

## Verständnis der newtonschen Mechanik bei bayerischen Elftklässlern – Ergebnisse beim Test „Force Concept Inventory“ in herkömmlichen Klassen und im Würzburger Kinematik-/Dynamikunterricht

Thomas Wilhelm

Lehrstuhl Didaktik der Physik, Am Hubland, 97074 Würzburg

(Eingegangen: 17.03.2005; Angenommen: 25.07.2005)

### Kurzfassung

Der Test „Force Concept Inventory“ (FCI) ist in den USA weit verbreitet und wird an Schulen und Hochschulen zur Bestimmung des qualitativen Verständnisses der newtonschen Mechanik eingesetzt. In Deutschland gibt es zwar Ergebnisse von Hochschulen, aber kaum Ergebnisse von Gymnasien.

In der vorgestellten Untersuchung wurde der Test 13 elften Klassen in Bayern als Vor- und Nachtest vorgelegt (kein Kurssystem). Deren Ergebnisse in Vor- und Nachtest werden vorgestellt, wobei auch auf besonders schwere bzw. lernresistente Aufgaben eingegangen wird. Die Leistungszuwächse der ganzen Klassen sind deutlich geringer als in Leistungskursen oder bei Studenten.

Schließlich wird ein Vergleich mit den Ergebnissen aus einem Projekt („Würzburger Kinematik-/Dynamikunterricht“) durchgeführt, in dem qualitatives Verständnis durch den Einsatz dynamisch ikonischer Repräsentationen und graphischer Modellbildung gefördert wurde und die Lehrer besonders trainiert wurden. Hier zeigen die Schüler (aus 7 Klassen) einen signifikant besseren Leistungszuwachs.

### 1. Der Test

Der bekannte Test „Force Concept Inventory“ (FCI) wurde 1992 von Hestenes, Wells und Swackhamer (1992, S. 142) [1] vorgestellt und war zunächst als Hilfe für Lehrer gedacht, das Denken ihrer Schüler zu erforschen. Der Test ist mittlerweile aber zu einem standardisierten Diagnoseinstrument geworden, das in den USA an Highschools, Colleges und Universitäten in breitem Umfang eingesetzt wird (Girwicz, Kurz et al., 2003, S. 1) [2] und zum Vergleich der Lernwirkung unterschiedlicher Unterrichtskonzepte angewandt wird (Schecker, Gerdes, 1999, S. 75) [3]. Dazu hat insbesondere die Metaanalyse von Hake (1998) beigetragen [4]. Einen Überblick über die Geschichte des Tests, seine Struktur und die wichtigsten Forschungsergebnisse findet man bei Savinainen und Scott (2002) [5].

Der Test besteht aus Denkaufgaben zur newtonschen Mechanik (Multiple-Choice-Aufgaben mit je fünf Antwortalternativen), bei denen Antwortalternativen auf bekannte Schülervorstellungen abgestimmt sind und eine Entscheidung zwischen dem newtonschen Konzept und den Alltagsvorstellungen erzwungen wird (Hestenes et al., 1992, S. 142) [1]. Die ursprünglich von Hestenes et al. (1992, S. 154 – 158) [1] veröffentlichte Formulierung enthält 29 Items. Diese wurden von Schecker und Gerdes übersetzt und eingesetzt [3, 6, 7, 8]. Im Rahmen der hier vorliegenden Arbeit wurde die Formulierung kaum geändert, aber der Test neu formatiert und die Bilder neu in besserer Qualität erstellt. Die ursprüngliche erste Version des FCI-Testes wurde jedoch in den

USA aufgrund der gemachten Erfahrungen überarbeitet, klarer formuliert und auf 30 Items erweitert (Halloun et al., 1997) [9]. Kautz hat von dieser deutlich verbesserten zweiten Version eine deutsche Version hergestellt [2], die unter <http://modeling.asu.edu/R&E/Research.html> downloadbar ist wie auch die englischsprachige überarbeitete Version.

Die Validität und Reliabilität des FCI-Tests und insbesondere die Zuordnungen der Items zu sechs Inhaltsbereichen (= Subskalen des FCI-Tests) wurde häufig untersucht, kritisiert und in Frage gestellt [1, 3, 6, 8, 10 – 13]. Es stellte sich auch heraus, dass der Kontext einer Aufgabe beeinflusst, welche Schülervorstellungen aktiviert werden [3, 6, 14]. Ein Überblick über diese Diskussionen gibt Wilhelm [15].

### 2. Ergebnisse von Gymnasiasten

Im Schuljahr 2003/04 haben 13 herkömmlich unterrichtete elfte Klassen aus fünf bayerischen Gymnasien (aus Mittelfranken, Oberfranken und der Oberpfalz) den FCI-Test (1. Version) am Schuljahresbeginn und im letzten Schuljahresdrittel (einige Wochen nach Abschluss des relevanten Unterrichts) durchgeführt (acht Klassen aus dem mathematisch-naturwissenschaftlichen Zweig, in einem Fall Angabe unsicher. In diesem Zweig sind die Lehrer aus Zeitgründen eher zu einem Test bereit). Dabei haben jeweils alle elften Klassen der entsprechenden Schulen teilgenommen. Damit ist nicht nur eine Streuung über verschiedene Schulzweige, sondern auch über unterschiedliche Lehrer gewährleistet. In die Aus-

wertung wurden nur die 258 Schüler aufgenommen, von denen sowohl ein Vor- als auch ein Nachtest vorliegt.

**2.1 Beschreibungen der Gesamtergebnisse**

Da die Lehrer angehalten waren, den Schülern so viel Zeit zu geben, wie diese benötigen, trat nicht wie bei Girwidz, Kurz et al. (2003, S. 2) der Effekt auf, dass die späteren Items nicht von allen bearbeitet wurden [2]. Keine Antwort wurde von den Schülern nur dann gegeben, wenn sie sich für keine Lösungsmöglichkeit entscheiden konnten. Beim Nachtest hat im Durchschnitt ein Schüler 0,14 Items (= 0,5 %) nicht beantwortet. Beim Vortest, bei dem sich die Schüler schwerer taten, hat dagegen ein Schüler im Durchschnitt 0,55 Items (= 1,9 %) nicht beantwortet. Mit Abstand am häufigsten (6,6 % der Schüler) wurde das Item 28 nicht beantwortet, bei dem es um die Kräfte geht, sie auf eine Kiste wirken, die auf dem Boden mit Reibung und konstanter Geschwindigkeit gezogen wird (Kräftegleichgewicht). Sinnvollerweise wurden in der zweiten Version des FCI-Tests bei diesem Item die Antwortmöglichkeiten einfacher formuliert und ein Teilaspekt in einem getrennten Item abgefragt.

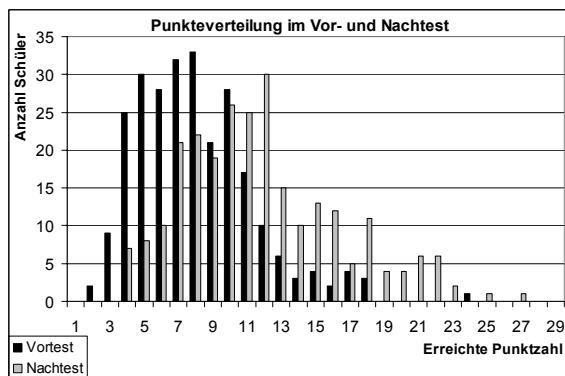


Abb. 1: Histogramm der Punkteverteilung in Vor- und Nachtest (Anzahl Schüler, die die jeweilige Anzahl von Items richtig lösten) (258 konventionell unterrichtete bayerische Elftklässler)

