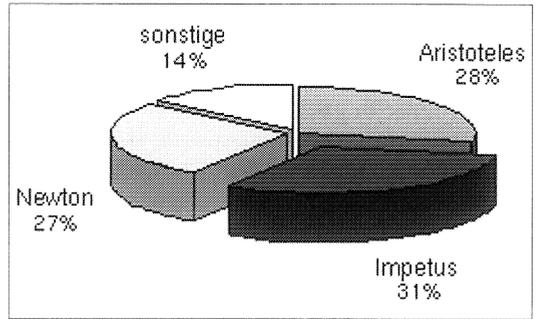


Abb. 7: Aufteilung der Antworten zu den sieben Aufgaben der Subskala „Kraftverständnis“ auf die Vorstellungskategorien (Nachtest, N=49 Schüler).

Vergleicht man diese Ergebnisse mit den Antwortmustern aus der Vorerhebung, dann zeigt sich, daß es nach dem Unterricht bei einer etwa gleichen Verteilung zwischen starken und schwachen Tendenzen doch eine größere Anzahl Schüler gibt, die eine Tendenz in Richtung einer newtonschen Vorstellungskategorie aufweisen.

Bei aller Vorsicht, die sich aus der ungünstigen Konstruktion der FCI Aufgaben und den damit verbundenen Auswertungsproblemen ergibt - wechselnde Kombinationen in den Kategorien der Auswahlantworten -, lassen sich einige Schlüsse ziehen:

- Auch nach dem Mechanikkurs findet man nur selten konsequent newtonsche Denker unter den Schülern.

	Vorstellungskategorie			Tendenz	nicht entscheidbar
	aristotelisch	impetustheoretisch	newtonsch		
VK1	0%	0%	60%	stark	0%
	0%	40%	0%	schwach	
VK2	0%	89%	0%	stark	0%
	0%	11%	0%	schwach	
KK1	0%	20%	13%	stark	13%
	0%	53%	0%	schwach	
KK2	0%	36%	0%	stark	27%
	0%	18%	18%	schwach	

Tab. 5: Kategorisierung der Antwortmuster Nachtest.

- Es herrschen Mischzustände aus allen drei Dispositionen vor. In den Mischzuständen dominiert die Impetuskonzeption, während aristotelische bzw. naive Herangehensweisen untergeordnet sind. Eine positive Wertung dieses Befunds bestünde darin, ihn als Zwischenzustand auf dem Weg zu einem Newtonschen Konzept zu betrachten (vgl. die „intermediate states“ bei Niedderer & Goldberg, 1995).
- Die qualitative Analyse der Fehlantworten bestätigt die quantitative Auswertung der Lösungshäufigkeit in der Feststellung, daß die Aufgaben der Subskala „Kraftverständnis“ im Rahmen des FCI besonders schwierig sind.
- Der Unterrichtseinsatz von Modellbildungssystemen hat bezüglich der Verfestigung newtonscher Denkmuster keinen nachweisbaren Effekt.

6. Zusammenfassung und Diskussion

6.1 Zur Wirkung von Modellbildungssystemen

Die zentrale Wirkungshypothese der Studie zur Arbeit mit einem grafikorientierten Modellbildungssystem im Mechanikunterricht lautete:

Die Schüler der Versuchsgruppe entwickeln in höherem Maße die Fähigkeit zur halbquantitativen Beschreibung von Bewegungsvorgängen unter dem Einfluß von Kräften.

Die Untersuchungen auf Basis des FCI bieten hierfür keine Bestätigung. Die Schüler der Versuchsgruppe schnitten beim FCI-Gesamtergebnis im Nachtest signifikant schlechter ab als die Kontrollgruppe. Absolut gesehen ist der Unterschied jedoch gering. Der gewichtete Lernzuwachs war in Versuchskurs 1 dagegen höher als in der Kontrollgruppe. (Für Versuchskurs 2 kann kein Hake-Faktor g berechnet werden.)

