

## Computereinsatz in der Newton'schen Mechanik – Vergleich von Modellbildung und Videoanalyse –

Jannis Weber, Thomas Wilhelm

Max-von-Laue-Str. 1, 60438 Frankfurt am Main

[weber@physik.uni-frankfurt.de](mailto:weber@physik.uni-frankfurt.de)

(Eingegangen: 08.06.2020; Angenommen: 25.11.2020)

### Kurzfassung

Die Grundaussagen der Newton'schen Mechanik sind aufgrund vielfältiger und dazu scheinbar widersprüchlicher Alltagswahrnehmungen für Schülerinnen und Schüler nur schwer zu erlernen. Dies zeigt sich in einem nicht angemessenen Kraftverständnis und hartnäckigen Schülervorstellungen im Bereich der Mechanik. Hier wird eine Vergleichsstudie vorgestellt, bei der die durchgeführten Interventionen die Verbesserung des Konzeptverständnisses der Newton'schen Mechanik zum Ziel haben. Es wird mit realen Experimenten gearbeitet und Reibungseinflüsse werden bewusst thematisiert und diskutiert. Vertieft werden dabei die grundlegenden Gesetze der Mechanik. Es wird einerseits der Ansatz gewählt, dass Lernende reale Bewegungen am Computer selbst modellieren und mit der Realität vergleichen und diese andererseits mit Videoanalyseprogrammen analysieren. Die Ergebnisse deuten an, dass die Interventionen wirksam für das Verbessern des Konzeptverständnisses der Newton'schen Dynamik sind. Außerdem wird auf die spezifischen Unterschiede zwischen beiden Interventionen eingegangen.

### Abstract

Students experience many difficulties learning the fundamental relationships in Newtonian Mechanics, partly due to many pre-existing students' conceptions that originate from everyday life. These misconceptions often persist even after instruction in mechanics and lead to a supposed incompatibility between physics lessons and personal experiences. This article presents a comparative study that tries to enhance students' concept knowledge in the field of dynamics after their regular lessons in school by discussing real experiments with friction and by an intensive use of computers. One intervention uses modelling software whereas the students in the other intervention work with video analysis software. The results suggest that the interventions are both effective in teaching Newtonian mechanics. Additionally, differences between the interventions are discussed in the article.

### 1. Einleitung

Schülervorstellungen halten sich in der Mechanik besonders hartnäckig und sind sehr gut erforscht [1]. Ein Erklärungsansatz dafür ist die Tatsache, dass sich die Mechanik mit der direkten Lebenswelt der Schülerinnen und Schüler beschäftigt und sich die Schülervorstellungen somit im Laufe des Lebens der Lernenden festigen konnten.

Die Schulmechanik steht also vor der Aufgabe, den Schülerinnen und Schülern physikalisch korrekte und tragfähige Konzepte zu liefern, sodass sie neben ihren bestehenden Präkonzepten die physikalisch erwünschten Konzepte aufnehmen und diese in Aufgaben nutzen. Dies erweist sich erfahrungsgemäß als schwierig.

Schülerinnen und Schüler gehen zu einem Großteil davon aus, dass eine Kraft in Bewegungsrichtung wirken müsse, damit sich etwas bewegen kann. Dies ist aus Alltagssicht naheliegend, da man z. B. auf dem Fahrrad dauerhaft in die Pedale treten muss, um mit

einem bestimmten Tempo zu fahren, und auch im Auto wird das Gaspedal konstant betätigt, wenn es sich mit einer gleichbleibenden Geschwindigkeit bewegt. Ohne eine Thematisierung der Reibung ist für Schülerinnen und Schüler die Notwendigkeit nicht ersichtlich, von ihren Konzepten auf die der Newton'schen Mechanik überzugehen.

Eine Thematisierung von Reibung stellt den Physikunterricht aber vor mathematische Probleme. Es ist nicht möglich, reale Bewegungen, bei denen mehrere Kräfte wirken, die möglicherweise geschwindigkeits- oder ortsabhängig sind, ohne Hilfsmittel quantitativ zu betrachten. Der Computer kann allerdings ein solches Hilfsmittel sein.

Einerseits können mit dem Computer Bewegungen modelliert werden. Dabei muss sich der Nutzer bzw. die Nutzerin Gedanken über die auf einen Körper wirkenden Kräfte machen und so das zu einer Bewegung gehörende mathematische Modell erstellen. Der Nutzer bzw. die Nutzerin sollte selbst entscheiden, ob

Reibung in einem konkreten Fall einbezogen werden muss oder nicht. Es wird also ein theoretisches Modell erstellt, was dann mit der Realität verglichen werden kann.

Andererseits kann der Computer zur Messung und Auswertung von Experimenten genutzt werden. Dabei werden Messdaten dargestellt und interpretiert, wobei der Einfluss von Reibung qualitativ besprochen werden kann. Ein Beispiel hierfür ist die Videoanalyse, die mit einfachen Mitteln eine Analyse von alltäglichen Bewegungen ermöglicht. Hier wird also von Messdaten eines Experiments ausgegangen und diese mit der dahinterliegenden Theorie verknüpft und erklärt.

## 2. Theoretischer Hintergrund

### 2.1. Mathematische Modellbildung

Mathematische Modellbildung bezeichnet die „Konstruktion eines Netzwerks physikalischer Begriffe und Beziehungen, mit denen das Verhalten eines physikalischen Systems beschrieben und vorhergesagt werden kann“ [2]. Mathematische Modellbildung für den Physikunterricht ist keine neue Idee und wurde schon in den 1980er-Jahren eingesetzt. Die verwendete Software hat sich seitdem grundlegend verändert [3]. In den frühen Jahren wurde die Software nahezu ausschließlich von der Lehrkraft bedient. Dabei wurde meist auf die Nutzung von Programmiersprachen wie „Basic“ oder „Pascal“ zurückgegriffen. Möglich ist die mathematische Modellbildung auch mit Tabellenkalkulationsprogrammen wie Microsoft Excel, wobei in beiden Fällen die sogenannte „Methode der kleinen Schritte“ genutzt wird. Graphische Modellbildungsprogramme bieten hingegen den Vorteil, dass die zugrundeliegende Struktur des mathematischen Modells visuell dargestellt wird. Außerdem gibt es gleichungsbasierte Modellbildungssoftware, in der lediglich die wirkenden Kräfte, die Anfangsbedingungen und die Masse eines Objekts eingegeben werden müssen, was die Bedienung vereinfacht. Jede dieser Arten der Modellbildung bringt Vor- und Nachteile mit sich und sollte je nach Zielsetzung des Unterrichts eingesetzt werden.

Untersuchungen zum Einsatz der mathematischen Modellbildung zeigen, dass insbesondere in der Sekundarstufe die zum Teil weitreichenden Erwartungen bisher nicht erfüllt wurden [4; 5]. Dennoch konnte gezeigt werden, dass der Einsatz von graphischen Modellbildungsprogrammen im Vergleich zum normalen Unterricht eine Verbesserung der halbquantitativen Beschreibung von mechanischen Bewegungen bewirkt [6]. Benacka konnte einen motivierenden Effekt von Modellbildung mit Excel bei Oberstufenschülern zeigen [7-9]. In Universitäten hat sich „VPython“ [10] etabliert, eine Software, die auf der Programmiersprache Python beruht und mit der für die Modellbildung einfach 3D-Animationen erstellt werden können [11]. In Schulen scheint der Einsatz von VPython mit mehr Problemen daherzukommen [12].

Gerade wenn Lernende die Software selbstständig nutzen sollen, ist eine intuitive und einfache Bedienung wichtig. Daher wird hier der Einsatz des gleichungsbasierten Modellbildungsprogramms „Newton-II“ [13] erforscht, zu dem es bisher keine empirischen Untersuchungen gibt. Bei „Newton II“ steht der Zusammenhang zwischen Kräften und Bewegung im Mittelpunkt, während die kinematischen Zusammenhänge vom Programm automatisch berechnet werden.

### 2.2. Videoanalyse

Videoanalyse ist eine kontaktlose Messmethode, die es erlaubt, zweidimensionale Bewegungen auf eine einfache Art zu analysieren und die gewonnenen Daten auf vielfältige Weise darzustellen. Im Vergleich zu Messwerterfassungssystemen sind gerade diese einfach zu nutzenden Darstellungsmöglichkeiten ein großer Vorteil. Die Videoanalyse wird bereits seit vielen Jahren im Physikunterricht eingesetzt [14] und sie hat sich in den letzten Jahren, auch durch verbesserte Software, zunehmend etabliert [15; 16]. Videoanalyse kann an Laptops, Tablets und Smartphones durchgeführt werden, die unterschiedliche Vor- und Nachteile mit sich bringen.

Insgesamt haben Studien zur Videoanalyse gezeigt, dass Probandinnen und Probanden bei der Nutzung einen Vorteil beim Interpretieren von Bewegungsgraphen [17] und dem Konzeptverständnis zum Bremsvorgang haben [18]. Außerdem wurden positive Effekte auf Motivation und Neugierde gemessen [19]. Es wurde gezeigt, dass das Konzeptverständnis insbesondere bei anspruchsvollen Aufgaben in der Kinematik durch den Einsatz von Videoanalyse stärker als durch normalen Unterricht verbessert wird [20]. Die Nutzung von multiplen Repräsentationen [21], wie bspw. Graphen oder Vektorpfeilen in Kombination mit Videos, kann sich positiv auf den Lernzuwachs auswirken [22]. Insgesamt sind die Forschungsergebnisse zur Videoanalyse aber weitestgehend auf den Bereich der Kinematik beschränkt. In dieser Studie wird der Einsatz der Software „measure dynamics“ [23] für Laptops im Bereich der Dynamik erforscht.