Für jedes richtig gelöste der 29 Items gab es einen Punkt. Das Histogramm der Punkteverteilung beim Vor- und Nachtest zeigt Abb. 1. Beim Vortest ergab sich bei den 258 Schülern ein Mittelwert von 8,0 Punkten (= 28 % richtig gelöste Items) (Standardabweichung: 3,4 Punkte = 12 %), wobei sich die Verteilung von 2 bis 24 Punkten (= 7 % bis 83 %) erstreckt (mehr als sechs Standardabweichungen). Beim Nachtest ergab sich ein Mittelwert von 11,8 Punkten (= 41 %) (mit größerer Standardabweichung: 4,6 Punkte = 16 %), wobei die Verteilung von 4 bis 27 Punkten (= 14 % bis 93 %) reicht (fünf Standardabweichungen). Das ergibt einen relativen Zugewinn der ganzen Gruppe von nur  $g = 18\%$ . Ein Vergleich der  $g$ -Werte mit Girwidz, Kurz et al. (2003, S. 6) [2] ist nicht sinnvoll, da es sich dort um deutsche Studenten handelt. Ein Vergleich mit den Ergebnissen von Hake (1998, S. 66) [4] zeigt, dass

1113 amerikanische Highschools Schüler aus 14 Highschools auch auf einen Vortestwert von 28 % kommen. Die relativen Zugewinne von Highschools, Colleges und Universitäten liegen bei traditionellem Unterricht jeweils nahe bei 23 % (keine genaueren Angaben vorhanden). Im Gegensatz zum Vorgehen von Hake wird außer dem relativen Zugewinn der Gruppe (berechnet aus den Mittelwerten bei Vor- und Nachtest) noch zusätzlich der mittlere relative Zugewinn (berechnet aus den relativen Zugewinnen der einzelnen Schüler) berechnet (bei diesem Test immer fast der gleiche Wert), um auch beim Vergleich verschiedener Gruppen Signifikanztests durchführen zu können.

Interessant ist nicht nur die Streuung bei den Schülerergebnissen, sondern auch bei den Klassenergebnissen. Das mittlere Vortestergebnis der Klassen variiert zwischen 21 % und 39 %, das Nachtestergebnis zwischen 27 % und 53 %. Der relative Zugewinn der einzelnen Klassen reicht von 7 % bis 31 % (Mittelwert der Klassenergebnisse:  $18,1\% \pm 7,0\%$ ).

Da Hestenes und Halloun (1995, S. 505) [12] ein Testergebnis von 60 % richtigen Antworten als Schwelle für ein newtonsches Verständnis ansehen, wurde geschaut, wie viele Schüler 17 oder mehr der 29 Items richtig lösten ( $\geq 59\%$ ). Dieser Anteil stieg von 3 % der Schüler im Vortest auf lediglich 15 % der Schüler im Nachtest an. Das ist ein bedenkenswert schlechtes Gesamtergebnis.

**2.2 Analyse auf Itemebene**

Für Lehrer, die ihren Unterrichtserfolg mit wenigen Items abprüfen wollen, ist es wichtig, wie die einzelnen Items beantwortet wurden, weshalb hier noch eine kurze Analyse auf Itemebene gemacht wird. In Abb. 2 sind die Vortestergebnisse der einzelnen Items an der Oberkante der schwarzen Säulen abzulesen. Die Säulen für absolute Zugewinne wurden auf die der Vortests gestellt, während die Säulen für die Verluste an die Grundlinie gestellt wurden. Die Ergebnisse des Nachtests erhält man bei Zugewinn als Summe (= Oberkante graue Säulen) und bei Verlust als Differenz der schwarzen und hellen Säulen.

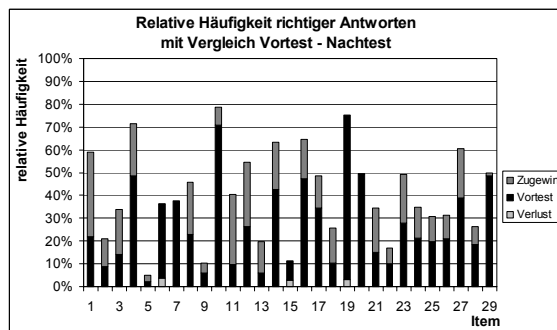


Abb. 2: Relative Häufigkeit richtiger Antworten bei den einzelnen Items mit Vergleich von Vor- und Nachtest (258 bayerische Elftklässler)

| Item erste (zweite) Version | richtig (Vortest/ Nachttest) | Thema der Aufgabe  | Häufigste falsche Antworten, Alltagsvorstellung                | Häufigkeit Vortest/ Nachttest |
|-----------------------------|------------------------------|--|--|-------------------------------|
| 5 (13)                      | 2% / 5%                      | Kräfte auf eine senkrecht nach oben geworfene Kugel                    | <b>D: I3. Allmählicher Verlust des Impetus (Dissipation)</b>   | 60% / 76%                     |
| 15 (-)                      | 11% / 9%                     | Fallender und zurückprallender Hartgummiball                           | <b>A: AF1. Nur aktive Körper üben Kräfte aus.</b>              | 33% / 38%                     |
| 9 (11)                      | 6% / 10%                     | Kräfte bei reibungsfreien Gleiten auf einer Oberfläche                 | <b>B und C: II. Impetus vermittelt durch „Schlag/Stoß“</b>     | 47% / 36%<br>38% / 48%        |
| 22 (30)                     | 10% / 17%                    | Kräfte auf einem fliegenden Golfball                                   | <b>C: II. Impetus vermittelt durch Schlag/Stoß</b>             | 81% / 72%                     |
| 13 (15)                     | 6% / 20%                     | Kräfte zwischen Fahrzeugen (leichtes schiebt schweres beschleunigt an) | <b>C: AR2. Der aktivste Körper produziert die größte Kraft</b> | 68% / 68%                     |
| 2 (4)                       | 9% / 21%                     | Kräfte beim Zusammenstoß zweier unterschiedlich schwerer Fahrzeuge     | <b>A: AR1. Größere Masse impliziert größere Kraft</b>          | 84% / 76%                     |
| 18 (17)                     | 10% / 26%                    | Kräfte auf einen Aufzug bei konst. Steiggeschwindigkeit                | <b>A: AF2. Bewegung impliziert aktive Kraft</b>                | 49% / 50%                     |
| 28 (-)                      | 19% / 26%                    | Kraft auf Kiste auf Boden mit Reibung bei $v = \text{konstant}$        | <b>D: AF2. Bewegung impliziert aktive Kraft</b>                | 41% / 52%                     |

Tab. 1: Die acht schwierigsten Aufgaben (im Nachttest weniger als 30 % der Schüler richtig) und die entsprechenden Fehlkonzepte (Index nach [1], Übersetzung von [6]) mit Angabe der Häufigkeiten der richtigen und häufig falschen Antworten in Vor- und Nachttest (13 bayerische elfte Klassen, 258 Schüler)

Beim Vortest reicht der Anteil richtiger Antworten von 2 % (Item 5) bis 75 % (Item 19). Die schwierigsten Items (unter 12 % richtig) waren die Items 5, 9, 13, 2, 11, 22, 18 und 15 (siehe Tab. 1). Über 50 % richtige Aufgaben gab es nur bei Item 19 (75 %, Bewegungsrichtung bei zwei unterschiedlichen Kräften) und bei Item 10 (Bewegungsrichtung eines Balles, der aus einem kreisförmigen Kanal austritt). Item 19 ist eigentlich schlecht formuliert, denn gemeint ist, in welche Richtung sich die zunächst ruhende Kiste anfangs bewegt; es wird aber der Eindruck erweckt, es ging um die Bahnkurve während einer längeren Bewegung. Dieses Item wurde in der zweiten Version des FCI-Testes weggelassen. Interessant ist, dass das zu Item 10 physikalisch äquivalente Item 4 (Ball wird an Faden herum geschwungen und Faden reißt) deutlich schlechter als das Item 10 ausfällt (48 %).

Beim Nachttest reichte der Anteil richtiger Antworten von 5 % (Item 5) bis 79 % (Item 10). Interessant sind insbesondere die am schlechtesten bearbeiteten Aufgaben und die hinter den falschen Antworten stehenden Fehlvorstellungen. Die Items, die von weniger als 30 % der Schüler richtig beantwortet wurden, findet man in Tab. 1 einschließlich der relevanten Fehlvorstellungen. Über 70 % richtige Antworten gab es wieder bei Item 10 und 19 sowie nun bei dem zu Item 10 äquivalentem Item 4.

