

Mechanik in der Sekundarstufe I mittels Neuer Medien entdecken

VON DER FAKULTÄT FÜR MATHEMATIK UND
PHYSIK
DER GOTTFRIED WILHELM LEIBNIZ
UNIVERSITÄT HANNOVER ZUR ERLANGUNG DES
AKADEMISCHEN GRADES

DOKTOR DER NATURWISSENSCHAFTEN
DR. RER. NAT.

GENEHMIGTE DISSERTATION VON

M.Ed. Ingmar Schneider

1. Gutachter: Prof. Dr. G. Friege
 2. Gutachter: Prof. Dr. M. Komorek
 3. Gutachterin: Prof. Dr. K. Kremer
- Tag der Disputation: 23.02.2021

Kurzzusammenfassung

Der Mechanikunterricht in der Sekundarstufe 1 bietet unterschiedliche didaktische Herangehensweisen für den ersten schulischen Kontakt mit der Kinematik und der Dynamik. Abgesehen von statischen Zugängen zur Mechanik werden Bewegungen klassischerweise in einer Dimension eingeführt, d. h. auf „Vor- und Rückwärtsbewegungen“ beschränkt. Alternative Konzepte (vgl. (Jung, Reul und Schwedes 1977)) betrachten Bewegungen von Beginn an als zweidimensional, was den Vektorcharakter, speziell der Geschwindigkeit, hervorhebt. Beide Herangehensweise haben eigene Vorzüge wie auch Schwächen. Bisherige Studien (z. B. Tobias (2010)) zeigen eine Überlegenheit der zweidimensional-dynamischen Zugänge gegenüber klassischer Konzepte auf.

In der vorliegenden Studie wird untersucht, inwiefern diese Zugänge im Laufe der Verbreitung von Tablets unterstützt durch Simulationen und Lernumgebungen mit automatischen Feedback in zwei Varianten (Textfeedback und Animationsfeedback) ihre Wirkung entfalten. In einer Feldstudie (N=275) wurde der Lernerfolg durch Bearbeitung einer Lernumgebung nach zweidimensionalem bzw. eindimensionalem Zugang mit quantitativen wie qualitativen Methoden ermittelt.

Dazu gehört etwa ein Fachtest, der mittels Raschanalyse ausgewertet wurde. Weiterhin wurden mit einzelnen Versuchspersonen Interviews geführt, in denen fachliche Aspekte thematisiert wurden, ebenso wie auch die Wahrnehmung der Arbeit mit der Lernumgebung und der Arbeitsform. Zu letzten Aspekten bearbeiteten alle Schüler Fragebögen. Begleitend wurde mittels Selbsteinschätzung auch die kognitive Belastung beim Bearbeiten der Simulationen erhoben.

Es zeigte sich dabei, dass die Lernumgebungen insgesamt lernwirksam sind, wobei im 2D-Zugang minimal höhere Effekte zu verzeichnen waren. Hinsichtlich der Feedbackvarianten zeigte das Feedback durch Animationen ebenso leicht größere Effekte. Die Lernumgebung als solche stieß überwiegend auf Akzeptanz und Interesse seitens der Schülerinnen und Schüler, wobei ähnliche Methoden z. T. häufiger gewünscht wurden. Die kognitive Belastung beim Bearbeiten der Simulationen wurde im Allgemeinen sehr niedrig bewertet, jedoch in der 2D-Lernumgebung als größer angesehen.

Schlagworte: Mechanik, Simulationen, Unterrichtskonzepte

Abstract

Mechanics lessons in secondary level 1 offer different didactic approaches for the first school contact with the kinematics and dynamics. Apart from static approaches to mechanics, movements are classically introduced in one dimension, i.e. limited to "forward and backward movements". Alternative concepts (see Jung, Reul and Schwedes (1977)) consider movements from the beginning as two-dimensional in order to emphasize the vector character, especially of the speed respectively velocity. Both approaches have their own advantages as well as weaknesses. Nevertheless previous studies (e.g. Tobias (2010)) show a superiority of the two-dimensional dynamic approaches against classical concepts.

In the present study it is examined to what extent these approaches in the course of tablet distribution supported by simulations and learning environments with automatic feedback in two variants (text feedback and animation feedback) can have their effect. In a field study (N=275), the learning success was assessed after working in a learning environment with two-dimensional or one-dimensional approach with quantitative and qualitative methods.

This includes, for example, a test, which is evaluated following Rasch analysis. Furthermore, interviews were conducted with participants in which technical aspects were discussed as well as the the perception of working with the learning environment and the form of work. On the last aspects all students worked on questionnaires. Accompanying self-assessment of the cognitive load during the processing of the simulations was evaluated.

It was shown that the learning environments as a whole were effective in learning with minimally higher effects in 2D access. Regarding the feedback variants, the animation feedback showed slightly larger effects. The learning environment as such met predominantly with acceptance and interest on the part of the students. Similar learning methods in the physic lessons were by all means requested more frequently. The cognitive load when editing the simulations was generally rated very low, but in the 2D learning environment viewed as larger.

Keywords: mechanics, simulations, education

Inhaltsverzeichnis

Abbildungsverzeichnis	iii
Tabellenverzeichnis	vii
I Überblick	1
1 Worum es geht	3
2 Abschnitte der Arbeit	5
II Theorie	7
3 Mechanik im Anfangsunterricht	9
3.1 Mechanik als Grundlage und Problemfeld	9
3.2 Sachanalyse	11
3.3 Alltagsvorstellungen zur Mechanik	15
3.4 Darlegung der Konzepte	24
3.5 Bestandsaufnahme: Was soll unterrichtet werden?	28
3.6 Empirische Befunde	34
3.7 Schlussfolgerungen für die empirische Studie	42
4 „Neue Medien“ im unterrichtlichen Einsatz	47
4.1 Von Computerraum und Whiteboard zum Tabletcomputer	47
4.2 Über Simulationen und Lernumgebungen	49
4.3 Feedback in neuen Medien	62
4.4 Der Cognitive Load in neuen Medien	70
III Materialentwicklung	77
5 Simulationen und Lernumgebung	79

5.1 Simulationen	79
5.2 Lernumgebung	81
5.3 Feedback	87
IV Methodik	95
6 Forschungsinteresse	97
6.1 Forschungsfragen	97
7 Studiendesgin	99
7.1 Anlage der Studie	99
8 Erhebungsinstrumente	105
8.1 Fachtests	105
8.2 Fragebögen	108
8.3 Fragebögen zum Cognitive Load	111
8.4 Interviewerhebung	112
8.5 Fragebogen für Lehrkräfte	117
9 Die Studien in der Unterrichtspraxis	119
9.1 Umsetzung der Studien	120
9.2 Beschreibung der Stichproben	125
10 Aufbereitung der Daten	129
10.1 Quantitative Daten	129
10.2 Qualitative Daten	130
10.3 Verwendete statistische Größen und Verfahren	134
V Ergebnisse	145
11 Ergebnisse der Vor- und Vergleichsstudien	147
11.1 Exploration	147
11.2 Vorstudie	157
12 Ergebnisse Hauptstudie	179
12.1 Qualitative Daten	179
12.2 Quantitative Daten	200
VI Zusammenfassung	237
13 Diskussion	239

13.1 Forschungsfragen	239
13.2 Kritik	244
13.3 Ausblick	246

VII Anhang 261

A Erhebungsinstrumente 263

A.1 Fragebögen und Tests	263
A.2 Tests	270
A.3 Interviewleitfaden	288
A.4 Ratingkategorien offene Frage	290

B Simulationen 293

C Weiteres Material 301

D Ergänzungen zu Ergebnissen 303

D.1 Codierung der fachlichen Interviewinhalte	305
---	-----

Abbildungsverzeichnis

3.1	Bahnkurve im dreidimensionalen Raum	11
3.2	Vergleich von zurückgelegtem Pfad (Schlangenlinie) und zugehörigem Ortsverschiebungsvektor $\Delta\vec{x}_{21}$	18
3.3	Vergleich derselben Bewegung, eingezeichnet im Zeit-Weg-Diagramm (links) und Zeit-Ort-Diagramm (rechts).	19
3.4	Race-Car-Problem	22
3.5	Darstellung einer „Graph-als-Bild“-Vorstellung: Der Graph stellt vermeintlich ein „Gebirge“ dar (Nach: Wiesner, Schecker und Hopf (2011)).	22
3.6	Kartesisches Koordinatensystem (A) vs. Koordinatenachse entlang des Weges (B). Die Punkte markieren äquidistante Teilstücke auf den jeweiligen Achsen. (Darstellung angeregt durch Schmit (2014).)	27
4.1	Verknüpfung mit unterschiedlichen Darstellungsformen (Darstellung angelehnt an: Phet-Simulations)	52
4.2	Feedback-Modell nach Hattie und Timperley (2007), eigene Übersetzung	67
5.1	Simulation „Mias Kart“	80
5.2	Beispielseite für thematischen Einstieg in Themenelement (Bild CC0 Lizenz, Quelle: Pixabay; z. T. unkenntlich gemacht)	85
5.3	Beispielseite für thematische Zusammenfassung eines Themenelements.	86
5.4	Beispielseite für Multiple-Choice-Fragen zur Feedbackgenerierung	87
5.5	Beispiel für Textfeedback nach falscher Antwort	91
5.6	Beispiel für Feedback nach korrekter Antwort	92
5.7	Beispiel für Animationsfeedback nach falscher Antwort	92
9.1	Darstellung des zeitlichen Ablaufs der Studie (1 DS = 1 Doppelstunde). Der zeitliche Maßstab dient nur der Orientierung. In der Umsetzung schwankt die Dauer je nach Lerngruppe und Zugang stärker.	120

10.1	Schematische Darstellung der Item Characteristic Curves dreier Aufgaben unterschiedlicher Schwierigkeit entsprechend des Rasch-Modells	138
10.2	Beispielhafte Person-Item-Map	139
11.1	Die Multiple-Choice-Aufgaben F1 und F2	149
11.2	Ergebnisse bzgl. Verständlichkeit	150
11.3	Ergebnisse bzgl. Bedienung	150
11.4	Ergebnisse bzgl. Umgang mit Simulationen	151
11.5	Ergebnisse bzgl. Umgang mit Feedback	151
11.6	Ergebnisse zu weiteren Aspekten	152
11.7	Ergebnisse der Fachwissensaufgaben F1 und F2	153
11.8	Kumulierte Häufigkeiten der gewählten Antworten zu zwei Beispiel MC-Aufgaben. Korrektes Item jeweils am häufigsten gewählt.	154
11.9	Kumulierte Häufigkeiten der gewählten Antworten zu zwei Beispiel MC-Aufgaben. Korrektes Item jeweils am häufigsten gewählt.	154
11.10	Mittelwerte aller Items aus beiden Fragebögen der Pilotierung	159
11.11	Nutzung digitaler Medien.	160
11.12	Vergleich der Skalen(mittel)werte zum Selbstkonzept	162
11.13	Vergleich der Skalen(mittel)werte zum Interesse	163
11.14	Vergleich der Skalen(mittel)werte zum Flow	164
11.15	Überblick über die weiteren Items im Vorfragebogen	165
11.16	Überblick über die weiteren Items im Nachfragebogen.	166
11.17	Überblick über die relativen Lösungshäufigkeiten im Vor- und Nachtest (oben bzw. unten)	172
11.18	Überblick der Personenfähigkeiten aller SuS	174
11.19	Darstellung der Personenfähigkeiten in der 1D- und 2D-Gruppe (links bzw. rechts)	174
11.20	Darstellung der Personenfähigkeiten von Mädchen und Jungen (links bzw. rechts)	175
11.21	Darstellung der Personenfähigkeiten je nach Versuchsbedingung	176
11.22	Darstellung der Personenfähigkeiten von Schülern unterschiedlichen Vorwissens	176
12.1	Häufigkeiten der Zuordnungen pro Kategorie	183
12.2	Häufigkeit der Zuordnungen pro Kategorie 1D	197
12.3	Häufigkeiten der Zuordnungen pro Kategorie 2D	199
12.4	Mittelwerte aller Items aus beiden Fragebögen	201
12.5	Einsatz von Computer, Tablet oder Smartphone im privaten Bereich	203

12.6	Einsatz von Computer, Tablet oder Smartphone im jeweiligen Unterricht / von Simulationen im Physikunterricht	204
12.7	Mittelwert und Zuwachs des Interesses	205
12.8	Mittelwert und Zuwachs des Selbstkonzepts	207
12.9	Mittelwert und Zuwachs des Flows	209
12.10	Ergebnis der Fragen zur selbst eingeschätzten Bedienkompetenz von Touchscreens und Spielen.	209
12.11	Ergebnis der Fragen zur Fehlerkultur seitens der Lehrkraft.	210
12.12	Ergebnisse der feedbackbezogenen Fragen	211
12.13	Ergebnis der Fragen zur selbst eingeschätzten Bedienkompetenz von Simulationen und Lernprogramm.	212
12.14	Ergebnis der Fragen nach dem Wunsch zum häufigeren Einsatz von Lernprogramm oder Simulationen	212
12.15	Ergebnis der Fragen nach dem empfundenen Lernerfolg	214
12.16	Ergebnis der Fragen nach der empfundenen Klarheit	214
12.17	Verteilung der für das Lernprogramm inkl. Simulationen vergebenen Schulnoten (Post, alle SuS).	215
12.18	Überblick über die relativen Lösungshäufigkeiten im Vor- und Nachtest	218
12.19	Überblick der Personenfähigkeiten aller Schülerinnen und Schüler.	220
12.20	Personenfähigkeiten der SuS, aufgeschlüsselt nach Zugängen	220
12.21	Personenfähigkeiten, nach Feedback aufgesplittet.	221
12.22	Personenfähigkeiten aller SuS, aufgesplittet nach Geschlecht	222
12.23	Personenfähigkeiten aller SuS, aufgesplittet nach Vorwissen .	223
12.24	Überblick der Personenfähigkeiten der SuS nach Jahrgangsstufe	224
12.25	Überblick der Personenfähigkeiten nach Versuchsbedingung aufgeteilt	225
12.26	Überblick der Personenfähigkeiten, unterteilt nach Geschlecht und Zugang.	226
12.27	Personenfähigkeiten der SuS in Vergleichs- und Hauptstudie.	231
12.28	Gegenüberstellung Vergleichsgruppe und Hauptstudie mit jeweiligem Zugang	231
A.1	Fiktive Schüleräußerungen zur Vorlage im Interview, zwei Versionen (Bilderquelle: Pixabay, gemeinfrei)	292
C.1	Exemplarischer Notizzettel, der von Schülerinnen und Schülern während der Lerneinheit an unterschiedlichen Stellen ausgefüllt werden sollte.	302
D.1	Relative Lösungshäufigkeit aller Items der Vergleichsgruppe	304

Tabellenverzeichnis

5.1	Kurzübersicht der behandelten Inhalte in beiden Lernumgebungen	81
8.1	Übersicht der Fragen für die Lehrkräfte	117
9.1	Versuchspersonen der Vorstudie	125
9.2	Noten der Schüler der Vorstudie	126
9.3	Übersicht der Schüler der Vergleichsgruppe	126
9.4	Übersicht der Versuchspersonen der Hauptstudie	127
9.5	Übersicht über Interviewpartner/innen	127
9.6	Übersicht über Physiklehrkräfte der Klassen	128
10.1	Übersicht der verwendeten Transkriptionsregeln	131
10.2	Übersicht der genutzten Effektstärken (Bühner und Ziegler 2009; Janczyk und Pfister 2013)	136
11.1	Übersicht zu den Skalen der Vorstudie. Hinweis: Beim Anteil fehlender Werte werden nicht-anwesende SuS nicht berücksichtigt (siehe 10.1.1).	161
11.2	Übersicht der freien Schülerkommentare der Vorstudie	168
12.1	Top 3 Begriffe	185
12.2	Übersicht zur Skala Interesse	204
12.3	Übersicht zur Skala Selbstkonzept	206
12.4	Übersicht zur Skala Flow	208
12.5	Nennung von relevanten Antwortoptionen bei Tempo-Aufgabe (Anteil gültiger Nennungen in %)	227
12.6	Mittelwerte der empfundenen kognitiven Belastung im 2D-Ansatz	234
12.7	Mittelwerte der empfundenen kognitiven Belastung im 1D-Ansatz	235
A.1	Übersicht der Fragen zum Interesse in Pre- und Postfragebogen der Hauptstudie	264
A.2	Übersicht der Fragen zum Selbstkonzept in Pre- und Postfragebogen der Hauptstudie	265
A.3	Übersicht der Fragen zum Flow in Pre- und Postfragebogen der Hauptstudie	266

Tabellenverzeichnis

A.4	Übersicht der Fragen zur Fehlerkultur in Pre- und Posttest	. 267
A.5	Übersicht der Fragen zur sonstigen Aspekten im Prefragebogen	268
A.6	Übersicht der Fragen zur sonstigen Aspekten im Posttest	. . 269

Teil I.

Überblick

1. Worum es geht

In dieser Studie werden zwei Ansätze des Mechanikunterrichts in Jahrgang 7 und 8 verglichen. Das Lernangebot an die teilnehmenden Schüler und Schülerinnen¹ der Hauptstudie bestand dabei aus einer digitalen Lernumgebung, welche auch Simulationen und Feedback beinhaltet. Für jeden Zugang wurde eine spezifische Lernumgebung entwickelt. Der fachliche Lernzuwachs sowie affektive Merkmale werden mittels Fachtests, Fragebögen und Interviews untersucht.

Ohne jeden Zweifel ist die Fachdidaktik der Physik um *guten* Physikunterricht bemüht. Die weithin bekannten Kriterien nach Duit und C. Wodzinski (2010) geben dazu grundlegende Einblicke. Jedes Kriterium, jedes der dort beschriebenen „Puzzlestücke“, hat das Potential für tiefere Betrachtungen. Beispielhaft sind Schülervorstellungen und Alltagserfahrungen in vielen Bereichen der Schulphysik seit vielen Jahren Forschungsgegenstand (vgl. Schecker, Wilhelm u. a. 2018; R. Müller 2004). Auch sinnvolle und vielfältige Experimentierarrangements waren und sind Bestandteil von didaktischer Forschung (vgl. Tesch 2005; Barth 2014). Das Puzzlestück „Ist fachlich konsistent und schlüssig“ (Duit und C. Wodzinski 2010) mag man als selbstverständlich und notwendige Grundvoraussetzung sehen – was auch in Ordnung ist. Doch mag dem Leser oder der Leserin dabei unter Umständen eher die Fachsystematik vor dem geistigen Auge erscheinen. Der eigentliche Clou dieses Puzzlestücks liegt für den Autor der vorliegenden Arbeit aber eher darin, dass dieses in anderer Lesart auch dazu auffordert unterschiedliche Zugänge zu einem Thema gegeneinander abzuwägen. Physikunterricht kann unter vielfältigen Blickwinkeln „fachlich schlüssig und konsistent“ sein! Im naturwissenschaftlichen Unterricht stehen Lehrkräfte häufig vor der Entscheidung für ein Thema einen bestimmten Zugang auswählen zu müssen. Nicht zuletzt ist es auch Aufgabe der Fachdidaktik unterschiedliche Zugänge empirisch zu erproben und zu bewerten. Dies ist beileibe keine neuartige Aufgabe; wohl schon immer wurden unterrichtliche Zugänge erprobt. Im Bereich der Mechanik betrifft das z. B. unterschiedliche Zugänge zur Mechanik im Anfangsunterricht (etwa Jahrgangsstufe 7/8). Dort herrschten lange Zeit

¹Die differenzierte Ansprache beider Geschlechter erfolgt explizit garantiert nur in der Thematisierung von Geschlechterunterschieden. Ansonsten wird zu Gunsten einer vernünftigen Lesbarkeit eine Variation von Schülern, Lernenden, Lehrkräften, Lehrern, Schülerinnen und Schülern usw. verwendet.

statische Zugänge zum Kraftbegriff vor, aktuell aber zumeist dynamische Zugänge über Bewegungen in einer Dimension („1D-Zugänge“). Bereits in den 1970er Jahren (Jung, Reul und Schwedes 1977) begannen im deutschsprachigen Raum Untersuchungen von alternativen zweidimensional-dynamischen Konzepten, die auch in der Zwischenzeit in der Fachdidaktik aufgegriffen wurden – ihren Weg in die Schule fanden sie trotz scheinbar überzeugender Argumente nur über größere Stolpersteine. Eine letzte Abwägung zwischen traditionellem Mechanikunterricht und eines zweidimensionalen Zugangs fand in noch jüngerer Vergangenheit in 2010 statt (Tobias 2010). Nicht nur aus Gründen der Replikation stellt sich ein Vergleich von traditionellem und zweidimensionalem Unterricht nach wie vor als interessant dar.

Neben Argumenten innerhalb der Mechanik kann dazu auch die Kombination mit einem weiteren Puzzlestück beitragen. Das Puzzlestück „Bietet Methoden- und Medienvielfalt, aber keine Beliebigkeit“ (Duit und C. Wodzinski 2010) scheint ebenso klar zu sein, wie das zuvor beschriebene. Im Zuge der rasanten Verbreitung sogenannter „Neuer Medien“ bekommt aber auch dieses Puzzlestück eine sehr aktuelle Komponente. Die rasende Verbreitung von Tablets und damit einhergehende Möglichkeiten² dringen in die Schule – Konzepte zur Einbettung sind aber nicht immer gegeben (vgl. Aufenanger 2017). Wie können Tablets bzw. die damit einhergehenden Potentiale (Simulationen, Lernsoftware, Sensoren, adaptive Software, Kameras jeweils für jeden Schüler und jede Schülerin uvm. . .) in den Unterricht eingebunden werden? Auch dieser Frage nachzugehen ist Aufgabe der Fachdidaktik. Dabei geht es idealerweise nicht um die Geräte als solche. Der Einsatz von Tablet, Computer oder Kreidetafel soll kein Selbstzweck sein. Sie sind nach Ansicht des Autors nicht per se lernförderlich, sondern werden es womöglich durch die gebotenen Möglichkeiten. Im Verlauf der vorliegenden Arbeit wird erörtert, inwiefern die Möglichkeiten Neuer Medien wie Tablets ihren Beitrag zum Mechanikunterricht liefern können.

Somit verbindet diese Arbeit zwei zentrale Komponenten: die Untersuchung zweier Zugänge zum Mechanikunterricht, kombiniert mit dem Potential Neuer Medien (hier sei vorweg genommen, dass dieses Potential in Form von digitalen Lernumgebungen mit Simulationen und Feedback genutzt wird). Im Folgenden sei ein Überblick über die Gliederung dieser Arbeit gegeben.

²So war der Autor dieser Arbeit noch 2010 der mittlerweile widerlegten Ansicht, dass sich Tablets nicht durchsetzen werden, da sie „nichts Halbes und nichts Ganzes“ seien . . .

2. Abschnitte der Arbeit

Die Arbeit besteht aus mehreren Teilen, die jeweils fortlaufend in einzelne Kapitel unterteilt sind. **Teil 1** gibt einen Überblick über die Inhalte dieser Arbeit und umreißt dort die Problemstellung.

Teil 2 umfasst die theoretischen Überlegungen der Arbeit. In Kapitel 3 werden die relevanten Sachinhalte dargestellt. Darauf folgt ein Überblick über die Schülervorstellungen, die in diesen Themengebieten zu Tage treten. Die beiden Unterrichtskonzepte werden didaktisch analysiert. Auf dieser Basis werden Curricula in Deutschland darauf untersucht, inwiefern 1D- oder 2D-Konzepte zwingend im Anfangsunterricht vorgesehen sind. Im weiteren Verlauf werden empirische Befunde zu den Konzepten vorgestellt. Diese erwachsen aus unterschiedlichen Studien der letzten Jahre und Jahrzehnte im deutschsprachigen Raum (u. a. Jung, Reul und Schwedes 1977; R. Wodzinski 1996; Tobias 2010). Am Ende dieses Abschnitts wird erörtert, inwiefern eine weitere Forschung überhaupt von Nöten ist, da eine Vielzahl von Erkenntnissen bereits vorliegt. In Kapitel 4 wird die zweite Säule dieser Arbeit, das Lernen mit neuen Medien, theoretisch aufgearbeitet. Zunächst werden dabei die Vorzüge von Simulationen für das Lernen von Mechanik dargelegt, im gleichen Zuge wird das entdeckende Lernen mit Simulationen beschrieben. Daran schließt sich die Beschreibung von Lernumgebungen wie auch des allgemeinen und fachspezifischen Tableteinsatzes an. Die in den bisherigen Teilkapiteln aufgeführten Potentiale, aber auch Probleme, leiten zum Kapitel 4.3 über. In diesem wird nochmals dessen Notwendigkeit thematisiert und anschließend das Wesen von Feedback und Merkmale effektiven Feedbacks erörtert. Die zwei verwendeten Feedbackvarianten werden anschließend theoretisch fundiert.

Teil 3 befasst sich mit der Entwicklung des Lernmaterials. Angelehnt an die Struktur von Kapitel 4 werden in Kapitel 5 die konkreten Materialien vorgestellt. Dies umfasst zunächst die Charakterisierung der Simulationen und die Grundsätze derer Erstellung. Im folgenden Kapitel wird die konkrete Lernumgebung beschrieben, unterteilt in strukturelle sowie inhaltliche/didaktische Aspekte. Am Ende dieses Abschnitts werden Beispielseiten konkret vorgestellt. Das letzte Kapitel des Entwicklungsteils gibt einen Überblick über die Ausgestaltung des verwendeten Feedbacks.

Dazu gehört auch die Darstellung der Multiple-Choice-Fragen, die zur Generierung des Feedbacks dienen.

In **Teil 4** wird das methodische Vorgehen der Studie beschrieben, wobei zunächst auf dem bisherigen basierend das Forschungsinteresse und die Forschungsfragen dargestellt werden. Dies umfasst in Kapitel 7 beispielsweise die Erfordernisse und Charakteristika von Vergleichsstudien, wie auch die Abwägung zwischen Labor- und Feldstudie. Auch das Design der Studien (Vor-, Begleit- und Hauptstudien) ist hier beschrieben. In Kapitel 8 wird die Entwicklung der Erhebungsinstrumente wie Fragebögen und Tests aller Teilstudien erläutert. Dazu gehört auch eine Beschreibung der jeweils erhobenen Konstrukte. Ebenfalls ist hier der Interviewleitfaden im Vergleich zu anderen Interviewstudien beschrieben. In Kapitel 9 erfolgt eine Beschreibung der unterrichtlichen Umsetzung, was naturgemäß auch Schwierigkeiten bei der Umsetzung umfasst. Den Abschluss dieses Kapitels bildet die Beschreibung der Stichproben. Das letzte Kapitel widmet sich der Datenaufbereitung. Hier ist der Umgang mit den Rohdaten beschrieben, ebenso wie die grundlegenden statistischen Verfahren. Ein eigener Abschnitt widmet sich dabei der Rasch-Analyse als zentralem Instrument der Auswertung des Fachtests.

In **Teil 5** werden letztlich die Ergebnisse der Studie vorgestellt. Dieser Teil ist in die Kapitel 11 und 12 unterteilt, die die Ergebnisse der Vor- und Kontrollstudien sowie der Hauptstudie darlegen. Zu den Vorstudien gehört eine Explorationsuntersuchung mit einer Klasse, eine Vorstudie mit 2 Klassen sowie eine Kontrollgruppe mit 5 Klassen. In der Hauptstudie wurden 11 Klassen mit insgesamt $N=275$ Schülern untersucht. Die jeweilige Ergebnisdarstellung erfolgt untergliedert in quantitative und qualitative Daten sowie einem einleitenden Teil zur Qualitätssicherung. Zu den quantitativen Daten gehören dabei Daten aus Fragebögen, die die Zustimmung zu bestimmten Aussagen messen. Ebenso gehören dazu codierte Antworten aus fachwissenschaftlichen Tests (hier liegen jeweils nur Schülerdaten vor). Zu den qualitativen Daten gehören Daten aus offenen Antwortformaten (Schüler- und Lehrerdaten) wie auch Interviewmaterial (nur Schülerdaten). Neben der reinen Ergebnisdarstellung sind entsprechende Diskussionsteile an passender Stelle eingefügt.

Teil 6 fasst die Zielsetzung der Studie knapp zusammen und stellt die Ergebnisse unter Bezug auf die Theorie dar. Dieser Teil untergliedert sich inhaltlich in die Ergebnisdarstellung im Bezug auf die Forschungsfragen, Kritik wie einen Ausblick auf die Fortsetzung der Forschungsarbeit.

Teil II.

Theorie

3. Mechanik im Anfangsunterricht

3.1. Mechanik als Grundlage und Problemfeld

Mechanik gilt als „ältestes Teilgebiet der Physik“ (Scheck 1990; Gross u. a. 2016). Zahllose Anwendungen moderner Ingenieurskunst oder Naturwissenschaft beruhen auf Erkenntnissen der Mechanik. Entsprechend ist sie ein klassischer Grundpfeiler nicht nur der Hochschulphysik sondern ebenso der Schulphysik. Als erster Zugang zum Thema soll im Folgenden dazu die fachwissenschaftliche und fachdidaktische Sichtweise eingenommen werden.

FACHWISSENSCHAFTLICHE SICHTWEISE

Die Auffassung der Mechanik als eine der Grundlagen der Physik sei durch die folgenden Ausführungen illustriert. Die Mechanik ermöglicht das Beschreiben und Erklären des (Bewegungs)verhaltens von punktförmigen Massen oder ausgedehnten Körpern auch unter Einfluss von Kräften. Dazu gehören gerade und krummlinige – und damit beschleunigte – Bewegungen (vgl. Meschede 2015; Demtröder 2018). Weiterhin bildet sie die Grundlage für die Betrachtung von Schwingungen und Wellen und damit auch der Akustik. Auch Bewegungen von Teilchen in elektrischen oder magnetischen Feldern lassen sich auf mechanische Betrachtungen zurückführen. Astronomische Fragen wie auch der Übergang von klassischer Mechanik zu spezieller und allgemeiner Relativitätstheorie bieten nur einen kleinen Ausblick auf die Relevanz der Mechanik in den modernen Naturwissenschaften. Scheck (1990) geht sogar so weit, die Mechanik als Grundlage für die komplette theoretische Physik zu bezeichnen. Im Bereich der Ingenieursfächer tritt die Mechanik unter Anderem als technische Mechanik im Maschinenbau oder Bauingenieurwesen auf, wo es beispielsweise darum geht „Brücken, Kräne, Gebäude, Maschinen, Fahrzeuge oder Komponenten von Mikrosystemen statisch und dynamisch so zu analysieren, dass sie bestimmte Belastungen ertragen oder bestimmte Bewegungen ausführen können“ (Gross u. a. 2016). Dieses kurze Schlaglicht kann die Bedeutung der Mechanik selbstverständlich nicht komplett

darstellen. Es lässt aber die besondere Bedeutung dieses Teilgebiets der Physik und den Charakter als *Grundlage* erahnen.

FACHDIDAKTISCHE SICHTWEISE

Ob der besonderen Stellung der Mechanik und trotz der großen Lernschwierigkeiten sind unterschiedliche Themengebiete der Mechanik in sämtlichen Curricula der einzelnen Ländern verankert. Im Anfangsunterricht (in der Regel Jahrgangsstufe 7 oder 8) sind dabei klassischerweise das Beschreiben von Bewegungen sowie das dynamische und – teils untergeordnet – statische Wirken von Kräften Thema (Näheres in 3.5). Nichtsdestotrotz gilt Mechanik weithin als schwierig zu lernen, was sich im geringen Erfolg des Mechanikunterrichts widerspiegelt (Tobias 2010; Wilhelm 2005). Alltagsvorstellungen tragen dazu ihren Teil bei (Duit 2004b), da „bei den Schülern bereits vorhandene Denk- und Wissensstrukturen aktiv in die Informationsaufnahme und -verarbeitung eingreifen“ (Wiesner 2004), (siehe auch Kapitel 3.3). Alltagsvorstellungen aber sind – wie so oft – im täglichen Leben bei oberflächlicher Betrachtung mitunter plausibel (Duit 2004a). Denn auch in der Realität wird für eine Fahrradfahrt in der Ebene eine mehr oder weniger konstante Kraft benötigt. Da aber der Grund dafür – die zu kompensierenden Einflüsse aus Reibung und Luftwiderstand – im Allgemeinen nicht als Kräfte angesehen werden, entsteht auch so keine Vorstellung von einer kräftefreien, gleichförmigen Bewegung. Auch jene in diesem Beispiel skizzierte komplexe, nicht-intuitive Zugänglichkeit von Randbedingungen (der Reibung) erschwert das Lernen. Diese erforderliche „erhebliche Abstraktion von der alltäglichen Erfahrung“ (Tobias 2010) gehört zu den sachbezogenen Lernschwierigkeiten (Jung, Reul und Schwedes 1977). Weitere Kategorien sind, illustriert durch Tobias (2010), die innen- und außenbedingten Lernschwierigkeiten. Zu ersten gehören dabei „Interesse und Motivation“ wie auch konkrete Alltagsvorstellungen. Zu letzteren zum Beispiel ein „unangemessenes Unterrichtsangebot“, wozu nach Tobias (ebd.) u. a. eine Überbetonung statischer Aspekte gehöre.

Des Weiteren macht auch die übliche Belastung des Physikunterrichts durch seine relativ große Unbeliebtheit, die sich mit zunehmendem Alter und mitunter auch geschlechtsspezifisch nur noch weiter manifestiert und verschärft, vor dem Mechanikunterricht nicht halt. Diese wohlbekannte Problematik soll aber an dieser Stelle nicht weiter ausgeführt werden (siehe hierzu zum Beispiel Zwiorek (2006), Hoffmann, Häußler und Lehrke (1998) oder Wiesner, Schecker und Hopf (2011)).

In diesem Abschnitt wurde die grundlegende Bedeutung der Mechanik für Naturwissenschaft und Technik dargelegt. Weiterhin wurde sie als komplexer Fachinhalt charakterisiert, der Bestandteil sämtlicher Curricula ist. Für die Bereitstellung der fachlichen Inhalte dieser Arbeit wird

im folgenden Abschnitt eine Sachanalyse vorgenommen. Diese ist durch die typischen Inhalte des Anfangsunterrichts bzw. dieser Studie begrenzt. Daran schließen sich Beschreibungen von Konzepten an, die den Umgang mit diesen Fachinhalten ermöglichen sollen.

3.2. Sachanalyse

In diesem Abschnitt werden die für den Anfangsunterricht bzw. die vorliegende Studie relevanten Sachgebiete dargestellt. Das bedeutet auch, dass Aspekte wie Trägheit oder das dritte Newtonsche Axiom nicht thematisiert werden. Der Abschnitt gliedert sich in kinematische und dynamische Aspekte.

KINEMATIK

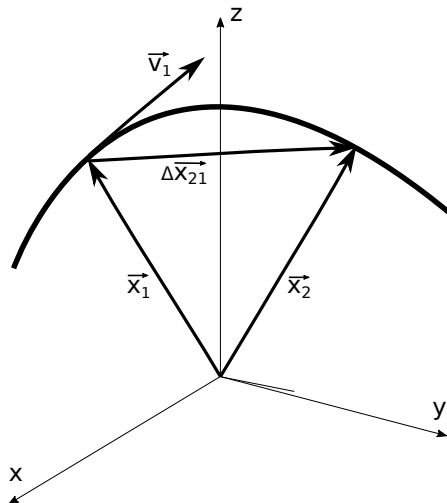


Abbildung 3.1.: Bahnkurve im dreidimensionalen Raum

Die Lage eines Punktes im dreidimensionalen Raum kann durch den Vektor \vec{x} beschrieben werden, dessen Komponenten die Koordinaten entlang der drei Achsen x, y und z eines Orthogonalsystems darstellen (Abb. 3.1). Die Wahl dieses Bezugssystems ist dabei willkürlich, ein absolutes,

ausgezeichnetes Bezugssystem existiert nicht. Die Lage des Punktes im Raum in Bezug auf das gewählte System bezeichnet man auch als *Ort* \vec{x}

$$\vec{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Die zeitliche Entwicklung der Koordinaten des Punktes ist entsprechend durch die Ortsfunktion (Demtröder 2018)

$$x\vec{x}(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}$$

gegeben. Die Spur der Bewegung des Punktes im Raum wird als Bahnkurve bezeichnet.

Die momentane Geschwindigkeit lässt sich anschaulich als infinitesimale Annäherung des Ortes x_2 an x_1 auffassen (vgl. Abb. 3.1). Der Differenzenquotient der durchschnittlichen Geschwindigkeit¹ (s. u.) wird so zum entsprechenden Differentialquotienten (ebd.).

$$\vec{v}(t) = \dot{\vec{x}}(t) = \lim_{\Delta t \rightarrow 0} \frac{\vec{x}(t + \Delta t) - \vec{x}(t)}{\Delta t}$$

Es wird deutlich, dass es sich bei der Geschwindigkeit um die zeitliche Ableitung der Ortsfunktion handelt. Somit lässt sich die Momentangeschwindigkeit zu einem bestimmten Zeitpunkt angeben. In Abbildung 3.1 ist diese o. B. d. A. beispielhaft durch den Vektor \vec{v}_1 dargestellt.

Die Differenz $\Delta\vec{x}_{21}$ zweier Ortsvektoren \vec{x}_2 und \vec{x}_1 wird als *Ortsverschiebung* bezeichnet

$$\Delta\vec{x}_{21} = \vec{x}_2 - \vec{x}_1$$

Diese muss nicht dem tatsächlich zurückgelegten Weg zwischen den Punkten entsprechen, der (wesentlich) länger sein kann als die Ortsverschiebung (siehe Abb. 3.1).

Als durchschnittliche Geschwindigkeit \hat{v} im Intervall Δt wird

$$\hat{v} = \frac{\Delta\vec{x}_{21}}{\Delta t}$$

bezeichnet. Dabei ist Δt die Differenz der Zeiten des Punktes am Ort \vec{x}_1 und \vec{x}_2 , also ein Zeitintervall.

Geläufiger ist hingegen die Betrachtung der Durchschnittsgeschwindigkeit als gemittelter Betrag der Geschwindigkeit entlang des Weges (der

¹o. B. d. A. mit $\vec{x}_2 = \vec{x}(t + \Delta t)$ und $\vec{x}_1 = \vec{x}(t)$

auch nicht gleich dem Betrag der oben beschriebenen Durchschnittsgeschwindigkeit sein muss, da wie beschrieben der tatsächlich zurückgelegte Weg nicht der Ortsverschiebung entsprechen muss).

DYNAMIK

Um Bewegungsänderungen von Körpern, d. h. die Veränderung der Geschwindigkeit in Richtung oder Betrag, herbeizuführen, bedarf es der Wechselwirkung zwischen mindestens zwei Körpern. Diese Wechselwirkungen werden mit dem Konzept der *Kraft* beschrieben. Während die Kinematik sich der Beschreibung von Bewegungen widmet, ist es also Aufgabe der Dynamik, diesen Einfluss von Kräften auf die Bewegung von Körpern zu beschreiben und zu erklären. Dass die Kraft auch eine Vektorgröße sein muss, folgt aus der Einsicht, dass sie den Geschwindigkeitsvektor in Betrag *und* Richtung verändern kann (Demtröder 2018; Tipler, Mosca und Wagner 2015). Das Formelzeichen der Kraft ist dabei \vec{F} und ihre Einheit ist *Newton* (N). Definitionsgemäß wirkt eine Kraft von 1 N, wenn ein Körper einer Masse von 1 kg eine Beschleunigung² von $1 \frac{\text{m}}{\text{s}^2}$ erfährt. Die Tatsache, dass „keine Proportionalitätskonstante auftritt, ist der Wahl der Einheiten zu verdanken: Man wählt die Einheitskraft so, dass sie der Einheitsmasse die Einheitsbeschleunigung mitteilt.“ (Meschede 2015).

Aus der oben beschriebenen Annahme folgt automatisch, dass sich die Bewegung eines Körpers, der nicht unter dem Einfluss einer Kraft steht, nicht verändert. Dabei ist unerheblich, ob tatsächlich keine Kraft auf ihn wirkt³ oder ob sich die angreifenden Kräfte derart kompensieren, dass die resultierende Kraft gleich Null ist (Demtröder 2018). Diese Überlegung spiegelt das erste Newtonsche Axiom wider:

1. Newtonsches Axiom

Ein Körper bewegt sich ohne den Einfluss von Kräften gleichförmig-geradlinig oder verharrt in Ruhe.

Die Grundannahme *Geschwindigkeitsänderung erfordert Kraft* wird im zweiten Newtonschen Axiom konkretisiert, in welchem der Zusammenhang zwischen Kraft und Impulsveränderung⁴ in einen mathematischen Zusammenhang gestellt wird.

²Beschleunigung als Geschwindigkeitsänderung pro dafür benötigter Zeit. Im Beispiel ändere sich die Geschwindigkeit (bzw. deren Betrag) in 1s als um 1m/s

³Dieser idealisierte Fall kommt nicht vor, eher sind die Wechselwirkungen so gering, dass sie vernachlässigbar sind (Demtröder 2018).

⁴Zum Impuls siehe 2. Axiom.

2. Newtonsches Axiom

Die Änderung des Impulses ist der Einwirkung der bewegenden Kraft proportional und geschieht in Richtung der Kraft. ^a

^aNolting (2013)

Die Bewegungsgröße, der Impuls \vec{p} ist dabei definiert als $\vec{p} = m \cdot \vec{v}$. Daher ergibt sich $\vec{F} = \dot{\vec{p}} = \frac{d}{dt}(m \cdot \vec{v})$. Daraus folgt unmittelbar, dass die Richtung der Kraft gleich der Richtung der Impulsänderung ist. Die gängige Formel $\vec{F} = m \cdot \vec{a}$ wurde in dieser Form aber erst von Euler propagiert. Dabei ist sie nicht mehr derartig allgemein wie die Newtonsche Form, da die Veränderung der Masse nicht mehr als Grund für eine Bewegungsveränderung erkannt werden kann – sie gilt also *nur* für den Fall $m = const.$. Nicht nur in der Schulphysik stellt sie aber die gängige und bekannteste Form dar. Die Newton'sche Originalfassung, die durch Jung, Reul und Schwedes (1977) mit „Stoß = $\Delta(m\vec{v})$ “ übersetzt wurde, weicht von der heute verwendeten Form ab, da sie sich nur auf den Vergleich zweier Zustände bezieht.

IM WANDEL: KRAFTBEGRIFF UND GRUNDGLEICHUNG DER MECHANIK
Anhand der folgenden Betrachtungen lässt sich u. a. der historische Weg wie auch unterschiedlich nutzbare Erscheinungsformen der Grundgleichung der Mechanik nachvollziehen:

$$\overrightarrow{Sto\beta} = \Delta(m\vec{v}) \quad (3.1)$$

$$\frac{\overrightarrow{Sto\beta}}{dt} = \frac{\Delta(m\vec{v})}{dt}$$

$$\vec{F} = \frac{d\vec{p}}{dt} = \dot{\vec{p}} \quad (3.2)$$

$$\int_{t_1}^{t_2} \vec{F} dt = \int_{t_1}^{t_2} \frac{d\vec{p}}{dt} \cdot dt$$

$$\int_{t_1}^{t_2} \vec{F} dt = \int_{\vec{p}_1}^{\vec{p}_2} 1 \cdot d\vec{p}$$

$$\int_{t_1}^{t_2} \vec{F} dt = [\vec{p}]_{\vec{p}_1}^{\vec{p}_2}$$

$$\int_{t_1}^{t_2} \vec{F} dt = \vec{p}_2 - \vec{p}_1$$

$$\int_{t_1}^{t_2} \vec{F} dt = \Delta\vec{p} \quad (3.3)$$

Gleichung (3.1) kommt nach Jung (1980) der Newtonschen Originalformulierung am nächsten. Der folgende Schritt, die Betrachtung der zeitlichen Änderung, zeigt das Verständnis von Kraft als Stoßrate. Gleichung (3.2) repräsentiert die auf universitärer Ebene verwendete Form Grundgleichung. Das Integral $\int \vec{F} dt$ (3.3) wird als *Kraftstoß* bezeichnet (H. Hammer und K. Hammer 1995). Bei Betrachtung der konstanten (oder durchschnittlich wirkenden Kraft) ergibt sich die anschauliche Vorstellung der *über eine Zeitdauer wirkenden Kraft*.