### 2.3. Cognitive Load Theory

Die Cognitive Load Theory (CLT) [24; 25] postuliert eine limitierte Kapazität des für Lernprozesse wichtigen Arbeitsgedächtnisses. Zudem ist die Zeit begrenzt, in der die zu verarbeitenden Informationen zur Verfügung stehen. Das Arbeitsgedächtnis ist für die Konstruktion von neuen Schemata wichtig und sollte daher nicht überlastet sein. Dies sollte bei der Erstellung von Lernumgebungen beachtet werden.

Es werden drei Arten der kognitiven Belastung unterschieden [26]. Die intrinsische kognitive Belastung (intrinsic cognitive load/ICL) hängt von der Komplexität der Informationen ab, die verarbeitet werden müssen, und ist damit von der Lernaufgabe und dem Vorwissen des Lernenden abhängig. Sie kann durch die Lernumgebung also nicht beeinflusst werden. Die extrinsische kognitive Belastung (extraneous cognitive load/ECL) wird maßgeblich durch die Gestaltung

der Lernumgebung beeinflusst und sollte nach der CLT möglichst reduziert werden, da sie für den Lernprozess nicht förderlich ist. Die lernbezogene kognitive Belastung (germane cognitive load/GCL) ist die für den Lernprozess relevante Lernbelastung, die für die Konstruktion und Automatisierung neuer Schemata verantwortlich ist und möglichst groß sein sollte.

Die CLT ist eine Theorie, die eine Begründung dafür liefern kann, welche Lernumgebungen lernförderlich sind und welche zu einer unnötigen kognitiven Belastung führen, die eine Konstruktion von neuen Schemata behindert. Im Kontext der Videoanalyse wurde gezeigt, dass eine Reduktion des ECL mit einem höheren Lernzuwachs einhergeht [27; 28].

### 3. Zielsetzung der Studie und Forschungsfragen

Die Studie betrachtet Schülerinnen und Schüler, die im normalen Schulunterricht die Newton'schen Gesetze bereits behandelt haben. Sie soll untersuchen, inwiefern eine anschließende Intervention zur Newton'schen Dynamik mit mathematischer Modellbildung bzw. Videoanalyse wirksam für das weitere Erlernen bzw. Vertiefen der Dynamik ist:

- Führt eine computergestützte Intervention mit mathematischer Modellbildung oder Videoanalyse nach dem Erlernen der Newton'schen Gesetze zu einem messbaren Zuwachs im Konzeptverständnis?

Neben der Wirksamkeit der Interventionen ist außerdem eine differenziertere Betrachtung von Interesse, bei der es darum geht, ob sich die Interventionen im Lernzuwachs unterscheiden und ob es Unterschiede in verschiedenen inhaltlichen Bereichen oder vorhandenen Schülervorstellungen gibt:

- Welche Unterschiede lassen sich zwischen einem modellierenden Vorgehen mit mathematischer Modellbildung und einem messenden Vorgehen mit Videoanalyse beim Erlernen der Newton'schen Mechanik erkennen?

Die Ergebnisse sollen dabei in die Cognitive Load Theory eingebettet werden.

Da zum Vertiefen der Grundaussagen der Dynamik ein von der Theorie ausgehender (Modellierung) und ein vom Experiment und dessen Auswertung ausgehender (Videoanalyse) Zugang gewählt wird, besteht die Hypothese, dass Lernende, die eher an theoretischen Zusammenhängen interessiert sind, einen größeren Lernzuwachs bei der Modellbildung erzielen, während Lernende, die am Experimentieren interessiert sind, bei der Videoanalyse mehr lernen. Daraus könnten sich folglich Empfehlungen für den Einsatz der beiden Methoden im Schulunterricht ableiten, weshalb gefragt wird:

- Ist es für den Lernzuwachs in einer jeweiligen Intervention entscheidend, wie groß das Interesse an Theorie bzw. Experiment bei den Probandinnen und Probanden ist?

Da Modellieren im Mechanikunterricht eher selten Anwendung findet [16] und das Modellverständnis bei vielen Schülerinnen und Schülern nicht zufriedenstellend ist [29], wird untersucht, ob eine Intervention zum mathematischen Modellieren eine Veränderung der Einstellung zu Modellen in der Physik mit sich bringt, ohne dass dies in der Intervention explizit thematisiert wird:

- Kann durch das aktive Modellieren eine Veränderung der Einstellungen gegenüber Modellen erreicht werden, ohne dass die Eigenschaften von Modellen aktiv thematisiert werden?

### 4. Studiendesign

Um die Forschungsfragen zu beantworten, wurde eine quasi-experimentelle Studie im Prä-Post-Design durchgeführt [30]. Dazu wurde eine Intervention zur Dynamik mit mathematischer Modellbildung (mit „Newton-II“) und eine mit Videoanalyse (mit „measure dynamics“) erstellt, welche sich nur im Einsatz der Software unterscheiden. Dabei werden jeweils vier Experimente durchgeführt (Fallbewegung mit Reibung, abschnittsweise beschleunigte Bewegung eines Wagens, schiefer Wurf mit und ohne Reibung, Kreisbewegung), die daraufhin in Partnerarbeit mit dem Computer modelliert oder analysiert werden. Die Probandinnen und Probanden arbeiten nach einer Einführung in das Programm selbstständig. Die Klassen (E-Phasen/11. Klassen) besuchen dazu das Schülerlabor der Goethe-Universität Frankfurt und werden klassenweise zu einer der beiden Interventionen zugeordnet. Der Vortest wird etwa eine Woche vor der Intervention in der Schule durchgeführt, während der Nachtest direkt im Anschluss an die Intervention folgt. Die Interventionen nehmen ca. 3,5 Stunden in Anspruch, wonach sich direkt der Nachtest anschließt (Abb. 1).

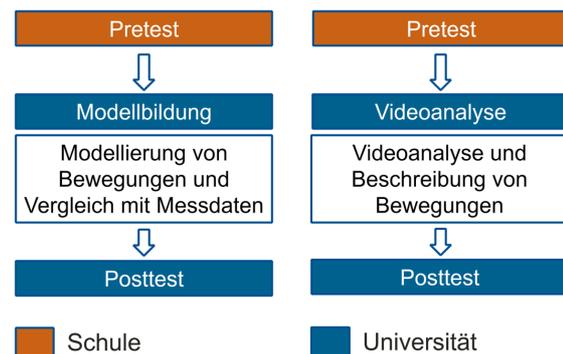


Abb. 1: Ablauf der Vergleichsstudie. Die Zuordnung erfolgt klassenweise.

### 5. Testinstrument

Für die Studie wurde ein Testinstrument mit verschiedenen Abschnitten erstellt und pilotiert.

#### 5.1. Konzeptverständnis

Der Hauptteil des Testinstruments besteht in der Testung des Konzeptverständnisses. Von den insgesamt 23 Items (davon vier Kontrollitems) des Tests sind 13

aus dem FCI-Test [31]. Davon wurden sieben original übernommen, vier geändert und zwei durch eine zweite Stufe (Begründung der Antwort) ergänzt. Vier Items sind aus dem FMCE-Test von Thornton & Sokoloff [32] in der Version von Wilhelm [21] (siehe auch [33]), drei Items sind eigens erstellt, zwei von Wilhelm [34] nach Flores et al. [35], und eins ist von Warren [36] in veränderter Form übernommen. Alle Items werden dichotom ausgewertet und ein zweistufiges Item gilt als richtig, wenn auch die Begründung richtig gewählt wurde.

Der FCI-Test ist der bekannteste Test für das Kraftverständnis. Er testet reliabel das Kraftverständnis [37] und bietet eine hohe Vergleichbarkeit zu anderen Studien [38]. Allerdings lassen sich die Items im FCI-Test nicht reliabel verschiedenen Inhaltsbereichen zuordnen, die eine differenziertere Betrachtung erlauben würden [39]. Da der FCI-Test aber geeignete Items enthält, diente er auch hier als Grundlage für die Erstellung des Tests.

Nach einer Pilotierung mit  $N = 85$  Schülerinnen und Schülern wurden durch einen Scree-Test vier inhaltliche Dimensionen mit guten Cronbachs Alphas identifiziert (siehe Tab. 1).

Inhaltsbereiche	Items	$\alpha_c$
Beschleunigungsdiagramme	4	0,91
1. Newton'sches Gesetz mit Kräftekompensation	4	0,75
Kraft bei bekannter Bewegung	5	0,74
Bewegung bei bekannter Kraft	6	0,70

**Tab. 1:** Dimensionen zum Konzeptverständnis

Aus dem FCI-Test wird zudem der Inhaltsbereich „drittes Newton'sches Gesetz“ mit vier Items als Kontrollskala hinzugenommen, die bereits im FCI-Test eine hohe interne Konsistenz aufweist [40; 41]. Das dritte Newton'sche Gesetz ist nicht Teil der Intervention, weswegen hier keine Änderung zu erwarten ist. Der Test weist also bezüglich der Art der Items eine große Ähnlichkeit zum FCI-Test auf.

## 5.2. Modellverständnis

Um die Einstellungen der Probandinnen und Probanden gegenüber Modellen in der Physik messen zu können, wurden eigene Items erstellt und pilotiert. Die Items decken dabei verschiedene Eigenschaften ab, die Modelle innerhalb der Physik annehmen können und testen deklaratives Modellverständnis. Die Antworten erfolgen auf einer fünfstufigen Likert-Skala. Diese Skala mit sieben Items über verschiedene Eigenschaften von Modellen wurde mit  $N = 116$  Probandinnen und Probanden pilotiert, wobei sich ein Cronbachs Alpha von 0,72 ergab.

## 5.3. Bewertung der Intervention

Um die affektiven Merkmale zur Intervention testen zu können, wurden Items von Laukenmann et al. [42] übernommen. Zusätzlich wurden eigene Items zur Bewertung der verwendeten Software erstellt. In diesen Items wird die Nutzerfreundlichkeit, die Eignung

zum Lernen von Physik und der Wunsch des Einsatzes im Schulunterricht abgefragt. Die Skala zur Relevanz des behandelten Themas enthält Items, die die Wichtigkeit des Gelernten und den Wunsch, mehr zu erfahren, abfragen, während es in der Skala „Spaß an der Intervention“ um die erlebte Freude an dem Tag geht. Die Items werden auf einer fünfstufigen Likert-Skala beantwortet. Die Pilotierung fand hier ebenfalls mit  $N = 85$  Schülerinnen und Schülern statt (siehe Tab. 2).