In Abb. 2 sind auch die absoluten Veränderungen durch den Unterricht ablesbar. Überraschend ist, dass es vier Items gibt, bei denen sich die Schüler sogar leicht verschlechtert haben (Nr. 6, 19, 15 und 7). Der Zugewinn ist damit bei den einzelnen Items zwischen -4 Prozentpunkten (Nr. 6) und 37 Prozentpunkten (Nr. 1).

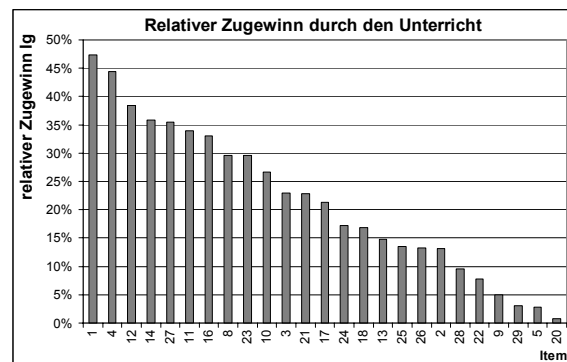


Abb. 3: Relativer Zugewinn durch herkömmlichen Unterricht (N = 258) bei den Items mit positivem Zugewinn

Da aber der absolute Zugewinn nur in Zusammenhang mit dem Vortestwert aussagekräftig ist, wurde wie bei Girwidz, Kurz et al. (2003) [2] ein relativer Zugewinn  $I_g$  pro Item berechnet (siehe Abb. 3). Der  $g$ -Wert von Hake (1998, S. 65) [4] bezieht sich auf den relativen Zugewinn an Punkten einer Gruppe bei allen Items: Zugewinn an Punkten oder Prozentpunkten durch möglichen Zugewinn. Der  $I_g$ -Wert von Girwidz et al. [2] bezieht sich auf den relativen Zugewinn an richtig antwortenden Schülern bei einem Item: Zugewinn an Schülern durch möglichen Zugewinn.

Größere relative Zugewinne von über 40 % gab es nur bei dem Item 4 ( $I_g = 44$  %, siehe Beschreibung oben) und Item 1 ( $I_g = 47$  %). Bei Item 1 wurde nach den Fallzeiten zweier gleich großer, aber unterschiedlich schwerer (Faktor 2) Metallkugeln gefragt, die von einem Gebäudedach fallen (ein Rechenbeispiel mit Luftreibung (und Auftrieb) ergibt für 4 m

| Item erste (zweite) Version | richtig (Vortest/ Nachttest) | Thema der Aufgabe                                      | Häufigste falsche Antworten, Alltagsvorstellung   | Häufigkeit Vortest/ Nachttest |
|-----------------------------|------------------------------|--|---|-------------------------------|
| 19 (-)                      | 75% / 72%                    | Bewegungsrichtung bei unterschiedlichen Kräften        | A: C11. Größte Kraft bestimmt Bewegung<br>C: C12. Kraft-Kompromiss bestimmt Bewegung                        | 13% / 12%<br>7% / 12%         |
| 6 (8)                       | 36% / 33%                    | Bahnkurve einer Scheibe nach einem senkrechten Kick    | D: I4. gradueller/verzögerter Impulsaufbau/-gewinn  | 34% / 36%                     |
| 15 (-)                      | 11% / 9%                     | Fallender und zurückprallender Hartgummiball           | A: AF1. Nur aktive Körper üben Kräfte aus.  | 33% / 38%                     |
| 7 (9)                       | 38% / 37%                    | Tempo einer Scheibe nach einem senkrechten Kick        | B und C: K3. Keine vektorielle Zusammensetzung von Geschwindigkeit  | 27% / 17%<br>13% / 23%        |
| 20 (19)                     | 49% / 50%                    | Geschwindigkeitsvergleich anhand von Stroposkopbildern | D: K1. Ort-Geschwindigkeit nicht unterschieden<br>A: K2. Geschwindigkeit-Beschleunigung nicht unterschieden | 19% / 23%<br>21% / 17%        |
| 5 (13)                      | 2% / 5%                      | Kräfte auf eine senkrecht nach oben geworfene Kugel    | D: I3. Allmählicher Verlust des Impetus (Dissipation)   | 60% / 76%                     |
| 29 (27)                     | 48% / 50%                    | Bewegungsart nur unter Einfluss von konstanter Reibung | B: I3. allmählicher Verlust des Impetus (Dissipation)   | 29% / 33%                     |

Tab. 2: Die sieben resistentesten Aufgaben (relativer Zugewinn  $I_g$  unter 4 %) und die entsprechenden Fehlkonzepte (Index nach [1], Übersetzung von [6]) mit Angabe der Häufigkeiten der richtigen und häufig falschen Antworten in Vor- und Nachttest (13 bayerische elfte Klassen, 258 Schüler)

Fallhöhe nur sehr geringe Differenzen in der Fallzeit). Beim Vortest antworteten 60 % entsprechend der Alltagserfahrung, dass die schwerere Kugel deutlich weniger Zeit braucht, aber nur 16 % behaupten es sei die Hälfte (24 %: etwa gleiche Zeit). Beim Nachttest geben nun 59 % die hier richtige Antwort, dass beide ungefähr gleich lange brauchen (24 %: schwere deutlich weniger, 16%: schwere die Hälfte). Unklar ist, ob die Schüler nun meinen, alle Körper fallen gleich schnell, oder ob sie eine Vorstellung haben, wann sich die Luftreibung wie stark auswirkt.

Die Items, bei denen der relative Zugewinn  $I_g$  kleiner als 4 % ist, findet man in Tab. 2 einschließlich der relevanten Fehlvorstellungen. Es handelt sich dabei um die Items, die resistent gegenüber dem herkömmlichen Unterricht sind, wobei es große Unterschiede zwischen den Häufigkeiten gibt. Die Items 19, 20 und 29 haben zwar nur sehr geringe Zugewinne, schneiden aber beim Vortest mit am Besten ab, während die Items 5 und 15 außerdem bei Vor- und Nachttest sehr schlecht abschneiden. Bei einem anderen Test [15, 16], in dem die Schüler bei gleicher Aufgabenstellung wie in Item 5 die Kraft-richtung angeben sollten, ohne dass attraktive Textantworten vorgegeben wurden, konnten 22 % der bayerischen Gymnasiasten ( $N = 188$ ) die richtige Lösung angeben, d.h. die Distraktoren dieser Aufgabe des FCI-Tests sind sehr attraktiv formuliert.

### 2.3 Analyse von Subskalen

Will man verschiedene Klassen oder Unterrichtskonzepte vergleichen, ist der Vergleich bei den einzelnen Items schwierig und nicht notwendigerweise aussagekräftig. Um nicht nur das Gesamtergebnis

vergleichen zu können, könnte man bei aller gebotener Vorsicht die von Hestenes et al. (1992, S. 142) [1] vorgeschlagenen Subskalen, die von Schecker et al. (1999, S. 84) [3] vorgeschlagene Subskala oder andere Subskalen analysieren. Wie bei Huffman und Heller (1995) [11] bzw. Heller und Huffman (1995) [13] findet man beim Nachttest aber keine oder sehr schwache Korrelationen zwischen den einzelnen Items (von den 406 Korrelationskoeffizienten liegen nur sieben über 0,3 und nur 29 weitere über 0,2) und bei einer Faktorenanalyse nach der Hauptkomponentenmethode mit Varimax-Rotation findet man nur wenige Faktoren, die außerdem nicht den Subskalen entsprechen. Eine Faktorenanalyse bei den nur 258 Schülern (Nachttest, dichotomen Auswertung) ergibt (sowohl nach einem Scree-Test als auch, wenn ein Faktor 5 % der Gesamtvarianz erklären muss) vier Faktoren: Die Faktoren 1 und 3 sind nicht interpretierbar, Faktor 2 (Item 2, 3, 11 und 13) entspricht fast der Subskala „3. newtonsches Axiom“ und Faktor 4 besteht aus den physikalisch äquivalenten Items 6 und 26. Werden zu der Faktorenanalyse weit mehr Schülerantworten hinzugenommen (Vor- und Nachttest, Vergleichs- und Treatmentgruppen) ergeben sich auch bei 970 Schülerantworten nur drei nicht interpretierbare Faktoren. Dennoch werden im Folgenden die Subskalen gemäß Hestenes et al. [1] und Schecker et al. [3] analysiert.