Bei konstanter (oder Betrachtung der durchschnittlichen) Kraft und konstanter Masse vereinfacht sich die Gleichung zu:

$$\begin{aligned}\vec{F} \cdot \Delta t &= \Delta \vec{p} \\ \vec{F} \cdot \Delta t &= m \cdot \Delta \vec{v}\end{aligned}$$

Das Δt lässt sich in dieser Form als *Dauer der Geschwindigkeitsänderung* interpretieren. Mit dem Wissen, dass sich Geschwindigkeiten nur ändern sofern Kräfte einwirken, lässt sich daraus auch die Sprechweise als *Einwirkungsdauer* herleiten. Inhaltlich sind beide Sprechweisen äquivalent. Das \vec{F} steht bekanntermaßen für die Kraft, m für die (träge) Masse sowie $\Delta \vec{v}$ für die Zusatzgeschwindigkeit, also die Differenz der Geschwindigkeit zwischen Anfangs- und Endzustand.

Die Verwendung dieser integralen Form bedingt keinesfalls eine zweidimensionale Mechanik. Viele gängige Anwendungsaufgaben lassen sich auch bei der Reduktion auf skalare Größen durchführen.

3.3. Alltagsvorstellungen zur Mechanik

In diesem Abschnitt sollen ergänzend zur Darstellung der Sache die gängigen Alltagsvorstellungen⁵ zur Mechanik dargelegt werden. Dabei handelt es sich um keine vollständige Aufzählung, sondern um eine aspekt-hafte Darstellung von für diese Arbeit relevanten Vorstellungen.

Bewegung Was bedeutet es, in die gleiche Richtung zu fahren? Die umgangssprachliche Belegung des Ausdrucks „in die gleiche Richtung“ fahren wird eher als „zum gleichen Ziel fahren“ oder „zum selben Punkt zeigen“ verstanden (Wiesner 2011). Im physikalischen Sinne aber hat ein Geschwindigkeitsvektor genau dann die gleiche Richtung, wenn er

⁵Bewusst wird der Begriff „Fehlvorstellung“ oder „Schülvorstellung“ gemieden. Denkt man beispielsweise an den Geschwindigkeitsbegriff, so ist die Belegung des Begriffs im *Alltag* der Schüler (aber eben auch der Erwachsenen und Physiker...) anders als in der Physik, aber eben nicht grundsätzlich *falsch*. Vgl. hierzu auch Jung, Reul und Schwedes (1977).

ein Ein- oder Vielfaches des anderen darstellt – die Komponenten also bis auf jeweils ein Vielfaches übereinstimmen. Der alltagssprachliche Gebrauch zielt also mehr auf die Bewegung zum gleichen Ziel ab – was physikalisch nicht der Bewegung in die gleiche Richtung entspricht. Bezüglich des Beschreibens von Bewegungen nennt R. Wodzinski (2004) die „Ganzheitlichkeit“, mit der Schülerinnen und Schüler Bewegungen beschreiben. So seien diese weniger auf die zeitliche Veränderung der momentanen Bewegung fixiert, sondern auf das Gesamtbild. Daher kann eine Bewegung als „schlangenlinienförmig“, „im Kreis“ oder „ruckartig“ beschrieben werden, was der eher sequentiellen Beschreibung physikalischer Natur eher entgegensteht. Näheres zu Vorstellungen zu Bewegungen an sich und zur Rolle von Bezugssystemen z. B. bei Jung (1980).

Geschwindigkeitsbegriff Wie bereits in Kapitel 3.2 dargelegt, ist die Geschwindigkeit die Ableitung des Ortsverschiebungsvektors nach der Zeit. Entsprechend ist also auch die Geschwindigkeit ein Vektor. Betrag und Richtung sind anschaulich als „Schnelligkeit“ oder „Tempo“ und Bewegungsrichtung zu deuten. In der Alltagssprache ist der Begriff *Geschwindigkeit* jedoch anders belegt, und zwar insofern, als dass der Richtungscharakter der physikalischen Auslegung nicht berücksichtigt wird. So antworteten beispielsweise in der Untersuchung von Tobias (2010) im Vortest nur etwa 14% aller Befragten Schüler korrekt auf eine Frage, die dieses physikalische Verständnis des Vektorcharakters überprüft. Weiter würde im Alltag zwei Fahrzeugen, die *gleich schnell* aber z. B. in entgegengesetzte Richtung oder hintereinander durch eine enge Kurve fahren, eine gleiche Geschwindigkeit attestiert werden. Auch das synonyme Verwenden der Begriffe Höchstgeschwindigkeit und Tempolimit zeigt einmal mehr die – nicht überraschende und im Alltag tragfähige – gleiche Belegung der Begriffe (Jung 1980).

Anzumerken ist weiterhin, dass in der deutschen Sprache eine Unterscheidung der Betragsgröße und der Vektorgröße nicht scharf vorgenommen werden kann. Zwar lassen sich für die Betragsgröße Begriffe wie Schnelligkeit oder Tempo definieren – diese werden aber vor jeglicher Definition synonym zum Geschwindigkeitsbegriff genutzt. Andere Sprachen erlauben diese Unterscheidung eher. Als Beispiel wird im Englischen *velocity* (=Geschwindigkeit) und *speed* (= etwa Betrag der Geschwindigkeit) genutzt, oder im Französischen entsprechend *vélocité* und *vitesse*. Wie bereits erwähnt ist bei aller physikalischen Korrektheit zu beachten, dass die Physik eben nur *ein* Normensystem (ebd.) darstellt, der Alltag stellt ein *weiteres* dar. Wo nun Fachleute wissen, welches System wann wie anzuwenden ist (vgl. ebd.), sollte es das Ziel sein, Schüler für die Unterschiede in der Bedeutung zu sensibilisieren und nicht die physikali-

sche Sichtweise als die einzig richtige darzustellen. Auf diese Weise wäre dies nicht mehr als das „Ersetzen eines Dogmas durch ein anderes.“ (ebd.). Die beschriebenen Alltagsvorstellungen zum Geschwindigkeitsbegriff sind innerphysikalisch nicht vollständig korrekt, wären für sich alleine betrachtet nach Einschätzung des Autors jedoch nicht schwerwiegend. Die Auswirkungen der Alltagsvorstellung resultiert aber in Problemen bei dynamischen Betrachtungen. Dazu gehört als klassisches Beispiel die gleichförmige Kreisbewegung, bei der sich die Geschwindigkeit im physikalischen Sinne laufend ändert, das Tempo aber nicht. Daher ist es für Schüler schwer zu begreifen, dass eine Veränderung der Geschwindigkeit stattfindet und somit eine Kraft nötig ist (Wilhelm 2005; Tobias 2010).

Schmit (2014) merkt zusätzlich an, dass sprachlich beim Geschwindigkeitsbegriff nur selten zwischen Momentan-, Durchschnitts- oder mittlerer Geschwindigkeit unterschieden wird, es sei „i. d. R. allgemein von Schnelligkeit oder Geschwindigkeit die Rede“. Zusätzlich erwähnt er, dass sich die Geschwindigkeit immer auf ein Bezugssystem bezieht und somit „keine absolute Größe darstellt“ – das Bezugssystem aber i. d. R. keine Erwähnung findet.

Ort und Weg Dieser Abschnitt bezieht sich weniger auf eine wohlbekannte Alltagsvorstellung, sondern mehr auf eine didaktische Hürde, die durch den unscharfen Gebrauch dieser Begriffe einhergeht und somit Lernschwierigkeiten auslösen kann. Amenda und Schecker (2014) beklagen in einem Schlaglicht-Artikel die uneinheitliche und darüber hinaus problematische und unintuitive Verwendung der Begriffe Ort, Ortsverschiebung, Weg und Wegstrecke. Der *Weg* bezeichnet fachphysikalisch die Länge der Bahnkurve eines Körpers, während *Ort* die Position des Körpers bezüglich eines Koordinatensystems festlegt. Die Begriffe Ort und Weg werden inkonsistent genutzt und darüber hinaus auch noch häufig mit demselben Formelzeichen s versehen. Als prototypische Beispiele für das Dilemma werden von Amenda und Schecker (ebd.) folgende Zitate angeführt:

- „Der Wurf ist beendet, wenn der Körper wieder den Boden erreicht: Der Weg s ist null.“
- „Befindet sich der Körper zur Zeit t_1 am Ort $s_1 = s(t_1)$ und bewegt er sich von s_1 zu einem zweiten Ort $s_2 = s(t_2)$, dann legt er den Weg $\Delta s = s(t_2) - s(t_1)$ zurück [...]“ (jeweils ebd.).

Anhand der obigen Definition von Weg und Ort wird die Kuriosität des ersten Zitats deutlich. Dieses sagt aus, dass der Weg nach der Beendigung einer Bewegung wieder „0“ werden kann, was aber im Widerspruch zur Definition steht. Nach dieser kann der Weg nur zum (willkürlichen) Beginn

der Bewegung/der Messung „0“ betragen, danach nur noch ansteigen oder gleich bleiben.

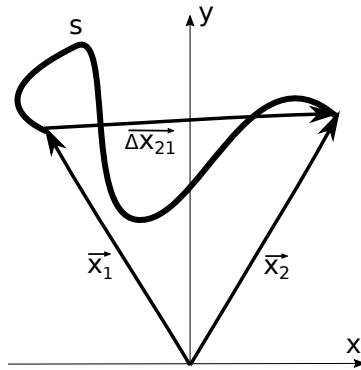


Abbildung 3.2.: Vergleich von zurückgelegtem Pfad (Schlangenlinie) und zugehörigem Ortsverschiebungsvektor $\Delta \vec{x}_{21}$.

Die Problematik des zweiten Zitats lässt sich an folgendem Beispiel erkennen. Angenommen ein Körper bewegt sich entlang des in Abb. 3.2 eingezeichneten, krummlinigen Wegs. Es ist einleuchtend, dass der Körper letztlich eine andere bzw. längere Strecke zurückgelegt haben muss, als wenn er der kürzest möglichen Verbindung von Start- und Endpunkt – dem Vektor der Ortsverschiebung – gefolgt wäre. In diesem Fall sind also Weg und Ortsverschiebung nicht identisch (siehe auch Kapitel 3.2). Die unmittelbare Bedeutung für den Betrag der Durchschnittsgeschwindigkeit ergibt sich wie folgt: In diesem Beispiel ergäbe die Berechnung nach $|\vec{v}| = \frac{|\Delta \vec{x}|}{\Delta t}$ nicht den selben Wert wie $v = \frac{\Delta s}{\Delta t}$, sofern Δs als Wegelement betrachtet wird. Die einzige Äquivalenz dieser beiden Berechnungsweisen ergibt sich⁶ für Momentangeschwindigkeiten – d. h. für den Fall dass sich die der Ortsverschiebung zugrundeliegende Zeitdifferenz dem Wert 0 s annähert. Somit wären dann Wegelement und Ortsverschiebung näherungsweise identisch. Aus diesen Überlegungen folgt ebenso, dass auch das Durchschnittstempo nicht einfach nur der Betrag der Durchschnittsgeschwindigkeit ist.

Die angemahnte Vermischung von Ort und Weg ist zunächst nicht augenfällig, wenn hauptsächlich Betrachtungen in einer Dimension angestellt werden. In diesem Fall sind die Unterschiede tatsächlich in den meisten Fällen nicht entscheidend. Denn bewegt sich ein Körper in einer Dimension vom Ursprung weg, so entspricht die Ortsverschiebung exakt

⁶außerhalb der eindimensionalen Betrachtungen

dem Weg (allenfalls der Betrag des Ortes). Problematisch wird erst die Fahrt in die umgekehrte Richtung, wenn der Weg naturgemäß weiterhin zunimmt, der Ort bzw. dessen Koordinaten sich in Richtung Ursprung wieder verkleinern (und somit der Ortsverschiebungsvektor sich je nach Bewegungsrichtung im Vorzeichen unterscheidet). Als Vorteil lässt sich dies interpretieren, als dass bei Verwendung des Ortes die Bewegungsrichtung unterscheiden lässt (vgl. Amenda und Schecker 2014).

In diesem Zusammenhang soll auch nochmals der Unterschied zwischen Zeit-Ort- und Zeit-Weg-Diagrammen betont werden. Entsprechend der vorherigen Ausführungen ist der Graph im Zeit-Weg-Diagramm monoton steigend, während dieser im Zeit-Ort-Diagramm sowohl steigen als auch fallen kann und so auch Informationen über die Bewegungsrichtung enthalten sind. In diesem Sinne wäre ein Weg-Zeit-Diagramm auch für eindimensionale Bewegungen nicht eindeutig, sofern Richtungswechsel vorgenommen werden (Beispiel siehe Abb. 3.3).

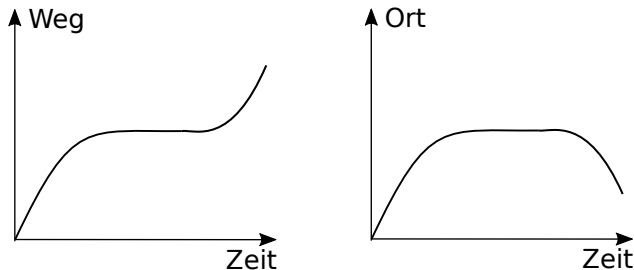


Abbildung 3.3.: Vergleich derselben Bewegung, eingezeichnet im Zeit-Weg-Diagramm (links) und Zeit-Ort-Diagramm (rechts).

Entsprechend wurde von einer Expertengruppe aus Fachdidaktikern und Lehrkräften im Rahmen einer GDCP-Schwerpunkttagung empfohlen, den Begriff *Weg* zugunsten des Begriffs *Ort* nicht weiter zu verwenden.

Kraftbegriff In diesem Abschnitt wird eine Auswahl an vielfältigen Alltagsvorstellungen zum Kraftbegriff dargestellt. Diese orientiert sich am Bedarf dieser Arbeit.

“Der Kran hat ganz schön viel Kraft!“ oder “Mein Papa ist sehr kräftig!“: Diese im alltäglichen Kontext gebräuchlichen und dort nicht falschen Aussagen zeigen, durch die physikalische Brille betrachtet, exemplarisch jedoch zwei Problematiken. Zum einen zeigt sich die häufig verankerte Vorstellung, dass nur aktive Körper (Menschen, Tiere, Maschinen etc.) Kräfte ausüben können. (Jung 1980; Jung, Reul und Schwedes 1977; R. Wodzinski 2004). Dass auch sogn. passive Körper (z. B. Straßenbelag,

ruhende Felsbrocken, Baumstämme) Kräfte ausüben⁷ können, kommt für Schüler dabei nicht zwingend in Frage⁸ (Wiesner 2004). Zum anderen wird nahegelegt, dass Kraft etwas ist, dass man *haben* kann. Kraft wird also „als Eigenschaft eines Körpers oder Systems und nicht als Wechselwirkungsgröße angesehen“ (ebd.). Beide Aspekte dieser Vorstellung lassen sich mitunter recht einfach verknüpfen, denn nur wer Kraft hat, kann auch welche ausüben (vgl. R. Wodzinski 2004).

Die Verbindung von Bewegungen und Kräften aus Schülersicht soll zunächst scheinbar paradoxerweise von der Statik ausgehend erläutert werden. Statik handelt von ruhenden Körpern. Typische Aufgaben („Laternen an Stahlseilen in Straßenmitte“⁹) legen die Vorstellung nahe, dass bei herrschendem Kräftegleichgewicht mit $F_{\text{res}} = 0 \text{ N}$ am Körper eben dieser ruht. Aus fachlicher Sicht ist dies nicht falsch, jedoch nicht der einzig denkbare Fall. Entsprechend des 1. Newtonschen Axioms kann sich ein Körper im Kräftegleichgewicht ebenso gleichförmig und geradlinig fortbewegen. Beide Fälle abdeckend könnte man also allgemein aus fachlicher Sicht sagen, dass im Kräftegleichgewicht keine Änderung des Bewegungszustandes des betreffenden Körpers eintritt. Am Beispiel der gleichförmigen Kreisbewegung sei hierbei eine weitere Problematik aufgezeigt: Aus Schülersicht verändert sich die Bewegung nicht, da sich der Körper ganz gleichmäßig im Kreis bewegt (vgl. Abschnitt zu Bewegungen, speziell zur ganzheitlichen Beschreibung von Bewegungen) (ebd.).

Wie schon die Vorstellung zum Stillstand bei Kräftegleichgewicht, ist auch die Vorstellung, dass Kräfte grundsätzlich in Bewegungsrichtung wirken, nur in Spezialfällen korrekt. Bei geradlinigen, „positiv“ beschleunigten Bewegungen ist dies nämlich tatsächlich richtig. Ein Beispiel dafür ist z. B. das senkrechte Fallen eines Balles von einem Turm. Ebenso können statische Betrachtungen diese Sichtweise verstärken, da sich ein Körper aus dem Ruhezustand heraus tatsächlich in Richtung der nun resultierenden Kraft bewegt, sofern eine weitere Kraft wirkt (Wiesner 2004; Tobias 2010). Allerdings lassen sich schon mit leichten Veränderungen zumindest zum ersten Beispiel Gegenargumente konstruieren. Wird der fallende Ball beispielsweise nicht einfach fallen gelassen, sondern zunächst senkrecht nach oben geworfen wirkt mit der Erdanziehungskraft noch während der Aufwärtsbewegung dauerhaft eine Kraft entgegengesetzt zur Bewegungsrichtung (vgl. hier auch die bekannte Aufgabe von Viennot

⁷Schon die Begriffe „ausüben“ oder „bewirken“ legen verständlicherweise ein aktives Einwirken nahe!

⁸R. Wodzinski (2004) betont hier die so schon genuine Ungleichbehandlung scheinbarer zweier Arten von Körpern, sodass Probleme beim Verständnis des Wechselwirkungsprinzips wahrscheinlich sind.

⁹vgl. https://de.wikipedia.org/wiki/Seilstatik#Seil_unter_Einzellast; zuletzt abgerufen am 14.6.2019

(1979)). Ebenso wirkt bei einer gleichförmigen Kreisbewegung dauerhaft eine Kraft in Richtung des Mittelpunkts des Kreises, also eher senkrecht zur Bewegungsrichtung. Weiterhin rollt bei einer Kraft senkrecht zur Bewegungsrichtung rollenden Kugel, diese anschließend nicht in Richtung dieser Kraft – um nur einige Beispiele zu nennen.

Die Verbindung zwischen Kraft und Geschwindigkeit/Schnelligkeit einer Bewegung auf Ebene der Alltagsvorstellungen lässt sich zusammenfassend durch das folgende Zitat charakterisieren:

„Kraft bewirkt Bewegung, und zwar je größer die Kraft, desto schneller die Bewegung. Aber man kann auch umgekehrt argumentieren: ein Körper, der sich bewegt, kann etwas bewirken, d. h. er hat Kraft, und zwar umso mehr, je schneller er sich bewegt.“

(R. Wodzinski 2004)

So wird mitunter angenommen, dass für die Größe der Endgeschwindigkeit eines Körpers ausschließlich die Stärke der ausgeübten Kraft verantwortlich ist (und die Dauer der Kraftwirkung vernachlässigt wird). Viennot (1979) wie auch R. Wodzinski (2004) schlussfolgern daraus eine angenommene Proportionalität zwischen Kraft und Geschwindigkeit. Viennot (1979) ergänzt dazu, dass bei diesen Betrachtungen eine Berücksichtigung der Beschleunigung nicht vorgenommen wird (z. B. „If $v = 0$, then $F = 0$, even if the acceleration a is not zero“ und „If $v \neq 0$, then $F \neq 0$, even if $a = 0$ “)

Leisen (2005) ergänzt als weitere Schwierigkeit beim Lernen des physikalischen Kraftbegriffs, dass die Kraft „keine Erhaltungsgröße und keine mengenartige Größe“ ist. Wären sie eine solche, könnte man über diese womöglich einfach bzw. relativ intuitiv nachdenken und rechnen.

Diagramme Eine falsche Vorstellung Diagramme betreffend ist der „Graph-als-Bild-Fehler“ (Nitsch 2015). Dabei wird der Graph nicht als Repräsentant eines funktionalen Zusammenhangs (oder hier eher als Darstellung von Messwerten) verstanden, sondern als Abbildung einer realen Situation.

Die in Abbildung 3.4 gezeigte Situation könnte für Schüler also den realen Verlauf der Strecken aus der Vogelperspektive zeigen, mit neun Kurven, drei davon Haarnadelkurven. Diese Interpretation wurde in diversen Untersuchungen speziell auch bei Schülern der Sekundarstufe I nachgewiesen. Betroffen davon waren in der Regel immer Zeit-Geschwindigkeits- oder Zeit-Weg-Diagramme (ebd.). Andere Interpretationsfehler können bei der Betrachtung des Graphen als eine Art Profil oder Querschnitt durch eine Landschaft auftreten. Dabei werden die Abschnitte als Berge und

II. Theorie

Täler interpretiert. Mit welcher Häufigkeit diese Vorstellung auftritt ist nicht geklärt, sie gehört jedoch zu den am besten bekannten.

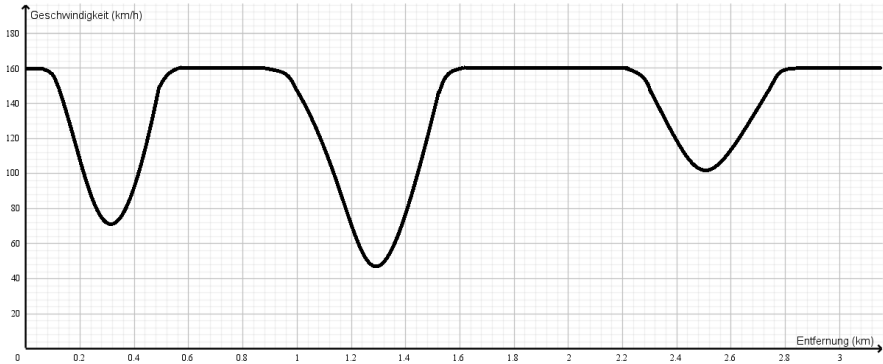


Abbildung 3.4.: Graph zum sogenannten Race-Car-Problem (nach Nitsch (2015) bzw. Janvier (1981))

Die in Abbildung 3.5 gezeigte Gegenüberstellung macht eine mögliche Interpretation eines Graphen nach der „Graph-als-Bild“-Vorstellung deutlich. Dabei wird der Verlauf des Graphen, wie schon oben erwähnt, als ein Querschnitt durch eine Landschaft gesehen, auf dem ein Fahrradfahrer nun entlang fährt. Wiesner, Schecker und Hopf (2011) sehen dies generell als einen Aspekt diverser Fehlinterpretationen von Abbildungen.

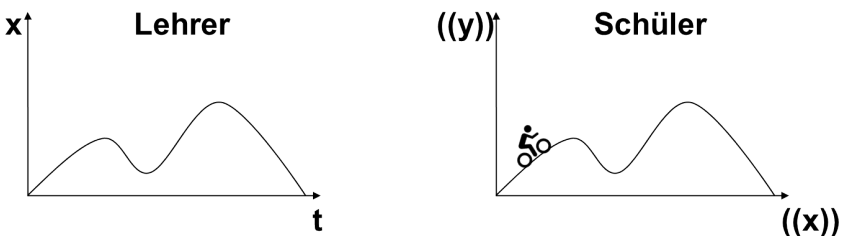


Abbildung 3.5.: Darstellung einer „Graph-als-Bild“-Vorstellung: Der Graph stellt vermeintlich ein „Gebirge“ dar (Nach: Wiesner, Schecker und Hopf (2011)).

Korrekt wäre in Abbildung 3.4 (und 3.5, die Achsenbeschriftung als Konvention vorausgesetzt) die Interpretation als Zeit-Ort- bzw. Zeit-Geschwindigkeits-Diagramm.

Beschleunigung Auch wenn die Beschleunigung kein Teil des hier verwendeten Lehrgangs ist, werden zentrale Alltagsvorstellungen dargelegt, da sich Inhalte des Lehrgangs zu einem späteren Zeitpunkt auf die Beschleunigung beziehen können.

Beschleunigung wird in der Alltagssprache als Synonym für das Schnellerwerden verwendet. Nach Schecker (1985) kann die Beschleunigung als „Bilanzgröße“ aufgefasst werden. Wilhelm (2005) unterscheidet zwischen vier Stufen der Vereinfachung. Dabei wird vor allem deutlich, dass – wie es schon bei der Geschwindigkeit der Fall ist – die Beschleunigung erst in der höchsten Stufe als vektorielle Größe erkannt wird¹⁰.

- 1. Stufe: Auf dieser Stufe wird *Beschleunigung* synonym mit *Geschwindigkeit* verwendet, da unter einer „beschleunigten Bewegung ‘nur’ eine schnelle Bewegung“ verstanden wird (ebd.).
- 2. Stufe: Nach Schecker (1985) bedeutet Beschleunigung „im Alltagsgebrauch das Schnellerwerden eines Körpers. Primär wird, physikalisch gesprochen, die Zunahme des Geschwindigkeitsbetrages betrachtet. Diese bilanziert den Beschleunigungsvorgang.“. Hier kommt somit auch zum Tragen, dass nur große Änderungen des Geschwindigkeitsbetrags einer großen Beschleunigung entsprechen. Wilhelm (2005) ergänzt diese „positive“ Beschleunigung noch um die „negative“, womit das bilanzierte Langsamerwerden gemeint ist.
- 3. Stufe: Auf dieser Stufe wird die Beschleunigung als Änderung des Geschwindigkeitsbetrags pro Zeit aufgefasst. Für eindimensionale Bewegungen sei dies zunächst eher unproblematisch (ebd.). Schwierigkeiten ergäben sich erst bei Bewegungen in negative Richtung¹¹, sodass eine negative Beschleunigung nicht *langsamer werden* bedeutet, sondern das Gegenteil: *schneller werden*.
- 4. Stufe: Beschleunigung als Änderung der Geschwindigkeit (im vektoriellen Sinne) pro Zeit. Entscheidender Unterschied zu den vorangegangenen Stufen ist hier das erstmalige Auftreten des vektoriellen Charakters. Die Beschleunigung hat hier also auch eine Richtung.

Im Sinne einer durchschnittlichen Beschleunigung entspricht dies der physikalischen Vorstellung.

Nach Wilhelm (ebd.) treten unterschiedliche Vorstellungen gleichzeitig in einer Klasse auf (vgl. auch Abschnitt 3.6.1). So sei es von der Aufgabenstellung abhängig, auf welche Vorstellung zurückgegriffen wird.

¹⁰Warum sollte auch eine in irgendeiner Form geartete Verwandte der skalaren Geschwindigkeit nun eine Vektorgröße sein?!

¹¹In diesem Bezug auch schon auf Stufe 2.

3.4. Darlegung der Konzepte

Im folgenden Abschnitt sollen die zwei zentralen Zugänge zur Mechanik in der Sekundarstufe I dargestellt werden. Dabei werden neben den charakteristischen Grundgedanken auch die Vorzüge und Problematiken der Zugänge erläutert. In Kapitel 3.6 wird jeweils die Lage an gegenwärtigen empirischen Daten dargelegt.

3.4.1. Klassisches Konzept: Eindimensionale Mechanik

Der Begriff „Klassisches Konzept“ mag zu Verwirrung führen. Denkt man an klassischen Mechanikunterricht, so kommen einem Federn, Kraftmesser, Waagen (statischer Kraftbegriff) oder sog. „einfache Maschinen“ wie Flaschenzüge in das Gedächtnis (vgl. Brenneke und Schuster 1965).

Die Problematiken von statischen Zugängen zum Kraftbegriff sind vielfältig. Als Beispiel soll nochmal der Stillstand eines Körpers genannt werden, an welchem sich die angreifenden Kräfte zu 0N addieren. Das sich Körper unter selben (!) Bedingungen auch gleichförmig bewegen können bleibt so völlig uneinsichtig (vgl. Tobias 2010). Leisen (2005) stellt weiterhin zusammenfassend fest:

„In jedem noch so einfachen statischen Problem treten derart viele Kräfte auf, dass eine fachlich korrekte Behandlung auf Schulniveau so gut wie ausgeschlossen ist. Dazu stellen die Fragen zu den Angriffspunkten, Wechselwirkungskräften, Gleichgewichtskräften und zu den Systemgrenzen fachdidaktische Hürden und methodische Stolpersteine dar. Jeder Lehrende ist klug beraten, mit den Bewegungen zu starten und die Statik als Appendix zu behandeln.“

(ebd.)

Im Folgenden soll als klassisches Konzept der *dynamische 1D-Zugang* aufgefasst werden. Dieser ist Bestandteil der aktuellen Curricula (siehe Abschnitt 3.5) und wird im folgenden Abschnitt charakterisiert.

GRUNDIDEE

Der 1D-Zugang zur Mechanik behandelt zweifelsohne die einfachsten Formen von Bewegungen. Hinsichtlich der Elementarisierung im Sinne von Vereinfachungen (Reinhold 2006) bietet die Bewegung entlang eines geradlinigen Weges nur wenig Möglichkeiten zur weiteren Simplifizierung. Mit diesen Voraussetzungen sind Richtungswechsel ausschließlich auf Vorwärts- und Rückwärtsbewegungen beschränkt. Diese Einschränkung auf eine Dimension macht es möglich, die Vektorgröße *Geschwindigkeit* auf

eine Komponente zu beschränken und somit als skalare Größe zu behandeln. Die Richtungsänderung äußert sich dadurch lediglich im Vorzeichen der Geschwindigkeit/des Geschwindigkeitsbetrages (Wilhelm 2005).

Ein zunächst überzeugender Vorteil des 1D-Zugangs liegt in seiner Einfachheit. Die Vernachlässigung anderer Bewegungen außer der Vor- und Rückwärtsbewegung macht die Behandlung der Sache schlicht. Darüber hinaus ist das Erfassen von Messgrößen, zum Beispiel des Ortes oder der Geschwindigkeit, in einer Dimension angenehm einfach. Weiterhin kann die erforderliche Mathematik zur Berechnung diverser Größen relativ einfach gehalten werden – bestimmte Spezialfälle betrachtet kann die (Momentan)geschwindigkeit über $v = \frac{s}{t}$ berechnet werden (sofern der Geschwindigkeitsbetrag konstant ist und der Bezugspunkt zur Messung des Weges im Koordinatenursprung liegt wodurch die jeweilige s-Koordinate dem zurückgelegten Weg entspricht). Ebenso sind relativ einfache Experimente durchführbar, die wiederum eine einfache Aufnahme von Messwerten wie Ort und Zeit ermöglichen. Auch mit neuerer Sensorik sind Werte wie der Geschwindigkeitsbetrag wie auch die Beschleunigung leicht zu erhalten.

Zu den offensichtlichen Schwachpunkten des 1D-Konzepts gehört die mangelnde Anschlussfähigkeit. Aus Sicht der Lernenden gibt es nach diesem Konzept keinen Vernünftigen Grund anzunehmen, dass eine Kreisbewegung eine nach innen gerichtete Kraft erfordert, da die Geschwindigkeit zu jedem Zeitpunkt gleich scheint (Tobias 2010). Ebenso wird *Beschleunigung* lediglich als „Schnellerwerden“ (wenn überhaupt, auch als „Langsamerwerden“) verstanden – der Aspekt des Richtungswechsels wird vernachlässigt (Wilhelm 2005). Das Erfassen der gleichen Richtung von Kraftwirkung und Beschleunigung ist weiterhin eher schwach, da Kräfte im Eindimensionalen nur entlang einer, bzw. sogar derselben, geraden Linie wirken.

Die Benutzung des Begriffs „Geschwindigkeit“ erfolgt im 1D-Zugang für eine skalare Größe, was im Deutschen aufgrund fehlender, gebräuchlicher sprachlicher Differenzierung nachvollziehbar ist, aber ebenfalls zur erwähnten mangelnden Anschlussfähigkeit beitragen kann.

3.4.2. Neues Konzept: Zweidimensionale Mechanik

Zunächst stellt sich die Frage, inwiefern konsequent zweidimensionale Konzepte als „neu“ bezeichnet werden sollten. Zugegebenermaßen ist der Titel provokant formuliert, denn die Anfänge dieses Konzepts reichen weit in das 20. Jahrhundert hinein – was aufgrund der Fachlogik nicht weiter verwunderlich ist. „Neu“ bezieht sich vielmehr darauf, als dass das

Konzept in der schulischen Realität bislang eine untergeordnete Rolle spielt (vgl. Kapitel 3.5).

GRUNDIDEE

Grundlegende Idee der zweidimensionalen Zugänge ist das Betrachten von zweidimensionalen Bewegungen von Anfang an und somit das Verwenden eines vektoriellen Geschwindigkeitsbegriffs. Ziel ist eine erhöhte Anschlussfähigkeit. Im 2D-Ansatz wird nur eine Komponente des dreidimensionalen Ortsvektors vernachlässigt. Auf diese Weise können Bewegungen in der Ebene sinnvoll betrachtet werden. Ebenso ist es sinnvoll die Geschwindigkeit als Vektorgröße einzuführen – auch mit der Hilfe von Pfeilen – und dabei zwischen Geschwindigkeit, Tempo/Schnelligkeit (also dem Geschwindigkeitsbetrag) und der Richtung zu unterscheiden.

Ein Vorteil dieses Konzepts ist die erhöhte Anschlussfähigkeit durch die Einführung eines vektoriellen Geschwindigkeitsbegriffs. Diese zahlt sich bei der Behandlung von Kreisbewegungen und krummlinigen Bewegungen aus. Weiterhin ist die Verknüpfung von Richtung der Kraftwirkung und Richtung der Beschleunigung deutlicher erkennbar und ist nicht auf Spezialfälle beschränkt. Ganzheitlich betrachtet werden Pfeile als hilfreiches Werkzeug in der Physik kennengelernt, dass es ermöglicht Probleme zu lösen. Weiterhin würde eine mögliche Erweiterung auf 3 (4,5,...) Dimensionen keine bedeutsame Änderung der Begrifflichkeiten mehr erfordern.

Als womöglich problematisch – bzw. komplexer als im 1D-Ansatz – könnte (!) die Vektoraddition angesehen werden. Speziell betrifft dies mitunter eindimensionale Spezialfälle (vgl. Wiesner 2011; R. Wodzinski 1996). Tatsächlich herausfordernd ist die Konfrontation mit zahlreichen Alltagsvorstellungen. Ohne das bisher Belege vorliegen, steht zu vermuten, dass das spätere Differenzieren zwischen *Kraft*- und *Geschwindigkeits*pfeilen zu Schwierigkeiten führen könnte.¹²

Zu bisher erfolgreichen Umsetzungen des 2D-Konzepts (u. a. durch Jung, Reul und Schwedes (1977), R. Wodzinski (1996) und Tobias (2010)) gehörte die frühe Behandlung dynamischer Zusammenhänge, als die Ursache von Bewegungs(änderungen) betreffend. Dazu wird auf die Begriffe Einwirkung (einer wirkenden Kraft) und Zusatzgeschwindigkeit (siehe (3.2)) und die Einsicht „Von Nichts kommt Nichts“ zurückgegriffen. Auf den Beschleunigungsbegriff wird im Anfangsunterricht noch konsequent verzichtet.

¹²Wobei fairerweise auch berücksichtigt werden muss, dass bei erstmaligem Kontakt mit Pfeilen im Kontext „Kraftpfeile“ später solche Probleme zu vermuten sind („Jeder Pfeil ist ein Kraftpfeil“).

3.4.3. Eindimensional - wirklich?

Als Anmerkung soll im folgenden Abschnitt eine regelmäßig vorgenommene Elementarisierung aus Sicht des 2D-Konzepts erläutert werden. In gängigen Lehrwerken werden in der Regel eindimensionale Bewegungen betrachtet. Dies geschieht anhand von allerlei Beispielen – anhand von 100m-Läufern, Spielzeugautos oder Zügen. Betrachtet man jedoch zum Beispiel die Fahrt eines Autos auf der Autobahn, so ist diese in der Realität nur seltenst eindimensional – selbst wenn man Einflussfaktoren wie böigen Seitenwind oder Bodenwellen großzügig außer Acht lassen würde. Die Fahrt auf der Autobahn – ebenso wie der gerne herangezogene Weg zur Schule – sind in der Realität dreidimensionale Bewegungen. Fast niemals dürfte es keine Kurven oder Höhenveränderungen geben. Wie dieser Widerspruch zu eindimensionalen Bewegungen gelöst wird/werden kann, schlägt Schmit (2014) vor, dessen Ausführungen hier dargestellt werden.

Im Regelfall würde man nämlich für dreidimensionale Bewegungen ein kartesisches Koordinatensystem festlegen und die Bewegung entlang der Bahnkurve durch die drei Koordinaten (bzw. zwei für Bewegungen in der Ebene) des Ortsvektors beschreiben (siehe System A, Abb. 3.6).

Bei der Reduktion dieser selben Bewegung auf eine Dimension bietet es sich an, die dafür erforderliche einzige Koordinatenachse direkt an der Bahnkurve der Bewegung auszurichten (siehe System B, Abb. 3.6). So erfolgt die Reduktion auf eine „scheinbare“ Dimension – und der Unterschied zwischen Ort und Weg fällt nicht mehr ins Gewicht.

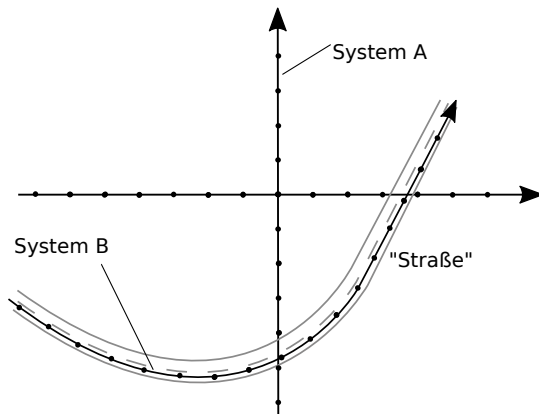


Abbildung 3.6.: Kartesisches Koordinatensystem (A) vs. Koordinatenachse entlang des Weges (B). Die Punkte markieren äquidistante Teilstücke auf den jeweiligen Achsen. (Darstellung angeregt durch Schmit (2014).)

Zur betrachteten Bewegung entlang der beispielhaften Straße „Straße“ (siehe Abb. 3.6) können die Koordinatensysteme jedoch auf unterschiedliche Art und Weise gewählt werden. Mittels dieser Methode kann die äußerlich betrachtete gleiche Bewegung einmal zweidimensional und einmal eindimensional, eher *scheinbar* eindimensional, betrachtet werden. Schmit (2014) weist aber darauf hin, dass sich mittels einer scheinbar eindimensionalen Koordinatenachse zwar der Weg, den ein Körper zurücklegt, prognostizieren lässt aber nicht sein tatsächlicher Ort, da die Bahnkurve prinzipiell nicht vorhersehbar ist.

3.5. Bestandsaufnahme: Was soll unterrichtet werden?

In diesem Abschnitt wird anhand der Curricula einzelner Bundesländer dargelegt, welche Kompetenzen in Teilen des Mechanikunterrichts hinsichtlich Bewegungen und Kräften im Anfangsunterricht erworben werden sollen. Das Ziel ist es zu untersuchen, ob vorgegebene Konzepte verfolgt werden sollen, die Einführung bestimmter Begrifflichkeiten vorgesehen ist oder sonstige Restriktionen oder Empfehlungen gegeben werden. Mangels Bezug zu dieser Arbeit werden traditionell vertretene Aspekte wie Trägheit und das Wechselwirkungsprinzip (sofern jeweils nicht untrennbar mit der Einführung des Kraftbegriffs verbunden) nicht in die Betrachtung der Lehrpläne mit einbezogen. Gleiches gilt auch für Kontexte (z. B. Verkehrssicherheit) sowie für weitere Aspekte aus den Bereichen der prozessbezogenen Kompetenzen. Speziell untersucht werden daher relevante folgende Aspekte:

- Beschreiben von Bewegungen (auch: Orts/Weg-Begriff und zugehörige Diagramme)
- Zugang zum Geschwindigkeitsbegriff
- Zugang zum Beschleunigungsbegriff (wenn vorgesehen)
- Zugang zum Kraftbegriff: dynamisch vs. statisch

Als unter anderem entscheidender Faktor beim Erlernen der Kinematik und Dynamik wurden Probleme im Zusammenhang mit skalarem und vektoriellem Geschwindigkeitsbegriff dargelegt, der zum Verständnis der Dynamik der Kreisbewegung oder allg. krummlinigen Bewegungen beiträgt. Aus diesem Grund wird nachfolgend ebenso untersucht, wie dieser Übergang – wenn er notwendig ist – angebahnt wird.

■ Transformation des Geschwindigkeitsbegriffs

Da die Analyse sämtlicher 16 Lehrpläne den Rahmen dieser Arbeit sprengen würde, erfolgt die Orientierung dabei an den vier einwohnerstärksten Bundesländer. Dies sind per 12/2016 mit insgesamt ca. 50 Mio. Einwohnern Nordrhein-Westfalen, Bayern, Baden-Württemberg und Niedersachsen¹³.

3.5.1. Niedersachsen

Die folgenden Ausführungen beziehen sich auf die Ausgabe des Kerncurriculums für die Naturwissenschaften, Bereich Physik, für das G9-Gymnasium aus 2015 (Niedersächsischen Kultusministerium 2015).

BEWEGUNGSBESCHREIBUNG

Die Bewegungsbeschreibung soll sich an geradlinigen Bewegungen orientieren. Dabei wird in diesem Zusammenhang besonders auf das Interpretieren von t-v und t-s-Diagrammen hingewiesen. Eine Unterscheidung oder Vorgabe der Begriffe Ort oder Weg wird dabei nicht vorgenommen.

GESCHWINDIGKEIT

Eine spezielle Einführung des Geschwindigkeitsbegriffs wird nicht gefordert. Die Interpretation von linearen Graphen in Zeit-Weg / Ort-Diagrammen mittels Steigungsdreiecken soll aber mit Hilfe dieses Begriffs erfolgen.

BESCHLEUNIGUNG

Ebenso wie bei der Geschwindigkeit werden keine besonderen Anforderungen an die Einführung gestellt – lediglich die Interpretation von Diagrammen soll mithilfe dieser Begriffe erfolgen.

KRAFT

Bewegungs- und Energieänderungen sowie Verformungen sollen phänomenologisch als die Wirkung von Kräften dargestellt werden. Über die Verformung ist eine Verknüpfung zum Hookeschen Gesetz möglich, an dem proportionale Zusammenhänge untersucht werden sollen (hier Statik-Anleihen)

TRANSFORMATION

Es wird explizit keine Transformation des Geschwindigkeitsbegriffs vorgegeben, was nahelegt, dass der Begriff im eigenen Ermessen der Lehrkraft als Vektorgröße eingeführt werden kann/muss.

¹³<https://de.statista.com/statistik/daten/studie/71085/umfrage/verteilung-der-einwohnerzahl-nach-bundeslaendern/>, abgerufen am 24.4.2018. Niedersachsen als Bundesland der Studie ohnehin gesetzt.

3.5.2. Baden-Württemberg

Die folgenden Ausführungen beziehen sich auf den Bildungsplan des Gymnasiums für das Fach Physik aus dem Jahr 2016 (Ministerium für Kultus, Jugend und Sport Baden Württemberg 2016).

BEWEGUNGSBESCHREIBUNG

Bewegungen sollen mündlich und mithilfe von Diagrammen in ein vorgegebenes Kategoriensystem eingeordnet werden und mit vorgegebenen Begriffen beschrieben werden. Dabei wird ausdrücklich der Ortsbegriff anstelle des Wegbegriffs gefordert – während an anderer Stelle aber wieder von s-t-Diagrammen geschrieben wird. Unklar bleibt in dem Zusammenhang mit *Bewegungsdiagrammen*, inwiefern dadurch, wie gefordert, die Richtung der Bewegung interpretiert werden kann.