Skalenbezeichnung	Items	$\alpha_c$
Bewertung der Software	5	0,85
Relevanz des behandelten Themas	3	0,78
Spaß an der Intervention	4	0,85

**Tab. 2:** Skalen zur Bewertung der Intervention

## 5.4. Cognitive Load

Um die kognitive Belastung zu messen, wurde ein Testinstrument von Becker et al. [27] übernommen, das auf einem englischsprachigen Test von Leppink et al. [43] basiert. Die Items werden auf einer sechsstufigen Likert-Skala beantwortet. Tabelle 3 gibt die Reliabilitäten nach Becker et al. [44] an.

Skalenbezeichnung	Items	$\alpha_c$
Intrinsic Cognitive Load (ICL)	3	0,86
Extraneous Cognitive Load (ECL)	3	0,55
Germane Cognitive Load (GCL)	4	0,93

**Tab. 3:** Skalen zum Cognitive Load

## 5.5. Kontrollvariablen

Um die Gleichheit der Gruppen in möglicherweise lernrelevanten Variablen zu überprüfen, wurden zudem verschiedene Kontrollvariablen erhoben. Dazu gehören die Mathematik- und Physiknote, das fachspezifische Selbstkonzept [45], das Fachinteresse [46], das Interesse an theoretischen Zusammenhängen (eigene Items), das Interesse am Experimentieren (eigene Items) und das Interesse an Computern [47] (siehe Tab. 4). Alle diese Items werden auf einer fünfstufigen Likert-Skala beantwortet.

Skalenbezeichnung	Items	$\alpha_c$
Fachspezifisches Selbstkonzept	4	0,93
Fachinteresse	4	0,87
Interesse an theoretischen Zusammenhängen	5	0,76
Interesse am Experimentieren	3	0,78
Computeraffinität	4	0,72

**Tab. 4:** Kontrollvariablen

## 6. Interventionen

Die Interventionen bestehen jeweils aus vier Experimenten, die die Grundaussagen der Newton'schen Mechanik verdeutlichen sollen. Der gesamte Prozess wird in beiden Gruppen von einem Arbeitsheft begleitet [48]. Allen gemeinsam ist, dass Reibungseinflüsse bewusst thematisiert werden und damit der Unterschied zwischen einer idealisierten Situation und einer realen Bewegung herausgearbeitet wird. Es

wurden dazu bewusst verschiedene Bewegungen ausgewählt, die sich ergänzen und damit den Kern der Newton'schen Mechanik verdeutlichen sollen. In beiden Interventionen stellt die Durchführung des Experiments den Ausgangspunkt jeder Arbeitsphase dar. Die Arbeitsphase selbst findet anschließend in Zweiergruppen am Computer statt.

In der mathematischen Modellbildung werden daraufhin Hypothesen formuliert, welche Kräfte während der Bewegung auf das sich bewegende Objekt gewirkt haben könnten. Diese Hypothesen werden dann mithilfe der Software expliziert und ausprobiert. Nach einer ersten Prüfung des Modells können die Hypothesen angepasst werden. Danach findet ein Vergleich des Modells mit Realdaten statt, die in das Programm eingefügt werden. Nach einer erneuten Anpassung des Modells beantworten die Modellierer mit dem Modell Fragen, die die Zusammenhänge zwischen den vorkommenden Größen betreffen.

Die Gruppe der Videoanalyse geht von dem Video des Experiments aus, aus dem mithilfe von „measure dynamics“ Daten gewonnen werden. Diese Daten können in unterschiedlicher Weise aufbereitet und angezeigt werden (Geschwindigkeits- und Beschleunigungspfeile im Video, synchron zum Video entstehende Graphen und Stroboskopbilder). Sie verdeutlichen den Zusammenhang innerhalb der kinematischen Größen und erlauben auch, von der Beschleunigung auf die wirkende Gesamtkraft zu schließen.

Der fundamentale Unterschied zwischen beiden Gruppen besteht neben der verwendeten Software in der Argumentationsrichtung. Eine Gruppe nähert sich von der Theorie und modelliert die Realität, die andere Gruppe geht von den Daten aus und analysiert diese. Dieses Vorgehen ist für alle in der Intervention durchgeführten Experimente identisch.

### 6.1. Fallkegel

Der Einstiegsversuch ist ein Fallkegel. Die relevanten Kräfte, die auf den Fallkegel wirken, sind eine konstante, nach unten wirkende Gravitationskraft und eine zunehmende, nach oben wirkende Luftreibungskraft. Verdeutlicht werden soll vor allem, dass sich der Fallkegel am Ende der Bewegung mit konstanter Geschwindigkeit bewegt, da die Luftreibungskraft so lange zunimmt, bis ein Kräftegleichgewicht besteht und die Summe aller auf den Körper wirkenden Kräfte null ist. Dadurch soll die Proportionalität zwischen Kraft und Beschleunigung verdeutlicht werden, da Schülerinnen und Schüler häufig von einer Proportionalität zwischen Kraft und Geschwindigkeit ausgehen. Der Versuch wird zudem zum Kennenlernen des jeweiligen Programms genutzt, wobei die Zweiergruppen zwar an ihrem eigenen PC arbeiten, die Lehrkraft aber über einen Beamer zeigt, wie mit dem Programm umzugehen ist.

### 6.2. Wagen

Nach der Einführung wird bei allen Experimenten selbstständig gearbeitet. Der zweite Versuch ist ein Wagen, der auf einer geradlinigen Fahrbahn in der

horizontalen Ebene fährt und von einem Gewicht, das über ein Seil mit dem Wagen verbunden ist, zunächst beschleunigt wird. Etwa bei der Hälfte der Bewegung kommt das Gewicht allerdings auf dem Boden auf, wodurch nur noch Reibungskräfte auf den Wagen wirken. Auch ohne eine Kraft in Bewegungsrichtung bewegt sich der Wagen weiter und wird lediglich durch Reibungskräfte gebremst.

### 6.3. Schiefer Wurf

Die erste zweidimensionale Bewegung ist ein schiefer Wurf, der von einem Katapult ausgeführt wird. Dabei ist von besonderer Bedeutung, dass der Flug einer Stahlkugel, bei der ein Modell ohne Luftreibung die Realität hinreichend genau beschreibt, mit dem Flug einer Styroporkugel verglichen wird, deren Flugbahn stark von Luftreibung beeinflusst wird. Die Form der Flugbahn ist sekundär. Der Fokus liegt nach wie vor auf dem Zusammenhang zwischen Beschleunigung und der Summe der wirkenden Kräfte.

### 6.4. Kreisbewegung

Der letzte Versuch ist eine Kreisbewegung, bei der eine Stahlkugel von einem Elektromagneten festgehalten und so im Kreis beschleunigt wird. Nach einer gewissen Zeit wird der Elektromagnet ausgeschaltet – die Kugel rollt geradeaus weiter. Typische Schülervorstellungen sind hier, dass die Kugel durch die Zentrifugalkraft nach außen gedrückt werde oder auch, dass während der gleichförmigen Kreisbewegung keine Beschleunigung vorhanden sei. Die Lernenden sollen vor dem Ausprobieren die Bahnkurve nach Ausschalten des Elektromagneten voraussagen. Das geradlinige Weiterrollen mit (nahezu) konstantem Tempo soll im Laufe der Arbeitsphase durch die (nahezu) Kräftefreiheit der Kugel (in der Ebene) erklärt werden.

## 7. Ergebnisse

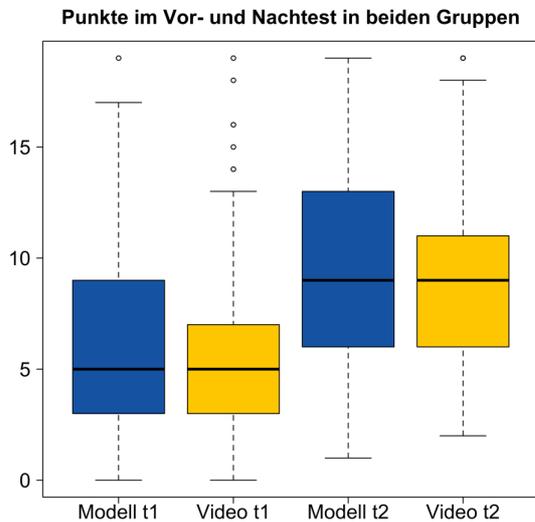
Für die Durchführung der Studie wurde für das erste Halbjahr 2020 eine Teilnahme von 33 Klassen mit insgesamt ca. 630 Schülerinnen und Schülern organisiert. Als die Hälfte der Klassen teilgenommen hatte, gab es infolge der Corona-Krise eine Schulschließung und die Vorgabe, dass im ganzen Schuljahr keine außerschulischen Aktivitäten mehr stattfinden dürfen.

Nach der Durchführung der Intervention mit 17 Klassen haben  $N = 246$  Lernende (davon 106 weiblich) an Vortest, Intervention und Nachtest teilgenommen. Mit diesen Daten sollen einige Ergebnisse vorgestellt werden.

Um die Vergleichbarkeit der beiden Gruppen zu überprüfen, wurden neben dem Konzeptverständnis im Vortest weitere Kontrollvariablen erhoben. Dabei handelt es sich um das fachspezifische Selbstkonzept, das Fachinteresse, das Interesse am Experimentieren, das Interesse an theoretischen Zusammenhängen, das Interesse an Computern und die Noten in Mathematik und Physik. In keiner dieser Skalen gab es einen Unterschied zwischen den Gruppen.