Wie bei Gerdes, Schecker (1999, S. 287) [6] nehmen die Reliabilitäten der Subskalen (Cronbachs  $\alpha$ ) von Vor- zu Nachttest zu, sind aber in der Regel unter 0,7 (siehe Tab. 3, linker Teil). Wie Hestenes und Halloun (1995) [12] bemerken, haben Schüler, die nicht über das newtonsche Konzept verfügen, keine konsistente Alltagstheorie. Deshalb liegen die Reliabilitä-

ten beim Vortest noch recht niedrig und steigen mit zunehmendem newtonschen Verständnis in der untersuchten Gruppe. Es ergeben sich bei den Nachtestwerten in der Regel leicht geringere Werte als bei Schecker et al. (1999, S. 82) [3], nur bei den Subskalen „3. Newton“ und „Kraftverständnis“ sind die Nachtestwerte deutlich geringer. Bei den im Abschnitt 3 diskutierten Nachtestwerten der Treatmentgruppe ergeben sich allerdings außer beim Gesamtwert auch bei den Subskalen „1. Newton“, „Arten von Kräften“ und „Kraftverständnis“ Reliabilitäten in der Größe von  $\alpha \approx 0,7$ .

| Subskala          | Reliabilität $\alpha$ |              |                     | Vor-test       | Nach-test      | relativer Zugewinn g |
|-------------------|-----------------------|--------------|---------------------|----------------|----------------|----------------------|
|                   | im Vor-test           | im Nach-test | bei Treatmentgruppe |                |                |                      |
| Gesamt            | 0,61                  | 0,75         | 0,80                | 28 %<br>± 12 % | 41 %<br>± 16 % | 18 %<br>(± 18%)      |
| Kinematik         | 0,38                  | 0,50         | 0,57                | 28 %           | 39 %           | 15 %                 |
| 1. Newton         | 0,46                  | 0,54         | <b>0,67</b>         | 33 %           | 47 %           | 20 %                 |
| 2. Newton         | 0,17                  | 0,31         | 0,41                | 29 %           | 34 %           | 7 %                  |
| 3. Newton         | 0,21                  | 0,51         | 0,61                | 17 %           | 36 %           | 23 %                 |
| Superposition     | 0,26                  | 0,18         | 0,34                | 28 %           | 34 %           | 8 %                  |
| Arten von Kräften | 0,37                  | 0,51         | <b>0,67</b>         | 22 %           | 36 %           | 18 %                 |
| Kraftverständnis  | 0,23                  | 0,35         | <b>0,68</b>         | 14 %           | 25 %           | 13 %                 |

Tab. 3: Mittelwerte von Vortest, Nachtest und relativen Zugewinn bei den einzelnen Subskalen bei traditioneller Unterricht an bayerischen Gymnasien im Schuljahr 2003/04 (N = 258, 13 Klassen) sowie Cronbachs  $\alpha$  bei Vor- und Nachtest und bei der Treatmentgruppe (N = 138, 7 Klassen)

Wie man in Tab. 3 (rechter Teil) sieht, sind die relativen Zugewinne bei allen Subskalen recht gering. Am Geringsten sind die relativen Zugewinne bei der Subskala „2. newtonsches Axiom“ (7 %), obwohl das zweite newtonsche Gesetz den Kern des newtonschen Kraftkonzeptes bildet. Ebenso sind die relativen Zugewinne bei der Subskala „Superposition“ sehr gering (8 %), was dadurch mitbedingt sein könnte, dass im herkömmlichen Unterricht fast nur eindimensionale Bewegungen betrachtet werden.

### 2.4 Subgruppenanalysen:

| Subgruppen nach Vorwissen |                     |                | Mittelwert      |                  |                    |
|---------------------------|---------------------|----------------|-----------------|------------------|--------------------|
| Vortest                   | erreichte Punktzahl | Anzahl Schüler | Vortest richtig | Nachtest richtig | relativer Zugewinn |
| schlecht                  | ≤ 6                 | 94             | 16 %            | 33 %             | 19 %               |
| mittel                    | 7 bis 9             | 86             | 27 %            | 39 %             | 16 %               |
| gut                       | ≥ 10                | 78             | 42 %            | 52 %             | 19 %               |

Tab. 4: Ergebnisse der im Vortest schwachen, mittleren und guten Schüler bei konventionellem Unterricht in Bayern (N = 258, 13 Klassen). Relativer Zugewinn hier als Mittelwert der individuellen relativen Zugewinne.

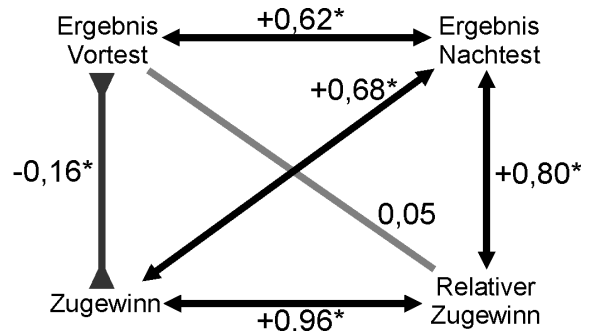


Abb. 4: Korrelationen zwischen verschiedenen Werten (N = 258, 13 Klassen)

Aus der Lernpsychologie weiß man, dass derjenige leichter und schneller bedeutungshaltig lernt, der schon mehr relevantes Vorwissen mitbringt (Renkl, 2002, S. 592) [17] (Wissens-Paradox). Deshalb wurden zum einen die Schüler aufgrund ihrer Vortestergebnisse in drei ungefähr gleich große Gruppen eingeteilt und deren Mittelwerte berechnet und zum anderen Korrelationen in der Gesamtgruppe zwischen verschiedenen Werten berechnet. Wie man Tab. 4 und Abb. 4 entnehmen kann, korrelieren die Vortestergebnisse mittelstark mit den Nachtestergebnissen. Allerdings korrelieren die Vortestergebnisse schwach negativ mit dem Zugewinn. Dies wird verständlich, wenn man bedenkt, dass es bei guten Vortestergebnissen schwieriger als bei schlechten Vortestergebnissen ist (evtl. sogar unmöglich), den gleichen absoluten Zugewinn zu erreichen. Deshalb wird der relative Zugewinn der einzelnen Schüler (absoluter Zugewinn bezogen auf den noch möglichen Zugewinn) betrachtet. Hier ist keine Korrelation zwischen Vortestergebnis und relativen Zugewinn zu erkennen und ein Kruskal-Wallis-Test ergibt entsprechend, dass die Hypothese, dass die Mittelwerte der relativen Zugewinne in den drei Gruppen gleich sind, nicht abgelehnt werden kann. D.h. es besteht kein linearer Zusammenhang zwischen Vortestergebnissen und dem relativen Zugewinn.

| Subgruppe                 | Vor-test         | Nach-test        | relativer Zugewinn g |
|---------------------------|------------------|------------------|----------------------|
| Gesamt weiblich (N = 122) | 24 % *<br>± 9 %  | 37 % *<br>± 15 % | 17 %<br>± 17 %       |
| Gesamt männlich (N = 81)  | 32 % *<br>± 13 % | 45 % *<br>± 17 % | 19 %<br>± 20 %       |

Tab. 5: Vergleich der FCI-Ergebnisse bei den beiden Geschlechtern. Ein \* gibt an, dass sich der Wert signifikant von dem entsprechenden Werte des anderen Geschlechtes unterscheidet, ein ° gibt an, dass sich Vor- und Nachtest signifikant unterscheiden (0,001-Niveau).