GESCHWINDIGKEIT

Die Geschwindigkeit soll als Quotient aus Strecke und Zeitspanne behandelt werden. Weder der Orts- noch der Wegbegriff wird explizit dabei aufgegriffen.

BESCHLEUNIGUNG

Beschleunigte Bewegungen sollen „verbal und mithilfe von Diagrammen“ beschrieben und klassifiziert werden. Eine verbindliche Art der Einführung des Beschleunigungsbegriffs ist nicht vorgegeben.

KRAFT

Kräfte sollen als Ursachen von Änderungen von Bewegungszuständen (Betrag und Richtung explizit erwähnt) wie auch als Ursache von Verformungen beschrieben werden. Zusätzlich sollen die Newtonschen Prinzipien der Mechanik auf Alltagsbeispiele angewandt werden – mögliche Reduktionen werden dabei nicht vorgegeben.

TRANSFORMATION

Mit Hinblick auf die Kreisbewegung wird für den Jahrgang 9/10 explizit vorgegeben, dass der Vektorcharakter der Geschwindigkeit klar gemacht werden muss. Dies soll entsprechend des Bildungsplans anhand von zusammengesetzten Bewegungen, wie etwa einer Flussüberquerung oder dem waagerechten Wurf, geschehen (z. B. „Förderbandaufgaben“ oder „Flussaufgaben“ Jung, Reul und Schwedes 1977; Wiesner 2011).

3.5.3. Nordrhein-Westfalen

Die folgenden Ausführungen beziehen sich auf den Kernlehrplan für das G8-Gymnasium in Nordrhein-Westfalen (Ministerium für Schule und Weiterbildung NRW 2008). Zu bemerken ist hier, dass dieser Lehrplan

vergleichsweise eng an die Basiskonzepte der Bildungsstandards Physik (Kultusministerkonferenz 2004) angelehnt ist. Dies geht über eine bloße tabellarische Aufstellung hinaus, als dass jedes Konzept nochmals ausführlich kommentiert wird.

BEWEGUNGSBESCHREIBUNG

Zwar wird im kommentierten Teil erwähnt, dass die Kinder Veränderungen von Bewegungen beobachten sollen, jedoch wird an keiner weiteren Stelle Bezug zu einer systematischen Bewegungsbeschreibung genommen. Entsprechend gibt es auch keine Vorgaben zur Verwendung der Begriffe *Ort* oder *Weg*. Ebenso finden t-s- oder t-v-Diagramm an keiner Stelle Erwähnung¹⁴.

GESCHWINDIGKEIT

Ein besonderer Zugang zum Geschwindigkeitsbegriff ist nicht vorgesehen, jedoch soll der Vektorcharakter der Geschwindigkeit durch die Schüler und Schülerinnen bis zum Ende der 9. Jahrgangsstufe beschrieben werden können. In der Aufzählung der *Inhaltsfelder und fachlichen Kontexte* erscheint diese Voraussetzung des Vektorcharakters nicht mehr explizit.

BESCHLEUNIGUNG

Keine explizite Erwähnung im Lehrplan

KRAFT

Bewegungsänderungen und Verformungen sollen auf das Wirken von Kräften zurückgeführt werden, wobei die Kraft ausdrücklich als Vektorgröße einzuführen ist. Im kommentierten Teil des Lehrplans ist darüber hinaus näher ausgeführt, dass auch körperliche Erfahrungen beim Erkennen von Wechselwirkungen zwischen zwei Körpern eine erhebliche Rolle beim Erlernen eines propädeutischen Kraftbegriffs spielen.

TRANSFORMATION

Eine explizite Veränderung des Geschwindigkeitsbegriff wird auch nach Jahrgangsstufe 9 nicht gefordert, was angesichts der verbindlichen vektorriellen Einführung nicht weiter verwunderlich ist.

3.5.4. Bayern

Die Ausführungen beziehen sich auf das Fach „Natur und Technik“, da Physik als grundständiges Fach erst ab Jahrgang 8 unterrichtet wird. Der Lehrplan ist gültig seit dem Schuljahr 2008/09 (Staatsinstitut für Schulqualität und Bildungsforschung 2008).

¹⁴Wohingegen der Umgang mit Diagrammen im Allgemeinen an anderer Stelle gefordert wird!

BEWEGUNGSBESCHREIBUNG

Vorgaben zur Bewegungsbeschreibung werden nicht gemacht, so auch keine zur Verwendung von *Ort* oder *Weg*. Mit kinematischen Grundgrößen zur Bewegungsbeschreibung sollen die Schüler aber vertraut sein – mehr dazu in den folgenden Abschnitten. Diagramme werden für den Anfangsunterricht nicht zwingend vorgesehen.

GESCHWINDIGKEIT

Die Geschwindigkeit soll zunächst eindimensional eingeführt werden. Im Anfangsunterricht soll dabei die Richtungsproblematik (Vorzeichen) nicht „systematisch“ angegangen werden. Die Delta-Notation sollte eingeführt werden, um den Differenzcharakter herauszustellen.

BESCHLEUNIGUNG

Ebenso wie bei der Geschwindigkeit soll bei der Beschleunigung die Delta-Notation verwendet werden. Die Diskussion der Vorzeichen kann mitunter auf das „negative Beschleunigen“ als Abbremsen beschränkt bleiben.

KRAFT

Ausdrücklich an Newton orientiert soll die Kraft auf ihre beschleunigende Wirkung zurückgeführt werden, wobei der Charakter des Axioms betont wird. So soll also eine – scheinbare – Herleitung über Messreihen nicht erfolgen. Das bereits vorgestellte Unterrichtskonzept zur zweidimensional-dynamischen Mechanik aus München / Frankfurt / Würzburg wird erwähnt und zur Einführung des Kraftbegriffs als Möglichkeit angeboten – ergänzt wird aber, dass im Falle der Verwendung dieses Konzepts noch die Beschleunigung für den eindimensionalen Fall eingeführt werden muss.

TRANSFORMATION

Eine Transformation eines potentiell eindimensionalen Geschwindigkeitsbegriffs wird nicht nahegelegt.

Anmerkung: Im Rahmen des LehrplanPlus soll in Bayern das schon beschriebene Konzept der zweidimensional-dynamischen Mechanik unterrichtet werden. Der LehrplanPlus gilt ab Jahrgang 5 ab 2017 aufsteigend. Auf eine erneute Beschreibung wird hier verzichtet.

3.5.5. Zusammenfassung

Die Einführung des Geschwindigkeitsbegriffs im Physikunterricht erfolgt unter relativ uneinheitlichen Maßgaben. Vorgesehen sind zum einen Zugänge über die Quotientenbildung aus Strecke und Zeitspanne, sowie auf gleiche Weise aber in anderer Gestalt über die Steigung in t-s-Diagrammen. In der Regel wird eine eindimensionale Einführung des Geschwindigkeitsbegriff nahegelegt, eine zweidimensionale aber niemals explizit ausgeschlossen. Explizit als Vektor soll die Geschwindigkeit in

Nordrhein-Westfalen eingeführt werden¹⁵. Die Beschleunigung wird in den meisten betrachteten Lehrplänen berücksichtigt, der Richtungscharakter wird aber niemals erwähnt. Daher ist auch hier von einer skalaren Einführung auszugehen, wobei auch vektorielle Zugänge nicht ausgeschlossen zu sein scheinen. Zur Verwendung von *Ort* oder *Weg* im Rahmen der Bewegungsbeschreibung werden nur einmal Vorgaben gemacht (*Ort* in Baden-Württemberg), ansonsten lassen sich die Präferenzen allenfalls über die Diagrammbezeichnung herleiten, was aber aufgrund der mangelnden Konventionen mehr als gewagt erscheint. Da die Unterscheidung zwischen *Weg* und *Ort* erst im zweidimensionalen tatsächlich deutlich wird, zeigt sich hier der eher skalare Charakter der erwarteten Fähigkeiten und Fertigkeiten.

Bei der Einführung des Kraftbegriffs dominieren Vorgaben, die Kräfte als Ursachen von Bewegungsänderungen betrachten. Dies ist aber nicht gleichbedeutend mit der Erforderlichkeit eines vektoriellen Geschwindigkeitsbegriffs. Vielmehr kann die Bewegungsänderung als *schneller oder langsamer werden oder die Bewegungsrichtung ändern* eingeführt werden. Im Endeffekt kommt dies oberflächlich auf einen vektoriellen Geschwindigkeitsbegriff hinaus, kann aber die entsprechende Begrifflichkeit der Geschwindigkeit als Vektorgröße vermeiden. Darüber hinaus verweist Bayern im Lehrplan (Staatsinstitut für Schulqualität und Bildungsforschung 2008) aber ausdrücklich auf das vorgestellte zweidimensionaldynamische Konzept aus München. Mit der Einführung des LehrplanPlus ab 2017 soll dies verbindlich unterrichtet werden.

Die verformende Wirkung von Kräften, die am ehesten mit dem statischen Kraftbegriff harmoniert, wird ebenfalls in den meisten betrachteten Lehrplänen – mal mehr und mal weniger ausdifferenziert (z. B. bis hin zum Hooke'schen Gesetz) – gefordert. An keiner Stelle steht dieser Aspekt allerdings exklusiv als einziger Weg zum Kraftbegriff verankert. Daraus lässt sich schließen, dass der statische Zugang zum Kraftbegriff nicht mehr als *der* klassische Zugang zur Kraft gesehen werden sollte.

Zusammenfassend werden kaum scharfe Begriffsbildungen vorgegeben. Die Einführung des Geschwindigkeitsbegriffs als Betragsgröße wird zwar durch Formulierungen nahegelegt, niemals wird jedoch die vektorielle Einführung abgelehnt – lediglich einmal explizit gefordert. Solange also keine Inhalte des Curriculums zu Gunsten eines anderen vernachlässigt werden, kann entsprechend unterrichtet werden (Curriculum als Regelstandard)! Die dynamische Einführung des Kraftbegriffs ist generell vorgesehen, auch wenn sich statische Aspekte nach wie vor identifizieren lassen.

¹⁵Auch wenn diese Vorgabe an anderer Stelle im Lehrplan nicht mehr explizit genannt wird

3.6. Empirische Befunde

In diesem Abschnitt wird die Befundlage von maßgeblichen, empirischen Arbeiten auf dem Gebiet des (Anfangs)unterrichts zur Mechanik im deutschsprachigen Raum dargestellt.

3.6.1. Kurzdarstellung von Studien und Ergebnissen

JUNG

Die folgenden Ausführungen beziehen sich ausnahmslos auf die Studie von Jung, Reul und Schwedes (1977). Sie sind der entsprechenden Arbeit entnommen. Angaben, die nicht Jung, Reul und Schwedes (ebd.) entstammen, sind somit die einzigen gekennzeichneten.

Studie „Jung“

Studie in 3. – 6. Klassen. Erprobung einer ausgiebig pilotierten Einheit zur zweidimensionalen Kinematik/Dynamik. Erhebung durch Fragebögen und einzelne Interviews. Schulung der Lehrer; inkl. Lehrer- und Schülerheft.

Der Unterricht zog sich über einen langen Zeitraum von ca. 20 Unterrichtsstunden, was für damalige Grundschulverhältnisse als außergewöhnlich lang galt. Dies resultierte mitunter in Motivationsproblemen seitens der Schülerinnen und Schüler. Da das Thema *Mechanik* darüber hinaus inhaltlich eigentlich nicht in der Grundschule verankert war, bedeutete dies einen erheblichen Aufwand in der Durchführung – auch durch die Unterweisung der Lehrkräfte. Der Lehrgang, unterstützt durch Schüler- und Lehrerhandbuch, gliederte sich inhaltlich in die Themengebiete

- Geschwindigkeit
- Bezugssystem und Zusatzgeschwindigkeit
- Geschwindigkeitsänderung in Kurven
- Geschwindigkeitsänderung und Stoß

Der Wissenszuwachs in diesen Gebieten wurde durch einen eigens konzipierten Mechanik-Test gemessen, wobei Vor- und Nachtest praktisch identisch waren. Überraschenderweise erzielten die jüngeren Schülerinnen und Schüler deutlich bessere Ergebnisse als die älteren. Allen voran überzeugten die Viertklässler, gefolgt von den Drittklässlern. Unterschiede bei den Geschlechtern zeigten sich durch etwas schlechtere Leistungen bei den Mädchen. Dieser Effekt war umso ausgeprägter, je höher der

IQ der betroffenen Personen war. Auch insgesamt waren die Ergebnisse durchaus vom IQ abhängig. Die Allgemeinbildung, die Raumvorstellung und Denkfähigkeit haben Einfluss auf Ergebnis in erwarteter Richtung: je höher der jeweilige Faktor ist, desto besser das Ergebnis des Tests. Überraschenderweise hat die Raumvorstellung nicht den größten Einfluss bei diesen Faktoren. Bezüglich der unterschiedlichen Themengebiete wurden Aufgaben aus dem Bereich *Tempo* am besten gelöst. Aber auch der vektorielle Geschwindigkeitsbegriff sowie das Konzept der Zusatzgeschwindigkeit erschienen gut erlernbar.

Die geführten Interviews zeigten u. a. die folgenden Erkenntnisse. Bezüglich des Geschwindigkeitsbegriffs wurde festgestellt, dass z. T. Bedingungen aus gestellten Aufgaben nur schwerlich akzeptiert werden konnten. So wurde z. B. mitunter nicht anerkannt, dass eine Kurve mit konstantem Tempo zu durchfahren sein kann, da man ja bremsen müsse. Dies lässt sich nach Einschätzung des Autors der vorliegenden Arbeit aber auf das Alter bzw. Entwicklungsstand der Lernenden zurückführen. Wie bereits in den bekannten Schülervorstellungen dargelegt, wurden weiterhin Geschwindigkeiten/Bewegungen ganzheitlich beschrieben. So wird die (momentane) Bewegungsrichtung einer Kreisbewegung entsprechend ein gebogener Pfeil zugeordnet – und nicht aus momentaner Betrachtung ein Pfeil tangential zur Bewegungsrichtung. Schwer fiel ebenfalls das Loslösen des Geschwindigkeitspfeil vom sich bewegenden Objekt. So konnte bspw. ein Schüler einen vor sich liegenden Pfeil nicht synchron mit der Geschwindigkeit einer Murmelbahn bewegen, ohne den Pfeil direkt an die Murmelbahn – und damit dem sich bewegenden Objekt – näher zu bringen.

Die gleiche Richtung von Stoß und Zusatzgeschwindigkeit ist zunächst schwer einsichtig, da aufgrund von durchaus korrekt beobachteten Situationen des Alltags¹⁶ hier eine fachlich falsche Vorstellung vorlag. Diese konnte durch entsprechende Experimente einsichtig gemacht werden. Weitere Schwierigkeiten im Bezug auf die Zusatzgeschwindigkeit waren im Kontext *Kurvenfahrt* festzustellen, wobei die zugehörige *eine, gesamte* Zusatzgeschwindigkeit nur schwer identifiziert werden konnte¹⁷. Im Grunde fiel es also schwer, eine Kurvenfahrt durch nur einen Stoß zu modellieren.

Als Schlussfolgerung¹⁸ wird gesehen, dass der Mechanikunterricht zu spät beginnt. Er befände sich zu weit in der Mittelstufe – was bis heute relativ unverändert ist (siehe Kapitel 3.5). Daher ist der Rat, schon in der 4. Klasse mit der Mechanik zu beginnen, da sich bei zu weitem Aufschieben

¹⁶Zum Beispiel das Kicken eines sehr leichten Balles quer zur Bewegungsrichtung

¹⁷Man beachte, dass eine vergleichbare Aufgabe bei Tobias (2010) gestellt wurde und dort sogar Lernende des 7./8. Jahrgangs größere Probleme hatten.

¹⁸Und nicht als objektiv messbares Ergebnis! Jung, Reul und Schwedes (1977) legen großen Wert darauf

Lernbarrieren für die Schüler bilden. Darüber hinaus sollte das Lernen der Mechanik mit propädeutischen Betrachtungen zur Dynamik beginnen. In einer Konkretisierung wird eine Art Spiralcurriculum vorgeschlagen, bei dem über frühe Schuljahre hinweg Inhalte systematisch wiederholt werden und Neues darauf aufbauend erarbeitet wird.

WODZINSKI

Die folgenden Ausführungen beziehen sich ausnahmslos auf die Studie von R. Wodzinski (1996). Sie sind der entsprechenden Dissertation entnommen. Angaben, die nicht R. Wodzinski (ebd.) entstammen, sind somit die einzigen gekennzeichneten.

Studie „Wodzinski“

Studie zu dynamischer (zweidimensionaler) Unterrichtskonzeption, Real- und Gesamtschulen und Gymnasien, 9. – 10. Schuljahr, etwa 40 SuS. Akzeptanzbefragung/Interviewstudie/vereinzelte Fragebögen

Das Ziel der Untersuchung aus dem Anfang der 90er Jahre war es, die „Umsetzbarkeit und Wirksamkeit“ einer zweidimensional-dynamischen Unterrichtskonzeption sowie die Betrachtung von damit zusammenhängenden Lernschwierigkeiten zu untersuchen.

Aus ihrer abschließenden Gesamtbewertung der Studie geht hervor, dass sie das dynamisch orientierte

Unterrichtskonzept für tragfähig hält – auch vor dem Hintergrund der bekannten Lernschwierigkeiten. Als besonders hilfreich wird ein „klarer begrifflicher Rahmen“ erachtet, der besonders die Einwirkung und die daraus resultierenden Bewegungsänderungen thematisiert. Auf diese Weise soll erreicht werden, dass die typischen Alltagsvorstellungen dadurch „an Bedeutung verlieren“. Eine Verknüpfung des Kraftbegriffs sollte mit der Veränderung von Bewegung assoziiert werden – und nicht mit der Bewegung an sich. Dies sei konzeptunabhängig (1D oder 2D) empfehlenswert. Eine frühe Einführung des Kraftbegriffs bewertet Wodzinski positiv und kann aus ihren Untersuchungen auch keine negativen Auswirkungen der im Unterrichtskonzept verwendeten „eher formale[n] Sprache“ feststellen. Das Konzept der Zusatzgeschwindigkeit beim senkrechten Stoß wird durch die Schüler zwar ungewohnt aber dennoch verständlich wahrgenommen. Der besonderen Wahrnehmung eindimensionaler Bewegungen im Vergleich zu zweidimensionalen trägt sie damit Rechnung, als dass sie in der Konzeption mit speziellen Beispielen auf eindimensionale Bewegungen eingeht.

Insgesamt schien das 2D-Konzept nicht wesentlich komplizierter für die Schülerinnen und Schüler zu verstehen zu sein, als das klassische 1D-Konzept. Befürchtete Schwierigkeiten vektorielle Betrachtungen betreffend traten zwar auf, waren aber weit „weniger gravierend“ als erwartet.

Im Bezug auf bekannte Alltagsvorstellungen („Passive Körper können keine Kraft ausüben!“) konnte Wodzinski besondere Erfolge feststellen, und zwar insofern, als dass vergleichbare Äußerungen nicht mehr festgestellt wurden. Auch aus diesem Grund legt sie großen Wert auf einen begrifflichen Rahmen, der den Schülern Rückgriffe auf physikalische Konzepte statt auf naive Vorstellungen ermöglicht.

Abschließend stellt sie fest, dass die unterschiedlichen Lernertypen so vielfältig sind, dass es kaum *die* Strategie geben kann, „die allen Schülerinnen und Schülern gerecht wird“.

WILHELM

Die folgenden Ausführungen beziehen sich ausnahmslos auf die Studie von Wilhelm (2005). Sie sind der entsprechenden Dissertation entnommen. Angaben, die nicht Wilhelm (ebd.) entstammen, sind somit die einzigen gekennzeichneten.

Studie „Wilhelm“

Studie an bayerischen Gymnasien, 11. Jahrgang . Einführung von Beschleunigung und Kräften in 2D. Vergleich mit traditioneller Gruppe. Unterstützung durch dyn.-ikonische Repräsentationen/bzw. Modellbildungssoftware.

Das Ziel der Dissertation war „Ein Gesamtkonzept für einen veränderten Kinematik- und Dynamikunterricht ein- und zweidimensionaler Bewegungen in der Jahrgangsstufe 11 des Gymnasiums zu entwickeln, das möglichst vielen Schülern hilft, möglichst viele Fehlvorstellungen zur Mechanik aufzuarbeiten“¹⁹. Als unterstützendes Mittel wurden im

Rahmen der Studien in den Anfängen der 2000er Jahre computergestützte Experimente sowie dynamisch-ikonische Repräsentationen eingesetzt. Vergleiche wurden mit traditionellem Unterricht gezogen. Zur Feststellung des Leistungsvermögens der Schülerinnen und Schüler wurden vorhandene Tests wie das Force Concept Inventory (FCI) (Hestenes und Wells 1992) eingesetzt. So sollten auch Auswirkungen auf die bekannten Schülervorstellungen untersucht werden.

In der traditionell unterrichteten Kontrollgruppe waren diesbezüglich die Resultate von den typischen Vorstellungen von Kraft und Beschleunigung – wenn auch uneinheitlich – dominiert. t - v -Graphen konnten schon vor Beginn der Einheit sicher zu bestimmten Situationen zugeordnet werden, zu t - a -Graphen gelang dies erst nach der Unterrichtseinheit knapp der Hälfte der Schüler. Schlechter zeigte sich diese noch beim Angeben der Beschleunigung bei senkrechten Würfeln. Hier überwiegen auch Interpretationen als Geschwindigkeit sowie angemessenere Vorstellungen als

¹⁹Man beachte hierbei nochmals, dass sich die vorgestellten Erkenntnisse auf den Unterricht der Oberstufe beziehen

Geschwindigkeitsbetragsänderung. Beim Angeben von Beschleunigungspfeilen zeigten sich für eindimensionale Bewegungen gute Ergebnisse im Gegensatz zu Beispielen mit krummliniger Bewegung.

Der FCI wurde im Vorfeld von einem knappen Drittel der Lernenden korrekt bearbeitet, nach dem traditionellen Lehrgang waren es 41%, mit typischen Schwächen im erwartbaren Bereich (z. B. Superposition). Zum 2. Newtonschen Axiom wurden für den 1D-Fall spezielle Aufgaben erstellt, wobei diese aber ebenfalls nur von einem Drittel korrekt gelöst wurden und aristotelische Vorstellungen nach wie vor in großer Zahl vorkommen. Nach Abschluss der Einheit erstellte Begriffsnetze zeigen darüber hinaus eine mangelnde Verknüpfung zwischen dynamischen und kinematischen Begriffen.

Die Treatmentgruppe zeigte im Bezug auf die Graphen-Aufgaben vor- und nachher ähnliche Ergebnisse. Bei Fragen nach der Richtung der Beschleunigung zeigten sich allerdings signifikant bessere Ergebnisse – auch in eindimensionalen Fällen wie dem senkrechten Wurf. Auch bei krummlinigen Bewegungen zeigte die Treatmentgruppe erheblich bessere Kenntnisse über die jeweilige Richtung der Beschleunigung. Daraus wurde „gefolgert, dass in den nach diesem Unterrichtskonzept unterrichteten Klassen mehr Schüler ein physikalisch angemessenes Beschleunigungskonzept erreicht haben“. Insgesamt lassen sich in der Treatmentgruppe mehr Schüler mit Newtonschen Vorstellungen und differenzierteren Concept-Maps identifizieren.

Wilhelm konstatiert:

„Insgesamt kann man also folgern, dass im herkömmlichen Unterricht nur wenige Schüler ein physikalisch angemessenes Verständnis der Beschleunigung erreichen, das ihnen ermöglicht, qualitative Aufgaben zur Beschleunigung — auch bei Richtungsumkehr oder im zweidimensionalen Fall — konsequent richtig zu lösen“

(Wilhelm 2005)

„Die Einführung kinematischer Größen anhand zweidimensionaler Bewegungen, die nur mit ikonischen Repräsentationen in Form von Vektorpfeilen sinnvoll ist (geeignete Elementarisierung), führt zu einem physikalischeren Verständnis des Beschleunigungsbegriffes und vermeidet Fehlvorstellungen durch eine ungeeignete Reduktion auf den Spezialfall eindimensionaler Bewegungen.“

(ebd.)

TOBIAS, 2010

Die folgenden Ausführungen beziehen sich ausnahmslos auf die Studie von Tobias (2010). Sie sind der entsprechenden Dissertation entnommen. Angaben, die nicht Tobias (ebd.) entstammen, sind somit die einzigen gekennzeichneten.

Studie „Tobias“

Studie 2008/2009 an bayerischen Gymnasien. Zunächst traditionelles Konzept unterrichtet, dann kurze Einweisung der Lehrkräfte. Anschließend unterrichteten diese das 2D-Konzept. Evaluation über Fragebögen, Interviews.

Im Vergleich des traditionellen Konzepts mit dem 2D-dynamischen Konzept wurden kognitive wie nicht-kognitive Faktoren verglichen. Erste wurden weitgehend in 1D- oder 2D-konzeptspezifische sowie allgemeine Faktoren unterteilt.

Zum spezifischen Wissen zum 2D-Konzept lassen sich folgende

Ergebnisse festhalten, die zu einem großen Teil aus den geführten Interviews gewonnen wurden: Der vektorielle Geschwindigkeitsbegriff wurde durch die Lernenden gut erfasst. So erinnerte die Mehrheit, dass Änderungen in Tempo *oder* Bewegungsrichtung gleichermaßen eine Geschwindigkeitsveränderung bedeuten können. Auch aus Sicht der Lehrkräfte ist der 2D-Geschwindigkeitsbegriff für die Schüler verständlich. Das zeigt sich auch darin, dass das Anfertigen von Pfeilbildern zu gegebenen Bewegungen ebenso kaum Hindernisse darstellt. Das Konzept der Zusatzgeschwindigkeit ist auf Anhieb nicht mehr so klar zugänglich. Auf Nachfrage konnten (in Interviews) aber auch dazu Angaben gemacht werden. Auffällig war an manchen Stellen aber eine Vermischung zwischen „Kraft“ und „Zusatzgeschwindigkeit“. ²⁰ Das Konstruieren von Pfeilbildern zu Anfangs-, End- und Zusatzgeschwindigkeit gelingt nur etwa der Hälfte der Schüler, wohingegen das Interpretieren entsprechender Pfeilbilder wesentlich besser gelingt. Die Richtung der Zusatzgeschwindigkeit wird richtigerweise mit der Richtung der wirkenden Kraft gleichgesetzt. Die Newton'sche Bewegungsgleichung in der dargebotenen Form ist im Bezug auf Je-desto-Beziehungen leicht begreiflich. Ebenso ist die Gleichung durch die Lernenden mitunter reproduzierbar. Einige Schüler können mithilfe dieser Gleichung durch mathematische Betrachtungen Schlüsse aus der Formel ziehen. Aus Sicht der Lehrkräfte ist die Gleichung ansprechend – wenn sie sich mitunter auch erst nach einiger Zeit an sie gewöhnt hatten. Insgesamt habe eine Akkomodation von der vorunterrichtlichen

²⁰Als Anmerkung des Autors sei hier gesagt, dass dies evtl. aus der jeweiligen Darstellung als Pfeil resultieren mag, ebenso wie aus verwandten Eigenschaften, wie der gleichen Richtung oder sich ähnlich verhaltenden Parametern: *größere Kraft bewirkt größere Zusatzgeschwindigkeit (bzw. deren Tempo)*

Darstellung hin zum physikalischen Konzept stattgefunden, wobei aber wahrscheinlich beide nebeneinander existieren.

Im Bezug auf nicht-kognitive Variablen wurde der Einfluss der Zugänge auf das Interesse an Physik und das Selbstkonzept bzgl. Physik / Mechanik untersucht. Mit Hinblick auf das Interesse und Selbstkonzept zur Physik ist bei beiden Konzepten kein Einfluss beobachtbar. Gleiches gilt, wenn beides mit Blick auf die Mechanik betrachtet wird. Es ergeben sich einzig konzeptspezifische Auffälligkeiten. So sind die traditionell unterrichteten Schüler zuversichtlicher was das Zeichnen von Diagrammen und Berechnen von Sachverhalten angeht, während die 2D-unterrichteten Schüler zuversichtlicher Bewegungen voraussagen oder erklären können zu glauben. Im Allgemeinen seien diese Ergebnisse nicht verwunderlich, da das Unterrichtskonzept primär auf Wissenszuwachs auslegt sei.

Bezogen auf das allgemeine Mechanikverständnis wurde in der Treatmentgruppe entsprechend des identischen Vor- und Nachtests deutlich mehr dazugelernt. Es ergeben sich bei einigen Aufgaben (höchst) signifikante Unterschiede mittlerer Effektstärke. Im Bezug auf die wenigen, separat ausgewerteten Items, die sich speziell auf eins der Konzepte bezogen, ergab sich der zu erwartende Unterschied, dass Fragen zur Zusatzgeschwindigkeit / Pfeilbildern von traditionell unterrichteten Schülern schlechter beantwortet wurden. Bei Fragen zur Beschleunigung, welche zum klassischen Konzept zugeordnet wurden, ergab sich hingegen kaum ein Unterschied. Dies wurde damit begründet, dass die Lehrkräfte im 2D-Konzept den Begriff – obwohl nicht zwingend vorgesehen – dennoch einführten, um dem Lehrplan gerecht zu werden.

Im Bezug auf Geschlechterunterschiede konnten keine großen aber signifikante Effekte nachgewiesen werden. Es zeigte sich, dass die Jungen im Vortest durchweg besser abschnitten als die Mädchen. Diese Lücke konnte im traditionellen Konzept kaum geschlossen werden, mithilfe des 2D-dynamischen Ansatzes hingegen gelang eine Angleichung der Testergebnisse.

Der Einfluss der Lehrkräfte wurde in der hier beschriebenen Untersuchung insofern minimiert, als dass dieselben Lehrkräfte jeweils einmal das 1D- und einmal das 2D-Konzept unterrichten, um eine gewisse Konstanz zu gewährleisten. Dabei ist zu bemerken, dass eine klassisch unterrichtete Klasse in keinem Fall bessere Ergebnisse erzielte, unabhängig von der Einstellung der Lehrer gegenüber den Konzepten. Die Einstellung bzw. Akzeptanz der Lehrkräfte schlägt sich erst im Follow-up-Test nieder. In diesem werden bei hoher Akzeptanz der Lehrkraft gegenüber dem neuen Konzept bessere Ergebnisse erzielt. Tobias weist hier nochmals auf die Bedeutung dieses Umstandes hin. So sei es bemerkenswert, dass die Lehrer trotz der großen Vertrautheit mit dem traditionellen Konzept keine

besseren Ergebnisse erzielen, als mit dem neuen Konzept. Letztes sei ja sogar im Nachteil, da die Lehrer für dieses keine wertvolle Unterrichtserfahrung vorweisen konnten²¹. Die didaktische Kompetenz der Lehrkräfte hinsichtlich Fehlvorstellungen hatte nach dem traditionellen Konzept durchaus Einfluss auf den Lernerfolg der Schülerinnen und Schüler. Beim Unterrichten nach dem 2D-Konzept aber blieb dieser Effekt aus. Tobias begründet dies damit, als dass dieses Konzept ja bereits gängige Fehlvorstellungen berücksichtigt und so der Faktor *Lehrkraft* in dieser Hinsicht nicht mehr so stark zum Tragen kommen muss, als es bei traditionellem Unterricht der Fall wäre.

Der Faktor *zusätzliche Unterrichtszeit* wirkt sich nur im 2D-Konzept lernförderlich aus. Im traditionellen Unterricht wurde durch noch größere Unterrichtszeit kein Zugewinn mehr erreicht, was Tobias auch hinsichtlich der Lehrplangestaltung anmerkt.

AMENDA, 2017

Die folgenden Ausführungen beziehen sich ausnahmslos auf die Studie von Amenda (2017). Sie sind der entsprechenden Dissertation entnommen. Angaben, die nicht Amenda (ebd.) entstammen, sind somit die einzigen gekennzeichneten.

Studie „Amenda“

Schulbuchanalyse zur Darstellung der Kinematik im Bereich der Sek II. Danach Durchführung von drei Studien (Vergleichsstudie skalarer vs. vektorieller Ansatz, Laborstudie zum vektoriellen Ansatz, Feldstudie zum vektoriellen Ansatz), jeweils mit Neu-/Weiterentwicklung der Materialien.

Bei der Entwicklung des Unterrichtskonzepts für die Sekundarstufe II zur Kinematik wurde besonderes Augenmerk auf eine absolut konsistente und widerspruchsfreie Darstellung der Fachinhalte Wert gelegt. Dazu gehören exakte Definitionen, Unterscheidung zwischen Momentan- und Durchschnittsgrößen, Einführung vektorieller Größen und Pfeilbildern mit zugehörigem Formalismus.

Die häufig in Ansätzen „geforderte Entmathematisierung des Physikunterrichts“ führt nach Ansicht des Autors der Studie zum „hohen Preis der fachlichen Inkonsistenz“. Dies erklärt auch den als anspruchsvoll empfundenen Anteil an Mathematik.

In einer anfänglichen vierstündigen Unterrichtssequenz zum Vergleich von 1D- und 2D-Konzept unter Laborbedingungen wurde für eine Gruppe Material zu einem skalaren und für eine Gruppe Material zu einem vektoriellen Ansatz ausgearbeitet. Ein eigens entwickelter Test im Anschluss an die kurze Unterrichtssequenz zeigte keine signifikanten Unterschiede

²¹Als Anmerkung kann hierzu gesagt werden, dass man sich auf neue Unterrichtskonzepte evtl. akribischer vorbereitet, als auf Vertrautes.

zwischen beiden Gruppen, gleichwohl auch Spezifika des 2D-Ansatzes aus Fairnessgründen nicht getestet wurden. Die ähnlichen Ergebnisse seien unter anderem auf die kurze Interventionszeit wie auch auf fehlende unterrichtliche Bewertung des Fachtests zurückzuführen. Es zeigte sich aber, dass der scheinbar komplexere 2D-Ansatz in der kurzen Zeit zu unterrichten ist.

In einer weiteren Lernwirksamkeitsstudie ausschließlich zum vektoriel- len Ansatz, die ebenso durch den Studienleiter selbst durchgeführt wurde, zeigten sich vom Pre- zum Posttest deutliche Unterschiede mit großen Effektstärken. Besonders die Feinheiten zum Geschwindigkeitsbegriff (die konsequente Differenzierung zwischen Weg und Ortsverschiebung) konnten gut gelernt werden. Ebenso zeigte sich, dass die nötige Vektor- rechnung prinzipiell „integrativ“ vermitteln lässt und kaum separaten Unterricht erfordert.

In einem dritten Durchgang wurde der entwickelte Lehrgang in einer Feldstudie durch unterschiedliche Lehrkräfte im normalen Unterricht integriert. Auch hier zeigten sich deutliche Lernzuwächse mit großen Effektstärken, auch hier wieder im Bereich der Unterscheidung zwischen Ortsverschiebung und Weg. Im Vergleich zur zweiten Lernwirksamkeits- studie unter Laborbedingungen zeigten sich jedoch keine Unterschiede in den Endergebnissen. Dass die Vektorrechnung „on the fly“ zu unterrichten ist – auch im üblichen Unterricht – war ebenfalls Erkenntnis dieser dritten Teilstudie.

3.7. Schlussfolgerungen für die empirische Studie

3.7.1. 1D vs. 2D: Deshalb lohnt sich ein weiterer Vergleich

Es ist mittlerweile unstrittig, dass zweidimensionale Zugänge lernwirk- sam sind, da dies in unterschiedlichen Settings gezeigt wurde (u. a. To- bias 2010; Jung, Reul und Schwedes 1977; Wilhelm 2005). In diesem Abschnitt sollen dennoch Argumente dargelegt werden, weshalb ein wei- terer Vergleich der Konzepte als sinnvoll erachtet werden kann. Dabei wird nach inhaltlichen sowie forschungsmethodischen Aspekten unter- schieden. Zunächst soll nochmals konstatiert werden, dass als Vergleich zum 2D-Konzept mitunter häufig statische Konzepte als „klassisches“ Gegenstück dargestellt werden. Hier soll als solches aber stets das 1D- Konzept herangezogen werden, denn auch dort lassen sich dynamische Zugänge realisieren.

INHALTLICH

Die Newtonsche Bewegungsgleichung in der Form $\vec{F} \cdot \Delta t = m \cdot \Delta \vec{v}$ wird im dargestellten 2D-Konzept den Evaluationen nach erfolgreich genutzt. Dabei werden vor allem die dargestellten und für Schüler einleuchtenden Je-Desto-Zusammenhänge als gewinnbringend herausgestellt (Tobias 2010). Die ersichtliche gemeinsame, gleiche Richtung des Kraft- und Zusatzgeschwindigkeitsvektors tritt im Rahmen der Behandlung der Gleichung vor den dominierenden halbquantitativen Zusammenhängen jedoch eher in den Hintergrund. Somit verschwimmen die Gründe, die die Gleichung für einen Einsatz im zweidimensionalen prädestinieren. Die skalare Variante $F \cdot \Delta t = m \cdot \Delta v$ erlaubt ebenfalls die halbquantitative Betrachtung von Kräften²².

Die Wirkung einer Kraft als „Änderung der Bewegungsrichtung“ ist auch in aktuellen, klassischen Lehrwerken und Lehrgängen enthalten, die sich am eindimensionalen Geschwindigkeitsbegriff orientieren. Die sich so allerdings stellende Hürde könnte sich auf den Begriff der „Bewegungsrichtung“ beziehen, der sich aus dem Eindimensionalen nicht intuitiv auf Bewegungen in der Ebene übertragen lässt (siehe *Ziel* vs. *Richtung*, 3.3). Generell sind aber Richtungsänderungen auch in 1D darstellbar, wenn auch begrenzt auf die „Hin- und Rückrichtung“. In diesem Sinne kann die Vernachlässigung von Bewegungsrichtungen außerhalb der einen Dimension als Elementarisierung der *Kraftwirkung als Änderung der Bewegungsrichtung* durch Partikularisierung (Reinhold 2006) aufgefasst werden. Weiterhin wird das Vernachlässigen des relativ komplexen Beschleunigungsbegriffs als sinnvoll erachtet, auch wenn durch die curricularen, äußeren Zwänge dies kaum möglich scheint. Das aber mögliche, spätere separate Behandeln der Beschleunigung ist jedoch kein Merkmal von 1D- oder 2D-Mechanik, sondern lässt sich in beiden Lehrgängen gleichermaßen durchführen. Somit lässt sich der Beschleunigungsbegriff auch im Rahmen der eindimensionalen Mechanik vermeiden.

Weiterhin stellt sich die Frage nach der Relevanz der „vollständigen Beschreibung der Mechanik“ (Jung, Reul und Schwedes 1977), die nur durch zweidimensionale Mechanik zu erreichen sei. In diesem Zusammenhang wird neben der Überwindung von Fehlvorstellungen in der Regel das Verständnis der Kreisbewegung²³ als substanziell für den Unterricht angesehen (Jung, Reul und Schwedes 1977; Tobias 2010; Wilhelm 2005). Ist es jedoch eigentlich das Bestreben des 2D-Ansatzes gerade keine Reduktion auf Spezialfälle zu betreiben, so ist die Kreisbewegung doch wieder ein

²²Mit der Einschränkung sich bei der – auch in 1D-Lehrgängen – vektoriellen Kraft auf deren Betrag beschränken zu müssen.

²³Was streng genommen auch zur Überwindung von Fehlvorstellungen (hier: Kreisbewegung mit $|v| = \text{const.}$ benötigt keine Kraft) gehört.

Spezialfall. Viel entscheidender für das 2D-Konzept sind nach eigener Einschätzung folgende drei Aspekte:

- Krummlinige Bewegungen (allg. Fall!)
- Vektoren/Pfeile als universelles Hilfsmittel in der Physik
- geschicktes und bewusstes Auswählen von Präzisierungen zu bestimmten Zwecken

Zur Erläuterung: Kreisbewegungen sind ein Spezialfall von krummlinigen Bewegungen, zu denen die Mehrzahl der alltäglich zu beobachtenden Bewegungen ohne weitere Reduktion gehören dürften (Auto fährt über Autobahn, Eichhörnchen durchquert Garten . . .). Daher sollte der Fokus im Zuge des 2D-Ansatzes auch bei diesen liegen, was in der Argumentation selten der Fall ist (vgl. Wiesner 2011). Über den Pfeil als Hilfsmittel der Physik schreibt Boczianowski (2011) umfassender. In seiner Eigenschaft als Vektorpfeil kommt er als Geschwindigkeitspfeil, Kraftpfeil oder Pfeil zur Darstellung des elektrischen Feldes daher, um nur einige Beispiele in der Physik zu nennen. Im Rahmen der zweidimensionalen Mechanik kann so also wichtige Vorarbeit geleistet zu werden, um Pfeile von ihrer festen Verknüpfung mit *einer* physikalischen Größe (wie der Kraft) zu lösen (ebd.). Streng genommen ist es natürlich auch denkbar den Pfeil in der 1D-Mechanik als Hilfsmittel einzusetzen, wobei hier das Potential bei weitem nicht ausgeschöpft wird.

Das geschickte Auswählen von Präzisierungen (z. B. „Ist vektorieller Geschwindigkeitscharakter hier relevant oder nicht?“) kann – in beiden Ansätzen – sinnvoll sein, um sich bestimmte Fälle zu vereinfachen. Als Beispiel soll hier die Bewegung eines Körpers in der Ebene und dessen Darstellung in t-x Diagrammen genannt sein. Hier könnte der vektorielle Charakter des Ortes hinderlich sein, da so eine komponentenweise Darstellung erforderlich würde. Es könnte also eine gewisse Durchlässigkeit zwischen 1D und 2D-Ansatz erstrebenswert sein.

Diese drei Argumente könnten die vermeintliche Überbetonung der Kreisbewegung als Hauptargument abmildern und die Verwendung der 2D-Mechanik bzw. die Durchlässigkeit zur 1D-Mechanik zugänglicher machen.

Weiterhin wurden die Auswirkungen des unterrichteten Konzepts auf das Verständnis von Zeit-Ort- oder Zeit-Geschwindigkeits-Graphen bisher nur in Ansätzen untersucht. Tobias (2010) zeigt, dass 1D-unterrichtete Schüler sich in einer Selbsteinschätzung eine höhere Kompetenz im Interpretieren von Graphen zuschreiben. Die Vermutung liegt daher nahe, dass das 1D-Konzept bezüglich dieser Fähigkeit, die in allen Curricula verankert ist, erfolgreicher sein könnte. Dies kann ebenfalls in weiteren Untersuchungen durch Fachtests analysiert werden.

Als neuartiger Aspekt ist weiterhin relativ ungeklärt, inwiefern sich die Anzahl der betrachteten Dimensionen auf die kognitive Belastung (Cognitive Load) beim Bearbeiten der Inhalte und Aufgaben in der Mechanik auswirkt. Besonders interessant scheint diese Frage im Zusammenhang mit Low- und Highachievern.

Zusammenfassend lohnt sich ein weiterer Vergleich zwischen beiden Konzepten, durch das weitere Ausreizen des 1D-Konzepts. Die Behandlung der Newtonschen Bewegungsgleichung in der integralen Form bietet sich auch für dieses Konzept an. Ebenso können Richtungsänderungen in partikularisierter Form explizit behandelt werden. Weiterhin lohnt sich der Vergleich der Graphenkompetenz unterschiedlich unterrichteter Kinder. Ein potentiell unterschiedlicher Cognitive Load bedarf ebenfalls einer Untersuchung. Weiterhin bietet der 1D-Ansatz für viele Spezialfälle günstige Vereinfachungen, wohingegen der 2D-Ansatz ihm zu eigene Stärke beim umfänglichen Verständnis krummliniger Bewegungen ausspielen kann.

Was einen Vergleich aus forschungsmethodischer Sicht sinnvoll macht sei im folgenden Abschnitt dargelegt.