Eine Profilanalyse des Verständnisses des Zusammenhangs zwischen Kraft und Bewegungsänderung ist wegen der nicht befriedigenden inneren Konsistenz der zu bildenden Subskalen des FCI schwierig. Betrachtet man den-

noch die dafür inhaltlich relevante Aufgaben- gruppe, so erreichten Versuchs- und Kontroll- gruppe praktisch gleiche Ergebnisse. Der Lernerfolg kann insgesamt nicht befriedigen. 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 innewohnender Antrieb) und newtonschen Denkweisen, wobei die „Kraft speichern“-Vorstellung überwog. Konsequenterweise newtonsche Denker waren in beiden Gruppen rar.

Diese Befunde weichen von den Ergebnissen einer Interviewstudie ab, die in den gleichen Kursen durchgeführt wurde (Schecker & Gerdes, im Druck). Hier ergaben sich am Ende des Mechanikunterrichts signifikante Vorteile der Versuchsgruppe beim Entwurf eines $v(t)$ -Diagramms zu einem nicht-trivialen Fahrbah- nexperiment und dessen verbaler Erläuterung. Die Versuchsgruppenschüler argumentierten auf einem qualitativ höheren Niveau, insbesondere gelang es ihnen besser, den Bewegungs- verlauf - abnehmende Beschleunigung, bzw. immer geringerer Geschwindigkeitszu- wachs - mit seinen dynamischen Ursachen - abnehmende beschleunigende Kraft - zu ver- binden. Die Vorteile gingen allerdings bei einem späteren Interview zu nicht-mechani- schen Kräften verloren.

Eine Rangkorrelation der Ergebnisse von 33 Schülern, die sowohl am FCI als auch an den Interviews teilgenommen haben, ergibt einen statistisch signifikanten Zusammenhang (Kendalls $\tau_c=0,39$; Irrtumswahrscheinlichkeit $<1\%$) zwischen dem Abschneiden in den bei- den Tests. Die Korrelation sinkt auf $\tau_c=0,33$, wenn man die Subskala „Kraftverständnis“ heranzieht.

6.2 Zur Aussagekraft des FCI

Die von Huffman und Heller (1995a) gestellte Frage „What does the FCI actually measure?“ hat nach unseren Ergebnissen durchaus seine Berechtigung. Ausreichend reliabel ist praktisch nur das Gesamtergebnis. Dieses testet aller- dings, wie wir aufgrund des Ergebnisvergleichs zwischen der Subskala „Kraftverständnis“ und

dem restlichen Aufgabenpool sowie der inhaltlichen Fehleranalyse zeigen konnten, nicht spezifisch den Kern eines grundlegenden Verständnisses des Zusammenhangs zwischen Kraft und Bewegung. Was der FCI getestet, ist ein diffuser Lernerfolg im Themenbereich Mechanik. Unter psychometrischen Gesichtspunkten ist der FCI verbesserungsbedürftig, wenn man damit Schülervorstellungen diagnostizieren möchte. So sind z.B. die Subkategorien der Konzeptualisierung des Kraftbegriffes nicht ausgewogen in den Items repräsentiert; einige Vorstellungsmuster und Teilgebiete (wie z.B. das zweite Newtonsche Axiom) werden nicht häufig genug abgetestet, um ausreichend reliable Skalen zu bilden.

Unsere Ergebnisse sind unter dem Vorbehalt der geringen Probandenzahlen zu sehen. Es ist erstaunlich, daß eine detailliertere Analyse der Aussagekraft des FCI einschließlich einer Fehleranalyse bisher kaum erfolgt ist - obwohl z.B. in den USA Tausende von Datensätzen vorliegen. Angesichts der weiten Verbreitung des FCI wäre es wichtig, in Fortsetzung der Arbeit von Huffman und Heller auf Basis einer breiteren Stichprobe zu klären, ob man dem FCI - wie es sich anzudeuten scheint - tatsächlich ein Ein-Faktoren-Modell unterstellen muß, so daß eine Unterscheidung nach Subskalen keine bessere Anpassung des Strukturmodells an die Daten ergibt, oder doch ein Mehr-Faktoren-Modell entwickelbar ist, das von den Konstruktionsmodellen abweicht, die die Autoren des FCI angewandt haben.