Schließlich ist noch interessant, inwieweit es Unterschiede zwischen den Geschlechtern gibt (siehe Tab. 5). 122 Schülerinnen und 81 Schüler haben ihr Ge-

schlecht angegeben. Dabei sind die männlichen Schüler sowohl im Vor- als auch im Nachtest signifikant besser (unabhängiger t-Test, 0,001-Niveau). Beide Gruppen verbessern sich signifikant (abhängiger t-Test, 0,001-Niveau). Der relative Zugewinn ist bei den Jungen aber nur leicht und nicht signifikant höher.

### 3. Ergebnisse einer Treatmentgruppe im Vergleich

In einem von der Heraeus-Stiftung geförderten Forschungsprojekt wurde ein Kinematik-/Dynamikkonzept entwickelt und getestet [15, 18]. Um physikalische Größen und ihre Zusammenhänge leichter erschließbar zu machen, werden sie im Ablauf kontextnah mit piktogrammartigen Darstellungen wie Vektoren und Verbindungslinien dargestellt. Diese Darstellungen werden dynamisch ikonischen Repräsentationen genannt. Dadurch wird ein Einstieg in die Kinematik über allgemeine zweidimensionale Bewegungen möglich, an der die kinematischen Größen eingeführt werden [15, 19]. In der Dynamik kann mit dieser Darstellung und mit einer kontinuierlichen Messung gezeigt werden, dass das Grundgesetz der Mechanik selbst bei veränderlicher Kraft in jedem Augenblick gilt [15, 20]. Auch Versuche mit mehreren Kräften und Reibungskräften sind möglich sowie komplexere Versuche zum dritten newtonschen Gesetz [15, 21]. Ikonische Darstellungen können ebenso in qualitativen, Verständnis verlangenden Übungs- und Testaufgaben verwendet werden. Die Darstellungen ermöglichen auch, vor der Durchführung der Versuche von den Schülern viele genaue Vorhersagen zu fordern (Schüleraktivitäten), insbesondere über relevante Größen und über die sie darstellenden Pfeile mit ihren Richtungen und ihren Änderungen [15, 20]. Eine andere bildhafte Darstellung wird beim Erstellen von Wirkungsnetzen genutzt, um alle Abhängigkeiten der Größen voneinander und damit Strukturzusammenhänge bewusster zu machen [15, 20]. Das Erstellen des Modells mit Hilfe einer Software zur graphischen Modellbildung hilft, eigene Vorstellungen über die Strukturzusammenhänge zu klären und der Ablauf gibt dann zusammen mit den Vorhersagen eine Rückkopplung. Damit werden Fehlvorstellungen aufgearbeitet und die wenigen Grundgleichungen der Mechanik betont. Nach dieser Konzeption wurde ein Gesamt-Unterrichtskonzept entwickelt und insgesamt 13 Lehrer in 17 bayerische Klassen (ganze Klassen ohne Kurssystem) der elften Jahrgangsstufe haben danach unterrichtet, für die eine vorbereitende und begleitende Fortbildung angeboten wurde [15, 18]. Außerdem wurden viele Unterrichtsmaterialien nach Lehrerwünschen gestaltet.

Der FCI-Fragebogen wurde allen 13 Lehrern, die an der Evaluation teilnahmen, für ihre 17 Klassen gegeben, jedoch aus Zeitgründen nicht in allen Klassen als Vor- und Nachtest durchgeführt. Die folgenden Ergebnisse beziehen sich auf die sieben Klassen,

deren Lehrer am Begleitseminar teilnahmen und sich nach eigenen Aussagen an die Grundideen des Konzeptes hielten (davon fünf Klassen aus dem naturwissenschaftlichen Zweig, d.h. dieser Anteil ist nur minimal höher als in der Vergleichsgruppe). Es wurden dabei nur die 138 Schüler berücksichtigt, die an Vor- und Nachtest teilnahmen. Bei den sieben Klassen dieser Analyse (sechs Lehrer) wurde jedoch nicht überprüft, ob sie tatsächlich entscheidende Prinzipien des Unterrichtskonzeptes umsetzten (z.B. Nutzung von dynamisch ikonischen Repräsentationen und das Einfordern von Vorhersagen).

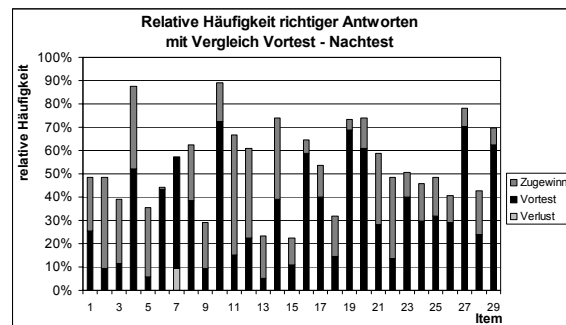


Abb. 5: Relative Häufigkeit richtiger Antworten bei den einzelnen Items mit Vergleich von Vor- und Nachtest (138 Schüler der Treatmentgruppe)

#### 3.1 Beschreibungen der Gesamtergebnisse

Alle 138 Schüler haben jeweils alle Items auch tatsächlich beantwortet. Das Histogramm der Punkteverteilung beim Vor- und Nachtest zeigt Abb. 5. Beim Vortest ergab sich bei den Schülern ein Mittelwert von 9,8 Punkten (= 34 % richtig gelöste Items) (Standardabweichung: 3,8 Punkte = 13 %), wobei die Verteilung von 2 bis 21 Punkten (= 7 % bis 72 %) geht (mehr als fünf Standardabweichungen). Damit sind diese Schüler im Mittel um 1,8 Punkte oder 6 Prozentpunkte besser als die Vergleichsgruppe, was auf dem Niveau von 0,01 signifikant ist (unabhängiger t-Test). Beim Nachtest ergab sich ein Mittelwert von 15,5 Punkten (= 53 %) (mit großer Standardabweichung: 5,3 Punkte = 18 %), wobei die Verteilung von 3 bis 29 Punkten (= 10 % bis 100 %) geht (fünf Standardabweichungen). Damit unterscheidet sich die Treatmentgruppe auf dem strengen Niveau von 0,001 von der Vergleichsgruppe (Mann-Whitney-U-Test, da nicht normalverteilt), die im Nachtest nur auf durchschnittlich 11,8 Punkte (= 41 %) kam. Die Treatmentgruppe erreicht damit einen relativen Zugewinn nach Hake von  $g = 29,7\%$ . Vergleicht man die individuellen relativen Zugewinne der Schüler, stellt man fest, dass der Unterschied der Mittelwerte zwischen der Treatmentgruppe (31,0 %) und der Vergleichsgruppe (17,8 %) ebenso signifikant (0,001-Niveau, t-Test) ist. Das Vortestergebnis der einzelnen Klassen geht in der Treatmentgruppe von 24 % bis 41 % und das Nachtestergebnis von 41 % bis 57 %. Damit variiert der relative Zugewinn der Klassen zwischen 22 % und 34 %.

Hier ist nun die Effektstärke (Verhältnis von Differenz der Testergebnisse der beiden Gruppen zu gewichteten Standardabweichungen) interessant. Die Treatment-Effektstärke bei dem Nachtstestergebnis bei allen Aufgaben beträgt  $d = 0,77$ . Betrachtet man statt den Nachtstestergebnissen den relativen Zugewinn der einzelnen Schüler in beiden Gruppen, ergibt sich eine Treatment-Effektstärke beim relativen Zugewinn bei allen Aufgaben von 0,66. Dies kann nach Häußler et al. (1998, S. 157 f.) [22] jeweils als ein relativ hoher Effekt bezeichnet werden. Nach Häußler et al. (1998, S. 151) [22] ist ein Lernerfolg ab einer Effektstärke von  $d = 0,46$  pädagogisch interessant.