### 7.1. Konzeptverständnis

Zum Testen des Konzeptverständnisses wurde ein Test mit 19 passenden Items erstellt und pilotiert (siehe Abschnitt 5) [49]. Wenn damit der Zuwachs im Konzeptverständnis betrachtet wird (Abb.2), können beide Interventionsformen als erfolgreich angesehen werden.



**Abb. 2:** Boxplots der erreichten Punktzahlen beider Gruppen in Vortest t1 und Nachtest t2 (maximal 19 Punkte möglich)

Da die Daten als ausreichend normalverteilt angesehen werden können, wurde mit dem t-Test für gepaarte Stichproben der Unterschied der Mittelwerte der Punktzahlen im Vor- und Nachtest auf Signifikanz überprüft. Dieser Test fällt in beiden Gruppen höchst signifikant ( $p < 0,001$ ) aus. Bortz [50] empfiehlt für diese Situation eine Effektstärke, in die neben der Streuung der Differenzen die Korrelation zwischen beiden Messwertreihen als Korrekturfaktor eingeht.<sup>1</sup> Dabei ergeben sich Effektstärken von  $d = 0,96$  (95 %-KI: [0,81;1,14], Modellbildung) bzw.  $d = 1,1$  [0,91;1,36] (Videoanalyse), was jeweils einem großen Effekt entspricht [51], obwohl es sich bei den Interventionen nur um eine Vertiefung nach dem eigentlichen Unterricht der Newton'schen Gesetze handelte.

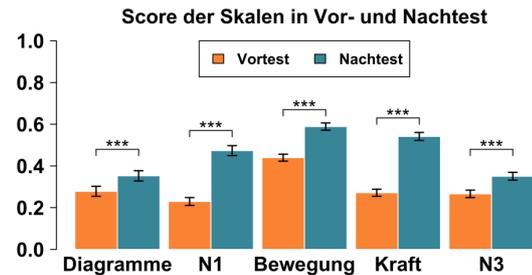
Dabei lassen sich die Test-Trainingseffekte als gering einstufen, da dieser Effekt bei Tests wie dem FCI gering ist [40], und da etwa eine Woche zwischen den Testzeitpunkten verging und sich außer bei einem Item keine Änderung in der Kontrollskala „3. Newton'sches Gesetz“ zeigte (siehe unten). Bei allen inhaltlichen Dimensionen des Tests gibt es höchst signifikante Zugewinne in beiden Teilgruppen (siehe Abb. 3). Es ergeben sich mit analoger Berechnung folgende Effektstärken in den Inhaltsbereichen:

- Beschleunigungsdiagramme:  $d = 0,23$  [0,1;0,36] (klein)

<sup>1</sup>  $d = \mu_D / \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$

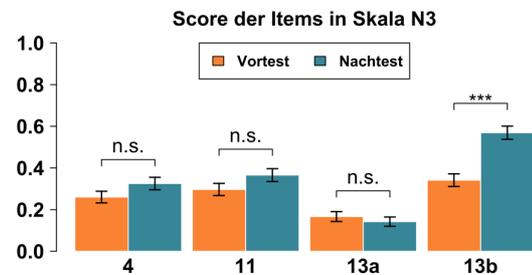
- 1. Newton'sches Gesetz:  $d = 0,66$  [0,55;0,78] (mittel)
- Bewegung bei bekannter Kraft:  $d = 0,61$  [0,49;0,75] (mittel)
- Kraft bei bekannter Bewegung:  $d = 0,89$  [0,76;1,02] (groß)

Dabei ist ersichtlich, dass die Konstrukte, die Inhalte der Dynamik überprüfen, alle einen Zuwachs mit mittlerem oder großem Effekt verzeichnen, während sich die einzige kinematische Skala mit geringerer Effektstärke verbessert.



**Abb. 3:** Anteil an richtigen Antworten für beide Gruppen in den Skalen in Vor- und Nachtest mit Standardfehler. Signifikanzniveau: \*\*\*  $p < 0,001$

Während der Zugewinn bei den ersten vier Skalen den Erwartungen entspricht, erstaunt zunächst der Zugewinn bei der Kontrollskala „3. Newton'sches Gesetz“ (Abb. 3). Wenn man sich die Items der Kontrollskala einzeln anschaut, ergibt sich ein differenzierteres Bild. Es ist zu erkennen, dass die Veränderung der Skala nur auf ein Item zurückzuführen ist (Abb. 4). In diesem Item wird nach den Kräften auf einen PKW und einen LKW gefragt, während der PKW den LKW mit konstanter Geschwindigkeit vor sich herschiebt (FCI-Item). Eine Hypothese zur Erklärung des größeren Anteils richtiger Antworten ist hier, dass die Probandinnen und Probanden durch die Intervention gelernt haben, dass eine konstante Geschwindigkeit immer mit einem Kräftegleichgewicht einhergeht, und dass sie diese Erkenntnis fälschlicherweise auf die Gleichheit von zwei an unterschiedlichen Körpern angreifenden Kräften übertragen haben (Wechselwirkungsprinzip), was eine typische Schülervorstellung ist [1].



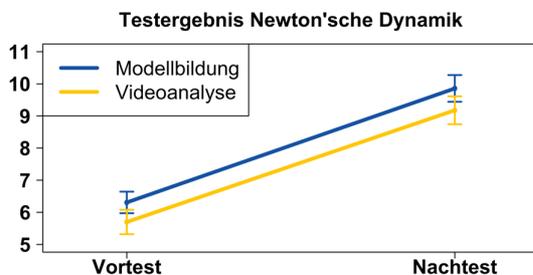
**Abb. 4:** Anteil an richtigen Antworten für beiden Gruppen in den Skalen in Vor- und Nachtest mit Standardfehler. Signifikanzniveau: \*\*\*  $p < 0,001$ , n. s. nicht signifikant

### 7.1.1. Vergleich der Lernzuwächse

In beiden Gruppen ergeben sich also Lernzuwächse, die wahrscheinlich auf die Intervention zurückzuführen sind. Um den Lernzuwachs zwischen den Gruppen vergleichen zu können, gibt es verschiedene Methoden. Eine Möglichkeit ist, den Lernzuwachs als Differenz zwischen Vor- und Nachtestergebnis anzusehen und mit einer Varianzanalyse zu überprüfen, ob diese von der Gruppenzugehörigkeit abhängt [50].

Die Varianzanalyse, die die Zeit als „within“-Faktor und die Gruppenzugehörigkeit als „between“-Faktor enthält, wird auch zweifaktorielle Varianzanalyse mit Messwiederholung (ANOVA) genannt. Für diese muss nach Bortz und Döring [50] eine Intervallskalierung der abhängigen Variablen, eine Normalverteilung der Messwerte in allen Stichproben sowie Varianzhomogenität und Sphärizität gegeben sein. Bis auf die Normalverteilung der Daten, welche nur bedingt gegeben ist, sind die Bedingungen in den vorliegenden Daten erfüllt. Nach Rasch et al. [52] sind Varianzanalysen aber robust gegenüber der Normalverteilungsannahme, weswegen eine ANOVA gerechnet werden kann.

Dabei hat nur der Faktor Zeit (also der Messzeitpunkt) einen signifikanten Einfluss auf die Punktzahl. In beiden Gruppen lernen die Probandinnen und Probanden also dazu, der Lernzuwachs unterscheidet sich aber nicht zwischen den Gruppen (Abb. 5).

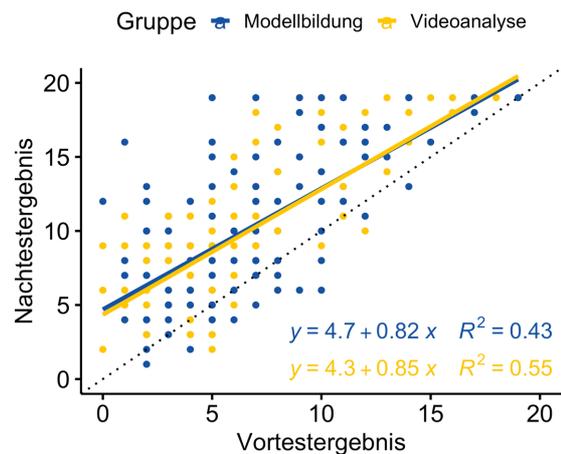


**Abb. 5:** Durchschnittliche Punktzahlen im gesamten Test (von 19) zur Newton'schen Dynamik in Vor- und Nachtest von beiden Gruppen mit Standardfehler

### 7.1.2. Kontrolle der Vortestergebnisse

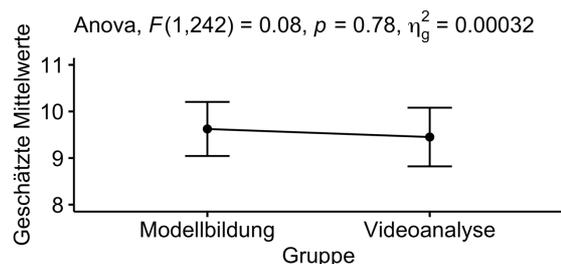
Die Betrachtung des absoluten Lernzuwachses ist problematisch, wenn eine Korrelation zwischen Vortestergebnis und Lernzuwachs vorliegt, was in der Regel der Fall ist. In diesem Fall liegt die Korrelation bei  $r = -0,19$ , was bedeutet, dass Probandinnen und Probanden, die im Vortest gut abschneiden, weniger dazu lernen (können). Beim Vergleich zwischen zwei Gruppen ist dies insbesondere problematisch, wenn es durch eine nicht vollständig randomisierte Zuteilung der Probandinnen und Probanden zu den Gruppen (die hier auch nur klassenweise möglich war) zu einer unterschiedlichen Verteilung verschiedener unabhängiger Variablen oder Kovariaten in den beiden

Gruppen kommt. Alternativ kann eine zweifaktorielle Kovarianzanalyse (ANCOVA) durchgeführt werden, die versucht, mögliche Unterschiede in den Vortestergebnissen (oder anderen unabhängigen Variablen) zu kontrollieren und damit zu berechnen, welche Nachtestergebnisse in den Gruppen zu erwarten wären, wenn diese mit einem gleichen Vortestergebnis gestartet wären. Die Bedingungen für eine Durchführung der zweifaktoriellen Varianzanalyse sind zu den bereits in Kapitel 7.1.1 überprüften Bedingungen zusätzlich noch die Homogenität der Regressionssteigungen und die Unabhängigkeit der Kovariate (= Vortestergebnis) vom Gruppeneffekt [53].