FORSCHUNGSDESIGN

Im deutschsprachigen Raum stellt die Studie von Tobias (ebd.) mit Stand von 2018 die einzige Untersuchung zur Wirkung von traditionellem und 2D-dynamischem Unterricht dar, die in größeren Umfängen den Anfangsunterricht untersucht. In diesem Rahmen stellt sich die Frage nach einer gewissen Replizierbarkeit der Ergebnisse. Das Forschungsfeld der Replikation genießt in der Erziehungswissenschaft nicht den Stellenwert, wie z. B. in naturwissenschaftlichen Disziplinen (Bienefeld 2017), obwohl es dazu beitragen kann, „Objektivität in einem höheren Maße zu gewährleisten.“ (ebd.) Eine Replikationsstudie ist dabei im weitesten Sinne eine Wiederholung einer bereits durchgeführten Studie, mit dem Ziel die Ergebnisse dieser zu überprüfen (vgl. Bienefeld 2017; **Eredfelder.2018**). Der Begriff der Replikationsstudie lässt sich dabei aber nicht scharf fassen. Nach Bienefeld (2017) gibt es ein großes Spektrum an „Präziserungsansätzen“ den Begriff betreffend. Als möglich erachtet er jedoch die Unterscheidung zwischen *direkter* und *systematischer Replikation*²⁴. Bei einer direkten Replikation wird dabei versucht „eine Untersuchung möglichst exakt zu wiederholen“ (ebd.). Im Gegensatz dazu werden bei systematischen Replikationen bestimmte Variablen verändert, wobei die Ausprägung und Anzahl dieser Variablen nicht unstrittig ist (Bienefeld 2017; vgl. auch Bortz und Döring 2006). Dies entspricht auch dem funktionalen Ansatz nach Schmidt (2012), dessen „grundlegende Idee es ist, [...] herauszuar-

²⁴Eine ähnliche Unterscheidung nimmt Eredfelder und Ulrich (2018) mit *konzeptuellen* und *direkten* Studien vor

beiten, welche Elemente in der Replikationsstudie konstant gehalten und welche verändert werden müssen.“ Entsprechend bietet sich hier an, die 1D- und 2D-Ansätze in einem weiteren Untersuchssetting im Rahmen einer Replikationsstudie, die funktionalen/systematischen Ansprüchen genügt, zu untersuchen. In solch einem anderen Setting „können auch Replikationen durchaus originell und spannend sein“ (Bortz und Döring 2006).

Ein weiterer Grund für weitergehende Untersuchungen liegen im Kontrollgruppendesign bisheriger Studien. So wurden bisher in der Regel als Kontrollgruppe ein „traditionelles“ nicht-modifiziertes Konzept gegen ein in einer spezifischen Art und Weise neues Konzept getestet (z. B. Tobias 2010; Wilhelm 2005). Argumentativ wurde der Erfolg des 2D-Konzepts zumindest im ersten Fall dadurch gestärkt, dass das „neue“ 2D-Konzept auch bei auf diesem Gebiet unerfahrenen Lehrkräften erfolgreich(er) unterrichtet werden konnte. Sicherlich ist dies durchaus bemerkenswert. Jedoch stellt sich die Frage, inwiefern eine besondere Hingabe beim Vorbereiten des „neuen“ Unterrichts im Vergleich zum Unterricht, der seit Jahr und Tag durchgeführt wird, erfolgte. Zur Abmilderung eines unsymmetrischen Neuheitseffektes (siehe 7.1.2) sollte daher besonderer Wert auf einen ausgeglichenen Vergleich gelegt werden. Denn nur in einer Studie wurden zu beiden Zugängen neue Materialien entwickelt (Amenda 2017). In manchen (bzw. Teilen mancher) Studien wurde auch lediglich ein neues Konzept getestet (auch Amenda 2017; Jung, Reul und Schwedes 1977). Des Weiteren wurden zum Erfassen des Lernzuwachses in der Regel gleiche Pre- und Posttests verwendet, was gerade bei unintuitiven Sachverhalten, wie sie die Mechanik darstellt, einen großen Lernzuwachs verspricht. Zur Abmilderung dieses Effekts können unterschiedliche Pre- und Posttests verwendet werden, die über eine Messung der Personenfähigkeit nach dem Rasch-Modell über bestimmte gemeinsame Items geankert sind (siehe Kapitel 10.3).

AUSBLICK

In diesem ersten Teil der Arbeit wurde die fachdidaktische Problemstellung inklusive zweier Unterrichtskonzepte aufgezeigt. Nach Zusammenfassung bisheriger Ergebnisse wurden zuletzt die Gründe für weitere Forschung dargelegt. Für eine weitere Ausschärfung dieses Vorhabens werden im nächsten Teil der Arbeit Aspekte von neuen Medien dargelegt, die ihrerseits einen Beitrag zum Lernen von Mechanik leisten können.

4. „Neue Medien“ im unterrichtlichen Einsatz

4.1. Von Computerraum und Whiteboard zum Tabletcomputer

„Wie schön war es noch vor 30 Jahren. Das höchste Maß an Medienkompetenz erreichte man durch einen Vorführschein für 16-mm-Projektoren, selbst der Videorekorder hatte noch keinen Einzug in die Schulen gehalten. Ein Film im Unterricht? Die Schüler freuten sich!“

(Ristic 2006)

Bezeichnet man einen Vermittler von Information als Medium (vgl. Girwidz 2010), verfügten Schulen über eine Medienausstattung, schon lange bevor elektrisch betriebene Medien im Unterricht angewandt wurden. Girwidz (ebd.) nennt dafür beispielhaft „vortechnische Medien“, wie Kreidetafel, Buch (siehe auch Bleichroth 1991) oder Modelle, welche jeweils schon Anfang des letzten Jahrhunderts verwendet wurden. Später folgte die Verbreitung von „technischen Medien“ (also im weitesten Sinn elektrisch betriebenen) wie „Ton-, Bild- oder visuelle[n] Medien“ (Girwidz 2010). Schon in den fünfziger Jahren wurden von Hahn (1955) neben Dias Lehrfilme empfohlen, um die „Anregung“ und die „Anschauung“ zu fördern oder „nicht beobachtbare Vorgänge anschaulich sichtbar zu machen“¹. Dabei wird sogar schon die Unterscheidung zwischen Realfilmen und Trickfilmen vorgenommen. Der dort in diesem Zusammenhang erwähnte 16 mm-Film war im Unterricht bis in die 90er Jahre präsent (vgl. hierzu Bleichroth 1991). Der später aufkommende Overheadprojektor ist je nach Schule bis heute ein (noch!) alltägliches Medium im Unterricht, auch wenn seine Bedeutung zu Gunsten von interaktiven Whiteboards (IWB) zurückgeht (vgl. Obst 2013). Schon lange Zeit vor Einführung dieser Whiteboards fanden Computer in Schulen Verbreitung (Welling 2017) – wenn auch in wesentlich kleinerer Stückzahl und zentraler als es später mit Tablets möglich sein würde. Schon Bleichroth (1991) erkennt zu dieser Zeit,

¹Also garnicht so unähnlich den heutigen Animationen und Simulationen!

dass „so viel Neues durch den Computer möglich wird und weil man soviel Vertrautes mit ihm ganz anders als bisher machen kann“ und der Computer einen daher zwingt „wie kaum ein Gerät zuvor, über Sinn und Ziele des Physikunterrichts nachzudenken“ (Bleichroth 1991). Darüber hinaus finden auch schon in Bleichroth (ebd.) die vielfältigen Einsatzmöglichkeiten des Computers Erwähnung, in dem er die Kombination unterschiedlicher Medien in einem Gerät hervorhebt.

Neue Medien

Whiteboards gelten gemeinsam mit Computern bzw. Tablets als Teil der gegenwärtigen „Neuen Medien“. Diese Begrifflichkeit unterliegt dem Wandel der Zeit (Ristic 2006; Berghaus 1999) und wird relativ unscharf gebraucht (Kircher, Girwitz und Häußler 2010), umfasst dabei aber Anwendungen von Computer und Internet, „virtuelle Schulen“ (ebd.) und allg. eine technische Verbesserung – vom schnelleren Zugriff auf Daten, Interaktivität bei der Nutzung als auch schlicht Verbesserung der Bild- und Tonqualität.

Sofern gerade im IT-Sektor Zukunftsprognosen in der Vergangenheit, auch mittelfristig, häufig falsch lagen, so prophezeite Ristic (2006) schon vor geraumer Zeit, dass „tragbare Mini-Rechner [...] in Zukunft zum Unterricht“ gehören. Dabei schließt sie schon ausdrücklich Tablet-PCs mit ein. Das, was bis dato noch visionär erschien, sollte schon wenige Jahre später den Einzug in die Klassenzimmer finden.

Dass der Einsatz von Medien und Methoden nie zum Selbstzweck verkommen sollte, ist Konsens im praktischen Unterrichten

wie auch in der Forschung (siehe auch Abschnitt 4.2.3). Entsprechend gibt es keinen Grund die neuen Medien davon auszunehmen, zumal an diese mitunter erhebliche Erwartungen gestellt werden (vgl. Deimann 2002). Exemplarisch schreibt Weidenmann (2009), dass das Lernen mit Multimedia „endlich die Einlösung des Versprechens von Comenius [...] zu sein scheint, «alle Menschen alles zu lehren; und zwar zuverlässig zu lehren, [...] nicht oberflächlich und nur zum Schein»“. Zunächst gibt es beim Lernen von Mechanik jedoch keinen Grund anzunehmen, dass durch bloßen Technologieeinsatz erhebliche Lernfortschritte gemacht werden, nicht zuletzt da auch Wenzel (2018) konstatiert, dass durch Einsatz neuer Medien sich weder die Unterrichtsqualität noch der Unterrichtsstil generell im wesentlichen ändere. Entscheidend ist also die gezielte Einbettung – nicht nur der Geräte sondern auch von entsprechender Software, wie z. B. digitalen Lernumgebungen oder Simulationen, welche über Tablets einfach zugänglich gemacht werden können. Es kann vermutet werden, dass durch gezielten Einsatz von Medien, speziell Simulationen, im Sektor Mechanik Lernfortschritte erzielen lassen (Wilhelm 2005, uvm. Kreiten 2010).

Im folgenden Abschnitt werden nach einer Begriffsklärung daher Potentiale und Befunde zum Einsatz von Simulationen, die dem Lernen von Mechanik begründeterweise zuträglich sein könnten, dargelegt.

4.2. Über Simulationen und Lernumgebungen

4.2.1. Simulationen

Begriffliches

Zunächst gilt zu klären, was eine Simulation² ist und welche Einsatzgebiete sich identifizieren lassen.

Nach Wissenschaftsrat (2014) sind Simulationen „Experimente mit einem formalen Modell, die Fragestellungen adressieren und Erkenntnisse erzielen, die nicht oder nur mit großem Aufwand über direkte Beobachtung und Messung oder über ein reales Experiment (wie z. B. ein in vitro- oder in vivo-Experiment) gewonnen werden können“. Auch Kramer und Neculau (2000) sehen Computersimulationen „im engeren Sinne“³ damit harmonisierend als „Nachahmung realer Prozesse mittels Computern auf der Grundlage formaler (analytischer, numerischer (sic!) oder grafischer) Modelle“. Claussen (1989) fasst unterschiedliche Definitionen der damaligen Zeit zusammen indem sie sagt, dass die Definition einer Simulation die „Beinhaltung eines zeitlichen Vorgangs sowie der Verwendung eines Modells, das wesentliche Elemente des simulierten Systems nachahmt“ einschließt. Ebenso betont sie den damit verbundenen Erkenntnisgewinn.

Doch welchem Zweck dienen Simulationen in Wissenschaft, Technik und Bildung allgemein? Aktuelle Einsatzgebiete von Simulationen in diesen Feldern lassen sich beispielhaft in sechs Teilgebiete unterteilen (Wissenschaftsrat 2014, auch genannte Beispiele).

- Prognosefunktion (z. B. Klimaentwicklung), Interventionen entwerfen, Bewusstsein schaffen
- Theoriebildung, verbesserte Modelle erstellen
- Optimierung Flughafenbetrieb, Materialeigenschaften
- Analyse großer Datenmengen, Mustererkennung, Big Data
- Präsentation und Kommunikation komplexer Zusammenhänge
- Echtzeitsimulationen: Pilotentraining, Anlagenfahrer im Kraftwerk

²Im Folgenden sei angenommen, dass es sich hierbei stets um Computersimulationen handelt

³D. h. eingeschränkt auf Simulationen per Computer (Kramer und Neculau 2000)

Nach Betrachtung dieser Punkte fällt auf, dass die ersten vier Aspekte dem Befriedigen eines allgemeinen Erkenntnisinteresses dienen, um für die Gemeinschaft bisher *Unbekanntes* zu erforschen. Die letzten beiden Aspekte verfolgen in erster Linie keinen Erkenntnisgewinn, sondern vielmehr das Nachahmen von bereits Bekanntem. Speziell die letzte Funktion von Simulationen hat „Unterstützungs-Charakter“, d.h. die Simulationen dienen Trainingszwecken und nicht dem Gewinnen von Informationen. Der vierte und fünfte Punkt ähnelt dem Einsatz im Klassenraum. Es deutet sich also an, dass bei der Klärung der Verwendung von Simulationen der Einsatz in der Schule vom Einsatz in der wissenschaftlichen Forschung/des technischen Alltags abgrenzen lässt. Grundlegend ist bei ersterem das Anstreben eines Erkenntnisgewinns *aller Beteiligten* bzgl. eines Sachverhalts. Bei letzterem, dem Einsatz von Simulationen im didaktischen Sinne, sind die fachlichen Inhalte der wissenschaftlichen Community wohlbekannt. Eher stehen dort die (interaktive) Präsentation und Kommunikation von Zusammenhängen (Beispiel: Das Kondensator-Labor der phet-Simulationen⁴) im Vordergrund. Mitunter sind aber auch die Echtzeitsimulationen, die zum Erlernen/Üben von Fertigkeiten dienen (Beispiel: Bedienung eines Oszilloskops) denkbar. Analog dazu verhält es sich mit auch mit traditionellen Schulexperimenten. Bei diesen sind die Ergebnisse der Versuche mitunter zwar nicht beliebig genau bekannt, es werden aber keine neuen Erkenntnisse für die wissenschaftliche Gemeinschaft gewonnen. So konstatiert auch Berger (2006b) „Im Gegensatz zur wissenschaftlichen Forschung wird es hier ein probates Mittel zum Heranführen an bereits gesicherte Erkenntnisse sein“. Daher haben Simulationen im hier vorliegenden Rahmen einen ähnlichen Pseudo-Versuchscharakter wie Realexperimente im Schulbetrieb.

Eine letztlich für den didaktischen Zweck zugängliche Definition liefert Rieber (2005) nach Niegemann u. a. (2008): „[Simulationen sind] Computerprogramme, die Phänomene oder Aktivitäten modellieren und die dafür vorgesehen sind, dass die Nutzer durch Interaktionen mit ihnen etwas über diese Phänomene und Aktivitäten lernen“. Passend ist diese Definition auch, da sie die Gründe für den Einsatz von Simulationen eher didaktisch fokussiert⁵ als die Definition/Charakterisierung nach Wissenschaftsrat (2014, s. o.), in der die *Undurchführbarkeit* realer Experimente dominiert.

ANIMATION UND SIMULATION

Animation und *Simulation* sind Begriffe, die häufig gemeinsam genannt werden. Da sich diese Arbeit auf eine mit Simulationen gestützte Lernum-

⁴<https://phet.colorado.edu/de/simulation/capacitor-lab> , abgerufen am 28.9.2018

⁵Genauere Gründe für den Einsatz von Simulationen werden im Verlauf genannt.

gebung bezieht, sollen diese beiden Begriffe nun voneinander abgegrenzt werden – sofern möglich.

Schnotz und Lowe (1998) verstehen unter einer Animation „a pictorial display that changes its structure or other properties over time and which triggers the perception of a continuous change“, also eine grafische Darstellung deren Erscheinungsbild/deren Eigenschaften sich ändern um die Wahrnehmung einer kontinuierlichen Veränderung zu erzeugen. In diese Definition reiht sich Berger (2006a) ein, der Animationen als „trickfilmartig[e], meist stark vereinfachte [...] Veranschaulichungen physikalischer Sachverhalte“ bezeichnet. Der trickfilmartige Aspekt wird auch von Be-trancourt und Tversky (2000) gestützt, die Animationen als Abfolge von Bildern sehen („any application which generates a series of frames“).

Diese Konkretisierungen sind bislang vom Grad der Interaktion der Nutzenden mit der Animation wie auch von der Verwendung von Modellen unabhängig. So können auch Darstellungen, die einen „hohen Grad an Interaktivität“ (Niegemann u. a. 2008) aufweisen als Animation bezeichnet werden. Auch Zang (2012) greift dies auf, indem er darlegt, dass „auch Programme, die solche Bilder direkt berechnen, [von der Begriffsbestimmung als Animation] umfasst [werden], unabhängig davon, welchen Einfluss der Nutzer des Programmes auf diese Berechnung nehmen kann“. Ist nun alles, „was sich zu bewegen scheint“ und aus einzelnen Bildern zusammengesetzt ist, eine Animation?

Zumindest im Falle von *Simulationen* liegt die Annahme nah. Nach Be-trancourt (2005), kann eine Animation zur Simulation werden, sobald sie einen bestimmten Grad an Interaktion aufweist. Wie groß dieser Grad ist, wird dabei nicht näher spezifiziert. Konsens herrscht jedoch darüber, dass eine Simulation Interaktion erlauben muss (Berger 2006a; Niegemann u. a. 2008). Simulationen können sich bei graphischer Visualisierung somit durchaus als Untergruppe von Animationen auffassen lassen.

Auch Berger (2006a) bemerkt eine unscharfe Grenze zwischen Animation und Simulation. So können Simulationen ihre Ergebnisse durchaus grafisch animieren⁶, während Animationen zur Darstellung Modelle benötigen.

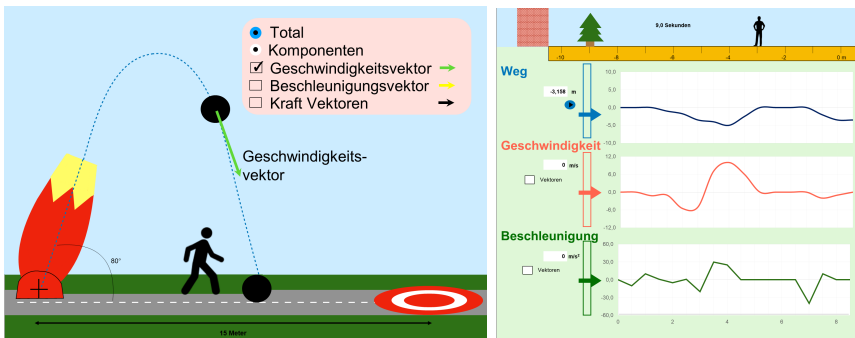
Lernen mit Simulationen: Potentiale

Doch weshalb sollen sich Simulationen, wie sie im obigen Abschnitt beschrieben wurden, zum Lernen von Mechanik eignen, welche Vorzüge liegen empirisch oder praktisch abgesichert vor?

⁶Das ist nicht selbstverständlich! Die Ergebnisse der Algorithmen könnten ohne weiteres z. B. auch einfach als Zahlenwerte ausgegeben werden.

An verschiedener Stelle nennen Girwidz (2015) und Kreiten (2010) vorteilhafte Einssatzszenarien für Simulationen. Dazu zählen in erster Linie Simulationen als Ersatz für Experimente, die in der Realität nur schwer oder nicht realisierbar wären. Dies kann an der Gefährlichkeit (hohe Spannungen, ionisierende Strahlung) liegen, wie auch an offensichtlichen Hindernissen (Simulation von Wetterphänomenen oder Stößen von Autos) oder an fehlendem Experimentiermaterial. Abgesehen von letztgenanntem Grund treffen die genannten, eher oberflächlichen, Aspekte auf Versuche der Mechanik je nach Kontext eher nicht zu. Hinsichtlich Einssatzszenarien betonen Kreiten (ebd.), dass „Simulationen [. . .] nie *anstelle* von realen Experimenten eingesetzt werden [sollen]“, nennen an dieser Stelle aber keine Begründung. Für den Autor dieser Arbeit jedoch soll die Wahl zwischen Simulation oder Realversuch i. d. R. keine dogmatische sein, sondern sich an Gegebenheiten und Zielen orientieren.

Interessanter stellt sich die Situation bei einem Blick durch die didaktische Brille dar. So verfügen Simulationen über umfangreiche Möglichkeiten der grafischen Darstellung und somit auch zur Verknüpfung von Darstellungsformen. Je nach verwendeter Plattform lassen sich die klassischen physikalischen Darstellungsformen Text, Bild, Formel, Graph oder Tabelle flexibel einbinden (Blake und Scanlon 2007).



(a) Verknüpfung mit symbolhaften Darstellungen

(b) Verknüpfung mit Graphen

Abbildung 4.1.: Verknüpfung mit unterschiedlichen Darstellungsformen (Darstellung angelehnt an: Phet-Simulations)

Entsprechend Abbildung 4.1 bedeutet dies beispielhaft, dass Vektorpfeile, die in der Realität natürlich nicht gesehen werden, dargestellt werden können – man also Bilder mit symbolhaften Darstellungen verknüpfen kann. Denkbar wäre zwar auch, im Realversuch „Papppfeile“ auf Spielzeugautos oder Bälle zu kleben, was jedoch das Ändern des Tempos wie auch

das Zu- und Abschalten der Darstellung fast unmöglich macht oder perspektivische Verzerrungen mit sich bringt. Die Verknüpfung von Graphen und Bildern zeigt die rechte Abbildung 4.1. Hier wird zur eigentlichen bildlichen Darstellung der Simulation die Darstellung des zugehörigen t-x-Diagramms eingeblendet. Auf diese Weise lassen sich zwischen der hier dargestellten Bewegung und dem Diagramm entsprechende Verknüpfungen herstellen. Mit der Verknüpfung von Darstellungsformen verbunden ist das Potential von Simulationen, auch unsichtbare Sachverhalte sichtbar darstellen zu können (Girwidz (2015) nach Rutten, van Joolingen und van der Veen (2012)). Girwidz (2015) zeigt als Beispiel eine interaktive Wellenmaschine oder die visualisierte Ausbreitung von elektromagnetischen Wellen.

Aus der Definition von Simulationen (Abschnitt 4.2.1) folgt fast direkt, dass je nach Auslegung vielfältige Variationen von Bedingungen durch den Bediener vorgenommen werden können, was „considerable potential for the enhancement of teaching“ bedeuten kann (Blake und Scanlon 2007). Variablen können also einfach verändert werden, was auch „Was wäre wenn“-Experimente ermöglicht (Kircher, Girwidz und Häußler 2015). So könnten im Bereich der Mechanik je nach Simulation, Massen, Geschwindigkeiten, Steigungen, Kräfte oder Wirkungen der Gravitation problemlos verändert werden. Zu dieser umfassenden Kontrolle über Bedingungen gehört auch der große Vorzug, dass – scheinbare – Störungen zu- oder ausgeschaltet werden können oder direkt bei der Erstellung vernachlässigt werden. Im Bereich der Mechanik betrifft dies beispielsweise Verluste durch Reibung o.Ä. (Kreiten 2010). Mitunter können auch darüber hinaus bestimmte Ungenauigkeiten wie z. B. vereinfachte Stoßmodelle, in Kauf genommen werden, sofern sich diese didaktisch einwandfrei rechtfertigen lassen (vgl. „Komplexität der Sachstruktur verringern (ebd.)“). Auch durch solche Reduzierungen kann die Aufmerksamkeit der Bediener auf wesentliche Aspekte gesteuert werden.

Ebenfalls speziell für den Bereich der Mechanik sind die von Urhahne u. a. (2000) genannten Vorzüge von Animationen⁷ zu nennen. Sie stellen fest, dass Animationen einen „fehlenden inneren Prozess, [...] die Imagination dynamischer Abläufe, ersetzen“ (ebd.), indem sie ihn visualisieren. Dies führe dazu, dass Schüler „dynamische Konzepte und Veränderungen über die Zeit schneller [...] begreifen“. Auch Weidenmann (2009) empfiehlt zur Unterstützung der Konstruktion von flexiblen, mentalen Modellen Darstellungen, „die Dynamik vorführen oder die sich durch die Lernenden auf Wunsch dynamisieren lassen“.

⁷Die enge Beziehung zwischen Animation und Simulation durch Darstellung bewegter Bilder ist in Abschnitt 4.2.1 dargelegt. Daher werden die Vorzüge von ersten hier ebenfalls eingebracht

Weiterhin unterstellen sie auch (auch wenn Evidenzen dazu unterschiedlich ausfallen) eine potentiell geringere kognitive Belastung beim Lernen mit Animationen, da Informationen statischer Bilder nicht erst verknüpft werden müssen⁸.

Entdeckendes Lernen mit Simulationen

Aufgrund der genannten Vorzüge, vor allem hinsichtlich der vielfältigen Darstellung und einfachen Veränderung diverser Einflüsse und Einstellungen und der „Interaktion mit dem Material“ (Girwidz 2015), bieten sich Simulationen an, Aspekte des *entdeckenden Lernens* zu fördern (vgl. Jong und van Joolingen 1998; Girwidz 2015). Die Eignung von entsprechenden Simulationen für diesen Zweck bestätigen auch Eckhardt u. a. (2013): „Computer simulations fit best with principles of scientific discovery learning. Learners are asked to infer the characteristics underlying the scenarios when they explore the principles, concepts, rules, and terms of simulated experiments.“ Doch was soll hier unter entdeckendem Lernen verstanden werden? Unter entdeckendem Lernen, welches begrifflich im Allgemeinen auf Bruner zurückgeführt wird, versteht Höttecke (2010) dabei das Freilegen von Erkenntnissen. Dies sei zwar unter Nature of Science Aspekten problematisch (denn warum sollten sich physikalische Erkenntnisse mit den Jahren wandeln, wenn sie lediglich „bereits vorgefertigt“ versteckt sind?). Da jedoch Schüler mit einem „schon verfügbaren und historisch gewachsenen Kanon an fachtypischen Konzepten, Begriffen und Problemlösestrategien konfrontiert“ (ebd.) werden, kann Entdecken als Aneignung von bereits Bekanntem bezeichnet werden. Die Betonung liegt hierbei eher auf dem Produkt des Lernens, das handlungsorientiert mit „Eigenverantwortung und Selbstbestimmung“ erarbeitet wird (ebd.). Die dazu erforderlichen kognitiven Prozesse Orientierung, Hypothesenbildung, Überprüfung der Hypothese und Schlussfolgerungen ziehen (Girwidz 2015, nach de Jong (2011)) können durch Simulationen angeregt werden (ebd.). Im Falle der Mechanik könnte dies bedeuten, dass die Schüler mittels der Simulationen bestimmte Beobachtungen machen und Zusammenhänge entdecken. Dies können Zusammenhänge zwischen Strecke und Zeit, Kraft und Geschwindigkeitsänderung oder Masse und Zusatzgeschwindigkeit sein.

Nach Lage der Literatur wäre es jedoch fatal anzunehmen, dass die Schüler sich absolut frei mit Simulationen beschäftigen und dabei die „versteckten“ Phänomene selbst entdecken. So konstatieren Eckhardt u. a.

⁸Mit Blick auf Simulationen formuliert Varnai (2006) mit Rückgriff auf Schnotz, „die Explorationsmöglichkeiten [Anm.: in Simulationen] bedeuten [...] zusätzliche kognitive Anforderungen“. Es liegt also die Annahme nah, dass ein ausgewogenes Verhältnis zwischen Explorationsmöglichkeiten und Darstellung gegeben sein sollte.

(2013), dass „there is ample empirical evidence to support the conclusion that pure and independent exploration of scientific principles, concepts, and rules does not lead to effective learning outcomes“. Entsprechend betont Girwidz (2015) die intensive Diskussion von „Notwendigkeit und Ausmaß von sinnvoll lenkenden und unterstützenden Maßnahmen.“

Begründet werden kann dies nach Eckhardt u. a. (2013) durch Probleme, die Jong und van Joolingen (1998) wie folgt ausführen:

Generieren von Hypothesen: Schülern ist mitunter unklar, wie eine Hypothese aufgebaut und formuliert wird. Weiterhin fällt es auf Basis der Beobachtungen an einer Simulation schwer, eine aufgestellte Hypothese abzuändern. Darüber hinaus könnten eher unwahrscheinlich zutreffende Hypothesen von vornherein nicht aufgestellt werden (ebd.).

Planen von Experimenten: Hier kommt der confirmation bias zum Tragen. Dies bedeutet, dass die Simulationen so ausgeführt oder beobachtet werden, dass die Vermutungen der Schüler in jedem Falle zutreffen („[students] search for evidence that support their current hypothesis“ (ebd.)). Weiterhin können unnütze Variablen variiert werden oder Versuche geplant werden, die Freude bereiten aber keine Modelle untersuchen („students created experiments that were “fun““ (ebd.))

Interpretation von Daten: In diesem Bereich betonen Jong und van Joolingen (ebd.) vor allem Schwierigkeiten beim Interpretieren von Graphen.

Selbstregulation: Erfolgreiche Lerner zeichnen sich nach Jong und van Joolingen (ebd.) gegenüber nicht erfolgreichen vor allem durch systematisches Vorgehen und planmäßiger Organisation und Zielsetzung des gesamten Prozesses aus – etwas, was nicht allen Schülern im Umgang mit Simulationen obliegt.

Eine weitere Problematik, die sich im Ursprung auf Lernen mit Medien/Multimedia bezieht, wird von Weidenmann (2009) vorgebracht. Er stellt mit Rückgriff auf unterschiedliche Literatur fest, dass Lernen mit Medien als einfach empfunden wird: „Manche Medien, wie z. B. Fernsehen bzw. Video, gelten bei Nutzern als «leicht»; diese sind also überzeugt, dass es wenig Anstrengung bedarf, um damit zu lernen. Andere, wie z. B. das Buch, gelten als «schwierig». Bilder werden „angenehm und interessant erlebt“ aber „weniger intensiv verarbeitet [...], weil zum einen bildhafte Darstellungen (besonders wenn sie [...] realitätsnah sind) eher als «leicht»perzipiert werden“. Um diesem zu entgehen, benötigt es Instructional Support.

Jong und van Joolingen (1998) schlagen mit Rückgriff auf unterschiedliche Studien diverse Gegenmaßnahmen vor. Hinsichtlich des experimentellen Umgangs mit Simulationen wird auf sogenannte „Experimentierhinweise“ hingewiesen, welche sich zwar nicht zwingend auf das Fachwissen, aber auf das Experimentieren an sich positiv auswirkten (Beispiel:

„Verändere grundsätzlich nur eine Variable auf einmal!“). Solche Hinweise können auch adaptiv gegeben werden, d. h. angepasst an die Eingaben des jeweiligen Schülers. Nach Rückgriff auf Leutner (1993) betonen Jong und van Joolingen (1998) die Auswirkungen unterschiedlich detailreich ausgeprägter Hinweise, wobei jedoch grundsätzlich positive Effekte vorzuweisen sind. Solche (adaptive) Hinweise lassen sich sowohl in automatischem Feedback realisieren, als auch im Vorfeld in den Instruktionen zu den jeweiligen Simulationen (vgl. auch Urhahne u. a. 2000).

Solche Instruktionen schlagen Jong und van Joolingen (1998) mit Bezug auf de Jong (1994) vor. Beispielsweise das entdeckende Lernen den Zusammenhang zweier Variablen betreffend können „investigation assignments“, also eine Art *Forschungsaufträge*, sinnvoll erscheinen. Nach unterschiedlichen Studien attestieren Jong und van Joolingen (ebd.) positive Effekte.

Weiterhin wird durch Jong und van Joolingen (ebd.) (vgl. auch Urhahne u. a. 2000) eine *model progression* diskutiert. Gemeint ist damit eine aufsteigende Komplexität der Inhalte (am Beispiel *freie Schwingung, gedämpfte Schwingung, erzwungene Schwingung*). Dabei ergibt sich anhand verschiedener Befunde eine unklare Lage. In einfacheren Kontexten zeigen sich keine positiven Effekte, anhand eines komplexeren Kontext (Multimeter) zeigten sich durchaus positive Effekte. Die Grundlagen der Newtonschen Mechanik sind in diesem Sinne wohl durchaus als komplex einzustufen.

Dass eine Lernumgebung und Feedback potentiell ein Gerüst und Struktur für die Schüler beim Lernen bieten und ebenso Raum für zuvor beleuchtete Vorschläge bieten, soll im nächsten Abschnitt zu digitalen Lernumgebungen *knapp* skizziert werden (Das Kapitel zu Feedback folgt im weiteren Verlauf).

4.2.2. Lernumgebungen

Aufgrund der in Kapitel 4.2.1 aufgeworfenen Sachverhalte und Lösungsvorschläge soll hier ein Konzept zur Integration von Simulationen vorgeschlagen werden. Unstrittig ist bisher, dass das Lernen mit Simulationen für Schüler eines geeigneten Maßes an instruktionaler Unterstützung bedarf. Selbstverständlich kann diese geschickte Einbettung im laufenden Unterricht geschehen. Einen Überblick zu Prinzipien, die diese Einsatzform nahelegt, gibt beispielsweise Girwidz (2015). Mit Simulationen wird in dieser Arbeit eine Anwendung von neuen Medien als Lernangebot für Schüler genutzt. Für deren Einbettung soll im gleichen Maße das Potential einer *digitalen* Einbettung genutzt werden – im Gegensatz zur direkten Implementierung in den Unterricht. Dazu bietet sich eine digitale Ler-

numgebung an. Im Folgenden sollen neben einer Begriffsbestimmung die Gründe dafür dargelegt werden.

Worum handelt es sich bei einer digitalen Lernumgebung? Eine Lernumgebung an sich besteht nach Reinmann und Mandl (2006) „aus einem Arrangement von Unterrichtsmethoden, Unterrichtstechniken, Lernmaterialien [...] Medien“, welches „durch Unterricht [hergestellt wird]“ (ebd.). Es handelt sich entsprechend dieser Definition also um ein ganzheitliches Ensemble, mit welchem ein Setting für den Unterricht generiert wird. Eine entsprechend adaptierte Version einer digitalen Lernumgebung bedeutet also ein Zusammenspiel von digital aufbereiteten Lernmaterialien auf einem ebenso digitalen Medium, welches durch bestimmte Unterrichtsmethoden genutzt wird. Ähnlichkeiten bestehen mithin zu einem Lernprogramm, wie auch Krummeck (2008) feststellt. Der Begriff Lernumgebung werde oft „im Zusammenhang mit einem Hypertext- bzw. Hypermediasystem verwendet. Insofern werden Lernumgebungen oft als eine spezielle Form von *Lernsoftware* definiert“. Dem entgegen hält der Autor dieser Arbeit, dass eine Zusammenstellung aus Text, Bildern, interaktiven HTML5-Elementen oder externen Simulationen kein geschlossenes Software-Paket darstellt, sodass in dieser Arbeit weiter von einer (digitalen) Lernumgebung geschrieben wird. Die Vorzüge einer digitalen Lernumgebung⁹ für dieses Projekt seien im Folgenden skizziert:

Aus praktischen Überlegungen bietet sich eine Lernumgebung an um die Verfügbarkeit von Materialien auf relativ einfach und ressourcenschonende Art und Weise bereit zu stellen. So sind keine zusätzlichen Ausdrücke für Arbeitsblätter, Informationstexte und Bilder nötig. Diese können alle eingebettet werden. Weiterhin kann einfach durch sämtliche Materialien navigiert werden, zudem kann kein Material verloren gehen¹⁰. Darüber hinaus ist sichergestellt, dass jeder Schüler sein Lerntempo selbst regulieren kann.

Zur Unterstützung des Lernens bietet eine Lernumgebung in besonderem Maße Strukturen, die Schüler als entlastend wahrnehmen könnten. Dazu kann zunächst ein ähnlicher struktureller Aufbau für jedes Thema mit jeweils wiederkehrenden Elementen gehören. In Kapitel 4.2.1 wurde bereits die Notwendigkeit eines Scaffoldings für die Lernenden betont. Zudem kann durch eine gebündelte Darstellung der Informationen – ohne zwischen Papier und Tablet wechseln zu müssen – eine Milderung von Split-Attention-Effekten vermutet werden, da es „ressourcensparender für das Arbeitsgedächtnis des Lernenden [ist], zusammengehörige visuelle Informationsquellen auch gemeinsam zu präsentieren“ (Niegemann 2008). Weiterhin können die Simulationen insofern in eine Lernumgebung einge-

⁹Im Laufe der Arbeit wird der Kürze halber schlicht *Lernumgebung* genutzt.

¹⁰Was sich im Sinne der Selbstständigkeit womöglich *langfristig* negativ auswirken mag...

bunden werden, als dass sie *adaptiv* mit dieser interagieren können und so flexibel auf bestimmte Inputs reagieren können. Dies bietet z. B. die Möglichkeit automatischen Feedbacks (vgl. Richtberg (2018), der in einer Online-Lernumgebung ebenfalls automatisches Feedback nutzt).

Zweifelsohne lassen sich viele der Vorzüge auch mit „analogen Methoden“ herstellen, jedoch bietet eine Lernumgebung eine Bereitstellung aus einer Hand. Zudem ist die Forschungslage zu digitalen Lernumgebungen momentan eher überschaubar. Darüber hinaus lassen sich digitale Lernumgebungen auf Tablets betreiben, welche in Schulen zunehmend Verbreitung finden. Aus diesem Grund soll in Kapitel 4.2.3 über den Stand der Forschung zum Tableteinsatz gegeben werden.

Ein ganz anders gelagerter – aber dennoch wichtiger – Grund, die Mechanikkonzepte mit eigenen Lernumgebungen miteinander zu vergleichen ist in der Forschungsmethodik begründet, siehe dazu Kapitel 7. Denn nur mithilfe einer Lernumgebung gelingt es, die Lehrervariable zu minimieren. Somit werden alle Schüler, aber vor allem jede Klasse, mit relativ gleichen Bedingungen bezüglich des Materials sowie des Unterrichtsgangs konfrontiert.

4.2.3. Tablets (im Physikunterricht)

Da sich das Betreiben einer Lernumgebung auf Tabletcomputern anbietet, soll hier kurz der aktuelle Stand des Tableteinsatzes in Schulen bzw. im naturwissenschaftlichen Unterricht/Physikunterricht skizziert werden. So ergibt sich die Möglichkeit, das vorliegende Projekt einzuordnen.

Aufenanger und Bastian (2017) berichten von einer zunehmenden Verbreitung von Tablets an Schulen, wobei dies weltweit gilt und nicht auf Deutschland beschränkt ist. Die große Verbreitung der Geräte scheint besonders bemerkenswert, als dass erste weithin angenommene¹¹ Tablet erst im Jahr 2010 erschien (das iPad). Zur genauen Verbreitung der Geräte in deutschen Schulen kann keine verlässliche Angabe gemacht werden, da entsprechende Zahlen nicht erhoben werden (vgl. Aufenanger 2017).

Nun sind Tablets in erster Linie handliche, relativ leistungsfähige Computer mit unterschiedlichen Schnittstellen, die sich per Touchscreen bedienen lassen. Worum geht es aber beim Einsatz von Tablet-Computern in der Schule? Welche (didaktischen?) Ziele werden unter dem Schlagwort der *Digitalisierung* verfolgt? Aufenanger und Bastian (2017) werfen diese Frage pointiert auf:

„Darüber hinaus muss diskutiert werden, welche Ergebnisse man überhaupt vom Tableteinsatz in Schule und Unterricht

¹¹Tablets ähnlicher Technologie gab es durchaus auch schon vorher

erwarten darf bzw. sollte. Geht es nur um bessere Leistungen, die sich etwa in Noten oder Testergebnissen ausdrücken, oder geht es allgemeiner um eine Verbesserung des Lehrens und Lernens? [...] Und was heißt es eigentlich, wenn davon die Rede ist, dass die Tablets einen Mehrwert für die Klasse bringen und worin drückt sich dieser aus? Was sind die Kriterien für eine Verbesserung des Lehrens und Lernens?“

(ebd.)

Zunächst sollte als Ansatzpunkt aber klargestellt werden, welche *Potentiale* der Einsatz dieser Geräte bereithalten kann. Welling (2017) nennt auf technischer Ebene folgende Vorzüge von Tablets:

- lange ohne Stromzufuhr nutzbar, effizientes Energiemanagement
- einfache „Gestekommunikation“ (Wischen, Tippen, Vergrößern mit zwei Fingern...)
- Kameras, auch in Verbindung mit Apps („Multifunktionsgerät für unterschiedliche Lernkontexte“)
- Internetfähigkeit

Relativ unscharf sind an gleicher Stelle (Welling 2017) hingegen didaktische Vorzüge aufgeführt. Er nennt dabei basierend auf weiteren Quellen:

- kollaborativeres Lernen
- „in konstruktivistischer Manier gemeinsam interagierend Wissen aneignen“
- „Lernprozesse adaptiver, heterogener und multimodaler“ gestalten, um individuellen Bedürfnissen entgegen zu kommen

Im folgenden Abschnitt wird daran anknüpfend anhand des Übersichtsartikels von Aufenanger (2017) knapp auf Befunde zum Tableteinsatz an Schulen eingegangen.

Eine der ersten in größerem Maße *national* beforschten Schulen war ein Hamburger Gymnasium. Die zuständige Projektgruppe (Autorengruppe Paducation 2014; vgl. auch Aufenanger 2017) kam zu dem Schluss, dass der „Unterricht in der Oberstufe sich erheblich verändert [habe]“. Auch die Lernprozesse hätten sich verändert. Weiterhin sei der Einsatz der Geräte stark von der jeweiligen Lehrkraft und deren Engagement abhängig gewesen. Es wurden durch Lehrkräfte digitale Arbeitsblätter,

Videos, Animationen und Präsentationen eingesetzt. Ein Fortbildungskonzept wurde aber dringend empfohlen. Seitens der Schüler erfolgte die Nutzung zu Dokumentations-, Recherche- und Präsentationszwecken (Aufenanger 2017). Autorengruppe Paducation (2014) konstatiert, dass die Tablets in der eingesetzten Form „zur Entstehung einer Lehr- und Lernkultur [beitragen], in der die Heranwachsenden angemessen auf das Leben in einer zunehmend von digitalen Medien geprägten Gesellschaft vorbereitet werden“. Abschwächend betont Aufenanger (2017) aber eingangs des Artikels die Kritik an Teilen der bisherigen Forschung. Die methodischen Kritikpunkte an einigen Untersuchungen sind dabei z. B. häufig Selbstauskünfte durch Schüler, fehlende fachdidaktische Einbindung, seltene Untersuchung realistischer Situationen des Tableteinsatzes sowie die Überbetonung der Messung kurzfristiger Effekte zu Ungunsten langfristiger Effekte. Als Quellen dienen darüber hinaus mitunter auch Reports oder Projektberichte, wie auch der folgende Bericht des Niedersächsischen Landesinstituts für schulische Qualitätsentwicklung (NLQ) einen darstellt.

Der NLQ-Bericht zur Niedersächsischen Initiative zum mobilen Lernen¹² stellt fest, dass „für Schülerinnen und Schüler [. . .] der Unterricht attraktiver geworden [ist]. Sie haben vor allem in den Bereichen Medienkompetenz, Informationsbeschaffung, Präsentation und selbstgesteuertes Lernen profitiert.“ (NLQ 2015). Weiterhin sah „die Hälfte der Lehrerinnen und Lehrer [. . .] durch den Einsatz der Tablets im Unterricht einen Zuwachs der fachbezogenen Kenntnisse“ (ebd.), wobei eine Vergleichsgruppe hier aber auch nicht untersucht wurde.