Besser ist, bei dem Test aus inhaltlichen Überlegungen heraus festzulegen, welcher Gesamtwert wünschenswert ist. So sehen Hestenes und Halloun (1995, S. 505) [11] ein Testergebnis von 60 % richtigen Antworten als Schwelle für ein newtonsches Verständnis an. Der Anteil der Schüler, die 17 oder mehr der 29 Items richtig lösten ( $\geq 58,6\%$ ), stieg von 7 % der Schüler im Vortest auf 42 % im Nachtstest an (siehe Tab. 6). Damit hat dieser Anteil von Schülern einen relativen Zugewinn von 38 % gegenüber nur 13 % bei herkömmlich unterrichteten Klassen, er ist also ca. dreimal so hoch.

|  | Traditioneller Unterricht 2003/04, N=259, 13 Klassen |      |          | Treatmentgruppe, 2002/03, N=138, 7 Klassen |      |          |
|--|--|------|----------|--|------|----------|
|  | vor  | nach | rel. Zu. | vor  | nach | rel. Zu. |
| Anteil Schüler mit $\geq 58,6\%$ richtig | 3 %  | 16 % | 13 %     | 7 %  | 42 % | 38 %     |
| Anteil Schüler mit $\geq 62,0\%$ richtig | 2 %  | 14 % | 12 %     | 4 %  | 38 % | 36 %     |

Tab. 6: Anteil Schüler, die 17 bzw. 18 oder mehr Fragen richtig beantwortet haben

Der gleiche FCI-Test (erste Version) wurde von Schecker und Gerdes (1999, S. 80 ff.) [3] bei einer Untersuchung mit Leistungskursen in Bremen eingesetzt. Die beiden konventionell unterrichteten Leistungskurse der Kontrollgruppe ( $N = 26$ ) erreichten dort mit 53,7 % so gut wie den gleichen Nachtstestwert wie die Treatmentgruppe der Würzburger Untersuchung (53,4 %). Die beiden Leistungskurse der Versuchsgruppe ( $N = 23$ ), bei der die Schüler in Gruppenarbeit Modellbildung durchführten, beantworteten allerdings nur 48,7 % der FCI-Items richtig. Da es sich in Bremen um Leistungskurse handelt (5 Unterrichtswochenstunden, 6-8 Wochen Kinematik, 4-5 Wochen Dynamik) (Schecker, 1998, S. 230) [23], ist ein Vergleich wegen dem größeren Interesse der Leistungskurschüler nur bedingt sinnvoll (Treatmentgruppe: 2 bzw. 3 Unterrichtswochenstunden, ca. 15 Std. Kinematik, ca. 20 Std. Dynamik). Dennoch ist es beachtlich, dass die Treatmentgruppe in Würzburg (ganze Klassen) auf gleiche bzw. höhere Werte als die Bremer Leistungskurse kommt.

### 3.2 Analyse auf Itemebene

Im Folgenden sollen noch kurz die Items genannt werden, bei welchen sich die relativen Zugewinne  $I_g$  bei Vergleichs- und Treatmentgruppe um mehr als 20 Prozentpunkte unterscheiden (siehe Tab. 7). Das sind fast alle Items (2, 11, 14) zum dritten newtonschen Gesetz, das anders als herkömmlich unterrichtet wurde; dabei spielte die Messwerterfassung mit dem PC und die dynamische Darstellung mit Graphen sowie (statische) ikonische Repräsentationen mit Vektorpfeilen eine wichtige Rolle [15, 21]. Verstehbar ist auch, dass die Schüler bei den kinematischen Aufgaben 20 und 21 deutlich besser sind, bei denen Geschwindigkeit und Beschleunigung aus Stroboskopbildern entnommen werden sollen: Neben einem besseren Verständnis der kinematischen Begriffe kann hier auch mitwirken, dass bei zweidimensionalen Bewegungen auch mit Zeitmarken (in anderer Darstellung) gearbeitet wurde [15, 19]. Des Weiteren sind die beiden äquivalenten Items 4 und 10 zum ersten newtonschen Gesetz bei vorausgehender Drehbewegung zu nennen, bei denen sich der Körper beim Loslassen mit der augenblicklichen tangentialen Geschwindigkeit (nicht Schnelligkeit) weiterbewegt. Bei Aufgabe 19, bei der die vektorielle Summe der beiden Kräfte gebildet werden muss, macht sich evtl. bemerkbar, dass bei der Kinematik und bei der eindimensionalen Dynamik auch häufig Vektoren mit ihren Richtungen addiert wurden [15, 20]. Die Items 22 und 5 betreffen Wurfbewegungen, wobei auch ein anderer Tests belegt [15], dass die Schüler der Treatmentgruppe bei einer Aufgabe wie Item 5 die Beschleunigung deutlich häufiger richtig angeben.

| Item erste (zweite) Version | Thema des Items             | $I_g$ Vergleichsgruppe | $I_g$ Treatmentgruppe |
|-----------------------------|-----------------------------|------------------------|-----------------------|
| 10 (6)                      | 1. Newton                   | 27%                    | 61%                   |
| 22 (30)                     | Wurf                        | 8%                     | 40%                   |
| 20 (19)                     | Kinematik                   | 1%                     | 33%                   |
| 2 (4)                       | 3. Newton                   | 13%                    | 43%                   |
| 4 (7)                       | 1. Newton                   | 44%                    | 74%                   |
| 5 (17)                      | Wurf                        | 3%                     | 32%                   |
| 11 (28)                     | 3. Newton                   | 34%                    | 61%                   |
| 19 (-)                      | Richtung der Resultierenden | -13%                   | 14%                   |
| 14 (16)                     | 3. Newton                   | 36%                    | 57%                   |
| 21 (20)                     | Kinematik                   | 23%                    | 42%                   |

Tab. 7: Die Aufgaben mit den größten Unterschieden im relativen Zugewinn  $I_g$  (Differenz über 20 Prozentpunkte) zwischen Vergleichs- und Treatmentgruppe

### 3.3. Analyse von Subskalen

Schließlich sollen noch die Unterschiede in den Subskalen betrachtet werden (Tab. 8). Dabei ist zu beachten, dass keine hohen Reliabilitäten vorliegen

| Subskala          | Traditioneller Unterricht<br>2003/04<br>(N = 258, 13 Klassen) |                       |                       | Unterricht nach Konzept,<br>Treatmentgruppe, 2002/03<br>(N = 138, 7 Klassen) |                        |                        | Effektstärke<br>beim Nachtest |
|-------------------|---|-----------------------|-----------------------|--|------------------------|------------------------|-------------------------------|
|                   | Vortest   | Nachtest              | rel. Zugewinn g       | Vortest  | Nachtest               | rel. Zugewinn g        |                               |
| Gesamt            | <b>28 %</b><br>± 12 %   | <b>41 %</b><br>± 16 % | <b>18 %</b><br>± 18 % | <b>34 %*</b><br>± 13 %   | <b>53 %*</b><br>± 18 % | <b>31 %*</b><br>± 22 % | <b>0,77</b>                   |
| Kinematik         | 28 %  | 39 %                  | 15 %                  | 40 %*  | 53 %*                  | 21 %                   | 0,50                          |
| 1. Newton         | 33 %  | 47 %                  | 20 %                  | 43 %*  | 60 %*                  | 29 %                   | 0,57                          |
| 2. Newton         | 29 %  | 34 %                  | 7 %                   | 38 %*  | 44 %*                  | 10 %                   | 0,37                          |
| 3. Newton         | 17 %  | 36 %                  | 23 %                  | 17 %   | 53 %*                  | 43 %*                  | 0,58                          |
| Superposition     | 28 %  | 34 %                  | 8 %                   | 29 %   | 44 %*                  | 21 %*                  | 0,44                          |
| Arten von Kräften | 22 %  | 36 %                  | 18 %                  | 26 %*  | 46 %*                  | 27 %*                  | 0,56                          |
| Kraftverständnis  | 14 %  | 25 %                  | 13 %                  | 17 %   | 42 %*                  | 30 %*                  | 0,77                          |

Tab. 8: Mittelwerte von Vortest, Nachtest und relativen Zugewinn bei den einzelnen Subskalen, Vergleich zwischen traditionellem Unterricht an bayerischen Gymnasien im Schuljahr 2003/04 (N = 258, 13 Klassen) und der Treatmentgruppe (N = 138, 7 Klassen) (Eigene Erhebungen). Ein \* gibt an, dass sich der Wert signifikant von dem entsprechenden Werte der Vergleichsgruppe unterscheidet (0,01-Niveau bei Vortest, 0,001-Niveau bei Nachtest und relativen Zugewinn) (in der Regel unabhängiger t-Test. Wenn nicht normalverteilt, auch Mann-Whitney-U-Test).

und die Subskalen aus Schülersicht nicht die Konzepte abtesten, die sie vorgeben. Dennoch zeigen sie, wie Schüler auf eine Gruppe spezieller FCI-Aufgaben antworten. Dabei zeigt sich, dass die Treatmentgruppe im Vortest bereits bei allen Subskalen (außer 3. newtonsches Axiom) etwas besser als die Vergleichsgruppe ist, wobei der Unterschied bei vier der sieben Subskalen signifikant (0,01-Niveau) ist.