**Abb. 6:** Lineare Regressionen für beide Gruppen

Abb. 6 zeigt die Homogenität der Regressionssteigungen. Das Vortestergebnis unterscheidet sich zwischen den Gruppen nicht signifikant. Die ANCOVA bestätigt das Ergebnis der zweifaktoriellen Varianzanalyse mit Messwiederholung (Abb. 7): Der Lernzuwachs unterscheidet sich nicht.



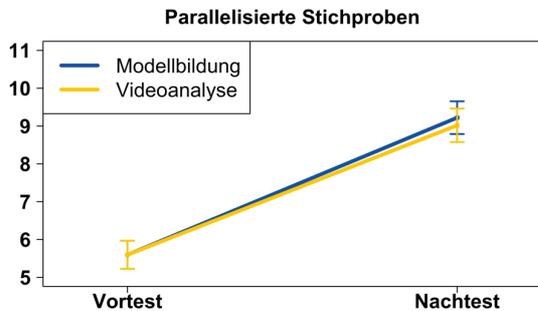
**Abb. 7:** Geschätzte Mittelwerte beider Gruppen bei einem Vortestergebnis von 6,03 Punkten mit 95%-Konfidenzintervall

### 7.1.3. Parallelisierung

Auch wenn die Vortestergebnisse durch Regression kontrolliert werden, kann eine „echte“ Gleichheit der Gruppen, wenn diese nicht durch das Setting gegeben ist, durch eine Parallelisierung erreicht werden.

Ziel der Parallelisierung ist, dass die Gruppen bezogen auf gewisse Merkmale übereinstimmen, um die interne Validität zu erhöhen. Es gibt verschiedene Möglichkeiten, parallelisierte Gruppen herzustellen.

Hier wurde durch Entfernung derjenigen Probandinnen und Probanden, die einen sehr hohen oder sehr niedrigen Lernzuwachs hatten, Gleichheit in beiden Gruppen bezüglich des Vortestergebnisses hergestellt. Nach dem Entfernen der nicht passenden Probandinnen und Probanden sind noch  $N = 99$  Lernende pro Gruppe in der Stichprobe. Auch nach einer Parallelisierung zeigt sich kein Unterschied zwischen den Gruppen (Abb. 8).

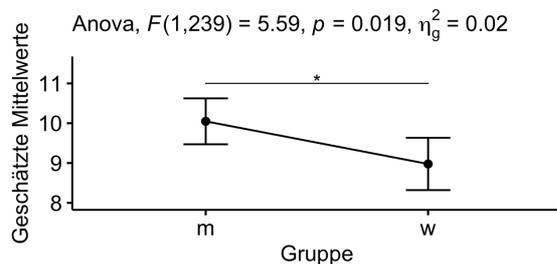


**Abb. 8:** Lernzuwachs nach Gruppen bei einer Parallelisierung der Stichproben durch Entfernung extremer Lernzuwächse und Standardfehler ( $N = 99$  Probandinnen und Probanden pro Gruppe)

In den Inhaltsbereichen „Beschleunigungsdiagramme“, „1. Newton’sches Gesetz“, „Kraft zu Bewegung“, „Bewegung zu Kraft“ und „3. Newton’sches Gesetz“ ergibt sich ebenfalls kein Unterschied zwischen den parallelisierten Gruppen.

#### 7.1.4. Einfluss des Geschlechts

Ebenfalls von Interesse ist, ob das Geschlecht einen Einfluss auf den Lernzuwachs hat. Eine zweifaktorielle Varianzanalyse ergibt auch unter Kontrolle des Vortestergebnisses ein höheres Nachtestergebnis der Jungen mit einem kleinen Effekt von  $\eta_g^2 = 0,02$  [0,00;0,07] [51] (Abb. 9), während sich in einer ANOVA keine Interaktion zwischen Geschlecht und Lernzuwachs zeigt.



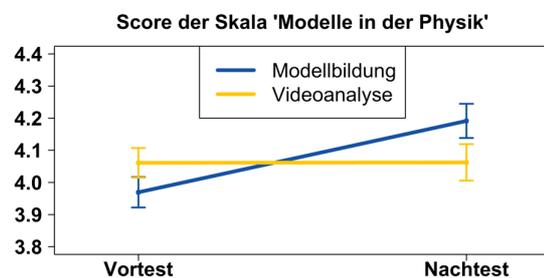
**Abb. 9:** Geschätzte Mittelwerte beider Geschlechter bei einem Vortestergebnis von 6,05 Punkten mit 95%-Konfidenzintervall. Signifikanzniveau \*  $p < 0,05$

Die unterschiedlichen Ergebnisse beider Verfahren (siehe Lords Paradoxon [54]) liegen in einer unterschiedlichen Konzeptualisierung von Änderung. Beide Verfahren haben Befürworter und Gegner [55-57] und es gibt eine lange Diskussion darüber, welches Verfahren in einer solchen Situation besser geeignet ist. Ohne diesen Konflikt abschließend auflösen zu können, lässt sich zusammenfassen, dass sich

der Lernzuwachs als Differenz zwischen Vor- und Nachtestergebnis zwischen den Geschlechtern nicht unterscheidet, die ANCOVA aber von einem höheren Nachtestergebnis bei den Jungen ausgeht, wenn beide Geschlechter mit demselben Vortestergebnis gestartet wären.

#### 7.2. Modellverständnis

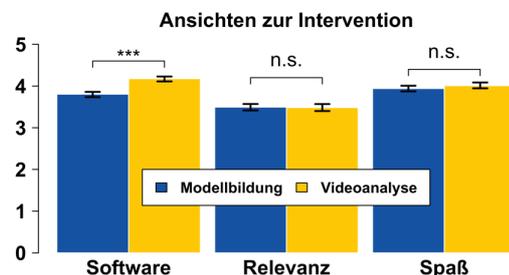
Der durchschnittliche Score in der Skala zum Modellverständnis ist bei der Gruppe der Modellbildung im Nachtest signifikant höher als im Vortest ( $p < 0,001$  mit Effektstärke von  $d = 0,45$  [0,27;0,67] [50]), während sich in der Gruppe der Videoanalyse nichts ändert (Abb. 10). Es lässt sich also sagen, dass das eigenständige Modellieren von Bewegungen zu einer Veränderung der Ansichten über Modelle führen kann, auch wenn die Eigenschaften von Modellen nicht explizit thematisiert werden.



**Abb. 10:** Durchschnittlicher Score (von 5) in der Skala "Modelle in der Physik" in Vor- und Nachtest von beiden Gruppen mit Standardfehler

#### 7.3. Einschätzung der Lernenden

Die Probandinnen und Probanden wurden im Nachtest zu verschiedenen Aspekten der Intervention befragt (siehe Abschnitt 5.3). Dabei wird die Software zur Videoanalyse („measure dynamics“) signifikant besser bewertet (mit Effektstärke von  $d = 0,55$  [0,3;0,82]) als die Software zur Modellbildung („Newton-II“). In den Skalen zur Relevanz des behandelten Themas und zum Spaß an der Intervention zeigen sich keine Unterschiede (siehe Abb. 11).



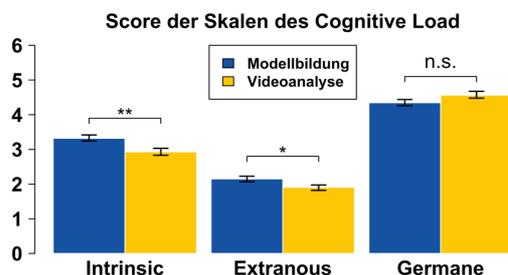
**Abb. 11:** Durchschnittlicher Score (von 5) in den Skalen zu Ansichten zur Intervention mit Standardfehler. Signifikanzniveau: \*\*\*  $p < 0,001$ ; n.s. nicht signifikant

#### 7.4. Cognitive Load

Wie im Abschnitt 2.3 dargelegt kann die Cognitive Load Theory einen Erklärungsansatz für den Erfolg einer gewissen Unterrichtsmethode liefern. Dabei wird zwischen dem Intrinsic Cognitive Load (ICL),

dem Extranous Cognitive Load (ECL) und dem Germane Cognitive Load (GCL) unterschieden.

Im Mittel erreichen die Probandinnen und Probanden in der Modellbildung einen Score von 3,33 im ICL, 2,15 im ECL und 4,35 im GCL (jeweils von 6), während in der Videoanalyse der Durchschnitt bei 2,93 im ICL, 1,90 im ECL und 4,58 im GCL liegt. Dabei unterscheiden sich die Werte im ICL und ECL nach dem t-Test für unabhängige Stichproben signifikant zwischen den Gruppen. Da nicht von der gleichen Streuung in beiden Gruppen ausgegangen werden kann, empfiehlt Bortz [50] hier mit gepoolter Standardabweichung zu rechnen. Dabei ergeben sich Effektstärken von  $d = 0,38$  [0,12;0,68] und  $d = 0,28$  [0,02;0,54], während im GCL kein Unterschied vorliegt (siehe Abb. 12). Das lässt vermuten, dass die Intervention zur Modellbildung für die Lernenden herausfordernder ist. Das äußert sich vor allem im Unterschied der beiden Gruppen in einem Item, in dem die Komplexität der verwendeten Formeln bewertet werden soll („Die Formeln zur Dynamik empfand ich als sehr komplex“;  $d = 0,52$  [0,26;0,77]). Das ist nicht verwunderlich, da hier verschiedene Reibungsarten (Luftreibung, Rollreibung) quantitativ modelliert werden, während sie bei der Videoanalyse nur qualitativ betrachtet werden. Der Unterschied im ICL ist fast ausschließlich auf dieses Item zurückzuführen. Auch die Lernumgebung scheint bei der mathematischen Modellbildung eine höhere kognitive Belastung ausgelöst zu haben, was nach den Lernenden vor allem an mehr „unklaren Begriffen“ lag („Die Aufgaben und Arbeitsaufträge waren voll von unklaren Begriffen“;  $d = 0,32$  [0,07;0,58]). Dies könnte daran liegen, dass bei der Modellierung u. a. mit bedingten Variablen gearbeitet werden muss. Der Germane Cognitive Load, der für die Konstruktion neuer Wissensschemata entscheidend ist, ist in beiden Gruppen gleichermaßen hoch und deckt sich mit dem Ergebnis, dass beide Gruppen gleichermaßen viel dazulernen.