Im *internationalen* Bereich zählen Karsenti und Fievez (2013) in einem Reviewartikel zusätzlich insgesamt 16 (positive) Effekte des Tableteinsatzes auf. Dazu gehören unter anderem eine erhöhte Motivation, Förderung von individuellem Lernen wie mehr Möglichkeiten für unterschiedliche Lehr-Strategien, Lernerfolgsmessung für Schüler, Vorzüge für SuS mit Lernschwierigkeiten sowie eine Förderung der Kreativität. Aufenanger (2017) kritisiert bezüglich dieser Überblicksstudie die Erhebungsmethoden (Selbstauskünfte) und das potentielle Auftreten von Neuigkeitseffekten – stellt aber fest, dass „Schülerinnen und Schüler motivierter im Unterricht dabei sind, ihre Lernprozesse individualisierter steuern, umfassende Medienkompetenz erwerben und ihren schulischen Alltag einfacher organisieren können.“

Zu beachten ist nach Welling (2017) jedoch, dass sich „die unterrichtliche Nutzung von Tablets [. . .] sich aber auch daran messen lassen [muss], in wie weit sie zur Förderung fachlicher Kompetenzen beiträgt. Antworten auf diese Frage sind Mangelware.“ (ebd.)

¹²Wobei im Detail unklar bleibt, inwiefern *Lernen* nun mobil ist.

Genau aus diesem Grund soll hier nach dem Einstieg und dem Überblick über die eher fachübergreifende Einführung, Verwendung und Evaluation von Tablets an Schulen eine fachspezifische Perspektive eingenommen werden. Die Vernachlässigung dieser Perspektive wie die wenigen Evidenzen zur Wirkung des Tableteinsatzes auf den fachspezifischen Lernzuwachs (siehe Welling (ebd.), Aufenanger (2017), auch Kapitel 4.2.3) geben dazu besondere Veranlassung.

Zu Beginn soll jedoch der Eindruck ausgeräumt werden, dass es keine fachspezifischen Anwendungsfelder der Tablets gäbe (Welling (2017) spricht ja von *Evidenzen* und nicht von *Einsatzfeldern*; s. o.). Selbstverständlich kann an dieser Stelle nicht für alle Fächer gesprochen werden – im Falle der Physik gibt es aber Anwendungsbeispiele, die den fachspezifischen Einsatz auszeichnen. Zunächst wird das Reichshofer Experimentierdesgin beschrieben, das das Tablet im weiteren Sinne zur Unterstützung des Experimentierens nutzt (s. u.). Daran schließen sich weitere Beispiele an, die aktuelle Beforschung von Tablets im Unterricht zeigen.

Einsatzbeispiele Im Rahmen des Projekts „Reichshofer Experimentierdesgin“ wurden Tablet-PCs als Arbeitsmittel in Experimentierphasen eingesetzt. Sie dienten dabei in erster Linie als Dokumentationswerkzeug für die jeweiligen Versuche (Bresges u. a. 2013), die die Schüler in Kleingruppen ohne zentrale Anleitung durchführten. Die „Beobachtungen und die Erkenntnisse“ wurden mittels „Kamera, Schnittsoftware, Textverarbeitungssoftware, Präsentationssoftware [dokumentiert]“ (ebd.), nachdem zuvor die Arbeitsanweisungen auf den Tablets erhalten wurden. Im Vergleich mit einer Kontrollgruppe, die ohne Tablet-PCs arbeitete, wurde dieses Setting als Design-Based-Research-Projekt auch wissenschaftlich begleitet und untersucht. Dabei erwies es sich nach dem ersten Durchlauf hinsichtlich der Ergebnisse des Post-Tests als unerheblich ob mit oder ohne Tablets gearbeitet wurde, jedoch war die Streuung der Ergebnisse in der Interventionsgruppe mit Tablets deutlich geringer, was sich vor allem für schwächere Schüler positiv auswirkte (Bresges u. a. 2013; Genz und Bresges 2017).

Blackmore, Wilkie und Manders (2015) untersuchten in einer Studie im englischen Raum den Einfluss von Tablets auf das Experimentierverhalten von 10-jährigen Kindern. Diese wurden beim Durchführen von 7 Experimenten beobachtet und in Nachhinein interviewt. Insgesamt wurde ein großer Einfluss der Tablets festgestellt, insbesondere hinsichtlich des Experimentierverhaltens „improved planning and experimental outcome recording“ (ebd.).

Pavla und Stepan (2016) zeigen die Möglichkeiten auf, das Tablet für die Erstellung von elektronischen Portfolios im naturwissenschaftlichen

Unterricht zu nutzen. Dabei betonen sie den „All-in-one“-Charakter der Geräte, wozu das Aufnehmen von Bildern, Videos und von Tonspuren (z. B. Tiergeräusche) gehört, wie auch das Anlegen von Mindmaps und Internetzugang im Allgemeinen (Pavla und Stepan 2016) gehören. Neben den allgemeinen Vorzügen eines Portfolios empfehlen sie den Einsatz von E-Portfolios, durchaus auch im Primarbereich.

Fridberg, Thulin und Redfors (2018) untersuchten das „Potential von Tablets im Vorschulunterricht, betreffend der Unterstützung von kollaborativem Lernen beim forschenden Lernen“ (Anm.: übersetzt). Sie stellten fokussiertes Argumentieren bei den Schülern fest, vor allem wird das Nutzen von Videoaufnahmen betont (auch Zeitlupenaufnahmen).

Zusammenfassend finden Tablets im Unterricht mittlerweile große Verbreitung, auch wenn keine genauen Daten darüber vorliegen. Unklar ist darüber hinaus, was überhaupt von Tablets erwartet wird. Bisherige Implementationsforschung zeigt Veränderungen des Unterrichts auf, wie z. B. den Einsatz von Videos oder digitalen Arbeitsblättern. Auch zur Dokumentation, Recherche und Präsentation werden die Geräte genutzt, womit die Schüler angeblich eine Vorbereitung auf eine digitalisierte Umwelt erfahren. Mitunter wird auch eine gesteigerte Motivation festgestellt, deren langfristige Wirkung für den Autor dieser Arbeit fraglich scheint (siehe auch Weidenmann (2009)). Darüber hinaus werden die Forschungsmethoden durchaus kritisiert. Abseits der allgemeinen Aspekte des Tableteinsatzes werden hinsichtlich des fachlichen Lernens mit dem Reichshofer Experimentierdesign und drei weiteren internationalen Fallbeispielen gezeigt, dass der Umgang mit Tablets auch fachspezifisch reizvoll sein kann. Alles in allem bleibt das Tablet aber „nur“ ein Werkzeug, das bestimmte technische Optionen bietet – diese müssen aber durch die Lehrkraft ausgewählt und bewertet werden, wie dies auch bei bisherigen Medien der Fall ist. Ein fachdidaktischer „Problemlöser“ scheint mit Tablets entsprechend erwartungsgemäß nicht gefunden. Im Sinne der Funktion der Tablets als Träger der Simulationen und der Lernumgebung wird das Tablet daher lediglich begleitend beforscht.

4.3. Feedback in neuen Medien

Ziel dieses Kapitels ist es, ein geeignetes theoretisches Fundament für Feedback in der tabletbasierten Mechanik-Lernumgebung zu legen. In diesem Kapitel wird zunächst die Notwendigkeit von Feedback für die geplante Lernumgebung dargelegt. Daraufhin wird der Begriff in spezifizierenderweise eingegrenzt und der *formative* Charakter des hier genutzten Feedback beschrieben. Darauf folgt die Beschreibung effektiven Feedbacks.

Die Schüler arbeiten in eigener Verantwortung mit der Lernumgebung. Das Tempo bestimmen sie dabei weitestgehend selbst. Dabei ist abzuschern, dass die Schüler korrekt fokussierte Beobachtungen anhand der Simulationen machen. Denkbar wäre nämlich durchaus, dass für manche Schüler nicht-fachliche Aspekte im Zentrum der Beobachtung stünden. Dies würde einem Lernerfolg entsprechend im Wege stehen. Aus diesem Grund ist ein System erforderlich, das in gewisser Regelmäßigkeit die Korrektheit der Beobachtungen der Schülerinnen und Schüler überprüft. Die – scheinbar – einfachste Möglichkeit bestünde darin, die Lehrkraft entsprechende Überprüfungen vornehmen zu lassen. Jedoch erfordert dies, engmaschig durchgeführt, erfahrungsgemäß einen erheblichen (!) Aufwand. Dies bestätigt auch Richtberg (2018): „Lehrkräfte können im laufenden Unterricht nicht allen Lernenden zeitgleich individuelles, auf den jeweiligen Lernprozess bezogenes Feedback anbieten. Hier kann computergestütztes Feedback ergänzend wirken und den Lernprozess positiv beeinflussen.“. Davon abgesehen wäre Feedback durch die Lehrkraft mit dem geplanten Forschungsdesign, das die Lehrkraft eher weniger berücksichtigt (vgl. späteres Kapitel 7) nicht vereinbar. Eine naheliegende Option bietet daher die Verwendung der tabletbasierten Lernumgebung für entsprechendes Feedback, denn „für den Schulunterricht bietet der Einsatz neuer Medien besonders im Bereich immediate Feedback Potential zur Unterstützung von Lehrkräften“ (ebd.).

4.3.1. (Formatives) Feedback

Summativ vs. Formativ

Prinzipiell versteht man unter *formativen* Assessment, auch *assessment for learning*, Beurteilungsverfahren, die durch ihre Anlage den Lernerfolg der Schüler unterstützen sollen (vgl. Black 2003). Davon abzugrenzen ist das *summative* Assessment, bei welchem die abschließende Bewertung zur Feststellung zur reinen Leistungsfeststellung dient (vgl. ebd.).

Das Feedback im Sinne der hier verwendeten Lernumgebung dient nicht der abschließenden Leistungsfeststellung. Aufgrund seiner Funktion als unterstützendes Moment im Umgang mit den Simulationen kann es hier auch als formatives Feedback bezeichnet werden.

In diesem Abschnitt soll der Feedbackbegriff¹³ im Sinne dieser Arbeit eingegrenzt werden. Es finden sich zahlreiche Möglichkeiten der Definition von *externem* Feedback – also Feedback, welches von außen an den Schüler herangetragen wird¹⁴ Es finden sich durchaus Definitionsversuche von Feedback in der Literatur, wobei für den jeweiligen Zweck unterschiedliche Ausschärfungen vorgenommen werden. Gemein ist den hier betrachteten Varianten jedoch, dass es bei Feedback im Sinne des Lernens um Informationen geht, die einer Person gegeben werden und sich auf ihr bisheriges Tun beziehen (vgl. Shute 2008; Hattie und Timperley 2007;

A. Müller und Ditton 2014; Niegemann 2008). Shute (2008) präzisiert diese Darstellung, indem sie fordert, dass Feedback das Denken bzw. Verhalten hinsichtlich des Lernerfolges verändern soll. Ähnlich ist dies bei A. Müller und Ditton (2014) gelagert, die mit Rückgriff auf Ramaprasad (1983) festlegen, dass Feedback erst dann Feedback ist, wenn es dazu „genutzt [wird] um Diskrepanz zwischen Ist- und Sollwert zu überwinden.“

Feedback kann von verschiedenen Akteuren gegeben werden (Lehrer, Mitschüler, Buch...) (vgl. Hattie und Timperley 2007). Während hierbei technische Systeme nicht explizit genannt werden, nennt diese Niegemann (2008) ausdrücklich, da er Feedback in elektronischen Lernumgebungen charakterisiert. So konstatiert auch Richtberg (2018), dass Feedback ebenso von technischer Seite, wie z. B. Webseiten gegeben werden kann: „Auch für Webseiten kann diese Art des externen Feedbacks genutzt werden, da Multiple-Choice Aufgaben durch den Computer automatisiert ausgewertet und bewertet werden können. Entsprechend dem Bewertungsergebnis

¹³A. Müller und Ditton (2014) grenzen Feedback zusätzlich von Rückmeldungen ab bzw. ordnen sie an den Enden eines Spektrums an. Erstes erfolge dabei eher zeitnah, konkret und auf Personen bezogen, während Rückmeldungen sich eher diffus und auf Systeme bzw. Systemeinheiten beziehen und zeitfern gegeben würden. In dieser Arbeit sollen die Begriffe i. S. einer guten Lesbarkeit synonym verwendet werden.

¹⁴Im Vergleich zu internem Feedback, welches sich der Schüler aus eigenen Überlegungen selbst gibt (vgl. A. Müller und Ditton 2014; Richtberg 2018).

kann anschließend direkt Feedback für den Lerner generiert und eingeblendet werden.“ (ebd.).

Entsprechend soll in dieser Arbeit in Anlehnung an die Literaturlage Feedback als *Information, die von einem technischen System gegeben wird und eine Personen über eine von ihr ausgeführte Aktion mit dem Ziel, die ausgeführte und gewünschte Aktion anzugleichen, in Kenntnis setzt* aufgefasst werden.

Feedbackebenen A. Müller und Ditton (2014) legen nach Narciss (2006) unterschiedliche Feedbackebenen¹⁵ dar, die sich in elaborierte und einfache Formen untergliedern. Zu den einfachen Formen der auf eine Aktion folgenden Rückmeldung gehören dabei schlichte Rückmeldungen, die nur Richtig/Falsch-Meldungen enthalten können (abgekürzt KR: Knowledge of result). Denkbar wären auf diesem Niveau auch Angaben zur Anzahl/zum Anteil der korrekt gelösten Aufgaben (KP). Ebenso kann die korrekte Antwort im Anschluss angezeigt werden (KCR) (A. Müller und Ditton 2014).

Zu den elaborierteren Formen des Feedbacks gehören Hinweise zum Knowledge of constraints, concepts und mistakes (KTC, KC, KM) (Narciss 2006). Zu den ersteren gehören etwa Hinweise auf Bearbeitungsregeln oder Aufgabenarten; zu den zweiten etwa Hinweise auf Fachbegriffe oder Kontexte. Zu letzterem gehören dabei Angaben über die Anzahl von Fehlern wie auch Hinweise zu Ort und Art der gemachten Fehler (A. Müller und Ditton 2014).

Die ausdifferenzierteste Form des Feedbacks bilden die Stufen *KH* (Knowledge of How to Proceed) und *KMC* (Knowledge of Meta-Cognition). Im Falle von *KH* werden der Person „Fehlerspezifische Korrekturhinweise, Aufgabenspezifische Lösungshinweise, Hinweise auf Lösungsstrategien, Leitfragen [oder] Lösungsbeispiele gegeben.“ (ebd.) Im Falle von *KMC* wird Hilfestellung in Form von metakognitiven Strategien gegeben (ebd.). Richtberg (2018) merkt an, dass „elaboriertes Feedback häufig mit KR oder KCR Feedback kombiniert“ wird, gerade wenn es durch Computer generiert werde.

Effektives Feedback In diesem Abschnitt soll umrissen werden, welche Kriterien für Feedback angelegt werden, welches im formativen Sinn lernwirksam sein soll.

An den letzten Abschnitt anknüpfend soll aufgezeigt werden, welche Feedbackspezifität als effektiv angesehen wird. So betont Shute (2008) mit Rückgriff auf div. Literatur, dass „Feedback deutlich und signifikant

¹⁵Jeweils durch eine Abkürzung zu kennzeichnen

effektiver ist, wenn es Details zur Verbesserung der gegebenen Antwort bereitstellt – statt bloß die Information über die Korrektheit“ (frei übersetzt durch den Autor). Letzteres für sich alleine könne dabei als „überflüssig, frustrierend oder beides“ (s. o.) empfunden werden. Im selben Zuge stellt sie in ihrem Literaturreview dar, dass effektives Feedback aus zwei Informationen besteht, nämlich der eben gescholtenen Information zur Richtigkeit – aber eben auch weitere, elaboriertere Informationen, „um den Lerner zur richtigen Antwort zu leiten“ (übersetzt nach Shute 2008) (wie schon oben durch Richtberg (2018) mit der Kombination aus KC und elaboriertem Feedback angedeutet wird).

Neben der Feedbackebene lässt sich die Effektivität durch den Zeitpunkt der Feedbackgabe beeinflussen. So kann es direkt nach der jeweiligen Aktion gegeben werden, wie auch mit zeitlichem Abstand. Die Forschungslage stellt sich nach Shute (2008) hier nicht sehr eindeutig dar. Befürworter des verzögerten Feedbacks beziehen sich dabei auf die Theorie, dass Fehler eher wieder vergessen werden und somit nicht mit der Speicherung einer späteren Korrektur in Konflikt geraten. Erfolge wurden in einzelnen Studien hinsichtlich einer Langzeitwirkung festgestellt (ebd.). Befürworter des zeitnahen Feedbacks argumentieren, dass, je eher interveniert wird, die Chancen auf ein angepasstes Verständnis umso größer sind – was ebenfalls in div. Studien dargelegt wurde (siehe hierzu ebd.). Weiterhin zeigte sich, dass zeitnahes Feedback in Feldstudien ihre Stärken ausspielen konnte – verzögertes Feedback hingegen in Laborstudien. Letztlich konstatiert Shute (ebd.), dass beide Varianten ihre Vorzüge und Nachteile haben können. Nach Einschätzung des Autors kann der nach Mathan und Koedinger (2002) in Shute (2008) zitierte Hinweis „the effectiveness of the feedback depends not on the main effect of timing but on *the nature of the task* (Hervorhebung durch Autor)“ einen Hinweis zur Entscheidungsfindung liefern. Eigentlich selbstverständlich ist jedoch eine ergänzende Empfehlung von Shute (ebd.), die besagt, dass Schüler zunächst selbst nach Lösungen suchen sollen bevor sie Antwortmöglichkeiten einsehen können.

Ergänzend zu den bisherigen Ausführungen ist Feedback nach Hattie und Timperley (2007, „Effective Feedback must answer three major questions [. . .]“) besonders effektiv, wenn es drei Fragen beantwortet (siehe auch vollständiges Modell in Abb. 4.2¹⁶):

- Where am I going?
- How am I going?
- Where to next?

¹⁶Aus ökonomischen Gründen und entsprechend des Ziels dieses Kapitels werden hier nur ausgewählte Aspekte des Modells – und keine vollständige Beschreibung – gegeben.

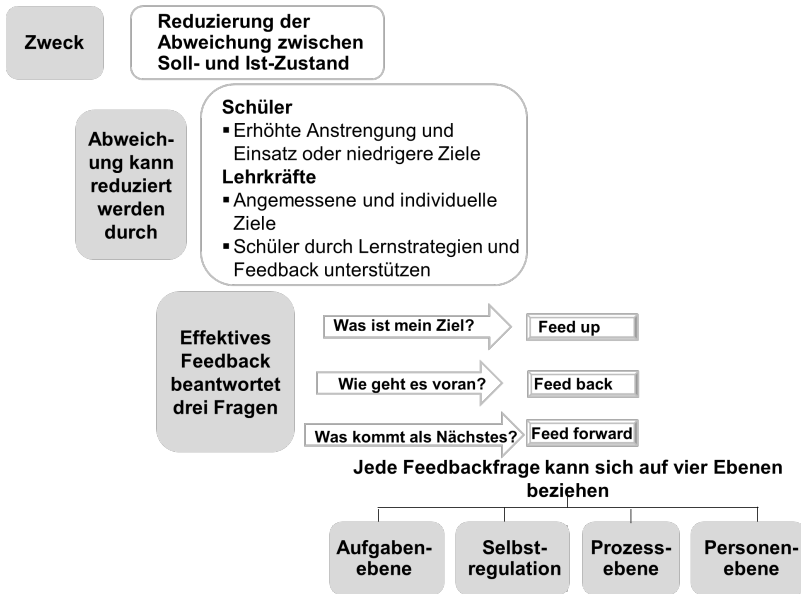


Abbildung 4.2.: Feedback-Modell nach Hattie und Timperley (2007), eigene Übersetzung

Die Zielklarheit herzustellen oder wiederholt zu verdeutlichen ist dabei Ziel der ersten Frage „Was ist mein Ziel?“. Dies ist nach Hattie und Timperley (ebd.) erforderlich, da „Feedback cannot lead to a reduction in this discrepancy if the goal is poorly defined, because the gap [...] is unlikely to be sufficiently clear to students“. Die Schüler müssen sich also dem Unterschied zwischen Soll- und Ist-Zustand bewusst sein. Das Feedback soll sich nach Hattie und Timperley (ebd.) dann auch auf dieses Ziel beziehen und nicht etwa auf die „Textlänge oder die Rechtschreibung, wenn es darum geht, im Text Stimmungen zu erzeugen“ (frei übersetzt nach ebd.)¹⁷.

Die zweite Frage, „Wie geht es voran?“, soll den Schülern Informationen über ihr Tun geben. Dies kann bspw. im Vergleich zu einem Standard oder einem gesetzten Ziel geschehen (ebd.). Dabei betonen sie, dass „Feedback is effective when it consists of information about progress, and/or about how to proceed.“ (ebd.). Dabei solle nicht das Gefühl eines „Tests“ entstehen.

Die dritte Frage, „Was kommt als nächstes?“, bietet die Chance innerhalb von Feedback, Ausblick auf das kommende zu schaffen. Z. B. können Anregungen für tiefergehende Überlegungen gegeben werden (ebd.), die das Lernen intensivieren. Ihnen nach ist die Antwort auf die Frage al-

¹⁷Was für die Schulpraxis sicher nur als zugespitztes Beispiel dienen kann.

so keineswegs bloß „Mehr!“ (Hattie und Timperley 2007). Insgesamt sei natürlich nicht erforderlich durch das Feedback alle drei Fragen streng separat von einander zu beantworten.

Die drei Fragen können jeweils auf vier Ebenen bezogen werden, von denen eine die Personen-Ebene ist. Diese Kombination ist ein Beispiel dafür, was effektives Feedback *nicht* ist. Ein einfaches „Gut gemacht!“ ist nach Hattie und Timperley (ebd.) nur ein einfaches Lob, was „sich nur schwerlich verstärkend auf vertiefte Auseinandersetzung, Hingabe oder Verstehen der Aufgabe auswirkt“ (frei übersetzt). So stellt persönliches, nicht-sachbezogenes Lob im Sinne eines Formativen Assessments kein sinnvolles Feedback dar. Niegemann (2008) legt zudem nahe, dass „überschwängliches Lob“ bzw. zu viel Lob für subjektiv als sehr einfach empfundene Aufgaben auch dazu führen kann, dass sich die lernende Person abgewertet fühlen kann. Ebenso empfiehlt auch Shute (2008), dass sich Feedback auf die Aufgabe und nicht auf den Lerner beziehen soll. Denkbar ist hingegen, auch nach Hattie und Timperley (2007), dass das persönliche Lob direkt mit aufgabenbezogenen Informationen verknüpft wird. Generell sollte vor allem mit persönlichem – und vor allem mündlichen Feedback – vorsichtig umgegangen werden. Hier besteht nach Shute (2008) die Gefahr eines zu großen Bias. Es sei daher eher empfehlenswert, auf geschriebenes Feedback auszuweichen. So wird beispielsweise auch computergestütztes Feedback empfohlen (vgl. auch Richtberg (2018)), da Feedback aus „vertrauenswürdigen Quellen“ durch Lernende ernster genommen wird und ein personenbezogener Bias hier kaum auftreten kann (Shute 2008).

4.3.2. Schlussfolgerung für die Studie: Simulationen und Feedback in der Mechanik-Lernumgebung

Was bedeuten die obigen Ausführungen für die Mechanik-Lernumgebung? Zunächst wurde in Kapitel 4.3 die Notwendigkeit des Feedbacks dargestellt. Da die Schüler am Tablet alleine arbeiten, schien eine Art der Rückmeldung zu dieser Arbeit zur Qualitätssicherung notwendig zu sein. *Feedback* im Sinne einer Information, die Lernenden gegeben wird, um eine Diskrepanz zwischen Ist- und Soll-Zustand zu überwinden, sollte dafür geeignet sein. Da elektronische Lernumgebungen potentiell die Möglichkeit bieten, Feedback je nach Eingaben automatisch zu generieren, soll diese Möglichkeit in der Lernumgebung auch genutzt werden – zumal computergestütztes Feedback als effektiv gilt (s. o.). Doch wo soll in der Lernumgebung Feedback eingesetzt werden? Im Kapitel zu Simulationen wurde deutlich herausgearbeitet, dass es nicht sinnvoll ist, die Schüler zu frei an Simulationen explorieren zu lassen. Daher bietet es

sich u. a. an, die Simulationen jeweils mit Feedback zu verknüpfen. So kann den Schülern im Umgang mit den Simulationen ein Instrument an die Hand gegeben werden, ihr bisheriges Tun und Beobachten zu überprüfen. Sowohl im Bezug auf Simulationen als auch im Bezug auf Feedback wurde dargelegt, dass die Orientierung an einem Ziel/einer Aufgabe erfolgen sollte, um das schon angesprochene freie Explorieren zu minimieren als auch das Feedback so effektiv und sachlich wie möglich zu gestalten. Daher bietet sich für die Lernumgebung ein Setting an, bei dem ein handlungsorientiertes Ziel ausgegeben wird (z. B. „Finde heraus, wie sich die Zusatzgeschwindigkeit verändert!“). Diese prototypische Formulierung reduziert mit genügend Freiraum das freie Explorieren und gibt einen inhaltlichen Zielbezugsrahmen für das Feedback. Hinsichtlich dieser Aufgabe bieten sich zur Generierung von Feedback¹⁸ beispielsweise Single-Choice-Aufgaben an, die sich eben auf die Aufgabe beziehen. So kann je nach ausgewählter Antwortoption ein gesondertes Feedback gegeben werden.

Weiterhin wurden unterschiedliche Feedbackebenen vorgestellt. Dabei wurde mehrfach dargestellt, dass elaboriertes Feedback gegenüber einfachem Feedback Vorzüge besitzt. Eine Kombination aus beidem aber ebenso effektiv für die Zwecke der Lernumgebung ist und soll daher auch hier verwendet werden. Dies bedeutet, dass ein KC und KH-Feedback kombiniert werden. So werden eine Falsch-Richtig-Meldung wie auch aufgabenspezifische Hinweise zur Korrektur kombiniert.

Bezüglich des Zeitpunkts des Feedbacks wurden unterschiedliche Befunde dargestellt. Jedoch ergibt sich hier aufgrund der *Natur der Aufgabe* (s. o.) ein unverzügliches Feedback, da ja in den Simulationen direkt Korrekturen vorgenommen werden sollen. Generell sollen die Antworten / Feedback-Inhalte auch erst – wie vorgeschlagen – dargestellt werden, nachdem die Schüler selbst die Aufgabe bearbeitet haben.

Auch die drei Fragen, die Feedback nach Hattie und Timperley (2007) beantworten soll (s. o.) können in gewisser Weise in das Feedback der Lernumgebung eingebunden werden. Anhand eines exemplarischen Beispiels soll dies verdeutlicht werden:

- *Woher komme ich? Du solltest herausfinden, wie sich die Zusatzgeschwindigkeit verändert.* Hier wird Bezug zur ursprünglichen Aufgabenstellung genommen, an der der Fortschritt nun bewertet werden kann.
- *Wie geht es voran? Leider nein!* Nun wird den Schülern verdeutlicht, „wie sie sich schlagen“ (vgl. ebd.). Gleichzeitig entspricht dies dem

¹⁸Das Feedback muss in elektronischen System ja durch etwas ausgelöst werden.

vorgesehenen KC-Feedback, d. h. der Information über die Korrektheit.

- *Wohin geht es? Probiere nochmals aus den Scooter direkt von unten zu rammen und achte auf die Richtung des Pfeils!* Dies entspricht weniger dem „Ausblick auf das Neue“ im Sinne Hatties, sondern mehr einer Aufforderung, den nächsten Schritt zu gehen. Dazu werden Handlungsanweisungen gegeben. Gleichzeitig entspricht dies dem KH-Feedback, also detaillierteren, aufgabenbezogenen Informationen zu weiteren Handlungen in den Simulationen.

FEEDBACK-ARTEN

Zur Erweiterung des Forschungsfeldes bietet es sich an, das oben beschriebene Feedback in zweierlei Art darzubieten. Die erste Art ist dabei klassischer Natur – in Textform – und ist so aufgebaut wie das obige Beispiel.

Ergänzend dazu kann sich Feedback in Form von kurzen Videos bzw. Animationen anbieten, um falsche Beobachtungen der Lernenden zu korrigieren bzw. zur Reflexion anzubieten. Dazu kann das KH-Feedback in diesen Animationen *dynamisch dargestellt* werden, statt es in Textform zu präsentieren. Hier könnte aus Perspektive der kognitiven Belastung eine Entlastung für den Betrachter und somit ein Vorteil für diesen vorliegen (vgl. auch Kapitel 4.4 und Kapitel 4.2.1). Andererseits kann auch, durch den so eher „passiven Konsum“ des Feedbacks (die Lernenden sehen zu, anstatt die KH-Hinweise selbst zu testen), eine potentiell geringere Verarbeitungstiefe unterstellt werden.

Hinweise für potentielle Unterschiede zwischen beiden Feedbackvarianten können in dieser Arbeit untersucht werden.

4.4. Der Cognitive Load in neuen Medien

Bei der Frage nach der Fairness der beiden Varianten der Lernumgebung wurde die folgende Fragestellung aufgeworfen: *Welchen Einfluss haben die unterschiedlichen Zugänge auf die kognitive Belastung der Lernenden im Umgang mit den Simulationen?* Daher wurde in Teilen der Studie dieser Frage nachgegangen¹⁹ Die nötigen theoretischen Grundlagen sollen hier knapp dargelegt werden. **Hinweis:** Im Rahmen dieser Arbeit kann nur eine aspekt-hafte Darstellung dieser interessanten Fragestellung – und keine umfassende Darstellung – erfolgen.

¹⁹Wichtige Arbeiten wurden dabei von Mitschker (2017) geleistet. Vielen Dank!

4.4.1. Über das Arbeitsgedächtnis, das Lernen und den Cognitive Load

Die Theorie des Cognitive Load bezieht sich auf die Auslastung des Arbeitsgedächtnisses. Mitschker (2017) bemüht sich um eine greifbare Arbeitsdefinition des Begriffs und findet resultierend nach Sweller (1988) und Renkl u. a. (2003) sowie Hagmann (2011), dass der Cognitive Load „als Beanspruchung des menschlichen kognitiven Systems bei der Verarbeitung von Informationen“ gesehen werden kann (vgl. ebd.), wobei die „wichtigste Einflussgröße dabei die Anzahl der Elemente, die zeitgleich beachtet werden müssen“ ist (Mitschker 2017).

Als zentral im Bereich des Cognitive Loads lassen sich nach Hagmann (2011) (zitiert nach Sweller und Chandler 1994) weiterhin die folgenden Annahmen sehen, die sich auf das Arbeitsgedächtnis beziehen.

”

- Die Kapazität des menschlichen Arbeitsgedächtnisses ist beim simultanen Behalten und Verarbeiten von Informationen sehr begrenzt.
- Die Kapazität des menschlichen Langzeitgedächtnisses ist dagegen (fast) unbeschränkt.
- Die Aneignung von Schemata zählt zu den primären Lernmechanismen. Automation und insbesondere der automatische Gebrauch von Schemata reduziert die Belastung des Arbeitsgedächtnisses.

Cooper (1998, S. 14) ergänzt diese Annahmen um die folgenden:

- Der Lernprozess erfordert, dass das Arbeitsgedächtnis aktiv die Unterweisungsmaterialien auffasst und verarbeitet, um die zu lernende Information zu dekodieren und im Langzeitgedächtnis zu speichern.
- Wenn die Verarbeitungsressourcen des Arbeitsgedächtnisses erschöpft sind, ist der Lernprozess erfolglos.

“

(ebd.)

Arten des Cognitive Load

Die Betrachtung des Cognitive Loads als eindimensionales Konstrukt beschränkt sich darauf, die Anzahl der zu verarbeitenden Elemente zu zählen (ebd.) und als Load abzubilden. Da aber nicht zwingend ein linearer Zusammenhang zwischen dieser Anzahl und dem Cognitive Load

herzustellen ist, wird eine mehrdimensionale Betrachtung im Allgemeinen als elaborierter angesehen (Krell 2017, vgl. Hagmann 2011). So sagen auch Paas und van Merriënboer (1994), dass “Cognitive load is generally considered a multidimensional construct that represents the load that performing a particular task imposes on the cognitive system of a learner”.

Im folgenden werden die Bestandteile des Cognitive Loads beschrieben.

INTRINSISCH²⁰

Betrachtet man den Cognitive Load als Beanspruchung des kognitiven Systems, so ist einsichtig, dass das Verarbeiten neuer Inhalte einen solchen Fall darstellen kann. Ebenso ist einsichtig, dass dem Inhalt eine bestimmte Schwierigkeit innewohnt, die für den Lernenden zunächst unabänderlich ist und relativ unabhängig von der Darstellungsform ist. Diese Art der Belastung wird als natürlicher, Intrinsic Cognitive Load bezeichnet (Hagmann 2011). Diese Belastung hängt von der Anzahl der zu lernenden Elemente ab. Ebenso mitbestimmend ist die Vernetzung dieser Elemente untereinander²¹, wie selbstverständlich auch das Vorwissen des Lerners.

IRRELEVANT²⁰

Der Extraneous Cognitive Load ist auf das Lernmaterial und nicht auf den Lernenden bezogen. Der gleiche Inhalt kann in unterschiedlichem Lernmaterial in anderer Form präsentiert werden. Daher kann er leichter oder schwieriger verständlich sein (Rey 2009). Der “unnötige Anteil dieser Belastungen” ist der Extraneous Load. Daher ist das Ziel bei der Gestaltung von Lernmaterial Effekte durch diesen Faktor abzumildern (Hagmann 2011; Rey 2009).

RELEVANT²⁰

Die Belastungen im dritten Teil des Cognitive Loads “widmen sich der Aneignung und Automatisierung von Schemata” und somit der “Auslastung zur Förderung von Lernen” (Hagmann 2011). Ein hoher Germane Cognitive Load ist also als positiv zu bewerten (Mitschker 2017).

Alle drei Formen des Cognitive Loads sind “additiv” (Hagmann 2011). Die Summe der Bestandteile sollte als die Kapazität des Arbeitsgedächtnisses nicht überschreiten (ebd.).

Im folgenden soll ein zweidimensionales Modell, das auch Paas und van Merriënboer (1994), Rey (2009) und Sweller, van Merriënboer und Paas (1998) nahelegen, vorgestellt werden. Dieses gliedert sich in den Mental Load und den Mental Effort. Dabei ist der Mental Effort einer Messung

²⁰Bezeichnungen im Englischen – intrinsisch: *intrinsic*; extraneous: *irrelevant* – eher: *zusätzlich*; germane: *relevant*

²¹Das Auswendiglernen von Größensymbolen hätte eine geringe Vernetzung und somit einen geringen Intrinsic Cognitive Load.

(siehe Kapitel 4.4.2) mitunter eher zugänglich als die oben beschriebenen Konstrukte.

MENTAL EFFORT

Nach Paas und van Merriënboer (1994) bezieht sich der Mental Effort auf das Maß der Ressourcen, die aufgewandt werden, um eine Aufgabe in einer bestimmten Umgebung zu lösen. Hagmann (2011) ergänt dazu:

„Zu diesen kognitiven Prozessen können u. a. auch Aufmerksamkeit, Motivation/Volition und Emotion als investierte Ressourcen verstanden werden (Hacker 1996, S. 2). Neben der kontrollierten, ressourcen-bindenden Verarbeitung kann diese auch unkontrolliert, automatisiert, implizit oder 'tacit' ohne mentale Anstrengung erfolgen (Hacker 1996, S. 5).“

(ebd.)

Der Mental Effort interagiert mit den Faktoren des Mental Loads. Als Beispiel beschreiben Paas und van Merriënboer (1994) mit Bezug auf die Labyrinth-Aufgabe einen Lerner, der viel mit Erfahrung mit dem komplexen Labyrinth und wenig Erfahrung mit dem einfachen hat. Daher sei der Mental Effort bei dem – scheinbar – einfacheren Labyrinth höher.

MENTAL LOAD

Nach Paas und van Merriënboer (ebd.) ist der Mental Load auf Anforderungen der Aufgabe bezogen und vom Lerner unabhängig. Als Beispiel führen sie zwei Labyrinth-Aufgaben an, von denen eine schwieriger zu lösen ist als die andere. Somit sei auch der Mental Load für jeden Lerner jeweils bei der schwierigeren Aufgabe höher. Hagmann (2011) ergänt dazu, auch in Anlehnung an Sweller, van Merriënboer und Paas (1998):

„Hierunter wird die Belastung, die durch die Anforderungen der Aufgabe bzw. des Aufgabenumfelds (wie Komplexität einer Aufgabe) an den Lerner gestellt werden, verstanden. Diese Anforderungen können sich sowohl auf den Grad der Element-Interaktivität (intrinsische Aspekte, die vom Instruktionsdesign nicht beeinflussbar sind), als auch auf Behinderungen bei der Aufgabenbearbeitung (irrelevante Aspekte, die vom Instruktionsdesign beeinflussbar sind) beziehen“

(Hagmann 2011)

4.4.2. Messung des Cognitive Load

Die Messung des Cognitive Load kann aus mehreren Gründen erforderlich sein (Krell 2017). Zunächst kann ein hoher Cognitive Load als hinderlich bei der Aneignung von neuen Fähigkeiten und Wissenskonstruktion gesehen werden, was mit Hinblick auf die Gestaltung von Lernumgebungen berücksichtigt werden sollte (ebd.). Weiterhin kann die Messung als Kontrollvariable fungieren, um die Ergebnisse eines Leistungstest besser einordnen zu können (ebd.). Problematisch an der Messung im Allgemeinen ist der “multi-dimensionale Charakter und die komplexe Wechselbeziehung zwischen Performance Mental Load und Mental Effort.” (übersetzt nach Sweller, van Merriënboer und Paas 1998)

Hagmann (2011) beschreibt drei grundsätzliche Möglichkeiten, den Cognitive Load zu erfassen.

LEISTUNGSBEZOGENE VERFAHREN

Als Beispiel für ein solches Verfahren führt Hagmann (ebd.) an, dass etwa parallel zur Arbeit an einem Computer in gewissen Abständen auf einem zweiten Bildschirm Buchstaben erscheinen. Die Aufgabe besteht nun darin, sich diese Buchstaben zu merken. Rückschlüsse auf den Cognitive Load werden dadurch gezogen, dass sich die Bearbeitung die Ressourcen des Arbeitsgedächtnisses aufgeteilt werden müssen. Daraus ergibt sich die Annahme „Je weniger die zusätzliche Aufgabe die Simultanbearbeitung belastet, je geringer sind die Anforderungen der Hauptaufgabe und vice versa“ (ebd.). Dieses Verfahren ist hier aus ökonomischen Gründen abzulehnen.

PHYSIOLOGISCHE VERFAHREN

Die Grundannahme hier ist, dass sich der Cognitive Load auf physiologische Reaktionen des Körpers auswirkt und damit bestimmbar ist. Zu solchen Funktionen gehören der Puls oder das Weiten der Pupillen (ebd.). Für den Zweck dieser Arbeit scheidet diese Methode aus diversen Gründen aus (Ausrüstung, Aufwand, Expertise, Genehmigung..).

SUBJEKTIVE VERFAHREN

Mithilfe subjektiver Verfahren beurteilen die Probanden ihre empfundene Belastung weitestgehend selbst. Es wurde dabei gezeigt, dass „dass dies möglich ist und es keine Schwierigkeiten bereitet, der erlebten Belastung numerische Werte zuzuordnen“ (ebd.). Die Messung erfolgt in der Regel anhand von mehrstufigen Likert-Skalen. Dabei ist die Stufigkeit und die Dimensionalität ebenso wie die Benennung der Pole nicht entscheidend (ebd.). Sweller, van Merriënboer und Paas (1998) berichten hier aber von sieben- oder neunstufigen Skalen.

Offensichtlich wird auf diese Weise der Belastungsmessung der Mental Effort bestimmt. Hagmann (2011) bezeichnet eine solche Messung des Mental Efforts (s. o.) die „favorisierende Technik innerhalb der Cognitive Load Theorie“. Dazu ergänzt er, dass „Nach Paas & Van Merriënboer (1993, S. 738) [...] die von Anwendern eingeschätzte Intensität der Anstrengung als die wesentliche Variable zur Erlangung einer reliablen Einschätzung des Cognitive Load angesehen werden [kann].“ (ebd.).

Die nötigen Grundlagen zur Beurteilung der eingangs aufgeworfenen Frage wurden dargelegt. Die an sich naheliegende Kognitive Theorie des multimedialen Lernens bietet zur Beantwortung der Forschungsfrage nur wenig neuen Input. Die dort breit diskutierte Verknüpfung von Bild und Text wie von audiokodierten Inhalten mit Betrachtung von zwei Aufnahmekanälen und separaten Verarbeitungskanälen, ist hier nicht Forschungsgegenstand. Da hier weder Audioinformationen verarbeitet werden und noch die Verknüpfung von Text und Bild im *Fokus* steht, erfolgte keine ausgiebige Darstellung der Kognitiven Theorie des multimedialen Lernens. Mitunter wird im folgenden Kapitel aber auf die dortige Grundannahmen zurückgegriffen, um Vermutungen zu stützen.

4.4.3. Der Cognitive Load beim Lernen mit Simulationen der Lernumgebung

Ziel ist es, die kognitive Belastung beim Bearbeiten der Simulationen zu untersuchen. Es geht nicht um die Belastung während der gesamten Arbeitszeit mit der Lernumgebung. Entsprechend knapp fällt dieser Abschnitt aus. Hier wird erläutert, welche Aspekte der Simulationen in 1D- oder 2D hinsichtlich des Cognitive Loads Schwierigkeiten darstellen *könnten*. Als zentraler wird dabei die begrenzte Kapazität des Arbeitsgedächtnisses angesehen.

Bei 2D-Simulationen werden im Vergleich zu den 1D-Simulationen vergleichsweise (und nicht absolut!) andere Darstellungsarten genutzt. Dazu gehören allem voran Vektorpfeile wie im Beispiel der Geschwindigkeit. Mitunter verändern diese Pfeile ihre Eigenschaften (Länge, Richtung) während der Bedienung einer Simulation. Der Zusammenhang zwischen Pfeil und dargestellter Bewegung muss dabei durch die Schüler aktiv konstruiert werden. Weiterhin erlauben die 2D-Simulationen Bewegungen in mehr als einer Dimension. Der so gewonnene Freiheitsgrad bietet komplexere Möglichkeiten für Bewegungen und Einflussnahme auf diese Bewegungen. Dazu gehören auch Aspekte der Koordination, wie z. B. im Rahmen der Scooter-Simulation (vgl. Anhang; dort muss zum passenden Zeitpunkt ein Stoß zwischen zwei Autoscootern ausgeführt werden.). Nach der kognitiven Theorie des Multimedialen Lernens ist das „Verknüpfen

(integrating) der Textrepräsentation mit der Bildrepräsentation“ ein Prozess beim Lernen mit Multimedia (Niegemann 2008). Für den Fall dieser Lernumgebung (siehe Kapitel 5.2) bedeutet dies, dass sich ebenfalls die zugehörigen Forschungsaufträge zur Bearbeitung der Simulationen (siehe Kapitel 5.2) schwierigkeitsgenerierend auswirken könnten, da womöglich mehr Informationen beachtet und mit den bildlichen Darstellungen der Simulation in Einklang gebracht werden müssen. Zu beachten ist bei allem Beschriebenen, dass alle Informationen visuell aufgenommen werden. Niegemann (ebd.) betont dabei, dass, wenn „zu viele Informationen gemeinsam über einen Kanal angeboten werden, [. . .] das Arbeitsgedächtnis überlastet und der Wissenserwerb behindert [ist].“.

Aufgrund der Darstellungsweisen, Forschungsaufträge und Freiheitsgrade der Simulationen, die in 1D und 2D-Zugang mitunter unterschiedlich auftreten können und dabei alle visuell aufgenommen werden, kann ein Unterschied der kognitiven Belastung vermutet werden. Diese wird über den Mental Load mittel subjektiver Befragungen (siehe dazu) erhoben.

Teil III.

Materialentwicklung

5. Simulationen und Lernumgebung

5.1. Simulationen

Programmierung

Die Verwendung bereits erstellter Simulationen schied aufgrund der geforderten Spezifität und durchgehenden Kontextorientierung aus. Eine erhebliche Aufgabe kam also dem Planen und Programmieren der benötigten Simulationen zu¹.