Beim Nachtest ist der Unterschied zur Vergleichsgruppe bei allen Subskalen noch größer, was auf dem strengen Niveau von 0,001 signifikant ist. Auch der relative Zugewinn (als Mittelwert der relativen Zugewinne der einzelnen Schüler) ist bei der Treatmentgruppe bei allen Subskalen größer und in vier der sieben Subskalen auf dem Niveau von 0,001 signifikant („1. newtonsches Axiom“ auf 0,01-Niveau signifikant); nur bei den Subskalen „Kinematik“ und „2. newtonsches Axiom“ ist der Unterschied beim relativen Zugewinn nicht signifikant.

Interessant ist auch der Unterschied beim Nachtest im Verhältnis zur Streuung. Die größte Effektstärke gibt es bei der von Schecker und Gerdes (1999, S. 84) [3] konzipierten Subskala „Kraftverständnis“, die mit  $d = 0,77$  hoch ist. Auch bei den Subskalen „3. newtonsches Gesetz“ ( $d = 0,58$ ) und „1. newtonsches Gesetz“ ( $d = 0,57$ ) ergibt sich noch eine relativ hohe Effektstärke. Am niedrigsten und schlechter als erwartet ist das Ergebnis bei der Subskala „2. newtonsches Gesetz“, bei dem sowohl der relative Zugewinn als auch die Effektstärke gering sind. Den Sinn dieser Subskala, die aus nur zwei Aufgabenpaaren besteht, kann man aber mit Schecker und Gerdes (1999, S. 83) [3] bezweifeln und stattdessen die Subskala „Kraftverständnis“ betrachten.

### 3.4 Andere Tests der Treatmentgruppe

Dass die Schüler der Treatmentgruppe mehr Verständnis des newtonschen Kraftbegriffes erreicht haben als konventionell unterrichtete Klassen, zeigt sich auch in Concept Maps, die in einigen Klassen angefertigt wurden [15].

Bei einem Test, der nur das Verständnis des zweiten newtonschen Gesetzes bei eindimensionalen Bewegungen abfragt [16], waren die Schüler der Treatmentgruppe allerdings nur wenig besser als die Vergleichsgruppe. Hier liegen nur schwache Effektstärken (0,2 – 0,3) vor [15].

Auch qualitative Tests zur Kinematik zeigen, dass deutlich mehr Schüler der Treatmentgruppe ein angemessenes Verständnis der Beschleunigung als gerichtete Größe haben. Bei Aufgaben zur Grapheninterpretation bei eindimensionalen Bewegungen gibt es dennoch sowohl bei Zeit-Geschwindigkeits- als auch bei Zeit-Beschleunigungs-Graphen keine signifikanten Unterschiede zur Vergleichsgruppe [15].

### 4. Diskussion der Ergebnisse

Der bekannte FCI-Test zum physikalischen Kraftkonzept wurde im Gegensatz zu den USA in Deutschland bisher nur in einzelnen Gymnasialklassen und einer großen Anzahl von Studenten gestellt, während er hier in der ersten Version erstmals einer größeren Anzahl von Klassen vorgelegt wurde. Diese Erhebung kann deshalb als Vergleichsbasis für Forschungsprojekte benutzt werden.

In Folge der Erkenntnisse der Schülervorstellungsforschung wurden Lehrpläne so abgeändert, dass in der Sekundarstufe I die Kraft nicht mehr statisch, sondern dynamisch eingeführt wurde und der Um-



fang der Statik zurückgedrängt wurde. Die Untersuchung zeigt, dass trotzdem nur wenige Schüler am Ende der Sekundarstufe I über newtonsche Vorstellungen verfügen, die sie auch bei qualitativen Verständnisaufgaben anwenden. Aber auch der Mechanikunterricht der elften Jahrgangsstufe, der sich sehr intensiv mit dieser Thematik beschäftigt, erreicht nur unzufriedenstellende Zugewinne. Die Schüler antworten auch nach dem Unterricht entsprechend den bekannten und bereits als stabil erkannten Schülervorstellungen. In Bayern möchte man im Rahmen des achtjährigen Gymnasiums die newtonsche Mechanik noch früher und intensiver behandeln. Studien werden zeigen müssen, ob damit die damit verbundenen Erwartungen erfüllt werden.

Die erwähnte Studie zeigt, dass es prinzipiell möglich ist, auch im Rahmen der engen Zeitvorgaben eines Lehrplans mehr Schüler zu mehr Verständnis des newtonschen Konzeptes zu führen. Da neue Elementarisierungen, neue Darstellungen und veränderte Unterrichtsstrategien genauso eingesetzt wurden wie eine intensive Lehrerschulung kann aber nicht gesagt werden, welche Anteile die einzelnen Maßnahmen an dem erreichten Lernerfolg haben. Denkbar ist, dass sich die Wirkungen dieser Maßnahmen nicht summativ addieren, sondern erst gemeinsam durch gegenseitige Wechselwirkung diesen Erfolg bewirken. Drei weitere Lehrer, die alle Unterrichtsmaterialien der Treatmentgruppe zur Verfügung hatten, aber an den vorbereitenden und begleitenden Lehrerfortbildung nicht teilnehmen konnten, erreichten in ihren drei Klassen nur den relativen Zugewinn der traditionell unterrichteten Klassen. Um daraus zu schließen, dass zu einem erfolgreichen Umgang mit den erstellten Materialien zusätzliche Kenntnisse und eine Veränderung von Einstellungen notwendig sind, die nur in einem Lehrerseminar erreichbar sind, müsste diese Gruppe größer sein.

Immer wieder sind junge Lehrer anzutreffen, die in ihrer Ausbildung kaum etwas über die bekannten Schülervorstellungen erfahren haben. Außerdem kann immer wieder festgestellt werden, dass auch Lehrern z.T. die Bedeutung physikalischer Begriffe und deren Zusammenhänge bei kinematischen und dynamischen Fragestellungen nicht ausreichend klar ist, was nicht verwundert, da die newtonsche Mechanik nur im ersten Semester kurz behandelt wird. Deshalb weisen die hier dargestellten Ergebnisse auch auf einen Bedarf in Lehreraus- und -fortbildung hin.