**Abb. 12:** Durchschnittlicher Score (von 6) in den Skalen des Cognitive Load von beiden Gruppen mit Standardfehler. Signifikanzniveau: \*\*  $p < 0,01$ ; \*  $p < 0,05$ ; n.s. nicht signifikant

### 7.5. Theoretisierende versus Experimentierende

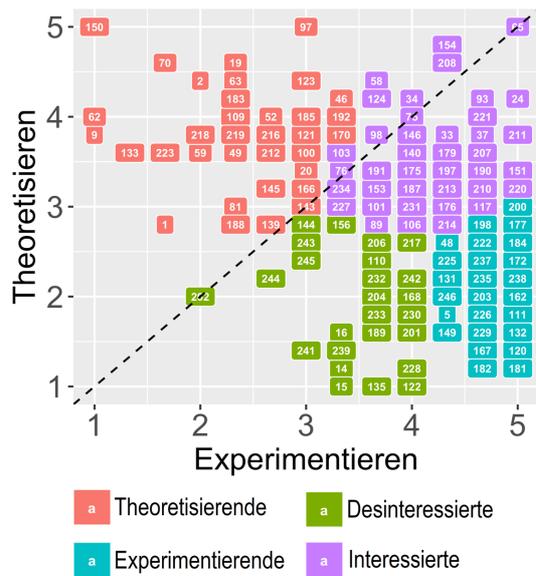
Mit den Scores im Interesse an theoretischen Zusammenhängen in der Physik und im Interesse am Experimentieren wurde eine Clusteranalyse durchgeführt, um Typen von Lernenden zu identifizieren und um

herauszufinden, ob ein primäres Interesse am Theoretisieren bzw. Experimentieren einen Einfluss auf den Lernzuwachs in den jeweiligen Interventionen hat.

Hierbei ist nicht zu erwarten, dass sich die Lernenden eindeutig in eine bestimmte Anzahl an voneinander verschiedenen Gruppen aufteilen. Dementsprechend werden für die Wahl der Anzahl an Clustern statistische Verfahren mit inhaltlichen Überlegungen kombiniert. Für die Wahl der optimalen Anzahl an Clustern gibt es kein eindeutiges Kriterium. Es stehen verschiedene Verfahren zur Bestimmung der optimalen Anzahl an Clustern zur Verfügung [z. B. 58-60], wobei neben den statistischen Verfahren auch die inhaltliche Interpretation eine Rolle spielt. Hier wurde das Ellenbogenkriterium [61] herangezogen. Dabei wird die Variabilität innerhalb der Cluster über der Anzahl an Clustern aufgetragen – die optimale Anzahl an Clustern befindet sich dort, wo der Graph einen Knick macht. Dieses Verfahren ist nicht eindeutig, verschafft aber einen ersten Einblick. Es deutet hier auf drei Cluster hin. Die sich ergebende Zuteilung zu drei Clustern ist aus inhaltlicher Sicht nicht sinnvoll, da so Personen, die in beiden Skalen sehr niedrig scoren, mit denjenigen Personen in eine Gruppe eingeteilt werden, die großes Interesse an theoretischen Zusammenhängen haben. Inhaltlich erscheint eine Wahl von vier Clustern sinnvoller.

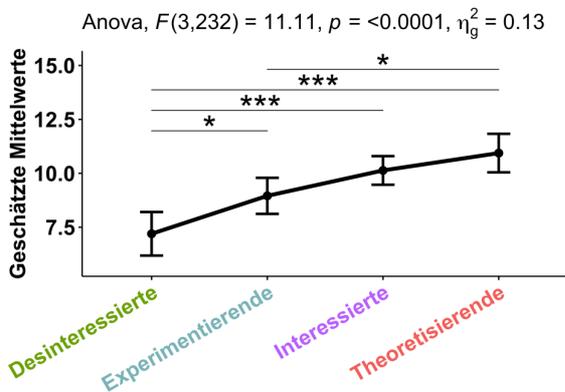
Die Zuteilung zu den Clustern erfolgte mit dem K-Means-Verfahren [62], das eine möglichst große Ähnlichkeit innerhalb der Cluster und eine möglichst kleine Ähnlichkeit zwischen den Clustern anstrebt. Erzeugt wurden die Cluster nach einem Algorithmus von Hartigan & Wong [63] mit euklidischer Distanzmetrik. Bei der Zuteilung zu vier Clustern ergibt sich ein Silhouettenkoeffizient von 0,37, was einer schwachen Strukturierung entspricht [64].

Ein Cluster besteht aus Schülerinnen und Schülern, die deutlich lieber Hypothesen aufstellen und über theoretische Zusammenhänge nachdenken, als zu experimentieren („Theoretisierende“). Ein Cluster besteht aus Schülerinnen und Schülern mit den gegenteiligen Einstellungen („Experimentierende“). In einem Cluster sind Schülerinnen und Schüler, die beides gerne mögen („Interessierte“), und ein weiteres Cluster besteht aus Schülerinnen und Schülern, die an beiden Bereichen ein gleichermaßen geringes Interesse zeigen („Desinteressierte“). Bei den „Interessierten“ und „Desinteressierten“ ist das Experimentieren beliebter als das Theoretisieren – es besteht also insgesamt eine Schiefe in den Daten in Richtung des Experimentierens (Abb. 13).



**Abb. 13:** Zuteilung der Probandinnen und Probanden zu den Clustern anhand des Scores der Skalen "Experimentieren" und "Theoretisieren"

Unter Kontrolle des Vortestergebnisses haben diese Cluster einen Einfluss auf das Nachtestergebnis mit einer mittleren bis großen Effektstärke von  $\eta_g^2 = 0,13$  [0,05;0,20] [51]. Dies kann man an den geschätzten Mittelwerten der Cluster im Nachtest sehen (Abb. 14).



**Abb. 14:** Geschätzte Mittelwerte der Cluster im Nachtest bei einem Vortest-Ergebnis von 5,99 Punkten mit 95 %-Konfidenzintervallen. Signifikanzniveaus: \*\*\*  $p < 0,001$ , \*  $p < 0,05$

Eine Hypothese war, dass die „Theoretisierenden“ in der Intervention zur mathematischen Modellbildung mehr lernen, während die „Experimentierenden“ in der Intervention zur Videoanalyse einen größeren Lernzuwachs verzeichnen. Diese Hypothese ließ sich auf der Basis der Daten nicht bestätigen.

### 8. Fazit und Diskussion

Es scheint, als seien beide Interventionen zum Vertiefen der Newton'schen Dynamik geeignet, wobei keine Unterschiede im Lernzuwachs zwischen den beiden Methoden gefunden werden konnten. Dazu

passt auch das Ergebnis, dass der Germane Cognitive Load in beiden Gruppen gleichermaßen hoch war.

Das bedeutet, dass hier im Gegensatz zu vielen bisherigen Studien mit Software zur mathematischen Modellbildung auch bei der selbstständigen Nutzung der Software keine großen, für den Lernprozess hinderlichen Schwierigkeiten beobachtet wurden [3], was wohl vor allem auf die Art der verwendeten Software und die damit verbundene Zielsetzung zurückzuführen ist. Das Programm „Newton-II“ mit der gleichungsbasierten Modellbildung scheint bei Bereitstellung der nötigen Hilfen für den Einsatz im Schulunterricht geeignet zu sein. Es bietet sich aber, im Gegensatz zu anderen Modellbildungsprogrammen, nur für das Lernen der Dynamik an. Nach kurzer Einarbeitungszeit waren die Schülerinnen und Schüler in der Lage, das Programm zu bedienen, wobei es bei der eigenständigen Bedienung scheinbar dennoch mehr Probleme gab als bei dem Videoanalyseprogramm „measure dynamics“. Gleichzeitig ist es möglich, die Einstellungen der Lernenden zu Modellen implizit zu verändern, wenn diese im Unterricht modellieren, was ein Argument für den Einsatz der mathematischen Modellbildung ist. Die mathematische Modellbildung wird aber als komplizierter wahrgenommen und besitzt einen höheren Intrinsic Cognitive Load als die Videoanalyse, da eine quantitative Betrachtung der wirkenden Kräfte unumgänglich ist.

Die Videoanalyse und deren Möglichkeiten zur Aufbereitung von Daten scheint für das Lernen der Dynamik ebenfalls geeignet zu sein. Auch wenn die Software die Verknüpfung zwischen Bewegung und Kräften nicht selbst herstellt, kann die Analyse der kinematischen Größen mit den richtigen Fragen zu einem besseren Kraftverständnis führen und Schüler(fehl)vorstellungen reduzieren. Die Software „measure dynamics“ wurde von den Schülerinnen und Schülern gelobt. Die Ergebnisse decken sich mit anderen Studien zur Videoanalyse, bei denen die lernhinderliche kognitive Belastung durch die Videoanalyse geringgehalten werden konnte. Während diese Studien auf mobile Endgeräte im Bereich der Kinematik setzten, scheinen die grundsätzlichen Ergebnisse auch auf die Nutzung von Laptops im Bereich der Dynamik übertragbar zu sein.