Da die Programmierumgebung Flash zur Erstellung von Simulationen als nicht mehr zeitgemäß erschien, wurden in angegliederten Masterarbeiten Alternativen ausgelotet. Maßgeblich dabei waren die u. a. Forderungen nach einer leistungsfähigen aber ressourcenschonenden Physikengine sowie ansprechenden grafischen Gestaltungsmöglichkeiten. Nachdem Optionen wie Algodoo auch aufgrund von Performanceproblemen verworfen wurden, fiel die Wahl letztlich auf die Umgebung Unity 3D. Diese Software auf Basis der elaborierten Programmiersprache C# bietet eine Lösung der Performanceprobleme. Dies ist darauf zurückzuführen, dass sie für zeitgemäße Spiele in drei Dimensionen ausgelegt ist und auch die Kompilation des Quellcodes ressourcenschonender erfolgt als bei Algodoo. Ebenso lassen sich die Simulationen in HTML implementieren oder in andere Formate exportieren.

Beispiel

Insgesamt wurden 21 unterschiedliche Simulationen erstellt. Eine Aufstellung aller Simulationen befindet sich im Anhang (siehe Kapitel B). Eine Simulation der 1D-Lernumgebung soll an dieser Stelle exemplarisch etwas ausführlicher dargestellt werden.

Abbildung 5.1 zeigt die Simulation „Mias Kart“. Inhaltlich wird dort der Zusammenhang zwischen der Änderung der Geschwindigkeit und wirkenden Kräften thematisiert. Dazu erhalten die Schüler auf der vorigen

¹An dieser Stelle herzlichen Dank an die Entwickler und Programmierer, ohne deren unermüdliche Mitarbeit und ihre Masterarbeiten dieses Projekt niemals möglich gewesen wäre.

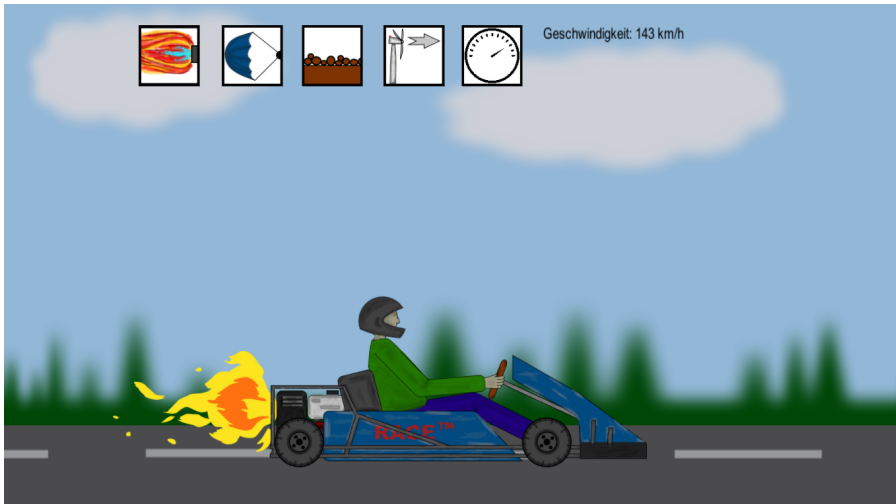


Abbildung 5.1.: Simulation „Mias Kart“

Seite der Lernumgebung einen entsprechenden Forschungsauftrag, um komplett freies Explorieren zu vermeiden (vgl. Kapitel 4.2.1 und Kapitel 5.2.2). Zunächst rollt das Kart antriebslos in das Bild hinein. Anschließend kann mit Hilfe der (nicht benannten) unterschiedlichen Wirkungen *Raketenantrieb*, *Bremfallschirm*, *Schotterweg* oder *Sturm*, bzw. jeweils einer derjenigen, auf das Kart eingewirkt werden. Dies resultiert in einer deutlichen Änderung der Geschwindigkeit. Speziell der Schotteruntergrund soll der Fehlvorstellung entgegenwirken, dass nur aktive Körper Kräfte ausüben können (vgl. Kapitel 3.3). Die Dauer des Einflusses ist dabei jeweils standardisiert (etwa 3 Sekunden) und lässt sich nicht beeinflussen. Ist kein Einfluss zugeschaltet ändert sich die Geschwindigkeit zunächst nicht spürbar, nimmt aber längerfristig ab (schließlich fährt das Kart ja auf einer Straße!). Diese „Minimalreibung“ hat den Vorteil, zunächst den Aufbau idealisierter, physikalischer Vorstellungen nicht zu behindern, aber letztlich genug realistische Aspekte zu bieten, um nicht als zweifelhaft eingestuft zu werden. Eine Anzeige der Geschwindigkeit kann zu- oder abgeschaltet werden.

Die optische und stilistische Gestaltung ist dabei ähnlich zu den zahlreichen weiteren Simulationen der Lernumgebung. Die Buttons/interaktiven Elemente müssen durch die Schüler in der Regel selbst identifiziert werden, sie sind gelegentlich aber auch durch Pfeile markiert. Die Anzahl der Parameter pro Simulation ist jeweils gering gehalten, um eine Überforderung der Schüler durch zu freies Experimentieren zu verhindern.

In einigen Fällen wird innerhalb der Simulationen auf eine Verknüpfung von Darstellungsformen zurückgegriffen (wie hier auch leicht durch die Anzeige der Geschwindigkeit angedeutet). Speziell betrifft dies z. B. die Verwendung von Vektorpfeilen in der 2D-Lernumgebung sowie die Anzeige von t-x- oder t-v-Diagrammen (vgl. Kapitel 4.2.1).

5.2. Lernumgebung

5.2.1. Inhaltliche Konzeption

Gestalterisch Als thematischer Rahmen für die „Geschichte“ der Lernumgebung dient ein Ausflug der beiden Protagonisten Max und Mia in einen fiktiven Freizeitpark. Die physikalischen Inhalte werden dabei anhand einzelner Attraktionen erkundet. Dazu gehören Achter- und Wildwasserbahnen, Freifalltürme, Autoscooter sowie Airhockeytische. Diese lassen sich in den Simulationen erkunden.

Fachlich Die Lernumgebungen sind in unterschiedliche Themenblöcke gegliedert (siehe Tabelle 5.1).

Inhalte 1D	Inhalte 2D
Bewegungsbeschreibung mittels Geschwindigkeit	Bewegungsbeschreibung & Einwirkung
Die Geschwindigkeit als zurückgel. Strecke pro Zeit	Geschwindigkeit als Richtung & Tempo (Tempo als zurückgel. Strecke pro Zeit)
x/v-t-Diagramme	x/v-t-Diagramme
Veränderungen des Bewegungszustandes (Geschwindigkeit und Richtung (1D) als Kraft)	Einwirkung & Zusatzgeschwindigkeit
Einflüsse auf die Geschwindigkeitsänderung („Stärke“, Masse und Wirkungsdauer)	Einflüsse auf die Zusatzgeschwindigkeit („Stärke“, Masse, Einwirkungsdauer)

Tabelle 5.1.: Kurzübersicht der behandelten Inhalte in beiden Lernumgebungen

Der 2D-Lehrgang orientiert sich dabei an Tobias (2010) bzw. Wiesner (2011). Der 1D-Lehrgang orientiert sich an einem klassischen 1D-Konzept, wobei Anleihen des 2D-Zugangs mit einbezogen werden (mehr dazu hier im Abschnitt *Fachdidaktisch*).

Fachdidaktisch DIDAKTISCHE GESTALTUNGSPRINZIPIEN

Der Lehrgang ist für insgesamt 2 – 3 Doppelstunden konzipiert. Dies resultiert aus der Überlegung, die Schülerinnen und Schüler der angesprochenen Altersstufe nicht zu lange mit der doch eher monotonen Arbeitsform zu beschäftigen (ausführlichere Betrachtungen, siehe Kapitel 7). Ein Einsatz von Realexperimenten, um die Monotonie aufzubrechen, hätte dem stringenten Verwenden digitaler Medien widersprochen. Weiterhin scheint es unmöglich, die in Tabelle 5.1 dargestellten Inhalte innerhalb von 2 – 3 Doppelstunden in vollem Umfang zu unterrichten. Daher werden generell Grundkenntnisse vermittelt, die für das Konzept bzw. für die Weiterarbeit in der Lernumgebung notwendig sind. Eine wiederholungs- oder vertiefungslose Weiterarbeit im späteren Unterricht ist somit nicht vorgesehen.

Zu diesem Zweck wurden die Inhalte in einzelne Einheiten zerlegt (angelehnt an *Zerlegung in methodische Schritte*, Reinhold (2006)). Eine thematische Einheit bildet dabei zum Beispiel das Ablesen eines t-x-Diagramms. Den strukturellen Aufbau beschreibt dabei der Abschnitt 5.2.2. Die sprachliche Gestaltung ist zunächst an der Umgangssprache und gfs. Unterrichtssprache orientiert (vgl. Starauschek 2006), somit erfolgt ein sparsamer aber gezielter Gebrauch von Fachtermini. An entsprechender Stelle werden Unterschiede zwischen Fachsprache und Alltagssprache explizit thematisiert.

FAIRNESS

Grundsätzlich wurde versucht eine große Fairness zwischen den Lernumgebungen herzustellen, so dass sich lediglich die Kernaspekte der Lehrgänge unterscheiden. Ansonsten wird größtmögliche Parallelität der Inhalte angestrebt, wie es auch Tabelle 5.1 schon andeutet. Kernaspekte der Unterschiede und Gemeinsamkeiten sind im Folgenden aufgeführt. Zum Thema Bewegungsbeschreibung ist beiden Lernumgebungen gemein, dass die Beschreibung einer Bewegung durch die Geschwindigkeit als eine praktikable Methode dargestellt wird. Fachlich betrachtet wäre eine vollständige Beschreibung der Bewegung erst durch weitere Randbedingungen oder die zeitabhängige Darstellung des Ortsvektors zu realisieren. Diese wird in beiden Lernumgebungen als „Wann-Wo-Darstellung“ erwähnt, aber im Folgenden wegen der Dominanz des Geschwindigkeitsbegriffs nicht vertieft. Zum Thema Geschwindigkeit gehört

zu den Unterschieden insbesondere die Vektordarstellung der Geschwindigkeit sowie das Konzept der Zusatzgeschwindigkeit. Diese kommen in der 1D-Lernumgebung nicht vor². Unterschieden wird daher der Geschwindigkeitsbegriff, der je nach Konzept entsprechend ausgelegt wird.

Die Inhalte zum Thema Diagramme unterscheiden sich prinzipiell kaum, da auch in der 2D-Umgebung in diesem Bereich nur eindimensionale Bewegungen betrachtet werden. Die Aspekte der Reduktion auf eine Dimension bzw. von unterschiedlichen Orts- und Geschwindigkeitskomponenten wird an einem Beispiel aufgezeigt, aber anschließend nicht vertieft. In der 1D-Umgebung werden mit einem zusätzlichen Beispiel „negative Geschwindigkeiten“ eingeführt.

Der Begriff der Kraft wird in beiden Lehrgängen eingeführt, jeweils nachdem Änderungen des Bewegungszustandes erkundet wurden (im 2D-Fall anhand der Zusatzgeschwindigkeit). Im 2D-Konzept wird hierbei die Definition um den Begriff der *Einwirkung* ergänzt, wie auch schon zuvor das Konzept der Zusatzgeschwindigkeit eingeführt wurde (auch mit eindimensionalen Beispielen wie einem Frontalzusammenstoß). Ebenso im 1D-Lehrgang wird die Änderung der Bewegungsrichtung als eine Wirkung von Kräften eingeführt, hier durch das „Zurückprallen“ eines Achterbahnwagens an einem Prellbock.

Zum Thema Einfluss auf die Geschwindigkeitsänderung werden in beiden Lernumgebungen Änderungen der Parameter Masse, Dauer der Kraftwirkung und Stärke der Kraftwirkung vorgenommen und erkundet. Im 2D-Lehrgang werden die Veränderungen anhand der Zusatzgeschwindigkeit beobachtet, im 1D-Lehrgang anhand der Geschwindigkeit.

DIDAKTISCHE REDUKTIONEN

Masse - Gewicht - schwer - leicht: In der Lernumgebung wird der Unterschied nicht inhaltlich thematisiert, es findet nur Erwähnung, dass ein solcher existiert. Im Sinne der sprachlichen Vereinfachung wird mitunter der Begriff Gewicht anstatt Masse verwendet, sofern dies in letzter Konsequenz keinen Unterschied macht. Ebenso wird wegen leichter Zugänglichkeit von schwer und leicht anstatt von größerer bzw. kleinerer Masse geschrieben.

Momentan-/Durchschnittsgeschwindigkeit: Der Unterschied wird ebenfalls an einer Stelle erwähnt, jedoch nicht weiter ausgeführt. So wird prinzipiell die Momentangeschwindigkeit betrachtet, welche bei Konstanz eben der Durchschnittsgeschwindigkeit auf einem Intervall entsprechen kann.

Kraft in/gegen Bewegungsrichtung bewirkt beschleunigen/bremsen: Im 1D-Falle ist die Betrachtung absolut korrekt, falsch kann sie bei Bewe-

²auch wenn dies durchaus möglich wäre!

gungen in zwei oder mehr Dimensionen werden. Wird daher so nur in 1D-Lernumgebung beschrieben.

Ort, Weg, Strecke: Ausdifferenzierte Unterscheidungen werden hier für den Anfangsunterricht nicht vorgenommen – sind aber auch kaum nötig, da hauptsächlich auf den Ort zurückgegriffen wird und der Weg nicht thematisiert wird.

Zusatzgeschwindigkeit vs. Tempo der Zusatzgeschwindigkeit (Länge des Pfeils der Zusatzgeschwindigkeit): Auch hier wird eine sprachliche Vereinfachung vorgenommen, um die Unhandlichkeit des Ausdrucks „größeres Tempo der Zusatzgeschwindigkeit“ durch „größere Zusatzgeschwindigkeit“ zu umgehen. Mitunter wird jedoch darauf hingewiesen, dass es sich bei „größerer Zusatzgeschwindigkeit“ um die Zusatzgeschwindigkeit mit dem größeren *Tempo* handelt (wie sollte auch die Richtung *größer* werden . . .). Auch bei Tobias (2010) wie ebenso in aktuellen Schulbüchern zum 2D-Konzept (R. Müller 2017) wird in der Beschreibung der Unterrichtskonzeption diese Verkürzung genutzt.

5.2.2. Strukturelle Konzeption

Die Lernumgebung besteht aus ca. 40 einzelnen Seiten, die fast ohne zu scrollen gelesen werden können. Für die Nutzung des Lernprogramms ist die Eingabe eines prinzipiell beliebigen alphanumerischen Codes erforderlich. So können sich einzelne Nutzer identifizieren und ihren Fortschritt speichern, sodass später an gleicher Stelle weitergearbeitet werden kann.

Die Navigation erfolgt über Vor- und Zurück-Buttons am unteren Ende der Seite. Ein Zurücknavigieren ist dabei grenzenlos möglich, das Vorwärtsnavigieren setzt das Ausfüllen der Multiple-Choice-Fragen voraus.

Der grundsätzliche Aufbau der Lernumgebung ist für jeden der einzelnen Themenbereiche in etwa gleich und im Folgenden beschrieben: Zunächst erfolgt eine Einleitung auf erzählerischer Ebene der „Geschichte“. Dabei wird die jeweilige Attraktion eingeführt. Daran schließt sich die Formulierung des zur Simulation gehörigen Forschungsauftrages an (vgl. Kapitel 4.2.1). Die Forschungsaufträge sind dabei im Grunde ähnlich formuliert. Dabei ist der Ausdruck *Finde heraus* zentral, beispielsweise „Finde heraus, was die Mitarbeiterin auf dem Diagramm ablesen kann“. Dadurch erfolgt eine Eingrenzung des Handlungsspielraums, lässt aber den Weg offen. Die Simulation muss dabei extern geöffnet werden³. Die Notwendigkeit des Feedbacks, welche als nächster Schritt folgt, wurde bereits in Kapitel 4.3 herausgearbeitet. Auf die konkrete Gestaltung des

³Dies ist für eine einfache Bedienung sehr ungünstig, ließ sich aber wegen der Unmöglichkeit, *.exe-Dateien zu verlinken, nicht sinnvoll vermeiden

Feedbacks wird in Abschnitt 5.3 eingegangen. Im Anschluss an das Feedback folgt min. eine weitere Textseite auf der die Inhalte und Beobachtungen aufgegriffen, erläutert bzw. zusammengefasst werden.

5.2.3. Beispielseiten

Exemplarisch sollen hier drei Beispielseiten dargestellt und kurz erläutert werden. Abbildung 5.2 zeigt eine typische Einstiegsseite, wie sie beim



Max und Mia belassen es dabei, sich die Wildwasserbahn von oben anzuschauen. Lieber gehen sie direkt zur Achterbahn! Max steigt direkt ein, Mia wartet am Kassenhäuschen.

Jeder kennt die Fotos, die während einer Achterbahnfahrt automatisch von den schreienden Fahrgästen gemacht werden. Am Kassenhäuschen kann man sie sich ansehen - ein schönes Bild siehst du zum Beispiel hier links :) „So wild, wie einem auf den Fotos die Haare wehen, sieht man ja direkt wie schnell so eine Bahn ist“, denkt sich die wartende Mia.

Denken kann man sich das sicherlich!
Aber kann man es auch genauer sagen? Mit der Fotokamera an unserer Achterbahn kann man das!
Die Kamera ist nämlich defekt. Sie macht immer im Abstand von einer Sekunde ein neues Foto!
Dadurch ist auf einer Fotoserie derselbe Wagen mehrfach zu sehen, nur an anderen Stellen.

A3: In der Simulation ist die Bahn und die Kamera dargestellt. Finde heraus, wie **und warum** du anhand der Fotoserien – auch ohne wehende Haare – erkennen kannst, ob die Bahn langsam oder schnell gefahren ist.

Öffne die Simulation "ZIELFOTO"!

Simulation gründlich erforscht?? Dann klicke auf weiter!

zurück


weiter

Abbildung 5.2.: Beispielseite für thematischen Einstieg in Themenelement (Bild CC0 Lizenz, Quelle: Pixabay; z. T. unkenntlich gemacht)

Einstieg in ein neues Thema verwendet wird. Neben illustrierenden Bildern findet sich ein einleitender Text, der sich zumeist auf die „Storyline“ bezieht. Im Rahmen dessen wird mitunter eine Fragestellung aufgeworfen. Diese gipfelt für gewöhnlich in einem der Forschungsaufträge. Im Beispiel soll also der Zusammenhang zwischen den Fotos und der Geschwindigkeit/Tempo des Wagens hergestellt werden. Dabei erfolgt auch der Hinweis auf die zu verwendende Simulation. Nach dem Schließen der Simulation soll dann per „Weiter“-Button auf die nächste Seite, die der Multiple-Choice-Fragen, navigiert werden.

III. Materialentwicklung

Abbildung 5.3 zeigt eine typische Seite einer Zusammenfassung eines Themenelements, hier der Einführung der Zusatzgeschwindigkeit in der 2D-Lernumgebung. Vorausgegangen ist hier die entsprechende Feedback-Sequenz inklusive des Feedbacks. Sofern auf solchen Zusammenfassungen Abbildungen vorhanden sind, sind diese zumeist nicht illustrierend, sondern im Sinne einer multicodalen Aufbereitung an die Inhalte des erklärenden Textes angebunden. Die sprachliche Ebene wechselt zwischen der Umgangssprache und der Unterrichtssprache, gegebenenfalls auch zu relativ scharfer Verwendung der Fachsprache. Auf einigen Seiten werden explizit Alltagsvorstellungen aufgegriffen. Zum Beispiel wird die Verwendung des Kraftbegriffs in beiden Lernumgebungen zwischen alltäglicher Verwendung und physikalischer Verwendung abgegrenzt; in der 2D-Lernumgebung zusätzlich die Verwendung des Geschwindigkeitsbegriffs (vgl. Kapitel 3.3).

Oh je, das ist gar nicht so einfach! Versucht Mia ihn wie vorher mit dem gleichen „Schubs“ Max in die Lücke zu stoßen, so verfehlt er sein Ziel. Max fährt also nicht in die Richtung der Einwirkung von Mia! Er fährt schräg nach rechts oben! In die Richtung zeigte Mias Einwirkung aber doch nicht?! 

Was ist passiert? Max bewegt sich nämlich mit seiner „Anfangsgeschwindigkeit“ weiter von links nach rechts – und gleichzeitig durch die (senkrechte) Einwirkung durch Mia nach oben!

Und was passiert, wenn man sich gleichzeitig nach rechts und nach oben bewegt? Richtig! Man bewegt sich nach „schräg“ rechts oben!

Durch die Einwirkung von Mia erhält der Scooter also zusätzlich zu seiner „Anfangsgeschwindigkeit“ eine „Zusatz-Geschwindigkeit“, die nach oben zeigt (genau in die Richtung der Einwirkung durch Mia).

Schau dir die Simulation nochmals an, wenn du dir nicht sicher bist!



Hast du dir die Simulation nochmal angesehen? Gut! Durch die Einwirkung bekommt Max also eine Zusatzgeschwindigkeit. Sein Bewegungszustand verändert sich dadurch! Er fährt nach der Einwirkung in die Richtung von „Anfangsgeschwindigkeit“ und „Zusatzgeschwindigkeit“! Die Zusatzgeschwindigkeit und die Einwirkung zeigen in dieselbe Richtung.

zurück

weiter

Abbildung 5.3.: Beispielseite für thematische Zusammenfassung eines Themenelements.

Die in Abbildung 5.4 exemplarisch dargestellte Seite schließt sich an eine Simulation an und zeigt Multiple-Choice-Fragen zum Generieren von Feedback. Diese Feedback-Sequenzen tragen den Titel „Alles klar?“ und den Namen der zugehörigen Simulation. Darunter wird der Forschungsauftrag nochmals in Erinnerung gerufen (vgl. *Was ist mein Ziel?*, Kapitel 4.3). Daran schließen sich die Antwortoptionen an. Ist die korrekte Option

ausgewählt, folgt auf der nächsten Seite ein inhaltlicher Texthinweis (vgl. *Wie geht es voran? Was kommt als nächstes?*). Im Falle einer falschen Antwort folgt je nach Variante Text- oder Animationsfeedback. Ist keine Option ausgewählt erfolgt eine Fehlermeldung mit entsprechender Aufforderung.

Alles klar...?: Achterbahn-Kontrolle



1:

Du solltest herausfinden, was der Mitarbeiter aus dem Diagramm über die Bewegung der Bahn erfahren kann. Kreuze die bestmögliche Antwort an!

- Der Graph, der während der Fahrt im Diagramm gezeichnet wird, zeigt den Verlauf der Strecke. Mit diesen Informationen kann die Mitarbeiterin die sichere Fahrt überwachen.
- Der Graph sieht je nach Tempo der Bahn anders aus.
- Das Diagramm/der Graph sieht bei jedem Tempo der Bahn etwa gleich aus. Der Graph und das Tempo sind daher unabhängig voneinander.

zurück

weiter

Abbildung 5.4.: Beispielseite für Multiple-Choice-Fragen zur Feedbackgenerierung

5.3. Feedback

Das folgende Kapitel beschreibt die Entwicklung der Multiple-Choice-Aufgaben anhand derer das Feedback generiert wird. Daran schließt sich eine Beschreibung des Feedbacks in beiden Varianten an.

5.3.1. Multiple-Choice-Aufgaben

Weshalb Multiple-Choice-Aufgaben? Aus bereits dargelegten Gründen wird bei der Lernumgebung automatisches Feedback eingesetzt. Feedback erfolgt auf Handlungen des Nutzers oder sonstigen Input. Im Falle dieser Lernumgebung wären für diesen Zweck z. B. bestimmte Handlungsmuster bei der Bedienung der Simulationen denkbar. Im Rahmen dieser Arbeit ist eine Programmierung von dafür nötigen Schnittstellen für alle

Simulationen nicht zu ermöglichen – ebenso wenig wie eine Erkennung der Handlungsmuster. Aus diesem Grund wird auf die Selbsteinschätzung der Lernenden zurückgegriffen. Diese wird aus den Antworten zu den Multiple-Choice-Fragen generiert, die sich nach Mie (2002) nicht nur für Bewertungssituationen eignen, sondern auch für Erarbeitungsphasen – welche das Arbeiten mit Simulationen hier darstellt. In den Multiple-Choice-Fragen überprüfen die Lernenden ihre Beobachtungen aus den Simulationen (und nicht zwingend ihr Fachwissen). An jede Simulation schließt sich daher ein Zweier-Set von Multiple-Choice-Fragen⁴ an. Bei den Fragen eines Sets liegen Fragen zum gleichen Forschungsauftrag zugrunde. Die drei Antwortoptionen werden aber in der ersten Runde etwas allgemeiner gehalten und in der zweiten Fragerunde präzisiert. Näheres zur Konstruktion und Anzahl der Items folgt in den nächsten Abschnitten.

Konstruktion der Items Die Multiple-Choice-Frage wird mit einer Wiederholung des Forschungsauftrages eingeleitet („Du solltest herausfinden, wie...“). Obwohl es anschließend jeweils nur eine korrekte Antwort gibt und es somit keine *Best-Choice-Aufgaben* sind (vgl. Moosbrugger und Kellava 2012), wird dies in der Frage nahegelegt („Kreuze die Möglichkeit an, die am besten zu deinen Beobachtungen passt!“). Dies folgt aus der Abwägung, dass die Schüler sich bei dieser Formulierung nicht in der Situation befinden, eine korrekte Antwort finden zu müssen, da sie keine Leistungssituation darstellt, die für eine Benotung von Bedeutung wäre⁵. Die Lernenden sollen sich vielmehr auf ihre Beobachtungen verlassen und diesen vertrauen.

Bei der Formulierung von Antwortoptionen ist besondere Sorgfalt für das Formulieren der Alternativantworten, der Distraktoren, aufzuwenden (Mie 2002), um im Sinne einer zufriedenstellenden Reliabilität zu vermeiden, dass die korrekte Antwort für jeden Schüler von vornherein ersichtlich ist. Daher ist das Ziel, dass die Korrektheit aller Antwortmöglichkeiten auch von uninformierten Lernenden für ähnlich wahrscheinlich gehalten wird (ebd.). Zu vermeiden sind zunächst also allzu abwegige Distraktoren („Nur bei Vollmond ändert sich die Fahrtrichtung“). Naheliegende Antwortmöglichkeiten ergeben sich also aus bekannten Fehlvorstellungen zum jeweiligen Themengebiet, wie auch aus Gesprächen mit erfahrenen Lehrkräften, Dozenten oder Lernenden.

⁴Es handelt sich eigentlich um *Single-Choice-Fragen*, da stets nur eine Antwort korrekt ist. Im weiteren Verlauf wird aber aufgrund des gebräuchlicheren Sprachgebrauchs weiterhin von Multiple-Choice-Fragen geschrieben.

⁵Idealerweise wäre sie das auch außerhalb der Forschungssituation nicht, da auch im Unterricht zwischen Lern- und Leistungssituation getrennt werden sollte

Weiterhin sollte die Länge der Distraktoren keinen durchschaubaren Mustern unterliegen. So ist beispielsweise zu vermeiden, dass die längste Antwortoption stets die korrekte ist. Gleiches gilt für die Stilistik der Antworten, welche harmonisch und einheitlich und nicht durchschaubar erscheinen sollte. Ein Negativbeispiel wäre beispielsweise eine einzige Antwortoption, die von der Fachsprache Gebrauch macht, während die anderen sehr umgangssprachlich formuliert sind.

Wenn möglich sind Antwortoptionen zu vermeiden, die sich nur marginal – beispielsweise in einem kurzen, leicht zu überlesenen Wort – unterscheiden. Ebenso sollten aus diesem Grund eigentlich möglichst lange, sich wiederholende Passagen in den Optionen vermieden werden. Abhilfe schafft in solchen Fällen das Auslagern des sich wiederholenden Teils in die Fragestellung. Dies konnte in der vorliegenden Studie jedoch nicht immer berücksichtigt werden, da der Aufwand in der Programmierung bei eventuell vorzunehmenden Änderungen zu hoch gewesen wäre. Daher konnte diesem Kriterium nicht immer nachgekommen werden. Mitunter wurden die Unterschiede in den Antwortoptionen gesondert markiert.

Wenn praktikabel, wurden Antwortoptionen aus zueinander gehörenden Gruppen generiert. So konnten plausibel scheinende Paare gebildet werden (z. B. zwei Paare aus „Tempo ändert sich / ändert sich nicht“ und „Richtung ändert sich/ändert sich nicht“). Auf dieser Weise wären vier Optionen erzeugt, die plausibel, etwa gleichlang und ähnlich erscheinen. Zur Anzahl der Items, siehe nächster Abschnitt.

Anzahl der Items Die passende Anzahl von Multiple-Choice-Items ist Gegenstand der Forschung. Es sind Aspekte der Ratewahrscheinlichkeit, der Ökonomie der Erstellung und der Auswertung zu berücksichtigen. Weiterhin ist der Workload für die Lernenden einzubeziehen. In der Praxis üblich sind 3 - 5 Antwortalternativen (Tinnefeld 2013). Nach Moosbrugger und Kelava (2012) führt eine größere Anzahl von Distraktoren zu einer geringen Wahrscheinlichkeit „für das rein zufällige Auffinden der richtigen Lösung“. Das vollkommen zufällige Raten ist hingegen nach der Meta-Studie von Rodriguez (2005) eher unüblich. So würden Lernende i. d. R. dennoch das unpassendste Item ausschließen (Anm. d. A.: Sofern die theoretische Idealvorstellung...) und es so effektiv doch um eine Antwortalternative reduzieren. Aus dieser Perspektive erhält man beispielsweise durch die Verwendung von 3 statt 4 Items keinen Nachteil.

Based on this synthesis, MC items should consist of three options, one correct option and two plausible distractors. Using more options does little to improve item and test score statistics and typically results in implausible distractors. ebd.

Auch der hier zuletzt benannte Punkt war entscheidend, zumal die Simulationen und die zugehörigen Beobachtungen im Komplexitätsgrad recht unterschiedlich angesehen wurden. Daher gelingt es bei eher einfacheren Simulationen kaum, zwei Multiple-Choice-Fragen mit jeweils drei Distraktoren zu finden, ohne dass auch nur einer davon unplausibel erscheint. Bestätigt wird diese Annahme aus einer Voruntersuchung, wo noch vier Antwortoption dargeboten wurden. Dort zeigte sich bisweilen, dass jeweils eine Antwortoption seltener berücksichtigt wurde (siehe Kapitel 11.1). Einzig problematisch an der Entscheidung für drei Optionen ist naturgemäß das Anbieten von *zwei* Paaren von Antwortmöglichkeiten, was nun nicht mehr möglich ist (siehe S. 89). Dadurch ergibt sich folgende Problematik: Solange alle drei Antwortalternativen ähnlich oder paarweise unterschiedlich formuliert sind, werden den Lernenden auf der Oberflächenebene anhand der Formulierungen kaum Präferenzen nahegelegt. Bei drei Antwortalternativen ist es jedoch ungünstig zwei ähnliche Formulierungen anzubieten, da so die dritte Alternative als ausgezeichnet wahrgenommen wird (entweder als korrekt oder falsch). Daher wurde in diesen Fällen mitunter darauf zurückgegriffen, die Formulierungen von eigentlich gleichen Aussagen durch Satzbau oder Wortwahl so zu variieren, dass die Formulierungen aller Antwortalternativen gewisse Unterschiede aufweisen.

Je nach gewählter Antwort auf die Multiple-Choice-Fragen wurde nun für jeden Fall entsprechendes Feedback gegeben. Dessen Natur wurde bereits in Kapitel 4.3 dargelegt.

5.3.2. Feedbackvarianten

Wie bereits auf Seite 69 beschrieben, erfolgt das Feedback in zwei Varianten – Textfeedback und Animationsfeedback. In einer Variante wird bei einer falschen Antwort ein Hinweis gegeben, der in der Simulation nochmals auszuprobieren ist. In der anderen Variante wird bei falschen Antworten eine Animation der Hinweise aus der ersten Variante gezeigt. Bei korrekten Antworten erfolgt in jeder Variante ein Texthinweis.

Zu beachten ist, dass das Feedback dabei so angelegt sein sollte, dass es einen fairen Vergleich zwischen beiden Feedbackvarianten ermöglicht. So war bei der Formulierung der Instruktion darauf zu achten, dass diese nachvollziehbar als Animation gestaltet werden konnte. Daher sind prototypische Formulierungen wie „Stelle die Bremse so ein, dass...“ oder „Stoße den Autoscooter nochmals direkt von unten an!“ gut geeignet. An einigen Stellen waren Formulierungen, die prinzipiell schlecht darstellbare Operatoren wie *Beobachte* und *Vergleiche* enthalten jedoch nur schwer umgänglich. Hier wurde in der Erstellung der Animationen auf sogenann-

tes Mouse-Highlighting zurückgegriffen. Dabei werden die relevanten Teilausschnitte besonders hinterlegt, wodurch die Aufmerksamkeit auf diese Punkte gerichtet werden soll.

Im nächsten Teil werden exemplarisch zwei Varianten vorgestellt. Diese beziehen sich auf die bereits vorgestellte Simulation „Mias Kart“. Zunächst wird hier die erste zugehörige Multiple-Choice-Frage dargestellt:

Du solltest herausfinden, wie sich die Fahrt von Mia beeinflussen lässt! Kreuze die bestmögliche Antwort an!

- Nur der Untergrund kann **nicht** die Geschwindigkeit und somit die Bewegung des Karts verändern, da er nicht „aktiv“ wirkt.
- Die Geschwindigkeit und somit die Bewegung des Karts ändert sich nur *deutlich*, wenn es durch z. B. Wind beeinflusst wird.
- Nur der Fallschirm und die Rakete können das Kart antreiben oder bremsen (denn sie sind ja fest mit ihm verbunden).

Feedback durch Text

The image shows a screenshot of a simulation interface with a light gray background and a dark border. The text is centered and uses different colors and bolding for emphasis. At the top, it says 'Dein Auftrag war:' in bold, followed by the task description. Below that, 'Deine Antwort' is shown in bold, followed by the user's incorrect answer in quotes. A red cross is placed over the word 'nicht'. The feedback states that the answer is incorrect. At the bottom, it provides a tip to try the simulation again with a specific suggestion.

Dein Auftrag war:
Du solltest herausfinden, wie sich die Fahrt von Mia beeinflussen lässt! Kreuze die bestmögliche Antwort an!

Deine Antwort
"Nur der Untergrund kann **nicht** die Geschwindigkeit und somit die Bewegung des Karts verändern, da er nicht "aktiv" wirkt."
war leider ein Irrtum.

Probiere die Simulation mit diesem Tipp nochmal aus:
Stelle den Schotter-Untergrund ein und beobachte den Verlauf der Geschwindigkeit.

Abbildung 5.5.: Beispiel für Textfeedback nach falscher Antwort

Abbildung 5.5 zeigt ein typisches Textfeedback nach inkorrekturer Antwort. Gut zu erkennen ist die Wiederholung des Forschungsauftrages zu Beginn. In der Mitte wird die gegebene Antwort wiederholt und sogleich als korrekt oder falsch bewertet (siehe auch Kapitel 4.3). Im unteren Bereich folgt dann der deutliche Hinweis, den folgenden Tipp mittels der Simulation zu erproben.

In einer Voruntersuchung wurde die falsche Antwort direkt mit einem markanten roten Kreuz als solche gekennzeichnet. Nach Beobachtungen, dass Schüler sich bisweilen nur auf das Erkennen dieser Markierung als auf den Tipp konzentrierten, wurde diese entfernt und eine zwar nach wie vor eindeutige aber weniger augenfällige Variante eingeführt (wie im Bild gezeigt).

III. Materialentwicklung

Dein Auftrag war:
Du solltest herausfinden, wie sich die Fahrt von Mia beeinflussen lässt! Kreuze die bestmögliche Antwort an!

Ein Hinweis zu deiner Antwort
"Die Geschwindigkeit und somit die Bewegung des Karts ändert sich nur deutlich, wenn es durch z. B. Wind beeinflusst wird.":

Gut beobachtet! Jeder der Einflüsse, die du einschalten kannst, beeinflusst die Geschwindigkeit des Karts. Die Geschwindigkeit verändert sich auch nur dann deutlich, wenn diese Einflüsse wirken.

Abbildung 5.6.: Beispiel für Feedback nach korrekter Antwort

Abbildung 5.6 zeigt ein typisches Feedback nach korrekter Antwort. Wie schon im vorigen Fall wird dazu zunächst der Forschungsauftrag und die gegebene Antwort wiederholt. Letztere wird dabei erst im unteren Teil als korrekt bewertet. Anschließend erfolgt noch eine kurze Erklärung zu den betreffenden Inhalten.

Ähnlich zu den inkorrekten Antworten war hier zunächst eine markante grüne Markierung eingefügt, die aber augenscheinlich ebenfalls dazu führte, dass weitergehende Hinweise nicht mehr beachtet wurden (siehe auch Kapitel 11.2).

Feedback durch Animation



(a) Screenshot aus Animationsfeedback Teil 1 (b) Screenshot aus Animationsfeedback Teil 2

Abbildung 5.7.: Beispiel für Animationsfeedback nach falscher Antwort

Die beiden Screenshots zeigen ein typisches Beispiel für Animationsfeedback. Dargestellt ist die animierte Variante des Textfeedbacks (siehe oben). Auch hier ist zunächst wieder der Forschungsauftrag aufgeführt, worauf die Bewertung (Stichwort „Irrtum“) folgt. Im unteren Bereich ist letztlich das Video zu sehen, das beliebig oft betrachtet werden kann. Bei genauer

Betrachtung fällt hier das sogenannte Mouse-Highlighting auf, das der Blickführung dient.

Das Feedback zu korrekten Antworten wird wie beim Textfeedback gegeben (mangels Hinweis, der in einer Animation darzustellen wäre).

Teil IV.

Methodik

6. Forschungsinteresse

In diesem Teil wird zunächst das Forschungsinteresse inklusive der Forschungsfragen dargelegt. Daran schließt sich eine Beschreibung des zur Untersuchung verwendeten Studiendesigns und Erhebungsinstrumente an, worauf wiederum eine Beschreibung der Durchführung der Studien folgt. Letztlich wird die Aufbereitung und Auswertemethodik der so gewonnenen Daten thematisiert.

6.1. Forschungsfragen

In Kapitel 3 wurde dargelegt, dass ein Vergleich des Lernerfolgs durch das 1D- und 2D-Konzept im Mechanikunterricht sich trotz der bisherigen Forschung weiterhin lohnt. Im Zentrum steht dabei ein möglichst *ausgeglichen*er Vergleich zweier *neu* entwickelter, *digitaler* Lernumgebungen zu jedem Konzept. Auch der Aspekt der Replikation ist hinsichtlich des Vergleichs des Lernerfolgs durch die Konzepte bedeutsam.

1. Inwiefern hat der Einsatz unterschiedlicher Unterrichtsansätze (1D bzw. 2D) Einfluss auf den Lernerfolg?

Im Rahmen dieses Vergleichs werden technische Potentiale von Neuen Medien gezielt und methodisch einheitlich ausgenutzt (siehe Kapitel 4.2.1 und 4.3). Dazu gehören Simulationen innerhalb einer digitalen, tabletbasierten Lernumgebung. Die technische Realisierung dieser digitalen Lernumgebung ermöglicht so auch den Einsatz von Feedback, welches durch schülerseitig zu beantwortende Multiple-Choice-Fragen generiert wird (siehe Kapitel 4.3). Dabei werden zwei Arten von Feedback bereitgestellt, deren potentiell unterschiedliche Wirkung unklar ist, so eine solche überhaupt zu identifizieren ist.

2. Inwiefern hat Textfeedback oder Animationsfeedback Einfluss auf den Lernerfolg?

Denkbar ist hierbei auch eine Interaktion zwischen dieser Feedbackvariable und dem Mechanikkonzept (beispielsweise könnten sich Animatio-

nen entlastend auswirken), weshalb auch dies untersucht werden soll.

3. Inwiefern interagiert das Unterrichtskonzept mit der Feedbackvariante hinsichtlich des Lernerfolgs?

Weiterhin ist die Akzeptanz des gesamten Lernsettings zu evaluieren, auch um daraus Hinweise für die *fachspezifische* Verwendung von Tablets zu erhalten.

4. Wie ist die Akzeptanz von Lernumgebung und Simulationen für Schülerinnen und Schüler im Anfangsunterricht der Mechanik?

Generell wird ein Set von (Kontroll)variablen wie Geschlecht, Vorwissen, wie aber auch affektive Konstrukte wie Interesse, Flow oder das Selbstkonzept berücksichtigt.

5. Inwiefern beeinflussen sich affektive Merkmale (Interesse, Selbstkonzept, Flow), die Unterrichtsansätze und Feedback?

In Anlehnung an die Ausführungen zum Cognitive Load wird entsprechend die letzte Forschungsfrage formuliert.

6. Inwiefern beeinflusst die Arbeit mit den Simulationen *unterschiedlicher* Mechanik-Zugänge die kognitive Belastung der Schüler?

7. Studiendesgin

7.1. Anlage der Studie

„Das typische Ziel einer Vergleichsstudie ist die Gegenüberstellung kognitiver bzw. motivationaler Wirkungen unterschiedlicher Unterrichtsansätze oder Medien“ (Theyßen 2014). Um solch eine belastbare Gegenüberstellung zu ermöglichen, sind vielfältige Ansätze denkbar. Im folgenden Abschnitt sollen daher grundsätzliche Fragen der Gestalt der vorliegenden Studie geklärt werden. Dabei werden Aspekte wie die Kontrolle von Randbedingungen, die Problematiken von Vergleichsstudien sowie die daraus resultierenden Implikationen für diese Studie dargelegt.

7.1.1. Labor- oder Felduntersuchung?

Eine grundsätzlich zu erörternde Frage stellt sich im Zusammenhang mit zu kontrollierenden Randbedingungen. Daher ist zu klären, ob die Untersuchung als Labor- oder Felduntersuchung¹ durchgeführt wird.

Für Bortz und Döring (2006) ist die Unterscheidung zwischen Labor- und Felduntersuchung keine Frage des Untersuchungsortes (wie die Bezeichnung nahe legen könnte), sondern eine Frage des Ausmaßes der Kontrolle der Versuchsbedingungen – „Im Feld geht sozusagen das normale Leben ungestört seinen Gang, im Labor dagegen ist das gesamte Setting auf den Forschungsprozess zugeschnitten.“ Im Kontext der fachdidaktischen Forschung ist mit *Feld* in der Regel eine schulische Unterrichtssituation (Aufschnaiter 2014) gemeint. Mit *Labor* ist hingegen meist ein (präparierter) Raum gemeint, in dem Schüler außerhalb des Unterrichts an der Untersuchung teilnehmen. Es ist naheliegend, dass sich in letzterem Setting die Randbedingungen leichter kontrollieren lassen (Bortz und Döring 2006). Zur Klärung der Vorzüge und Problema-

¹Im engeren Sinne wird hierbei zwischen Feldstudie und Feldexperiment unterschieden (Bergius 2018). Grundlegender Unterschied ist das Eingreifen in die Bedingungen bei einem Feldexperiment, während bei einer Feldstudie in erster Linie eine beobachtende Position eingenommen wird, ohne dass Bedingungen variiert werden oder eine Kontrollgruppe existiert (ebd.). In der fachdidaktischen Forschung hat sich dennoch der Begriff Feldstudie etabliert. Um Missverständnisse zu vermeiden, wird im Folgenden daher der Begriff *Felduntersuchung* genutzt.

tiken dieser beiden prototypischen² Modelle soll erörtert werden, welche (scheinbar störenden) Versuchsbedingungen die klassischen abhängigen Variablen wie Lernerfolg, Selbstkonzept und Interesse denn (spez. im Feld) überhaupt beeinflussen. Ohne Gewähr auf Vollständigkeit sind dies aus eher unterrichtspraktischer Sicht zunächst:

- Einfluss der Lehrkraft (Ausgestaltung/Interpretation des Unterrichtsinhalts, „Unsystematische Hinweise und Rückmeldungen“ (Aufschnaiter 2014))
- Lage der Stunde im Jahres-, Wochen- und Tagesverlauf
- Klassenarbeiten vor oder nach der Stunde
- Einstellung der Klasse zum Physikunterricht
- provisorische bzw. nicht zweckmäßige Räumlichkeiten
- terminliche Probleme/Ausfälle
- Klassenklima/Lernbereitschaft

Eher aus Forschungssicht initiiert (aus Theyßen (2014)):

- fachliche Inhalte
- Methoden
- Lernmaterialien
- Interventionsdauer
- Interessantheit der Intervention

Zu weiteren Determinanten, speziell des Lernerfolgs, sei das Angebots-Nutzungs-Modell von Helmke (2017) herangezogen. Anhand dessen lassen sich weitere Aspekte identifizieren, die den “Ertrag” (ebd.) mitbestimmen, die aber nicht mehr unkompliziert beobachtbar sind. Dies sind z.B. der familiäre Hintergrund, die Erziehung, Sozialisation, kulturelle oder regionale Besonderheiten oder das Lernpotenzial der einzelnen Schüler wie auch entsprechenderweise die fachlichen, didaktischen und pädagogischen Fertigkeiten der Lehrkraft.