### 5. Zusammenfassung

Die bayerischen Gymnasiasten haben am Beginn der elften Jahrgangsstufe 28 % der qualitativen Items richtig gelöst und nach dem Mechanikunterricht 41 %, was ein relativer Zugewinn von nur 18 % ist. Besonders geringe relative Zugewinne liegen bei den Subskalen „zweites newtonsches Axiom“ und „Superposition“ vor. Hier macht sich vermutlich

bemerkbar, dass im herkömmlichen Unterricht fast nur eindimensionale Bewegungen betrachtet werden. Korrelationen, Faktorenanalyse und Reliabilitäten bestätigen allerdings die in der Literatur geäußerte Kritik an der Einteilung in die einzelnen Subskalen. In der erwähnten Studie mit verändertem Kinematik-/Dynamikunterricht („Würzburger Kinematik-/Dynamikunterricht“) und mit intensiver Lehrerschulung wurden im Nachtest 53 % der Aufgaben richtig gelöst im Gegensatz zu nur 41 % in der traditionell unterrichteten Vergleichsgruppe, was ein signifikanter Unterschied ist (0,001-Niveau) und eine relativ hohe Effektstärke von  $d = 0,77$  ergibt. Auch der relative Zugewinn der Schüler ist mit 31 % signifikant höher als in der Vergleichsgruppe mit 18 % (Effektstärke 0,66). Der Anteil der Schüler, die mehr als 58 % richtig beantworteten, was als Schwelle für ein newtonsches Verständnis angesehen werden kann, liegt mit 42 % deutlich höher als in der Vergleichsgruppe mit 16 %. Der größte relative Zugewinn wurde dabei in der Subskala „Drittes newtonsches Axiom“ erreicht (43%). Die größte Effektstärke im Vergleich zur Vergleichsgruppe wurde dagegen in der Subskala „Kraftverständnis“ erzielt ( $d = 0,77$ ).

### 6. Literatur

- [1] Hestenes, D.; Wells, M.; Swackhamer, G. (1992): Force Concept Inventory – In: *The physics teacher* 30, S. 141 – 158
- [2] Girwidz, R.; Kurz, G.; Kautz, C. (2003): Zum Verständnis der newtonschen Mechanik bei Studienanfängern – der Test ‚Force Concept Inventory – FCI‘ – In: Nordmeier, V. (Red.): *Didaktik der Physik. Beiträge der Frühjahrstagung der DPG – Augsburg 2003*, Berlin
- [3] Schecker, H.; Gerdes, J. (1999): Messung der Konzeptualisierungsfähigkeit in der Mechanik – Zur Aussagekraft des Force Concept Inventory – In: *Zeitschrift für Didaktik der Naturwissenschaften* 5, Heft 1, S. 75 – 89
- [4] Hake, R. (1998): Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses - In: *American Journal of Physics* 66, Nr. 1, S. 64 – 74
- [5] Savinainen, A.; Scott, P. (2002): The Force Concept Inventory: a tool for monitoring student learning – In: *Physics Education* 37 (1), S. 45 – 52
- [6] Gerdes, J.; Schecker, H. (1999): Der Force Concept Inventory - Ein diagnostischer Test zu Schülervorstellungen in der Mechanik - In: *Der Mathematische und Naturwissenschaftliche Unterricht* 52, Heft 5, S. 283 - 288
- [7] Gerdes, J.; Schecker, H. (1998): Physiklernen mit Modellbildungssystemen – Erste Ergebnisse - In: Behrendt, H. (Hrsg.): *Zur Didaktik der Physik und Chemie, Probleme und Perspektiven, Vorträge auf der Tagung für Didaktik der*

- Physik/Chemie in Potsdam, September 1997, Gesellschaft für Didaktik der Chemie und Physik (GDGP), Leuchtturm-Verlag, Alsbach, S. 227 – 229
- [8] Schecker, H.; Klieme, E.; Niedderer, H.; Ebach, J.; Gerdes, J. (1999): Physiklernen mit Modellbildungssystemen. Förderung physikalischer Kompetenz und systemischen Denkens durch computergestützte Modellbildungssysteme, Abschlussbericht zum DFG-Projekt, Institut für Didaktik der Physik an der Universität Bremen und Max-Planck-Institut für Bildungsforschung Berlin
- [9] Halloun, I.; Hake, R.; Mosca, E.; Hestenes D. (1997): Force Concept Inventory (revised 1995) - In: Mazur, E. (Hrsg.): Peer Instruction: A User's Manual, Englewood Cliffs, NJ, Prentice Hall, S. 45 - 58
- [10] Halloun, I.; Hestenes, D. (1985): The initial knowledge state of college physics students - In: American Journal of Physics 53, Nr. 11, S. 1043 – 1055
- [11] Huffman, D.; Heller, P. (1995): What Does the Force Concept Inventory Actually Measure? - In: The Physics Teacher 33, S. 138 - 143
- [12] Hestenes, D.; Halloun, I. (1995): Interpreting the Force Concept Inventory. A Response to March 1995 Critique by Huffman and Heller – In: The Physics Teacher 33, S. 502 + 504 -506
- [13] Heller, P.; Huffman, D. (1995): Interpreting the Force Concept Inventory. A Reply to Hestenes and Halloun – In: The Physics Teacher 33, S. 503 + 507 – 511
- [14] McCullough, L.; Meltzer, D. (2001): Differences in male/female response patterns on alternative-format versions of FCI items - In Franklin, S.; Marx, J.; Cummings, K. (Hrsg.): Proceedings of the 2001 Physics Education Research Conference, Rochester, NY, S. 103 – 106
- [15] Wilhelm, T. (2005): Konzeption und Evaluation eines Kinematik/Dynamik-Lehrgangs zur Veränderung von Schülervorstellungen mit Hilfe dynamisch ikonischer Repräsentationen und graphischer Modellbildung, Dissertation, Veröffentlichung geplant
- [16] Heuer, D.; Wilhelm, T. (1997): Aristoteles siegt immer noch über Newton. Unzulängliches Dynamikverstehen in Klasse 11 - In: Der mathematische und naturwissenschaftliche Unterricht 50, 1997, Nr. 5, S. 280 – 285
- [17] Renkl, A. (2002): Lehren und Lernen – In: Tippelt, R. (Hrsg.): Handbuch der Bildungsforschung, Leske + Budrich, Opladen, S. 589 – 602
- [18] Wilhelm, T.; Heuer, D. (2004): Implementation eines innovativen Kinematik-/Dynamik-Unterrichtskonzeptes - In: Pitton, A. (Hrsg.): Chemie- und physikdidaktische Forschung und naturwissenschaftliche Bildung, Jahrestagung der GDGP in Berlin 2003, Reihe: Gesellschaft für Didaktik der Chemie und Physik, Band 24, Lit-Verlag, Münster, S. 203 - 205
- [19] Wilhelm, T.; Heuer, D. (2002): Fehlvorstellungen in der Kinematik vermeiden - durch Beginn mit der zweidimensionalen Bewegung - In: Praxis der Naturwissenschaften – Physik in der Schule 51, Nr. 7, S. 29 - 34
- [20] Wilhelm, T.; Heuer, D. (2002): Interesse fördern, Fehlvorstellungen abbauen - dynamisch ikonische Repräsentationen in der Dynamik - In: Praxis der Naturwissenschaften – Physik in der Schule 51, Nr. 8, S. 2 - 11
- [21] Wilhelm, T.; Heuer, D. (2004): Experimente zum dritten Newtonschen Gesetz zur Veränderung von Schülervorstellungen - In: Praxis der Naturwissenschaften – Physik in der Schule 53, Nr. 3, S. 17 - 22
- [22] Häußler, P.; Bündler, W.; Duit, R.; Gräber, W.; Mayer, J. (1998): Naturwissenschaftsdidaktische Forschung. Perspektiven für die Unterrichtspraxis, Institut für die Pädagogik der Naturwissenschaften, Kiel
- [23] Schecker, H. (1998): Physiklernen mit Modellbildungssystemen - Forschungskonzeptionen - In: Behrendt, H. (Hrsg.): Zur Didaktik der Physik und Chemie, Probleme und Perspektiven, Vorträge auf der Tagung für Didaktik der Physik/Chemie in Potsdam, September 1997, Gesellschaft für Didaktik der Chemie und Physik (GDGP), Leuchtturm-Verlag, Alsbach, S. 230 – 232

## 7. Adresse

StR Thomas Wilhelm, Lehrstuhl für Didaktik der Physik, Physikalisches Institut der Universität Würzburg, Am Hubland, 97074 Würzburg, Tel. 0931/888-5788  
[wilhelm@physik.uni-wuerzburg.de](mailto:wilhelm@physik.uni-wuerzburg.de)