Das Ziel der Entwicklung reliabler diagnostischer Tests für Schülervorstellungen erscheint uns lohnenswert. Bei einer Überarbeitung des FCI müßte dieser diagnostische Aspekt, der gegenüber dem Ziel des globalen Vergleichs von Unterrichtskonzeptionen inzwischen in den Hintergrund getreten ist, wieder betont werden. Eine konkrete Forderung besteht z.B. darin, bei allen Kraftaufgaben mindestens die Auswahlalternativen „Newton“, „Impetus“ und „Aristoteles“ vorzusehen. Gezielte diagnostische Tests für Kraftverständnis, Kinematik usw. auf Basis des Konstrukts der „Schülervorstellungen“ wären ein wichtiges Instrument, um die Repräsentation von Wissen im kognitiven System der Schüler näher zu ergründen.

Literatur

- Arons, A.B. (1981): Thinking, reasoning and understanding in introductory physics courses. *The Physics Teacher*, 19, 166-172.
- Dartmouth College (1997): STELLA 2.0 - Software for Education. Dartmouth, NH: High Performance Systems.
- Gerdas, J. & Schecker, H. (1998): Der Force Concept Inventory - Ein diagnostischer Test zu Schülervorstellungen in der Mechanik. *Der mathematische und naturwissenschaftliche Unterricht (im Druck)*.
- Hake, R.R. (1998): Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66, 64-74.
- Halloun, I.A. & Hestenes, D. (1985a): Common sense concepts about motion. *American Journal of Physics*, 53, 1056-1065.
- Halloun, I.A. & Hestenes, D. (1985b): The initial knowledge state of college physics students. *American Journal of Physics*, 53, 1043.
- Hestenes, D., Wells, M. & Swackhamer, G. (1992): Force concept inventory. *The Physics Teacher*, 30, 141-158.
- Heller, P. & Huffman, D. (1995): Interpreting the force concept inventory - A reply to Hestenes and Halloun. *The Physics Teacher*, 33, 503-511.
- Hestenes, D. & Halloun, I. (1995): Interpreting the force concept inventory - A response to Huffman and Heller. *The Physics Teacher*, 33, 502-506.
- Huffman, D. & Heller, P. (1995): What does the force concept inventory actually measure? *The Physics Teacher*, 33, 138-143.
- Klieme, E. (1998): Stufen mathematisch-naturwissenschaftlicher Kompetenz. Vortrag auf der Tagung der Arbeitsgemeinschaft für Empirische Pädagogische Forschung in Mannheim, September 1998.
- Mazur, E. (1997): *Peer Instruction*. London: Prentice Hall.
- Niedderer, H. & Goldberg, F. (1995): Lernprozess beim elektrischen Stromkreis. *Zeitschrift für Didaktik der Naturwissenschaften*, 1, 73-86.
- Schecker, H. (1988): Denkaufgaben zum Kraftbegriff. *Naturwissenschaften im Unterricht - Physik*, 36, 36-39.
- Schecker, H. (1998a): Physiklernen mit Modellbildungssystemen - Forschungskonzeption. Behrendt, H. (Hrsg.): *Zur Didaktik der Physik und Chemie - Probleme und Perspektiven*. Alsbach: Leuchtturm, 230-232.

- Schecker, H. (1998b): Physik modellieren. Stuttgart: Klett.
- Schecker, H. & Gerdes, J. (1998): Interviews über Experimente zu Bewegungsvorgängen. Zeitschrift für Didaktik der Naturwissenschaften (im Druck).

Dr. Horst Schecker ist Privatdozent im Institut für Didaktik der Physik der Universität Bremen.

Jörn Gerdes ist Lehrer am Ökumenischen Gymnasium Bremen und wissenschaftlicher Mitarbeiter im Institut für Didaktik der Physik der Universität Bremen.

Dr. Horst Schecker
Universität Bremen
Institut für Didaktik der Physik
Postfach 330440
28334 Bremen
email: schecker@physik.uni-bremen.de