Nach Aufschnaiter (2014) bietet eine Laboruntersuchung die Vorzüge, “eine systematische Kontrolle der für das Lehren und Lernen relevanten

²Prototypisch deshalb, da Labor- und Feldstudie die Pole eines Kontinuums darstellen, zwischen denen auch Studien mit Merkmalen beider Extreme angeordnet sein können (Bortz und Döring 2006).

Parameter”, wozu auch “unsystematische Interaktionen mit der Lehrkraft” gehören. Die weiteren Vorzüge beziehen sich nach Aufschnaiter (ebd.) auf die Einfachheit der gezielten Beobachtung einzelner Schüler/innen oder kleiner Gruppen. Des weiteren zeigen sich Laborstudien nicht zuletzt auch deswegen prädestiniert dafür, „prozesshafte“ Untersuchungen durchzuführen, d.h. zu klären *wie* etwas getan wird. So gelingt es in der Laborsituation „Prozessmerkmale leichter identifizieren und spezifischen Parametern zu[zu]schreiben.“ (ebd.).

Die Vorzüge einer Felduntersuchung seien im Folgenden dargestellt: Zunächst ist es naheliegend, Lernen dort zu untersuchen, wo es stattfindet (ebd.). Somit lässt sich die Wirkung eines Konzept in authentischer Umgebung erforschen (vgl. DBR-Ansätze hier). Weiterhin lassen sich mit überschaubarem Aufwand größere Fallzahlen (>200) als in Laborstudien generieren, da in der Regel immer komplette Klassengemeinschaften angesprochen werden. Die Persönlichkeitsmerkmale der Teilnehmer/innen sind ebenso breiter gestreut als in einer Laborstudie, da zu vermuten ist, bei einer Studienteilnahme außerhalb der Unterrichtszeit vorwiegend intrinsisch motivierte und physikaffine Schüler/innen anzusprechen und die Stichprobe bzgl. dieser Merkmale nicht mehr zur Grundgesamtheit passen würde. Dem ließe sich nur mit einem Anreizsystem entgegenwirken. Somit kann eine Feldstudie zu einer durchaus höheren externen Validität beitragen. Dies bedeutet, dass bei hoher externer Validität Ergebnisse mitunter „über die besonderen Bedingungen der Untersuchungssituation und über die untersuchten Personen hinausgehend generalisierbar sind.“ (Bortz und Döring 2006).

Die Problematiken einer Feldstudie zeigen sich in den bereits aufgeführten Variablen, die zu kontrollieren nur schwerlich möglich ist. Dies kann wiederum in einer mangelnden internen Validität resultieren. Ist das der Fall bedeutet dies, dass die „Ergebnisse [nicht; hinzugefügt vom Verfasser] kausal eindeutig interpretierbar sind.“ (ebd.).

Schwierigkeiten von Laborstudien: „Die Unnatürlichkeit der Untersuchungsumgebung“ lässt es allerdings häufig fraglich erscheinen, ob die Ergebnisse auch auf andere, »natürlichere« Situationen generalisierbar sind.“ (ebd.). Dies sei gerade im Kontrast von *schulischer Realität* zu *nachmittäglichem Untersuchungssetting* nochmals betont. Weitere Problematiken einer Laborstudie ergeben sich z. B. direkt aus den formulierten Vorzügen der Feldstudien (z. B. Fallzahlen, Schülersauswahl). Dazu kommt, dass auch im Labor zahlreiche Variablen nicht zu eliminieren sind. Hierbei soll nochmals auf die Determinanten des Lernerfolgs nach den Angebots-Nutzungs-Modell verwiesen werden (Helmke 2017). Somit läuft man Gefahr einer trügerischen Vorstellung der umfassenden Kontrolle zu unterlaufen.

Für die vorliegende Untersuchung wurde aus dem Dargelegten resultierend die Felduntersuchung mit Ausschluss der Lehrervariable als Form gewählt.

Als Begründung zählt vorrangig die Untersuchung in einer relativ natürlichen Unterrichtssituation, mit einer authentischen Lerngruppe in gewohnter Schulumgebung. Auf diese Weise kann potentiell eine höhere externe Validität gewährleistet werden, auch wenn dadurch die zweifelsfreie Zuschreibung von einzelnen Effekten zu bestimmten Faktoren nicht mehr ohne weiteres möglich ist. Auch lassen sich die relativ großen Fallzahlen so relativ einfach generieren.

Aufgrund der mittlerweile fortschreitenden Verbreitung von Tablets an Schulen ist es weiterhin der Einschätzung des Autors nach ratsam, die Erprobung von Tablets unter Realbedingungen zu forcieren. Nur so kann auch tatsächlich auftretenden Problemen der Bedienung, Fehleranfälligkeit und Robustheit Rechnung getragen werden, wie sie auch im „echten Tableteinsatz“ im gegenwärtigen Unterricht zu vermuten sind.

Im besonderen Maße sei auf die Verbreitung von Feldstudien in ähnlich angelegten Untersuchungen verwiesen (vgl. Jung, Reul und Schwedes 1977; R. Wodzinski 1996; Wilhelm 2005; Tobias 2010), sodass im Sinne einer gewissen Vergleichbarkeit auch hier eine Felduntersuchung durchgeführt wird.

Der wiederkehrend angesprochene Aspekt der Lehrervariable wird in der vorliegenden Studie durch den Einsatz von vorgefertigten Lernumgebungen (siehe Kapitel 5.2) im Hinblick auf die Anordnung der Inhalte quasi eliminiert und in Bezug auf die „unsystematischen Interaktionen“ (Aufschnaiter 2014) in erheblichem Maße reduziert.

7.1.2. Vergleichsstudien

In diesem Abschnitt soll die Besonderheit einer Vergleichsstudie dargestellt werden. Im Groben orientiert sich der Abschnitt an Theyßen (2014), die am Beispiel einer Medienvergleichsstudie unterschiedliche Aspekte erläutert, die in der Planung zu berücksichtigen sind. Da in der vorliegenden Studie ebenfalls grundsätzlich zwei Zugänge medienunterstützt miteinander verglichen werden, orientiert sich der folgende Abschnitt daran.

Fairness

Ein wichtiger Faktor beim Vergleichen zweier Zugänge, Methoden oder Konzeptionen ist die Fairness. Am Beispiel des Medienvergleichs bedeutet dies, dass eine aufwändige Neuentwicklung gegen eine zufällig vorhandene, schon immer dagewesene Methode zwar gegeneinander getestet wer-

den kann. Möglich Effekte können dann weniger im Medium begründet sein, als in der aufwändigen Entwicklung (ebd.). Ebenso problematisch beschreibt Theyßen (ebd.) auch das Gegenteil, indem langfristig bewährte und etablierte Konzepte gegen zügig erstellte Konzepte getestet werden. Daraus lässt sich schließen, dass beiden zu vergleichenden Aspekten die gleiche Aufmerksamkeit und Sorgfalt in der Planungsphase zuzukommen ist.

Ebenso wichtig ist die bereits in Kapitel 7.1.1 angesprochene Kontrolle der Randbedingungen. Für beide Kohorten sollen also möglichst ähnliche Bedingungen geschaffen werden. Dies schließt sowohl Faktoren, die erst im Feld relevant werden (z. B. die Instruktion, Betreuung) als auch designbedingte Faktoren mit ein. Letzteres bezieht sich z. B. auch auf die Methodenwahl (vgl. ebd.). So könne die Zuschreibung von Effekten auf Medien oder Zugänge komplexer werden, wenn sich mit dem Medium auch die Methode ändert (Beispiel (nach Theyßen (ebd.)): *Simulation in Einzelarbeit* gegen *Demonstrationsexperiment durch Lehrkraft*).

Im Falle der vorliegenden Studie geht es nicht um den Vergleich zweier Medien, sondern um den mediengestützten Vergleich zweier Unterrichtszugänge zur Mechanik. Im Sinne der beschriebenen Aspekte zur Fairness werden beide Konzepte auf dem gleichen Fundament erprobt. Beiden Zugängen liegt eine Lernumgebung mit eingebetteten Simulationen zu Grunde, die ähnliche inhaltliche Grobstrukturen, Textlängen und Kontexte aufweisen³. Es handelt sich also für beide Zugänge um Neuentwicklungen des Lernmaterials. Weiterhin wird bei beiden Zugängen das Lernmaterial auf gleiche Weise im Unterricht eingesetzt.

Mängel bezüglich der Fairness können sich jedoch bzgl. der unterschiedlichen "Freiheitsgrade" in den Simulationen ergeben. Es ist nicht abwegig zu vermuten, dass 1D-Simulationen – mit Beschränkungen der Bewegungsrichtungen – als nicht gleichartig spannend wie 2D-Simulationen – die naturgemäß mehr Gestaltungsspielraum bieten – empfunden werden. Ein weiterer Aspekt diesbezüglich ist die Fairness hinsichtlich der kognitiven Tests. So ist offensichtlich, dass Schüler, die nach zwei Zugängen unterrichtet wurden, jeweils Wissen erworben haben, dass die anderen nicht *erwerben konnten* (und nicht etwa *nicht erworben haben*). Somit muss bei der Beurteilung von Lernzuwächsen in dieser Hinsicht große Vorsicht walten gelassen werden.

Neuigkeitseffekte

„Ein wesentlicher Unterschied zwischen neu eingeführten und etablierten Bildungstechnologien besteht darin, dass die Nutzung der neuen

³Streckenweise sind beide Lernumgebungen mitunter fast identisch.

Technologien und die entsprechenden Einstellungen der Nutzer vom Neugigkeitseffekt geprägt sind.“ Ein Effekt wie ihn so Nistor (2013) beschreibt (siehe auch Kapitel 4.2.1), ist beim Umsetzen dieser Studie mitunter zu befürchten – da die Nutzung von elektronischen Lernumgebungen, Simulationen und Tablets *in dem Maße* unüblich ist. Nistor (ebd.) ergänzt, dass Nutzer neuer Technologien in solchen Fällen eher neugierig und mit größerem Interesse agieren. Theyßen (2014) betont aber auch, dass der Neugigkeitseffekt „nicht originär mit der neuen Konzeption verbunden“ ist, und „mit der Gewöhnung“ abnimmt – wie es schon der Name nahelegt.

Zwei Aspekte könnten den erwarteten Neugigkeitseffekt dennoch leicht abmildern. Zum einen sind die Schüler/innen mit Tablets/Smartphones aus ihrem Alltag vertraut. Zum anderen empfiehlt Theyßen (ebd.): „Für Studien mit kurzer Interventionsdauer kann man aber zumindest versuchen, alle zu vergleichenden Konzeptionen in etwa gleich interessant und attraktiv für die Lernenden zu gestalten, den Neugigkeitseffekt also bewusst in Kauf zu nehmen, aber nicht nur einseitig.“ Dem wird in dieser Studie Rechnung getragen, indem beide Zugänge mit gleicher Technologie und ähnlichem Aufbau wie Inhalt getestet werden.

Lehrervariable

Schon Jung, Reul und Schwedes (1977) erwähnten bei der Durchführung ihrer damaligen Studie: „Selbstverständlich spielt die Lehrervariable keine zu vernachlässigende Rolle. [...] Die schulisch-realen Möglichkeiten schienen uns aber allerdings kaum einen Ansatzpunkt zu geben, diese Variable effektiv zu kontrollieren [...]“ (ebd.). Auch im Kapitel zum Feldstudiendesign wurde die Rolle der Lehrkraft schon dargelegt. Eine Alternative zum lehrerzentrierten Unterricht ist, dass „Lernmaterialien und Arbeitsanweisungen schriftlich an die Schüler herangetragen werden“ (Theyßen 2014). Dies ergibt zwar eine weitere unbekannte Variable (ebd.), allerdings kann wohl davon ausgegangen werden, dass sich diese Variable in unterschiedlichen Lerngruppen ähnlich verhält – im Gegensatz zu unterschiedlichen Lehrerpersönlichkeiten. In der vorliegenden Studie wird dies durch die vorgefertigte Lernumgebung sichergestellt.

8. Erhebungsinstrumente

8.1. Fachtests

8.1.1. Grundsätzliches zur Konstruktion und Testfairness

Zur Messung eines Lernfortschritts bei den SuS sollen vor und nach der Intervention durch jeweils einen der beiden Mechanik-Lehrgänge Fachtests geschrieben werden. Zu beachten ist dabei, dass die SuS bisher über keine unterrichtlichen Vorerfahrungen zu der unterrichteten Thematik verfügen. Bei der Auswertung des Fachtests wäre daher zu berücksichtigen, dass mitunter erhebliche Lernzuwächse zu erwarten sind. Ursächlich dafür könnten nach Ansicht des Autors Pretests¹ sein, bei denen die SuS kaum eine realistische Chance haben auf Basis ihres Vorwissen zu antworten.

Aus dieser Perspektive scheint es geschickter, Vor- und Nachtest so konstruieren, dass das latente Konstrukt „Mechanikkompetenz“ jeweils auf unterschiedlichem Anspruchsniveau, jedoch mit vergleichbaren fachlichen Inhalten gemessen werden. Die Anordnung der Ausprägung dieses latenten Konstrukts an einer gemeinsamen Skala (vgl. Kapitel zur Raschanalyse, 10.3) kann durch das Verwenden gemeinsamer Items sichergestellt werden. Denn nur so können Vergleiche von Pre- zum Posttest gezogen werden.

Ein weiterer Aspekt der Testfairness betrifft den Vergleich zwischen 1D- und 2D-Zugang. So müssen die Items es ermöglichen, dass beide Gruppierungen faire Chancen haben, die Aufgaben lösen zu können. Fragen nach Vektoraddition oder der Differenzierung innerhalb des Geschwindigkeitsbegriffs sollten somit nur insofern in die Auswertung einbezogen werden, wenn sie gesondert ausgewertet werden.

8.1.2. Itemauswahl

Zur Zusammenstellung eines Aufgabenpools wurden bekannte Tests gesucht, z. B. Beichner (1994), Tobias (2010), Hestenes und Wells (1992), Thornton (1998). Die einzelnen Aufgaben wurden auf inhaltliche Passung

¹In dieser Arbeit werden die Fragebögen zu fachlichen Belangen als *Tests* bezeichnet, bzw. die Tests vor der Intervention als Pre- oder Vortests und die Bögen im Anschluss an die Intervention Post- oder Nachtests.

geprüft, da Aufgaben und Inhalte der Lernumgebung in etwa aufeinander abgestimmt sein sollten.

Inhaltsvalidität: Zur Überprüfung der Aufgaben wurden Physiklehrkräfte, Physikdidaktiker und Physiker der Arbeitsgruppe herangezogen (N=4). Ähnlich zu Tobias (2010) wurden die Aufgaben als eher für 1D- bzw. 2D-Schüler gut lösbare Aufgaben, sofern zutreffend, geratet. Ferner wurde die Schwierigkeit der Aufgaben für Schüler ohne Mechanikunterricht und für Schüler nach dem Mechanikunterricht in 7/8 jeweils dreistufig geratet, um auch daraus Informationen für die Zusammenstellung der Vor- und Nachtests ziehen zu können. Alle Items sind im Anhang zu finden (siehe Anhang A.2.1).

Anmerkungen zur Itemauswahl

In diesem Abschnitt werden Quellen, aus denen Items entnommen wurden, eingeordnet.

Zu Beichner 1994: Der *Test of Understanding of Graphs in Kinematics* (TUG-K) nach Beichner (1994) wurde entwickelt, da Tests wie der FCI oder Mechanic Baseline Test das Interpretieren von Graphen in zu geringem / keinem Umfang anbieten. Somit konnte das Ziel, das Aufdecken von Problemen beim Interpretieren von Graphen, nur durch eine solche Neuentwicklung erreicht werden.

Die Auswahl der Aufgaben wurde vor allem durch die Verwendung des Beschleunigungsbegriffes eingeschränkt. Dieser wird im vorliegenden Lehrgang zunächst nicht eingeführt. Ebenso betrifft dies das Berechnen von Geschwindigkeitsbeträgen für ein gegebenes Zeitintervall aus t-x-Diagrammen. Daher wurden auch Aufgaben dieser Art ausgeschlossen. Die zwei gewählten Aufgaben scheinen inhaltlich geeignet.

Zu Bader 2008, Bredthauer 2010: Bei diesen Quellen (Bredthauer u. a. 2010; Bader und Dorn 2008) handelt es sich um Schulbücher. Entgegen anderer verwendeter Quellen liegt die Zielgruppe der verwendeten Aufgaben aus Bader (d. h. Schulbuch *Dorn/Bader*) und Bredthauer (d. h. Schulbuch *Impulse*) eindeutig in den Klassenstufen 7 und 8 – die thematische Passung ist darüber hinaus ebenfalls gegeben.

Zu Tobias 2010: Der vollständige Test nach Tobias (2010) wurde wegen fehlender inhaltlicher Kongruenz nicht übernommen. So enthält das Original Aufgaben zum 3. Newton'schen Axiom (hier nicht thematisiert), aber

nicht zum Ablesen und Interpretieren von Diagrammen. Entnommene Items sind gekennzeichnet.

Zu FCI-Items / Hestenes 1992 (FCI): Der FCI (Hestenes und Wells 1992) als Messinstrument ist nicht unumstritten, weshalb der Frage „Was misst der FCI eigentlich?“ mehrfach nachgegangen wurde. Sein ursprüngliches Versprechen kann er nicht einhalten, u. A. da sich keine typischen Antwortmuster identifizieren lassen und so nicht auf „das“ Kraftkonzept eines Schülers geschlossen werden kann (Schecker und Gerdes 1999) (vgl. auch „Was der FCI misst, ist ein diffuser Lernerfolg im Bereich der Mechanik“ (ebd.)). Zweifel an seiner Tauglichkeit als Instrument zur Untersuchung unterschiedlicher Unterrichtskonzepte besteht hingegen kaum (ebd.). Allerdings weist Ivanjek (2016) darauf hin, dass der FCI als Vor- und Nachtest deutlich unterschiedlich funktionieren kann. Nicht zuletzt könnte dies auch auf die geringe Testbreite (Mangel an sehr leichten und sehr schwierigen Aufgaben) zurückzuführen sein. So schlägt sie zwei Testvarianten mit Ankeritems vor. Weshalb der FCI in der vorliegenden Studie aber nicht komplett genutzt wird, begründet sich u. A. in der deutlich (!) engeren bzw. anderen Themenauswahl der Lernumgebung und der Länge des Tests.

Zu Jung 1977: Jung, Reul und Schwedes 1977 führten Mitte der 70er Jahre eine Studie mit Schülern der Primar- und frühen Sekundarstufe zur Einführung der Mechanik durch. Eine Aufgabe des dabei durchgeführten Nachtests wird hier genutzt, Thema Tempo bzw. Geschwindigkeit. Bzgl. dieser Aufgabe wurde in allen damals untersuchten Klassen ein Lernfortschritt festgestellt (mit der Einschränkung, dass es sich um sehr junge Schüler handelt, für Klasse 7/8 könnte die Aufgabe zu einfach sein – komplexere Varianten bieten sich aber an). Zwei weitere Aufgaben, die nicht aus den Tests sondern aus den Schülerarbeitsmaterialien stammen, wurden adaptiert.

Zu Thornton 1988: Aufgaben 1 und 2 der Force Sled Aufgabe wurden ausgewählt. Zu den Aufgaben selbst nimmt der Autor u. A. wie folgt Stellung: „both the Force Sled and the Force Graph questions explore the relationship between force and motion by asking about similar motions [...]. The Force Sled questions make no reference to graphs and no overt reference to a coordinate system. They use “natural” language as much as possible, and they explicitly describe the force acting on the moving object. The combination of Force Sled questions 1–4 and 7, on the other hand, indicate reliably in a statistical sense! the prevalence of non-Newtonian and Newtonian student views.“ (Thornton 1988)

8.2. Fragebögen

In den Fragebögen² werden Aspekte wie das Interesse, das Selbstkonzept wie auch der Umgang mit Feedback, der Lernumgebung und den Simulationen sowie deren Akzeptanz vor und nach der Intervention erhoben.

Zur Konstruktion In den Fragebögen der Pre- und Post-Version wurde jeweils eine ungerade Anzahl von Antwortoptionen erstellt. Der Nutzen einer mittleren Antwortmöglichkeit wird in der Literatur recht kontrovers diskutiert. Während Moosbrugger und Kelava 2012 darlegen, dass mittlere Antwortoptionen

„als eine Ausweichoption [dienen], wenn die Testperson den angegebenen Wortlaut als unpassend beurteilt, die Frage nicht verstanden hat, die Antwort verweigert oder aber diese nicht kennt. Umgekehrt wird diese Kategorie von besonders motivierten Probanden gemieden.“

(ebd.)

Die in den Fragebögen gestellten Fragen orientieren sich an einer schülergerechten Sprache, zu einem Teil handelt es sich auch um bereits erprobte Fragen. Unverständnis scheint daher ein untergeordneter Faktor zu sein. Die Nähe der Fragen zum schulischen Erleben schließt ebenso aus, dass die Antwort nicht gekannt wird. Weiterhin würde man bei einer geraden Anzahl von Items bei nicht verstandenen Fragen ebenso das Risiko eingehen, dass sogar eine tendenziöse Antwort erzwungen wird. Dass die Kategorie von „besonders motivierten“ Teilnehmern gemieden wird kann unter Umständen sogar erwünschtes Verhalten sein, da extremere Merkmalsausprägungen ja eben erfasst werden sollen.

Die von Moosbrugger und Kelava (ebd.) vorgeschlagene „Weiß nicht“-Kategorie zur Abmilderung der potentiellen Probleme wurde nicht eingeführt. Nach Franzen 2014 erhöht sich so der Anteil der Personen, die aus Vermeidung des Aufwandes diese Kategorie wählen.

Weiterhin kann eine mittlere Kategorie wertvolle Informationen zur Verfügung stellen, die bei erzwungener Tendenz verloren gingen. Somit wurde eine ungerade Anzahl von Antwortoptionen, hier fünf, gewählt. Die Beschriftung erfolgte je nach Sachzusammenhang mit *(Fast) nie – Selten – Manchmal – Häufig – (Fast) immer* bzw. *Trifft nicht zu – Trifft eher nicht zu – Teils / Teils – Trifft eher zu – Trifft voll zu*.

²In dieser Arbeit werden die Fragebögen zu nicht-kognitiven Belangen als *Fragebögen* bezeichnet, bzw. die Bögen vor der Intervention als Pre- oder Vorfragebögen und die Bögen im Anschluss an die Intervention Post- oder Nachfragebögen.

In einigen Fällen war es nötig, die Formulierung der Items im *Nachfragebogen* entsprechend der Intervention leicht anzupassen (z. B. Pre: „Mir macht Unterricht Spaß“; Post: „Mir hat der Unterricht mit Simulationen Spaß gemacht.“ (zu diesem Vorgehen, vgl. z. B. Rode (2016))).

8.2.1. Beschreibung der erhobenen Konstrukte

Die Pre- und Postfragebögen bestehen aus Items, die sich auf bestimmte Konstrukte wie das Interesse beziehen und Items, in denen einzelne interessierende Aspekte erfragt werden (letztere z. B. betreffen den Bedienkomfort der Tablets oder den empfundenen Lernzuwachs wie auch statistische Daten wie Geschlecht, Nutzungshäufigkeit von Tablets etc.). Das Interesse, Selbstkonzept und Flow-Erleben werden dabei als gängige Konstrukte erhoben. Die Fehlerkultur wird erhoben, um gegebenenfalls das formative Feedback im Vergleich zum sonstigen Unterricht hinsichtlich des Umgangs mit Fehlern beurteilen zu können. Aufgrund der Forschung im Feld ist es unbedingt erforderlich ein angemessenes Verhältnis zwischen den eigenen und den absolut berechtigten Interessen der Schule, Lehrkräfte, Schüler und Eltern zu wahren. Daher können bei der Gestaltung der Lernumgebung und der wissenschaftlichen Begleitung Forschungsinteressen nicht immer im Vordergrund stehen. Im Speziellen geht es hier um die Länge und damit um die zeitliche Ökonomie der Durchführung der Tests (vgl. Moosbrugger und Kelava (2012)). Daher wären sicherlich weitere zu testende Aspekte und Konstrukte denkbar (z. B. *Kognitive-Fähigkeiten-Test*), diesen ließen sich aber vernünftigerweise nicht mehr unterbringen.

Interesse

Hinsichtlich des Interesses wird hier das Interesse am fachlichen Gegenstand wie auch die Begeisterung am unterrichtlichen Umgang mit diesem Gegenstand abgefragt³. Im Vorfragebogen bezieht sich dies auf den Vorunterricht, im Nachfragebogen auf die Intervention. Es zeigen sich beim Interesse durchaus Unterschiede zwischen Mädchen und Jungen. Im Rahmen der Interessenstudie von Häußler und Hoffmann (1995) zeigt sich, dass das Interesse von Mädchen am Physikunterricht (nicht aber zwingend auch das Interesse an Physik) geringer ist, als von Jungen. Einer der Faktoren, die das Interesse am Fach beeinflussen ist das Selbstkonzept (Häußler und Hoffmann 1995; R. Wodzinski 2015), welches im nachfolgenden Teil beschrieben wird. Die Items zum Interesse sind

³Es handelt sich also um ein eher ganzheitliches Konstrukt zugunsten der Kompaktheit und zulasten des Auflösungsvermögens

adaptiert nach Baumert u. a. (1997) und Hoffmann, Häußler und Lehrke (1998) und sind in Tabelle A.1 (Seite 264) aufgeführt.

Selbstkonzept

Unter dem (physikalischen) Selbstkonzept versteht man die „globale[] Einschätzung zu dem, wer man ist, was man kann und unter Umständen auch wie man diese Eigenschaften bewertet.“ (Rabe, Meinhardt und Krey 2012). Das Selbstkonzept kann dabei auch „domänenspezifisch“ betrachtet werden (ebd.), hier also im Bezug auf Physik. Nach Shavelson, Hubner und Stanton (1976) steht das Selbstkonzept in wechselseitiger Wirkung zum eigenen Handeln („One’s perceptions of himself are thought to influence the ways in which he acts, and his acts in turn influence the ways in which he perceives himself“ (ebd.). Am Beispiel des Physikunterrichts könnte dies bedeuten, dass bei einer klassischen Selbstwahrnehmung eines Schülers als „schlecht in Physik“ sich dies auch in seinen Handlungen widerspiegelt. Anders herum könnten gewisse Erfolgserlebnisse zu einem besseren physikalischen Selbstkonzept beitragen. Zwiorek (2006) konstatiert dazu, dass dabei Mädchen im Grunde „weniger Vertrauen in ihre physikalischen Fähigkeiten [haben] als Jungen.“. Daraus resultiert auch die recht verbreitete Untersuchung dieses Konstrukts (siehe z. B. Tobias 2010; Winkelmann 2015; Amenda 2017). Hinsichtlich des Interesses an Physik sei noch angemerkt, dass dieses auch vom jeweiligen Selbstkonzept des Schülers abhängt (Zwiorek 2006). Die Items zum physikbezogenen Selbstkonzept sind adaptiert nach Wackermann (2008) (wiederum basierend auf Bos (2009)). Die Items für dieses Konstrukt sind in Tabelle A.2 (Seite 265) aufgeführt.

Flow

Als Flow bezeichnet man nach Rheinberg, Vollmeyer und Engeser (2003) das „reflexionsfreie, gänzliche Aufgehen in einer glatt laufenden Tätigkeit, die man trotz hoher Beanspruchung noch unter Kontrolle hat.“ Das Flow-Erleben wird durch die jeweilige Person als positiv empfunden, eine Hinderung an der Tätigkeit zieht „Beeinträchtigungen“ nach sich (ebd.). Die Motivation entsteht durch die Tätigkeit an sich, das Ziel steht also gegenüber dem Weg dorthin im Hintergrund. Grundsätzlich lassen sich sechs Komponenten des Flow-Erlebens ausmachen (entnommen aus Rheinberg, Vollmeyer und Engeser (ebd.) mit Rückgriff auf weitere Literatur). Dazu gehören die Klarheit, was zu tun ist, keine Unter- oder Überforderung, ein „gleitender“ Arbeitsablauf, volle und automatische Konzentration auf das Wesentliche, ein gewisser Verlust des Zeitgefühls wie ein „Verschmelzen“ mit der Tätigkeit. Nach Mézes, Erb und Schröter (2012) wechselwirken

Flow und Lernleistung so, als dass ein gesteigerter Flow zu besseren Lernleistungen führt, wie auch umgekehrt. Weiterhin könnte der Flow ein Indikator für eine gute Passung von Fähigkeiten einer Person und Anspruch einer Aufgabe sein. Drei Items zum Flow-Erleben wurden nach Mézes, Erb und Schröter (ebd.) adaptiert (basierend auf Rheinberg (2004)). Die Items für dieses Konstrukt sind in Tabelle A.3 (Seite 266) aufgeführt.

Fehlerkultur

Die Fehlerkultur (der Umgang der Lehrkraft mit Fehlern der Schüler) des Unterrichts wird mittels Items von Wackermann (2008) (basierend auf Clausen (2004)) erhoben. Die Items für diesen Inhaltsbereich sind in Tabelle A.4 267 aufgeführt.

8.2.2. Weitere Inhaltsbereiche

In diesem Bereich werden im Vorfragebogen (siehe Tabelle A.5, Seite 268) zunächst Kontrollvariablen der Versuchspersonen erhoben. Dazu gehören die Noten in Physik, Mathematik und Deutsch wie auch das Geschlecht – jeweils zur Bildung sinnvoller Untergruppen bei der Analyse der unabhängigen Variablen. Weiterhin wird das persönliche Nutzungsverhalten digitaler Medien erhoben; diese Fragen zum Einsatz von Computern und Tablets im privaten Umfeld entstammen PISA (vgl. auch Finkenberg 2018). Die Fähigkeiten im bisherigen Umgang mit Touchscreens oder Simulationen werden über selbst konstruierte Items erhoben.

Die Items speziell auch zur Usability und Akzeptanz der Lernumgebung im Postfragebogen (siehe Tabelle A.6, siehe 269) sind Eigenentwicklungen (dazu gehören etwa die Verständlichkeit der Texte, Klarheit der Instruktion, Bedienbarkeit der Simulationen und Steuerung der Lernumgebung). Weiterer Bestandteil des Postfragebogens sind Aspekte der Feedbacknutzung, wie z. B. der empfundene Nutzen dessen.

8.3. Fragebögen zum Cognitive Load

Sowohl für eine Vorstudie im Rahmen einer Masterarbeit (Mitschker 2017) als auch für die Erhebung zum Cognitive Load (siehe Kapitel 4.4) im Rahmen der Hauptstudie wurden Fragebögen entwickelt. Nach Mitschker (ebd.) ist die empfundene geistige Anstrengung ein guter Indikator für den Mental Effort. Daher stellt dort die Frage „Wie sehr musstest du dich während der Simulation geistig anstrengen?“ ein zentrales Item dar. Zur Beantwortung steht eine neuenstufige Skala von *sehr geringen Anstrengung* bis *sehr große Anstrengung* zur Verfügung. Diese Frage sollte nach

erstmaliger Bearbeitung sowie nach eventuellem Feedback nochmals bearbeitet werden. Weitere Items, die von Mitschker (2017) als Indikatoren für intrinsischen, extrinsischen und germane Cognitive Load genutzt werden, sind hier zwar aufgeführt, die Ergebnisse aber werden aufgrund der Fokussierung der vorliegenden Arbeit nicht aufgeführt:

- Die Simulation hat meine Aufmerksamkeit geweckt.
- Die Simulation empfand ich als zu komplex.
- Ich hatte Spaß beim Arbeiten mit der Simulation.
- Die Aufgabe war mithilfe der Simulation einfach zu bearbeiten.
- Ich empfand die Simulation als langweilig.

Für die Hauptstudie ist die Messung des Cognitive Load kein Schwerpunkt. Entsprechend wurde aus ökonomischen Gründen nur das zentrale Item der Vorstudie – minimal umformuliert – übernommen. Auch in der Hauptstudie soll dieses Item nach wiederholter Simulationsbearbeitung nochmals bearbeitet werden. Es lautet: „Wie sehr musstest du dich beim Bearbeiten des Forschungsauftrags mit dieser Simulation anstrengen?“

8.4. Interviewerhebung

8.4.1. Bisherige Interviews: Inhaltliche Aspekte

Bereits in vorrangegangenen Studien (z. B. R. Wodzinski 1996; Tobias 2010) wurden Interviews mit den Lernenden durchgeführt. Zunächst soll hier eine knappe inhaltliche Darstellung dieser erfolgen, um im Anschluss daran die eigenen inhaltlichen Aspekte der vorliegenden Interviews einzuordnen.

INTERVIEWERHEBUNG TOBIAS

Die Interviews wurden ausschließlich in den Treatment-Klassen, die nach dem 2D-dynamischen Ansatz unterrichtet wurden, durchgeführt. So sollten weitere Erkenntnisse zu diesem Ansatz gewonnen werden, die aus Gründen der Vergleichbarkeit im standardisierten MC-Fachwissenstest nicht erhoben werden konnten. In den Interviews wurde zu den Kategorien Interesse, Selbstkonzept und Sachverständnis gefragt. Letzteres bildete dabei den Schwerpunkt. Die Kernthemen dabei waren der persönliche Eindruck der Schüler zu den „gelernten Inhalten“, der Geschwindigkeitsbegriff, Konstruktionen sowie die Newton'sche Bewegungsgleichung in der dort genutzten Form. Typische Fragen beziehen sich auf Konstruktionen

von End- oder Zusatzgeschwindigkeiten aus gegebenen Pfeilen oder auf die Auswirkungen der Veränderung der Masse oder der Einwirkungsdauer auf die Zusatzgeschwindigkeit.

INTERVIEWERHEBUNG WODZINSKI

Die Hauptbefragung wurde thematisch sowohl in einer ein- als auch in einer zweidimensionalen Variante durchgeführt. In mehreren Sitzungen wurden dazu gemeinsam folgende Themen bearbeitet: Beschreibung von Bewegungen, Tempo, Geschwindigkeit und Beschreibung von Geschwindigkeitsveränderungen. Auch die Bestimmung von Geschwindigkeitsänderungen im 1D-Fall, die Behandlung von Knautschzonen sowie Aspekte von Reibung, Gravitation und freiem Fall waren Interviewbestandteil. Spezifisch für den 1D- und 2D-Fall war die Herleitung der Newtonschen Bewegungsgleichung in den jeweiligen Dimensionen. Die Bestimmung der Zusatzgeschwindigkeit erfolgte in beiden Gruppen für den zweidimensionalen Fall, jedoch zu unterschiedlichen Zeitpunkten.

Weitere Befragungen im Rahmen des Projekts hatten zwar andere Schwerpunkte, bezogen sich aber im Groben auf die beschriebenen Inhalte. Auch wenn zusätzlich schriftliche Befragungen durchgeführt wurden, so lässt sich ein fachlicher Schwerpunkt konstatieren.

Zusammenfassend haben beide Interviews einen deutlichen fachlichen Charakter. Welche Anknüpfungspunkte bestehen (oder gerade nicht bestehen), soll im folgenden Abschnitt, der die Erstellung des Leitfadens thematisiert, aufgegriffen werden.

8.4.2. Erstellung des Leitfadens

Der Interviewleitfaden stellt das Gerüst der Fragen bereit, die im Laufe des Interviews geklärt werden sollen (Gläser und Laudel 2010; Niebert und Gropengießer 2014). Dabei ist die Reihenfolge sowie exakte Formulierung nicht festgelegt (Gläser und Laudel 2010).

Die Fragen des Leitfadens sollten sprachlich einfach gehalten werden (hier mit Hinblick auf die junge Zielgruppe sowieso), an den Alltag der Befragten anknüpfen und nicht suggestiv sein (vgl. Gläser und Laudel 2010; Niebert und Gropengießer 2014). Im vorliegenden Fall werden Formulierungsvorschläge angeboten, nicht zuletzt auch um eine gewisse Vergleichbarkeit zwischen Interviews herzustellen oder Gewöhnungseffekten vorzubeugen (Gläser und Laudel 2010). Die genaue Formulierung wird aber dem Interviewer überlassen. Lediglich zentrale Formulierungen sollten in etwa eingehalten werden bzw. als Stütze dienen, sind im Leitfaden fett markiert (ebd.), um ihre Ankerfunktion deutlich zu machen und um

sich schnell und flexibel orientieren zu können (Niebert und Gropengießer 2014). Somit stellt der Leitfaden keine „unumstößliche Vorschrift“ dar (Gläser und Laudel 2010), da er „lenken, aber nicht einschränken“ soll (Niebert und Gropengießer 2014). Es kann also durchaus aufgrund von Nachfragen oder anderen sich ergebenden Themen sein, dass vom Leitfaden abgewichen wird.

Die Anzahl der Fragen „richtet sich natürlich nach den zu erhebenden Inhalten“, wobei auch „die zur Verfügung stehende Zeit“ mit einkalkuliert werden muss (Gläser und Laudel 2010). Diesbezüglich ist es für die vorliegenden Interviews relativ komplex ein ertragreiches Verhältnis herzustellen, da vermutlich nicht viel Zeit für die einzelnen Schülerinterviews zur Verfügung stehen wird. Da es sich bei den Interviews aber nicht um den Hauptbestandteil der vorliegenden Studie handelt, wird im Zweifel dem Inhalt zuungunsten der Tiefe der Vorrang gegeben, damit ein Überblick über die thematisierten Aspekte gestaltet werden kann (siehe auch Kapitel 9.1.1).

Insgesamt besteht das Interview aus vier Frageblöcken (zzgl. Einleitung und Abschluss). Diese werden durch 1 – 2 Hauptfragen bestimmt. Für den Fall einer nicht erschöpfenden Antwort sind zu jeder Frage Nachfragen und Ergänzungen formuliert, die nach eigenem Ermessen eingesetzt werden können.

Die Inhalte des Fragebogens werden im folgenden Kapitel dargestellt.

Inhalte

Hinweis: Der vollständige Leitfaden befindet sich im Anhang in Anhang A.3 (Seite 288ff).

Wie beschrieben zeigen die erläuterten Interviews einen deutlichen inhaltlichen Schwerpunkt. Zum Teil soll dieser in den hier durchgeführten Befragungen ebenfalls aufgegriffen werden. Ebenso soll das Interesse an Inhalten oder das Selbstkonzept auch hier nicht im Vordergrund stehen bzw. nicht gezielt erfragt werden. Wie auch bei Tobias (2010) und R. Wodzinski (1996) wird das Verständnis des Geschwindigkeitsbegriffs erfragt. Im Unterschied zu ersterer Studie werden hier Schüler unabhängig vom unterrichteten Konzept befragt.

Der Leitfaden ist für Schüler, die nach dem 1D und 2D-Konzept unterrichtet wurden, und für Schüler mit Text- oder Animationsfeedback geeignet. An entsprechenden Stellen wird daher die Option geboten, eine passende Frageformulierung für den Probanden auszuwählen. Der Leitfaden gliedert sich in mehrere thematische Abschnitte. Der inhaltliche Aufbau ist dabei für alle Versuchsbedingungen gleich:

- Einstieg

- Lernumgebung
- Umgang mit Simulationen
- Fachinhalt
- Feedback
- Sonstiges

Der **Einstieg** in die Befragung erfolgt angelehnt an Tobias (2010). Die Interviewpartner/innen werden aufgefordert, die „Top 3“ der Begriffe zu nennen, über die sie etwas gelernt haben.

Während der Studie arbeiten die Schüler und Schülerinnen in erster Näherung in Einzelarbeit. Da es realistisch ist anzunehmen, dass diese Arbeitsform für eine Dauer von mehreren Doppelstunden in Kombination mit dem Medium *Tablet* bisher nicht eingesetzt wurde, soll hier vor allem erfragt werden, wie das Empfinden während der Arbeit an der **Lernumgebung** und *Tablet* war. Neben der schon erwähnten Einzelarbeit⁴ bezieht sich dieser Fragenblock auch auf das Fehlen von Realexperimenten und wie dieser Umstand von den Kindern – auch im Bezug auf das Lernen – wahrgenommen wird. Hinsichtlich der Frage, inwiefern Lernen mit Neuen Medien bzw. Simulation ernst genommen wird (siehe Kapitel 4.2.1), wird von den Befragten ebenfalls eine Stellungnahme dazu erbeten. Dann werden ihnen zwei gegensätzliche Statements von imaginären Schülern vorgelegt, in deren Spannungsfeld sie sich selbst einordnen sollen (vgl. sparsamer Einsatz von Provokationen, Gläser und Laudel (2010)).

Zum **Umgang mit Simulationen** wird den Befragten zur Stellungnahme die Aufzeichnung einer Simulation gezeigt, die von einer virtuellen Person bedient wurde. Die Simulation ist den Schülern dabei bekannt, da sie Bestandteil der selbst bearbeiteten Lernumgebung war und somit auch von ihnen bearbeitet wurde. Die dargestellten Szenen zeigen eine unzureichende Variablenkontrolle auf, die von den betrachtenden Schülern erkannt werden soll. So können Hinweise auf die Kontrollstrategien der Schülerinnen und Schüler gewonnen werden⁵.

Ob Fehler des/der Interviewten bei den nun folgenden fachlichen Fragen korrigiert werden (ebenso wie auch im Teil *Umgang mit Simulationen*), sollte ebenfalls den Transkripten entnehmbar sein und muss bei der Auswertung berücksichtigt werden. In jedem Fall besteht aber die Möglichkeit

⁴Man beachte auch hier, dass sich die Strenge der Auslegung der Einzelarbeit im Laufe der Studie durchaus variierte (siehe 9)

⁵Die vage Idee zu diesem Verfahren entstammt dem Stimulated Recall Verfahren (vgl. Messmer (2015)), natürlich mit dem großen Unterschied, dass dort *eigene* Vignetten demonstriert werden, hier aber *fremde*.

mit wohldosierten Hinweisen Hilfestellung zu geben – in letzter Instanz ist diese Entscheidung dem Fingerspitzengefühl des Interviewers überlassen.

Inhaltliche Aspekte, die im Interview behandelt werden, berühren den Geschwindigkeits- und Kraftbegriff, so wie er den jeweiligen Schülerinnen und Schülern in der Lernumgebung nahegebracht wurde.

Im Falle eines 2D-Schülers soll zusätzlich eine bis dahin unbekannte, kompakte Simulation bearbeitet werden. In dieser sollen die Schüler zwei Autoscootern unterschiedliche Kombinationen von Richtungen, Tempi und Geschwindigkeiten zuweisen. Inhaltlich handelt es sich also um eine bislang häufiger in Interviews eingesetzte Frage (R. Wodzinski 1996; Tobias 2010). Das Ergebnis ist den Transkripten zu entnehmen und wurde nicht am Tablet aufgezeichnet.