Die großen Unterschiede zwischen Vor- und Nachtestergebnis in beiden Gruppen sprechen dafür, auch im Unterricht komplexe Bewegungen, welche auf Basis bekannter Schülervorstellungen ausgewählt werden, zu thematisieren und dazu den Computer als notwendiges Hilfsmittel heranzuziehen. Reibungseinflüsse sollten bewusst und aktiv thematisiert werden.

Es lassen sich zudem Typen von Lernenden unterscheiden, was das Interesse am Experimentieren bzw. an theoretischen Zusammenhängen angeht, wobei Experimentieren generell beliebter ist. Diese Typen lernen bei den Interventionen unterschiedlich viel dazu. Vor allem scheint das Interesse an theoretischen

Zusammenhängen den Lernerfolg positiv zu beeinflussen. Es konnte aber nicht bestätigt werden, dass Theorieinteressierte beim Modellieren mehr dazulernen, während „Experimentierer“ bei der Videoanalyse mehr dazulernen.

Da die Ergebnisse der Studie in einer intensiven einmaligen Intervention erzielt wurden, lässt sich die externe Validität der Studie diskutieren. Die Übertragbarkeit auf den Schulunterricht ist nicht unmittelbar gegeben und sollte durch Folgestudien überprüft werden. Insbesondere wäre eine Studie sinnvoll, die überprüft, inwiefern ein Curriculum, das konsequent auf komplexe und alltagsnahe Probleme setzt und dazu den Einsatz von gleichungsbasierter mathematischer Modellbildung verstärkt, im Vergleich zu anderen Unterrichtsformen abschneidet. Dies ergäbe auch die Möglichkeit, die beobachtete Steigerung im Modellverständnis aufzugreifen und die Rolle von Modellen in der Erkenntnisgewinnung intensiver zu diskutieren. Dazu gehört, dass Physiklehrkräfte in der Nutzung der Programme geschult werden müssen, um diese eigenständig im Unterricht einsetzen zu können.

Da sich beide untersuchten Methoden in gewisser Weise ergänzen und entsprechende Kurskonzepte an Universitäten bereits einigen Erfolg erzielen konnten [z. B. 65], erscheint eine Kombination aus beiden Methoden für den Physikunterricht vielversprechend. Verschiedene Autoren konnten bereits zeigen, dass Computereinsatz (sowohl mathematische Modellbildung als auch Videoanalyse) zu einem schüler- bzw. studentenzentrierten und alltagsnahen Physikunterricht führen kann, der erfolgreicher ist als herkömmliche Varianten [38]. Diesbezüglich kann diese Studie als ein Ansatz angesehen werden, der zeigt, dass gleichungsbasierte Modellbildungsprogramme wie „Newton-II“ die technischen Hürden so reduzieren, dass sie sinnvoll in ein Curriculum eingebaut werden können und sich für die Lernenden weniger Probleme ergeben als bei Programmiersprachen [12] oder graphischen Modellbildungsprogrammen [4; 5].

Zu kritisieren ist die geringe Probandenzahl, die durch die Unterbrechung der Erhebung zustande kam. Im Optimalfall sollten weitere Klassen an den Interventionen teilnehmen, um belastbare Daten zu erhalten. Insbesondere für die Analyse von Subgruppen ist die Zahl der Teilnehmerinnen und Teilnehmer als gering einzuschätzen. Eine größere Klassenzahl hätte auch den Vorteil, dass statistische Verfahren wie die Mehrebenenanalyse besser mit einer größeren Anzahl an Objekten auf der höheren Ebene durchführbar sind, und damit die hierarchische Struktur der Daten berücksichtigt werden könnte.

## 9. Ausblick

Neben den hier vorgestellten Ergebnissen, die auf einer quantitativen Erhebung mittels Fragebogen basieren, ist es von Interesse, wie Schülerinnen und Schüler mit dem Computer arbeiten. Es würde sich ein umfassenderes Bild für die Eignung und den geschickten

Einsatz der beiden Methoden im Physikunterricht ergeben. Dazu wird in einer weiteren Untersuchung die Arbeitsweise der Probandinnen und Probanden betrachtet: Sie werden mithilfe von Bildschirmvideos während der Arbeit beobachtet. Neben den auftretenden Schwierigkeiten ist vor allem von Interesse, wie beim Modellieren und bei der Videoanalyse vorgegangen wird, wie zielgerichtet gearbeitet wird, welche Schwierigkeiten auftreten und zu welcher Art von (physikalischen) Gesprächen die beiden Herangehensweisen anregen.

Wie bereits diskutiert, wäre ein möglicher nächster Schritt für eine weitere Studie, mit den gewonnenen Erkenntnissen in den Schulunterricht zu gehen und ein Studiendesign mit einer höheren externen Validität anzustreben und zu erforschen, welche Effekte durch den konsequenten Computereinsatz in der Newton'schen Mechanik im Unterricht erzielt werden.

## 10. Danksagung

Wir danken der gemeinnützigen Stiftung GIERSCHE für die Finanzierung des Goethe-Schülerlabors Physik, ohne das die vorgestellte Studie in dieser Form nicht durchführbar gewesen wäre.

## 11. Literatur

- [1] Schecker, H.; Wilhelm, T. (2018): Schülervorstellungen in der Mechanik. In: Schecker, H., Wilhelm, T., Hopf, M., Duit, R. (Hrsg.). Schülervorstellungen und Physikunterricht. Ein Lehrbuch für Studium, Referendariat und Unterrichtspraxis, Berlin: Springer Spektrum
- [2] Schecker, H. (1998): Physik modellieren. Grafikorientierte Modellbildungssysteme im Physikunterricht, Stuttgart: Klett-Verlag
- [3] Weber, J.; Wilhelm, T. (2020): The benefit of computational modelling in physics teaching: A historical overview. In: Eur. J. Phys. 41 034003. <https://doi.org/10.1088/1361-6404/ab7a7f>
- [4] Hücke, L. (1999): Handlungsregulation und Wissenserwerb in traditionellen und computer-gestützten Experimenten des physikalischen Praktikums. Berlin: Logos
- [5] Sander, F.; Schecker, H.; Niedderer, H. (2001): Wirkungen des Einsatzes grafikorientierter Modellbildung im physikalischen Praktikum. In: Zeitschrift für Didaktik der Naturwissenschaften, 7, 147-165
- [6] Schecker, H.; Klieme, E.; Niedderer, H.; Ebach, J.; Gerdes, J. (1999): Abschlussbericht zum DFG-Projekt "Physiklernen mit Modellbildungssystemen". Förderung physikalischer Kompetenz und systemischen Denkens durch computergestützte Modellbildungssysteme. Bremen: Universität Bremen
- [7] Benacka, J. (2015a): Projectile general motion in a vacuum and a spreadsheet simulation. In:

- Physics Education 50(1), S. 58-63.  
<https://doi.org/10.1088/0031-9120/50/1/58>
- [8] Benacka, J. (2015b): Spreadsheet application showing the proper elevation angle, points of shot and impact of a projectile. In: Physics Education 50(3), S. 342-347.  
<http://dx.doi.org/10.1088/0031-9120/50/3/342>
- [9] Benacka, J. (2016): Numerical Modelling with Spreadsheets as a Means to Promote STEM to High School Students. In: Eurasia Journal of Mathematics, Science & Technology Education 12(4), S. 947-964. <https://doi.org/10.12973/eurasia.2016.1236a>
- [10] Scherer, D.; Dubois, P.; Sherwood, B. (2000): VPython: 3D interactive scientific graphics for students. In: Computing in Science & Engineering 2(5), pp. 56-62.  
<https://doi.org/10.1109/5992.877397>
- [11] Caballero, M.; Kohlmyer, M.; Schatz, M. (2012): Implementing and assessing computational modeling in introductory mechanics. In: Phys. Rev. ST Phys. Educ. Res. 8, 020106.  
<https://doi.org/10.1103/PhysRevSTPER.8.020106>
- [12] Caballero, M.; Burk, J.; Aiken, J.; Thoms, B.; Douglas, S.; Scanlon, E.; Schatz, M. (2014): Integrating Numerical Computation into the Modelling Instruction Curriculum. In: The Physics Teacher 52, pp. 38-42. <https://arxiv.org/abs/1207.0844>
- [13] <https://did-apps.physik.uni-wuerzburg.de/Newton-II>
- [14] Laws, P.; Pfister, H. (1998): Using digital video analysis in introductory mechanics projects. In: The Physics Teacher 36, 282-287.  
<https://doi.org/10.1119/1.880068>
- [15] Wilhelm, T.; Trefzger, T. (2010): Erhebung zum Computereinsatz bei Physik-Gymnasiallehrern: Beiträge zur DPG-Frühjahrstagung, Hannover 2010. In: Phydid B, Didaktik der Physik, Beiträge zur DPG-Frühjahrstagung, Berlin.
- [16] Wenzel, M. (2018): Computereinsatz in Schule und Schülerlabor. Einstellungen von Physiklehrkräften zu Neuen Medien, Studien zum Physik- und Chemielernen, Vol. 251, Berlin: Logos-Verlag
- [17] Beichner, R. (1996): The impact of video motion analysis on kinematics graph interpretation skills. In: American Journal of Physics 64 (10), 1272-1277. <https://doi.org/10.1119/1.18390>
- [18] Hockicko, P.; Trpišová, B.; Ondruš, J. (2014): Correcting students' misconceptions about automobile braking distances and video analysis using interactive program tracker. In: Journal of Science Education and Technology, 23(6), 763-776. <http://dx.doi.org/10.1007/s10956-014-9510-z>
- [19] Zollman, D.; Escalada, L. (1996): Applications of interactive digital video in a physics classroom. In: Journal of Educational Multimedia and Hypermedia 5(1), 73-97
- [20] Becker, A.; Klein, P.; Gößling, A.; Kuhn, J. (2019): Förderung von Konzeptverständnis und Repräsentationskompetenz durch Tablet-PC-gestützte Videoanalyse – Empirische Untersuchung der Lernwirksamkeit eines digitalen Lernwerkzeugs im Mechanikunterricht der Sekundarsufe 2. In: Zeitschrift für Didaktik der Naturwissenschaften 5.  
<https://doi.org/10.1007/s40573-019-00089-4>
- [21] Wilhelm, T. (2005): Konzeption und Evaluation eines Kinematik/Dynamik-Lehrgangs zur Veränderung von Schülervorstellungen mit Hilfe dynamisch ikonischer Repräsentationen und graphischer Modellbildung. Studien zum Physik- und Chemielernen. Band 46. Berlin: Logos-Verlag.
- [22] Hochberg, K.; Becker, S.; Louis, M. Klein, P.; Kuhn, J. (2020) Using Smartphones as Experimental Tools—a Follow-up: Cognitive Effects by Video Analysis and Reduction of Cognitive Load by Multiple Representations. In: J. Sci. Educ. Technol. <https://doi.org/10.1007/s10956-020-09816-w>
- [23] <https://www.phywe.de/de/software-measure-dynamics-einzellenz.html>
- [24] Sweller, J. (1988): Cognitive load during problem solving: Effects on learning. Cognitive Science, 12, 257-285.  
[https://doi.org/10.1207/s15516709cog1202\\_4](https://doi.org/10.1207/s15516709cog1202_4)
- [25] van Merriënboer, J.; Sweller, J. (2005): Cognitive load theory and complex learning: Recent developments and future directions. In: Educational Psychology Review, 17(2), 147-177. <https://doi.org/10.1007/s10648-005-3951-0>
- [26] Leppink, J.; van den Heuvel, A. (2015): The evolution of cognitive load theory and its application to medical education. In: Perspectives on Medical Education, 4(3), 119-127.  
<https://doi.org/10.1007/s40037-015-0192-x>
- [27] Becker, S.; Klein, P.; Gößling, A.; Kuhn, J. (2019): Using Mobile Devices to Augment Inquiry-Based Learning Process with Multiple Representations. arXiv:1908.11281
- [28] Becker, S.; Klein, P.; Gößling, A.; Kuhn, J. (2020): Investigating Dynamic Visualizations of Multiple Representations Using Mobile Video Analysis in Physics Lessons. In: Zeitschrift für Didaktik der Naturwissenschaften. <https://doi.org/10.1007/s40573-020-00116-9>
- [29] Leisner-Bodenthin, A. (2006): Zur Entwicklung von Modellkompetenz im Physikunterricht. In: Zeitschrift für Didaktik der Naturwissenschaften 12, 91-109

- [30] Weber, J.; Wilhelm, T. (2019): Mathematische Modellbildung in einer vergleichenden Untersuchung. In: PhyDid-B - Didaktik der Physik – DPG-Frühjahrstagung, S. 323-329. <http://phydid.physik.fu-berlin.de/index.php/phydid-b/article/view/958>
- [31] Hestenes, D.; Wells, M.; Swackhamer, G. (1992): Force Concept Inventory. In: The Physics Teacher 30, 141-158. <https://doi.org/10.1119/1.2343497>
- [32] Thornton, R.; Sokoloff, D. (1997): Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. In: American Journal of Physics 66(4), 338-352
- [33] Heuer, D.; Wilhelm, T. (1997): Aristoteles siegt immer noch über Newton - Unzulängliches Dynamikverstehen in Klasse 11. In: Der mathematische und naturwissenschaftliche Unterricht 50(5), 280-285
- [34] Wilhelm, T. (2007): Vektorverständnis und vektoriell Kinematikverständnis von Studienanfängern. In: Nordmeier, V., Oberländer, A., Grötzebauch, H. (Hrsg.): Didaktik der Physik - Regensburg 2007, Berlin: Lehmanns Media LOB.de
- [35] Flores, S.; Kanim, S.; Kautz, C. (2004): Student use of vectors in introductory mechanics. In: American Journal of Physics 72(4), 460-468
- [36] Warren, J. (1979): Understanding Force. Übersetzt aus dem Englischen von Backhaus, U. & Schneider, T. (1998). Verständnisprobleme beim Kraftbegriff. Universität Koblenz
- [37] Savinainen A.; Scott, P. (2002): The Force Concept Inventory: a tool for monitoring student learning. In: Physics Education 37(1), 45-52
- [38] 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(1), 64-74. <https://doi.org/10.1119/1.18809>
- [39] Huffmann, D.; Heller, P. (1995): What does the force inventory actually measure? In: The Physics Teacher 33, 138-143. <https://doi.org/10.1119/1.2344171>
- [40] Schecker, H.; Gerdes, J. (1999): Messung und Konzeptualisierungsfähigkeit in der Mechanik – Zur Aussagekraft des Force Concept Inventory. In: Zeitschrift für Didaktik der Naturwissenschaften 5(1), 75-89
- [41] Gerdes, J.; Schecker, H. (1999): Der Force Concept Inventory – Ein diagnostischer Test zu Schülervorstellungen in der Mechanik. In: Der Mathematische und Naturwissenschaftliche Unterricht 52(5), 283-288
- [42] Laukenmann, M.; Bleicher, M.; Fuß, S.; Gläser-Zikuda, M.; Mayring, P.; von Rhöneck, C. (2000): Eine Untersuchung zum Einfluss emotionaler Faktoren auf das Lernen im Physikunterricht. In: Zeitschrift für Didaktik der Naturwissenschaften 6, 139-155
- [43] Leppink, J.; Paas, F.; van der Vleuten, C.; van Gog, T.; van Merriënboer, J. (2013): Development of an instrument for measuring different types of cognitive load. In: Behavior Research Methods, 45(4), 1058–1072.
- [44] Becker, S.; Klein, P.; Gößling, A.; Kuhn, J. (2020): Reduktion kognitiver Belastung und Förderung von Konzeptverständnis durch Tablet-PC-gestützte Videoanalyse: Hauptstudie zur Lernwirksamkeit im Mechanikunterricht der Sekundarstufe 2. ZfDN 26 (1), eingereicht
- [45] Helmke, A. (1992). Determinanten der Schulleistung: Forschungsstand und Forschungsdefizit. In G. Nold (Hrsg.), Lernbedingungen und Lernstrategien. Tübingen: Narr, 23-34
- [46] Kunter, M.; Schümer, G.; Artelt, C.; Baumert, J.; Klieme, E.; Neubrand, M.; Prenzel, M.; Schiefele, U.; Schneider, W.; Stanat, P.; Tillmann, K.-J.; Weiß, M. (2002): PISA 2000. Dokumentationen der Erhebungsinstrumente. In: Materialien aus der Bildungsforschung 72. Max-Planck-Institut für Bildungsforschung. Berlin.
- [47] Brell, C. (2008): Lernmedien und Lernerfolg – reale und virtuelle Materialien im Physikunterricht. Berlin: Logos-Verlag
- [48] <http://www.thomas-wilhelm.net/dynamik.htm>
- [49] Weber, J.; Wilhelm, T. (2020b): Eine vergleichende Untersuchung zur Newton'schen Mechanik. In: Habig, S. (Hrsg.): Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen, Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Wien 2019, Band 40
- [50] Bortz, J.; Döring, N. (2006): Forschungsmethoden und Evaluation – für Human- und Sozialwissenschaftler. 4. Auflage. Heidelberg: Springer-Verlag.
- [51] Cohen, J. (1988): Statistical Power Analysis for the Behavioral Sciences. New York, NY: Routledge Academic. DOI: 10.4324/9780203771587
- [52] Rasch B., Frieze, M., Hofman, W., Nauman, E. (2010): Quantitative Methoden 1 – Einführung in die Statistik für Psychologen und Sozialwissenschaftler. 3. Erweiterte Auflage. Berlin: Springer-Verlag.
- [53] Field, A. (2011): Discovering Statistics using SPSS (and sex and drugs and rock'n'roll). 3. Auflage. London: SAGE.
- [54] Lord, F. (1967): A paradox in the interpretation of group comparisons. In: Psychological bulletin 68 (5), S. 304-305. DOI: 10.1037/h0025105
- [55] Cronbach, L.; Furby, L. (1970): How we should measure "change": Or should we? In: Psychological bulletin 74 (1). S. 68-80. DOI: 10.1037/h0029382

- [56] Rogosa, D. (1988): Myths about longitudinal research. In: K. Schaie, R. Campbell, W. Meredith und S. Rawlings (Hg.): *Methodological issues in aging research*. New York, NY, US: Springer Publishing Company, S. 171-209
- [57] Lord, F. (1967): Elementary models for measuring change. In: C. Harris (Hg.): *Problems in measuring change*. Madison: University of Wisconsin Press, S. 199-211.
- [58] Tibshirani, R; Walther, G; Hastie, T. (2001): Estimating the number of clusters in a data set via the gap statistic. In: *J. R. Statist. Soc. B* 63, 411-423
- [59] Salvador, S.; Chan, P. (2004): Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms. In: *Proceedings – International Conference on Tools with Artificial Intelligence, ICTAI*. 576-584
- [60] Sugar, C.; James, G. (2003): Finding the Number of Clusters in a Dataset. In: *Journal of the American Statistical Association* 98(463), 750-763
- [61] Thorndike, R. (1953): Who belongs in the family? In: *Psychometrika* 18, 267-276
- [62] MacQueen, J. (1967): Some Methods for Classification an Analysis of Multivariate Observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, Berkeley: University of California Press, 281-297
- [63] Hartigan, J.; Wong, M. (1979): A K-Means Clustering Algorithm. In: *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 28(1), 100-108
- [64] Ester, M.; Sander J. (2000): *Knowledge Discovery in Databases: Techniken und Anwendungen*. Hamburg/Berlin: Springer-Verlag.
- Laws, P.; Willis, M.; Sokoloff, D. (2015): Workshop physics and related curricula: a 25-year-history of collaborative learning enhanced by computer tools for observation and analysis. In: *The Physics Teacher* 53, 401-406