Das **Feedback** wurde mit einem weiteren Fragenblock untersucht, wobei nach Text- und Animationsfeedback unterschieden wird. Da die Lernumgebung so konzipiert bzw. programmiert ist, dass es möglich ist, das Feedback inhaltlich schlicht zu ignorieren. So ist das Ziel dieses Fragenblocks primär, herauszufinden, inwiefern das Feedback in Form von Hinweisen als Text oder Animationen durch die Schülerinnen und Schüler genutzt und gegebenenfalls umgesetzt wurde. Weiterhin wird um Einschätzung von solch automatischen Feedback gebeten – im Vergleich zu Rückmeldung im normalen Unterricht, in dem das Feedback i.d.R. individuell von der Lehrkraft oder auch Mitschülern gegeben wird.

Zum Abschluss des Interviews wurden **ergänzende** Punkte angesprochen. Dazu gehört auch die obligatorische Frage nach Anregungen, Kritik, Verbesserungsvorschlägen und sonstigen Bemerkungen seitens der Schülerinnen und Schüler. Ebenso sollen eventuelle inhaltliche Unklarheiten durch die Befragten geäußert werden

8.5. Fragebogen für Lehrkräfte

Die Fragebogenerhebung der Lehrkräfte ist ausdrücklich als Begleiterhebung zu verstehen, da die Perspektive der Lehrenden nicht im Fokus dieser Arbeit steht. Ein übergeordneter Zweck ist es, besondere Beobachtungen bezüglich der Lerngruppe zu erfassen. Das Verhalten einer Klasse zu beurteilen wäre zwar für die Durchführenden der Studie möglich – nicht möglich ist es hingegen, das Verhalten im Kontext des eigentlichen Physikunterrichts zu sehen. Wichtig ist die Einschätzung zur Lerngruppe, um in der späteren Auswertung eventuelle Klasseneffekte beurteilen zu können. Weiterhin werden überblicksartig Kriterien für den Einsatz von Simulationen im eigenen Unterricht erfragt, ebenso wie die Einstellung zum in der Klasse jeweils erprobten Mechanikkonzept. Die Items sind aufgrund ihrer „Erkundungsfunktion“ (vgl. Bortz und Döring 2006) durchgehend als offene Items angelegt. Tabelle 8.1 gibt einen Überblick über die einzelnen Fragen. Die Ergebnisse befinden sich in Kapitel 12.1.1.

Tabelle 8.1.: Übersicht der Fragen für die Lehrkräfte

Abkürzung	Fragentext
[Auffällig]	Beschreiben Sie Auffälligkeiten im Verhalten der Lerngruppe während der Arbeit mit dem Lernprogramm (bitte OHNE Nennung von Namen)
[Schwierig]	Beschreiben Sie besondere Schwierigkeiten, die sich den Schülerinnen und Schülern der Lerngruppe während der Arbeit mit dem Lernprogramm/den Simulationen ergaben (falls Sie entsprechende Beobachtungen gemacht haben).
[2D]	Geben Sie an, inwiefern Sie die frühe Einführung einer zweidimensionalen Dynamik in Jahrgang 7/8 eher als Chance oder als Schwierigkeit für die Lernenden sehen (gerne an konkreten Beispielen).
[1D]	Beschreiben Sie, inwiefern Sie die traditionelle Einführung des Geschwindigkeitsbegriffs als reine Betragsgröße als langfristig lernhinderlich oder lernförderlich einschätzen (gerne an konkreten Beispielen).
[KritSim]	Geben Sie an, welche Kriterien den erfolgreichen Einsatz von Simulationen im Physikunterricht Ihrer Erfahrung nach ausmachen (z. B. Eigenschaften der Simulation, der Lerngruppe...).
[Sonst]	Was ich sonst noch sagen möchte

9. Die Studien in der Unterrichtspraxis

In diesem Kapitel wird die tatsächliche Umsetzung der einzelnen Studien inkl. assoziierter Erhebungen dargestellt.

Zu den durchgeführten Studien gehören:

- Vergleichsgruppenstudie/Pilotierung der Fachtests
- Vorstudie
- Hauptstudie
- Cognitive-Load-Studie

In der Vergleichsgruppenstudie wurden in Klassen ohne Intervention große Itemumfänge für Fachtests erprobt, nachdem die Schüler Mechanikunterricht hatten. Diese Daten werden als Vergleich zur Hauptstudie herangezogen, gleichzeitig wurden sie aber auch zur Itemauswahl für die Hauptstudie herangezogen (daher Pilotierung).

In der Vorstudie wurde die Lernumgebung in zwei Versionen (1D und 2D inkl. zweier Feedbackarten) in zwei zehnten Klassen eingesetzt, um deren Praxistauglichkeit zu testen, wie auch die Fragebögen (nicht die Fachtests!) hinsichtlich ihrer Funktionalität zu testen.

Die Hauptstudie wurde mit N=275 Schülerinnen und Schülern siebter und achter Klassen durchgeführt. Dabei wurden beide Varianten der Lernumgebung jeweils mit beiden Feedbackarten untersucht.

Die Studie zum Cognitive Load wurde im Rahmen der Hauptstudie in 4 Klassen begleitend durchgeführt, um die kognitive Belastung beim Bearbeiten der Simulationen zu untersuchen. Eine größere separate Studie zum Cognitive Load wurde separat im Rahmen einer Masterarbeit (siehe Mitschker (2017)) durchgeführt.

Während der ersten Entwicklungsarbeiten der Lernumgebung wurde mit einer Schulklasse eine Exploration durchgeführt, deren Ergebnisse später knapp dargestellt werden und dort an passender Stelle auch sehr knapp charakterisiert wird.

9.1. Umsetzung der Studien

9.1.1. Hauptstudie

Die Zuordnung der Lerngruppen zum 1D- oder 2D-Konzept erfolgte generell aus rein praktischen Überlegungen. Dazu gehören eine ausgewogene Aufteilung der Gesamtstichprobe nach Jahrgangsstufe und Zugang wie auch ein Entgegenkommen gegenüber den Lehrkräften. So wurde darauf geachtet einer Lehrkraft mit mehreren Lerngruppen nach Möglichkeit nicht unterschiedliche Zugänge zuzuordnen, um ihr ein effizientes Weiterarbeiten nach Abschluss der Studie zu ermöglichen. Inclusive der Einführung und dem Ausfüllen der Tests und Fragebögen nahm die Durchführung in einer 2D-Klasse 3 – 4 Doppelstunden und in einer 1D-Klasse etwa 3 Doppelstunden in Anspruch. Eine Übersicht des generellen Ablaufs befindet sich in Bild 9.1.

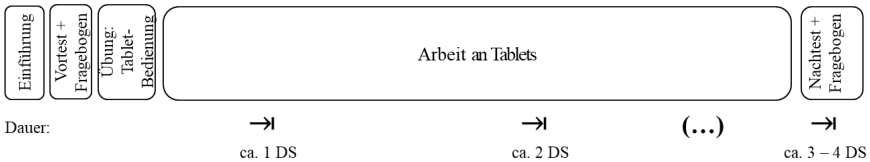


Abbildung 9.1.: Darstellung des zeitlichen Ablaufs der Studie (1 DS = 1 Doppelstunde). Der zeitliche Maßstab dient nur der Orientierung. In der Umsetzung schwankt die Dauer je nach Lerngruppe und Zugang stärker.

Um trotz des gewählten Feldstudiendesigns zu einer gewissen Standardisierung des Ablaufs beizutragen, war die Durchführung in allen Klassen ähnlich. Zunächst wurde eine Einführung per Präsentation vorgenommen. Diese beinhaltete einen strukturellen Überblick über die Lernumgebung sowie Hinweise auf zentrale Regeln. Dazu gehört auch der Hinweis darauf, dass es sich bei dem Stoff des Lernprogramms und der Simulationen um reguläre Unterrichtsinhalte handelt, an die im darauffolgenden Unterricht angeknüpft wird. Weiterhin wurde betont, dass die Unterrichtseinheit insofern bewertungsfrei ist, als dass die Vor- und Nachtests wie Fragebögen keinesfalls zur schulischen Leistungsfeststellung herangezogen und nur für Zwecke der Forschung verwendet werden. An die Präsentation schloss sich das Ausfüllen der Vortests bzw. Fragebögen an.

Für das Ausfüllen der Tests sowie der Fragebögen (auch Post) war für die Hilfs- und Lehrkräfte die Maßgabe, bei Fragen stets auf den Aufgabentext zu verweisen und keine inhaltlichen Hilfen zu geben. Deshalb waren die Aufgaben von den Schülern nach bestem Wissen zu beantworten, konnten aber auch übergangen werden. Durch dieses Auslassen von Aufgaben

soll das Erraten einer richtigen Antwort zumindest verringert werden. Dies ist unproblematisch möglich, da die Aufgaben nicht zur schulischen Leistungsfeststellung, sondern nur zu Forschungszwecken genutzt werden. Für das Ausfüllen der Tests und Fragen wurde keine Zeitvorgabe gestellt, d. h. jeder Schüler bearbeitete die Aufgaben im eigenen Tempo. I. d. R. nahm dies 15 – 25 Minuten in Anspruch. Nur in absoluten Ausnahmefällen konnten (Nach)tests aus Zeitgründen nicht fertig gestellt werden.

Im weiteren Verlauf wurde mit den Schülern gemeinsam die Bedienung des Tablets und der Software geübt. Jedes Tablet beinhaltet dabei ein Lernprogramm mit einer der beiden Feedbackvarianten. Die Zuordnung der Tablets zu den Schülern erfolgte in der Einführungsstunde. I. d. R. wurden dazu die Tablets mit abwechselnden Feedbackvarianten an nebeneinander sitzende Schüler ausgegeben. Eine systematische Zuordnung der Feedbackvariante zu einzelnen Schülern (z. B. je nach Leistungsstand) erfolgte somit nicht. Die Zuordnung von Tablet und Schüler blieb für die Dauer der Studie bestehen, auch wenn im Falle von Defekten das Gerät getauscht werden konnte. Die Schüler wurden in der Einführung dazu angehalten prinzipiell in Einzelarbeit an ihren Tabletcomputern zu arbeiten. Vor dem Hintergrund der Dauer der Unterrichtseinheit (ohne wissenschaftliche Begleitung) von etwa 2 – 3 Doppelstunden ist es erwartbar, dass die Schüler nicht durchgehend still an ihren Tablets arbeiten, sondern untereinander kommunizieren. Auch inhaltliche Kommunikation findet dabei sicherlich statt, zu starke Kooperation wurde bei Beobachtung auch durch die Lehr- oder Hilfskräfte eingeschränkt.

Während der selbstständigen Arbeit mit den Tablets in den regulären Stunden sollte bei inhaltlichen Fragen zunächst auf die Texte der Lernumgebung oder auf die Simulationen verwiesen werden. Aus der Verantwortung der Durchführenden den Lernenden, Lehrkräften und Schulen gegenüber, wurde mitunter bei größeren inhaltlichen Problemen jedoch mit Hinweisen angemessene Unterstützung geboten. Erfahrungsgemäß traten aber in allen Klassen kaum inhaltliche Fragen auf. Das Mitschreiben seitens der Schüler während der Arbeit am Tablet wurde ausdrücklich eingefordert, auch um die Nachhaltigkeit des Lehrgangs im weiteren Verlauf des Unterrichts zu sichern. An diversen Stellen der Lernumgebung wurden die Schüler zum Mitschreiben aufgefordert (siehe Vorlage dazu auf S. 302), aber auch darüber hinaus sollte mitgeschrieben werden, was den Schülern relevant erschien. Je nach Lerngruppe wurde der Hinweis ausgeschärft, indem konkret auf die Unterschiede zwischen relevanten (“Die Länge des Geschwindigkeitspfeils gibt die Schnelligkeit an”) und nicht-relevanten Inhalten (“Das Boot lässt sich mit den Pfeilen steuern”) beispielhaft verdeutlicht wurde. Das Mitschreiben wurde auch von fast allen Lehrkräften forciert. In betreffenden Fällen wurden die Lehrkräfte

jedoch darum gebeten eine Korrektur erst nach Abschluss der Tabletteinheit vorzunehmen. So sollte verhindert werden, dass vor Ausfüllen der Post-Tests weiterer schwer kontrollierbarer Einfluss auf den Lernerfolg genommen wird.

Die Bearbeitung der Nachtests sollte nicht zu knapp vor Stundenende angefangen werden, um Unterbrechungen bis zur Folgestunde zu vermeiden. Dies konnte aufgrund des Arbeitstempos in seltenen Fällen nicht gewährleistet werden. Generell wurde Schülern, die den Nachtest zu Beginn der Folgestunde begannen, angeboten, diesen erst nach einer kurzen Einarbeitungszeit anhand der eigenen Unterlagen und Simulationen zu beginnen, um einen gedanklichen Einstieg in das Thema zu ermöglichen. Das Annehmen dieses Angebots oblag der Entscheidung der einzelnen Schüler.

Abweichungen An dieser Stelle soll auf bemerkenswerte Abweichungen eingegangen werden, auch wenn – oder gerade weil – aus dem Alltag als Lehrkraft bekannt ist, dass jede Lerngruppe, jeder Raum und jede Stundenlage ihre Eigenheiten hat.

Zum Teil fand die Stunde in Schulen mit zwei Einzel- statt einer Doppelstunde pro Woche statt. Dies bot zwar insofern den Vorteil, dass das jeweils letztmalige Arbeiten mit den Inhalten der Lernumgebung nicht lange zurücklag. Jedoch fiel viel pauschale Zeit für Auf- und Abbau, das Hoch- und Herunterfahren und Anmelden im Lernprogramm an.

Größere Unterschiede waren auch in den räumlichen Gegebenheiten zu beobachten. Während einige Klassenräume eher eng waren, standen an anderen Schulen mehrere Räume zur Verfügung. Relevant werden diese Unterschiede vor allem beim Ausfüllen der Tests, so dass es nicht immer möglich war für entsprechende Schüler einen ruhigen Platz zum Bearbeiten zu finden.

In bestimmten Lerngruppen trat das Phänomen auf, dass Schüler eine geringe Verarbeitungstiefe zeigten. Denn obwohl unterschiedliche Strategien beim Durcharbeiten des Lernprogramms beobachtet werden konnten, ist ein abgeschlossenes Bearbeiten der Lernumgebung nach 45 Minuten selbst bei mäßiger Arbeitsweise unmöglich, da alleine das bloße „Durchklicken“ des Lernprogramms inkl. Öffnen der Simulationen und „Überfliegen“ der Texte über 30 Minuten dauert. In solchen Fällen folgte mit Hinweisen auf die Verarbeitungstiefe die Aufforderung von vorne zu beginnen oder Nennung einzelner Stichpunkte, zu denen die Schüler im Lernprogramm etwas Lernen können. Im weiteren Verlauf der Studie wurde dazu übergegangen, diese Stichpunkte und deren Zweck für alle sichtbar an die Tafel zu notieren und bei jedem Lernenden, der den Nachtest beginnen wollte, nochmals auf diese verwiesen. Weitere Verzögerungen

kamen durch Ausflüge oder Freizeiten von Teilen der Klassen zu Stande, so dass im Laufe der Durchführung größere Unterschiede im Bearbeitungsfortschritt der Schülerinnen und Schüler zu bemerken waren. Den jeweiligen Nachzögler*innen wurde separat eine kompakte Einführung in das Lernprogramm gegeben.

Umsetzung der Befragung der Lehrkräfte zur Hauptstudie

Der Lehrer-Fragebogen wurde den Lehrkräften gegen Ende oder nach der Unterrichtseinheit ausgeteilt. Eine weitere Option war das Verschicken des Fragebogens per E-Mail. Somit wurde der Fragebogen in der Regel direkt im Unterricht oder separat ausgefüllt und jeweils auf Papier oder elektronisch zurückgegeben.

Umsetzung der Interviews zur Hauptstudie

Mit einzelnen Schülern wurden nach Zustimmung aller erforderlichen Personen Interviews durchgeführt. In Absprache mit diesen und evtl. den Lehrkräften wurden Termine vereinbart. Die Interviews fanden in der Regel im Sammlungsraum oder nicht genutzten Klassenräumen statt. Schon während der Akquise der Interviewpartner*innen erfolgte die Aufklärung über den Zweck des Interviews und unmittelbar vor Beginn wurde u. A. der Datenschutz und die Möglichkeit Fragen nicht zu beantworten erläutert. Besonderheiten zum geführten Interview sowie die Zustimmung der Schüler*innen und deren Eltern wurden bei entsprechender Gelegenheit auf dem Interviewbogen protokolliert. Die Interviews wurden zur weiteren Transkription per Diktiergerät aufgezeichnet. Während des Interviews wurde auch über die Simulationen gesprochen und deren Bedienung ggf. aufgezeichnet. Entsprechend eines angemessenen Verhaltens des Interviewers wurde darauf geachtet die Atmosphäre nicht zu locker aber ebenso nicht „einschüchternd“ zu gestalten, indem man bspw. seine Forscher-Position überbetont! Es wurde daher versucht eine „natürliche Atmosphäre“ zu gestalten (Gläser und Laudel 2010). Dazu gehört nach Ansicht des Autors auch den Schülern mit Respekt zu begegnen und deren Einsatz wertzuschätzen. Dies gilt umso mehr, als dass auch über physikalische Sachverhalte gesprochen wird, die mitunter auch durch den Interviewer gewertet werden müssen¹. Bei Fragen nach Einstellungen und Bewertungen der Schüler wurde jedoch stets darauf geachtet, keine Wertungen abzugeben – im Zweifel wurden die Schüler sogar ermutigt,

¹Gläser und Laudel (2010) sehen zwar vor, dass in Interviews keinesfalls eine Bewertung vorgenommen werden sollte – im Rahmen dieses Interview, das nun auch fachliche Aspekte enthält, ist dies nicht immer vermeidlich.

ihre "wahre" Meinung zu äußern. Ein gewisses Maß an Feedback (z. B. Zeigen des aktiven Zuhörens/Verständnisses, (vgl. Gläser und Laudel 2010)) wurde aber gegeben – auch um den jungen Interviewten eine gewisse Sicherheit im Interviewprozess zu geben.

Der Leitfaden stellte sich als dominanteres Gerüst dar, als er im klassischen Falle hätte sein sollen. Dies ist letztlich vermutlich auf die knappe Interviewzeit und eine gewünschte hohe Vergleichbarkeit der Interviews aufgrund der relativ geringen Fallzahl zurückzuführen.

9.1.2. Umsetzung der Vor- und Vergleichsgruppenstudie

Die Fragebögen für die Pilotierungs- bzw. Vergleichsgruppen der Fragebögen wurden per Post versandt und ohne persönliche Anwesenheit von Hilfskräften oder des Studienleiters ausgefüllt. Zu diesem Zwecke wurde eine Anleitung erstellt, die die Schritte zur Durchführung und wichtige Hinweise enthält. Für das komplette Bearbeiten des Aufgabensets wurden maximal etwa 45 Minuten veranschlagt. Die ausgefüllten Fragebögen wurden anschließend i. d. R. persönlich übergeben oder postalisch zurückgesandt.

Die Umsetzung der Vorstudie erfolgte analog zur Umsetzung der Hauptstudie. Insgesamt wurden zwei Klassen des 10. Jahrgangs besucht (weshalb in den Lernumgebungen vereinzelt (!) Teile ausgespart wurden). Das höhere Arbeitstempo, das bereits vorhandene Vorwissen sowie die Lernumgebung sorgten dabei für eine schnellere Durchführung von etwa 2 Doppelstunden.

9.1.3. Umsetzung der Cognitive-Load-Studie

Die Cognitive Load Erhebung wurde im Rahmen der Hauptstudie durchgeführt, indem entsprechende Fragebögen in vier Klassen zusätzlich und begleitend zur Arbeit mit den Lernumgebungen ausgeteilt wurden. Hier soll jedoch noch der vorangegangene Teil der Cognitive-Load-Studie beschrieben werden, der als Block in einer Doppelstunde vor Beginn der Hauptstudie stattfand und im Rahmen einer Masterarbeit durchgeführt wurde (vgl. Mitschker 2017). Insgesamt wurden dabei 8 Klassen besucht, davon 4 des 7./8. Jahrgangs und 4 des 10. Jahrgangs. In den meisten Klassen wurde das Thema bereits unterrichtet, so dass mit schulischem Vorwissen gerechnet werden konnte (ebd.). Ähnlich zur Hauptstudie wurde zu Beginn – mit Tafelunterstützung – ein Überblick über den Zweck der Untersuchung wie über die Bedienung der Lernumgebung und wichtige Hinweise gegeben (wie etwa richtiges Ankreuzen der Fragebögen, genaues Ausprobieren und Beobachten, Tipps und Hinweise beachten). An die

Einweisung schloss sich eine Übungsphase zur Bedienung der Tablets an (ebd.).

Alle Schüler bearbeiteten 1D- und 2D-Simulationen in einer kurzen Version der Lernumgebung. Je die Hälfte der Tablets startete mit 1D- bzw. 2D-Simulationen. Durch diese Vertauschung sollte sichergestellt werden, dass alle Simulationen regelmäßig und keine deutlich seltener bearbeitet werden (z. B. wenn diese am Ende der kurzen Lernumgebung liegt) und Reihenfolgeeffekte vermieden werden. Die Fragebögen wurden parallel zu den Lernumgebungen gestaltet und während bzw. kurz nach der jeweiligen Bearbeitung der Simulation ausgefüllt.

9.2. Beschreibung der Stichproben

9.2.1. Teilnehmer/innen aus Vor- und Vergleichsgruppe

Vorstudien

Insgesamt nahmen zwei zehnte Klassen an der Vorstudie mit $N=45$ Schülerinnen und Schülern teil. Entsprechend verfügten die Schüler über entsprechendes Vorwissen und hatten auch in diesem Jahr bereits Mechanikunterricht. Insgesamt überwiegen sowohl die Mädchen als auch die Schüler die Schüler mit Textfeedback die jeweils andere Gruppe im Verhältnis 1 zu 2 (siehe Tabelle 9.1).

		Feedback		Geschlecht		Summe
		Text	Ani.	m	w	
Zugang	1D	15	7	8	14	
	2D	13	10	10	13	
Summe		28	17	18	27	45

Tabelle 9.1.: Versuchspersonen der Vorstudie

Hinsichtlich der Noten unterscheiden sich Mädchen und Jungen in Mathematik und Physik um etwa 0,5 Notenschritte, aber nur in Mathematik signifikant ($z(45) = -2,046; p = .041$). Im Fach Deutsch sind die Noten von Jungen und Mädchen ähnlich (siehe Tabelle 9.2).

Vergleichsgruppe

Insgesamt wurden fünf achte Klassen mit $N=117$ als Kontrollgruppe herangezogen. Diese Lerngruppen hatten zum Zeitpunkt des Tests den

Tabelle 9.2.: Noten der Schüler der Vorstudie

		Anzahl	Note PH	Note M	Note D
Geschlecht	m	18	2,22	2,11	2,61
	w	27	2,73	2,63	2,59
	m+w	45	2,52	2,42	2,60

anfänglichen, traditionellen Mechanikunterricht in Klasse 7/8 bereits – mal mehr mal weniger zeitnah² – abgeschlossen. Die Schüler bearbeiteten keine Fragebögen, sondern einmal einen Fachtest. Dabei waren auch alle Aufgaben enthalten, die bereits in der Pilotierung enthalten waren, wie somit auch alle Aufgaben des Nachttests der Hauptstudie. Eine Übersicht der Versuchspersonen der Kontrollgruppe liefert Tabelle 9.3.

Tabelle 9.3.: Übersicht der Schüler der Vergleichsgruppe

		Anzahl	Note PH	Note M	Note D
Geschlecht	m	52	3,13	3,31	3,38
	w	58	2,48	2,66	2,60
	N/A	7	2,43	2,57	3,00
	alle	117	2,77	2,94	2,97

Besonders auffällig zeigen sich hier die Geschlechterunterschiede in den Noten aller drei Fächer. Die Noten der Mädchen liegen im Mittel 0,6 bis 0,8 Noten besser als die der Jungen. Über alle Fächer zeigen sich daher mittels t-Test (siehe Kapitel 10) signifikante Unterschiede (PH: $(t(108) = -3,538; p = .001)$; M: $(t(108) = -3,414; p = .001)$; D: $(t(108) = -4,597; p < .001)$). Dabei treten kleine bzw. gerade noch kleine Effekte auf (PH: $d = .34$; M: $d = .33$; D: $d = .44$).

9.2.2. Hauptstudie

An der Hauptstudie nahmen insgesamt 275 Schülerinnen und Schüler teil. Eine genaue Aufschlüsselung nach Geschlecht, Jahrgang sowie Versuchsbedingung kann Tabelle 9.4 entnommen werden.

²sowohl direkt vorm Bearbeiten des Tests als auch wenige Wochen zuvor

		Feedback		Jahrgang		Geschlecht		Summe
		Text	Ani.	7	8	m	w	
Zugang	1D	76	66	45	97	78	60	
	2D	72	60	50	83	58	73	
Summe		148	126	95	180	133	136	275

Tabelle 9.4.: Übersicht der Versuchspersonen der Hauptstudie

Die einzige augenscheinliche Auffälligkeit ist die leichte Unterpräsenz des 7. Jahrgangs, in welchem je Zugang etwa 30 – 50 Personen weniger zu verzeichnen sind. Weiterhin blieb von 6 Personen das Geschlecht unbekannt.

Interviewte Schülerinnen und Schüler

Das Leistungsspektrum der interviewten Schüler (nicht in Tabelle 9.5 aufgeführt) ist etwas besser, als in der Gesamtstichprobe. So beträgt die Durchschnittsnote in Physik 2,1 (2,4 in der Gesamtstichprobe), der Median ist mit der Note 2 aber identisch. Insgesamt wären noch Schüler/innen im befriedigenden Notenspektrum wünschenswert gewesen. Die Auswahl erscheint aber dennoch genügend repräsentativ, auch wenn ein ausgeglicheneres Verhältnis in der Feedback-Variante wünschenswert gewesen wäre.

Tabelle 9.5.: Übersicht über Interviewpartner/innen

		Feedback		Geschlecht		Jahrgang	
		Text	Animation	m	w	7	8
Zugang	1D	3	2	4	1	2	3
	2D	4	1	2	3	3	2

Lehrkräfte der Hauptstudienklassen

Auch wenn die Lehrervariable im Rahmen dieser Studie keinen großen Einfluss hat werden die Daten dargestellt, da sie eventuell im Rahmen der Lehrerfragebögen zur Ergänzung herangezogen werden können.

Tabelle 9.6 zeigt, dass keine Lehrkraft über mehr als 11 Jahre Lehrerfahrung verfügt und sich eine Lehrkraft noch im Vorbereitungsdienst befindet. Die dazwischenliegende Unterrichtserfahrung deckt relativ gleichmäßig

Tabelle 9.6.: Übersicht über Physiklehrkräfte der Klassen

Lehrkraft	Geschlecht	Lehrerfahrung (in Jahren)
LK1	w	3
LK2	m	4,5
LK3	w	9
LK4	m	N/A
LK5	w	11
LK6	m	1

das gesamte Spektrum ab. Das Verhältnis von männlichen zu weiblichen Lehrkräften ist ausgewogen. Bei den weiteren Unterrichtsfächern dominiert das Fach Mathematik.

10. Aufbereitung der Daten

Im Laufe der Studie wurden in unterschiedlichen Settings qualitative und quantitative Daten erhoben. In diesem Kapitel wird beschrieben, wie diese Rohdaten für eine weitere Auswertung nutzbar gemacht wurden. Näheres zu den Erhebungsinstrumenten siehe Kapitel 8, die Fragen befinden sich im Anhang A.

Vorliegende quantitative Daten der Hauptstudie:

- Pre- und Posttests
- Pre- und Postfragebögen
- Fragebögen zum Cognitive Load

Vorliegende qualitative Daten der Hauptstudie:

- Schülerinterviews (Tonaufnahme / Transkript)
- Lehrerfragebögen
- Schüleräußerungen auf Post-Fragebogen

10.1. Quantitative Daten

10.1.1. Tests und Fragebögen

FACHTESTS

Die Daten aus den Vor- und Nachtests wurden zunächst in Tabellenkalkulationsprogramme und später in die Software SPSS übertragen. Dabei wurde die gewählte Antwortoption der Multiple-Choice-Items (a,b,c. . .) direkt übertragen, das „1-0-Coding“ (gelöst/nicht-gelöst) wurde später mithilfe von SPSS automatisch berechnet. Bei nicht-MC-Aufgaben wurden entsprechende Musterlösungen an die Codierer ausgegeben, anhand derer die Schülerantwort abgeglichen und entsprechend als richtig oder falsch vermerkt wurde. Bei MC-Fragen mit mehreren Antwortoptionen wurde nur die korrekte Kombination als gelöst gewertet.

Fehlende Angaben wurden direkt bei der Eingabe mit 99 codiert. Die Information, ob das Item aufgrund des Schwierigkeitsgrades, wegen Zeitproblemen nicht bearbeitet wurde oder ob es schlicht übersehen wurde

ist somit in der Codierung nicht enthalten, lag aber auch eindeutig nie vor. Fehlende Angaben bei einzelnen, komplett fehlenden Fragebögen oder Tests wurden mit 88 kodiert. Hier wird unterschieden, da unterstellt wird, dass die Schüler diese Bögen z. B. aus zeitlichen Gründen nicht mehr ausfüllen konnten (z. B. weil sie erst nach über einer Doppelstunde in die Studie eingestiegen sind und daher keinen Vortest mehr bearbeitet haben oder aber die Lernumgebung nicht abgeschlossen haben). So soll verhindert werden, dass diese "fehlenden" Werte als falsch oder fehlend gewertet werden, da die Schüler nicht einmal eine Chance hatten, die Fragen zu beantworten. Angaben zu fehlenden Werten finden sich an der jeweiligen Stelle der Auswertung. Ungültige Angaben (i. d. R. mehrere gegebene Antworten bei Fragen mit nur einer korrekten Option) wurden bei der Eingabe mit 999 kodiert.

Im weiteren Verlauf wurden die fehlenden und ungültigen Angaben der Fachtests als *nicht gelöst* und somit als 0 codiert. Besonders bei Items, die ungültig aufgrund von verbotenen Mehrfachkreuzen geratet wurden, kann dies durch widersprüchliche Antworten begründet werden.

FRAGEBÖGEN

Die Daten der Fragebögen wurden direkt als numerische Werte in Excel-Tabellen eingegeben. Dabei wurde der niedrigsten Zustimmung/dem seltensten Auftreten der Wert 0 zugewiesen und dem häufigsten Auftreten/der größten Zustimmung der Wert 4 (mit ganzzahligen Zwischenabstufungen). Die offene Antwort wurde wörtlich übernommen.

Aus unterschiedlichen Gründen fehlende Angaben (s. o.) wurden auch hier mit 99 bzw. 88 codiert. Als ungültige Werte mit 999er-Code galten doppelte Kreuze, sowie Kreuze, die sich nicht eindeutig einer der Stufen von 0 bis 4 zuordnen ließen. Ungültige Werte wurden in der Auswertung wie fehlende Werte behandelt.

10.2. Qualitative Daten

10.2.1. Schülerinterviews

Transkription Mit einzelnen Schülern wurden Interviews über die Lernumgebung, die Simulationen und die Inhalte geführt. Zur weiteren Verwendung wurden diese transkribiert. Dazu bemerken Dresing und Pehl (2010):

„Eine Transkription liefert kein vollständiges Abbild der aufgezeichneten Situation, sondern bedeutet immer eine Reduktion.

Den Grad der Reduktion gilt es abzuwägen und geeignet zu bestimmen. Die Reduktion betrifft dabei mehrere Bereiche.“

(ebd.)

Diese erwähnten Bereiche betreffen im methodischen Sinne die Entscheidung zwischen Voll- und Teiltranskript und den Detailgrad der Transkription bezüglich „Äußerungsmerkmalen“ (ebd.). Im vorliegenden Fall wurden die Interviews vollständig transkribiert. Dies beruht auf der inhaltlichen Dichte der Interviews, in denen es kaum Stellen mit nicht-relevantem Material zu geben schien. Der Detailgrad des Transkripts hängt der transkribierten (Laut)Äußerungen, der Pausen oder sprachlichen Glättung vom Untersuchungszweck ab (vgl. Gläser und Laudel 2010). Aus diesem Grund wird Lachen, Stöhnen oder Seufzen nur transkribiert, wenn es inhaltlich angebunden ist und beispielsweise ein ausgedehntes „Hmmm“ Hinweise auf besonders angestregtes Nachdenken gibt. Gleiches gilt auch für ausgeprägte Betonungen (Krüger, Parchmann und Schecker 2014). Die Interviews wurden von unterwiesenen Hilfskräften transkribiert. Da es für die Transkription keine verbindlichen Regeln gibt, mussten diese vorab individuell vereinbart werden (Gläser und Laudel 2010). Die so festgelegten Regeln sind in Tabelle 10.1 aufgeführt.

Tabelle 10.1: Übersicht der verwendeten Transkriptionsregeln

Beispiel	Erläuterung
IN	Interviewer
S1	Schüler/in (fortlaufend nummerieren)
aaaaaaalso	Dehnung
*	kurze Pause
3	längere Pause mit Angabe in s
hammwa=haben wir, kannse=kann sie	sprachliche Glättung
AchterBAHN	deutliche (!) Betonung
och, hm, naja (seufzt)(lacht)	Füllwörter übernehmen (inhaltlich angebunden) nicht sprachliche Lautäußerung, Übernehmen wenn inhaltlich angebunden (also bspw. i. d. R. kein Räuspern oder Husten)
Namen, auch von Schulen etc.	unkenntlich machen!!
[Anm.: XY]	Anmerkungen
01:00	Zeitmarke, in etwa jede Minute vor IN oder S1

Auswertung Die Auswertung der Interviews erfolgte angelehnt an die qualitative Inhaltsanalyse nach Mayring (2015). Die einzelnen Analyseblöcke (z. B. *Realexperimente* oder *Nutzung des Feedbacks*) sind bereits durch den Leitfaden vorgegeben. Die Auswertung orientiert sich daran und wird jeweils für diese Einheiten vorgenommen. Die Kodiereinheiten sind dabei einzelne Äußerungen der Schüler. Wenn es dem Verständnis des Zusammenhangs zuträglich ist, werden auch Äußerungen des Interviewers bzw. mehrere Äußerungen im Wechsel als eine Kodiereinheit betrachtet. Die so erhaltenen Textstellen zu einem Analyseblock wurden daraufhin nach folgenden Regeln paraphrasiert, generalisiert und reduziert.

Paraphrasierung

- Streichung von Stellen, die wenig oder nicht inhaltstragend sind
- Vereinheitlichung der Sprachebene (beide nach Mayring (ebd.))
- Inhaltliche Ergänzung der Schüleraussagen mit Aussagen des Interviewers, sofern für das Verständnis nötig oder der Ökonomie zuträglich. ^a

^aBeispiel: „I: Also fandest du die Simulationen nicht spannend, hast aber trotzdem etwas gelernt? S1: Ja, genau!“ würde zu „Ich fand die Simulationen nicht spannend, habe aber trotzdem etwas gelernt.“

Generalisierung

- Generalisierung auf höhere Abstraktionsebene, allgemeinere Formulierungen aber belassen
- Generalisierungen auf die gleichen Weise durchführen (beide nach Mayring (ebd.))

Reduktion

- Zusammenfassung bedeutungsgleicher Generalisierungen
- Nicht mehr bedeutungsvolle Paraphrasen werden weggelassen (beide nach Mayring (ebd.))

Ein weiterer Durchgang, in dem die Kategorien aus der letzten Bündelung nochmals auf höhere Abstraktionsniveaus gehoben werden, sei nach Mayring (ebd.) oft nötig, z. B. bis das gewünschte Abstraktionsniveau erreicht ist. In dieser Arbeit wird davon ausgegangen, dass das Niveau aus dem ersten Materialdurchgang bereits ausreichend abstrahiert ist. Weitere Verallgemeinerungen über die Ebene der Lernumgebung hinaus (z. B. auf elektronische Lernumgebungen als solche) scheinen hier nicht

sinnvoll. Weiterhin sind die Äußerungen mitunter nicht ergiebig genug und eine weitere Verallgemeinerung wäre nicht sinnvoll.

Auswertung der fachlichen Fragen Die Codierung der fachlichen Interviewinhalte erfolgt zum einen der Übersicht halber mittels Zuordnung zu einem Kategoriensystem. Zum anderen resultiert daraus eine einfachere und übersichtlichere Gegenüberstellung von Schülern des 1D und 2D Konzepts. Das Kategoriensystem wurde zunächst induktiv nach einem ersten Überblick am Material entwickelt und teils deduktiv aus vorliegenden Theoriebeständen (vgl. hierzu Kapitel 3.3). Eine Übersicht der endgültigen Kategorien findet sich in Kapitel D.1. Dort finden sich zu jedem Code auch hypothetische Äußerungen der Interviewten (vgl. Ankerbeispiele Bortz und Döring 2006). Pro Codiereinheit (s. o.) soll jeder Code nur einmal vergeben werden. Potentielle Hilfestellungen seitens des Interviewers sollen ebenfalls durch die Kennung *H* erfasst werden. Die Zuordnung der ca. 60 ausgewählten Codiereinheiten – ohne vorherige Paraphrase o. ä. – erfolgt dabei vom Autor und einem eingewiesenen Rater. Um dabei eine hohe Übereinstimmung der Rater (Bestimmung über Cohen's Kappa; Maß für die zufallskorrigierte Übereinstimmung von hier zwei Ratern.) zu erreichen, sind womöglich mehrere Codierungs-Durchgänge oder Anpassungen des Kategoriensystems notwendig.

10.2.2. Fragebögen Schüler: offene Frage

Die Daten der offenen Frage wurden in eine Excel-Tabelle eingetragen. Angelehnt an die qualitative Inhaltsanalyse nach Mayring (Mayring 2015) wurden die gegebenen Antworten in einem zweistufigen Verfahren zusammengefasst. Zunächst wurden die jeweilige Schüleraussagen auf ein allgemeineres Sprachniveau gehoben, um die Vergleichbarkeit unter den einzelnen Kommentaren zu gewährleisten. Im zweiten und letzten Schritt wurden diese Kategorien zugeordnet. Das Verfahren ist somit ähnlich zur Interviewauswertung. Die Kategorien wurden jedoch im Gegensatz dazu hier induktiv gebildet.

10.2.3. Fragebögen Lehrkräfte

Die Daten der Fragebögen der Lehrkräfte wurden in eine Excel-Tabelle zusammengetragen. Aufgrund der geringen Fallzahl und der sehr individuellen Antworten schien eine weitere Reduktion und Kategoriebildung nicht sinnvoll.

10.3. Verwendete statistische Größen und Verfahren

In diesem Abschnitt werden knapp die *zentralen* verwendeten Größen und Verfahren beschrieben. Dazu gehören die Hypothesentests, die Größen für Effektstärke und die Reliabilität wie die Rasch-Analyse.

Hypothesentests Hypothesentests überprüfen „formal zwei einander ausschließende statistische Hypothesen“ (Bortz und Döring 2006). Diese werden i. d. R. als Nullhypothese H_0 und Alternativhypothese H_1 bezeichnet. In dieser Arbeit werden diese als *Unterschiedshypothesen* oder *Veränderungshypothesen* ausgeführt sein. Entsprechend ist den Nullhypothesen inhaltlich die Bedeutung „Kein Unterschied zwischen den Gruppen“ bzw. „Keine Veränderung zwischen den Messungen“ zuzuordnen, den Alternativhypothesen jeweils das Gegenteil (vgl. ebd.).

Das Ergebnis eines Hypothesentests liefert die Wahrscheinlichkeit dafür, dass das gefundene Ergebnis auftritt, auch wenn „die Populationsverhältnisse der Nullhypothese entsprechen“ (ebd.). Im Prinzip wird also die Irrtumswahrscheinlichkeit dafür angegeben, fälschlicherweise der Alternativhypothese zuzustimmen. Dieser Wert wird üblicherweise als p -Wert bezeichnet. In der didaktischen und psychologischen Forschung ist für das Annehmen der Alternativhypothese im Allgemeinen ein Signifikanzniveau von $\alpha = 5\%$ akzeptiert. Ergebnisse mit p -Werten unter 5% werden daher als signifikant bezeichnet, unter 1% als hoch signifikant und unter 0,1% als höchst signifikant¹. Zu beachten ist aber dringend, dass ein (hoch/höchst) signifikantes Ergebnis in der didaktischen Forschung *nicht* automatisch ein praktisch bedeutendes Ergebnis darstellt.

Die in dieser Arbeit verwendeten Hypothesen-Tests werden im Folgenden kurz dargestellt

t-Test Der t-Test gehört zu den parametrischen Tests, d. h. im Test betrachteten Daten sollen der Normalverteilung unterliegen. Zusätzlich sollen sie intervallskaliert sein. Erstere Voraussetzung kann über Normalverteilungstests (Kolmogorov-Smirnow bzw. Shapiro-Wilk) überprüft werden. Im Allgemeinen gilt der t-Test jedoch speziell bei Stichproben von mehr als 30 Personen als robust gegenüber der Verletzung dieser Annahme (Janczyk und Pfister 2013).

Der Forderung nach intervallskalierten Daten kann in dieser Arbeit nicht immer nachgekommen werden, da beispielsweise die

¹auch mit ein, zwei oder drei *-Symbolen gekennzeichnet

Ergebnisse der Likert-Skalen einer Ordinalskala unterliegen und somit keine wohldefinierten Abstände vorliegen. Jedoch ist es gängige Praxis auch Daten solcher Skalenniveaus per t-Test zu untersuchen (vgl. Rode 2016).

Für diese Arbeit sind der t-Test für verbundene und unverbundene Stichproben relevant (vgl. Bühner und Ziegler 2009). Ersterer wird z. B. für Pre-Post-Betrachtungen innerhalb einer Gruppe herangezogen, letzterer für den Vergleich zweier unabhängiger Gruppen zu einem Zeitpunkt.

Angabe der Ergebnisse: ($t(54) = 5.34; p = 0.03$) (Angabe der Freiheitsgrade, des t-Werts und des p-Werts)

Wilcoxon-Test² Der Wilcoxon-Test ist ein Hypothesentest für nicht normalverteilte Daten (ebd.). Anstelle eines Vergleichs von Mittelwerten wie beim t-Test werden hier Median- bzw. Rangsummen-Vergleiche vorgenommen, um Unterschiede in den Gruppen zu untersuchen. Der Wilcoxon-Test ist daher das nicht-parametrische Äquivalent zum t-Test für verbundene Stichproben, also z. B. dem Pre-Post-Vergleich einer Gruppe (ebd.).

Mann-Whitney-U-Test³ Der Mann-Whitney-U-Test ist ebenso wie der Wilcoxon-Test eine nicht-parametrischer Test, d. h. die genutzten Daten müssen nicht der Normalverteilung unterliegen. Er ist das Äquivalent zum t-Test für unabhängige Stichproben, also z. B. dem Vergleich zweier Gruppen zu einem Messzeitpunkt (ebd.).

Hinweis: Aufgrund der Ausführungen wird dieser bei einer Anzahl von mehr als 50 Fällen der t-Test gerechnet. Bei Zweifeln an der Normalverteilung wird zusätzlich ein nicht-parametrischer Test gerechnet, dessen Ergebnis nur angegeben wird sofern er sich vom Ergebnis oder den resultierenden Effekten des t-Tests deutlich unterscheidet.

Effektgrößen Janczyk und Pfister (2013) betonen, dass auch minimalste Mittelwertsunterschiede (z. B. $\mu_A = 100, \mu_B = 100, 001$) zwischen Gruppen signifikant werden können, solange die Stichprobe nur groß genug ist. Dies bedeutet – wie schon oben angedeutet – mitnichten, dass dieser Unterschied auch eine praktische Relevanz besitzen muss. Für diese Arbeit würde dies beispielsweise bedeuten, dass ein fiktiver Unterschied

²hier durch den Autor mitunter auch Wx-Test abgekürzt

³hier durch den Autor mitunter auch MWU-Test abgekürzt

der relativen Lösungshäufigkeit von 0,5 % signifikant sein könnte. Dieser Unterschied wäre aber zu gering, um Empfehlungen für die Schulpraxis daraus abzuleiten.

Das bekannteste Maß für die Effektstärke ist *Cohen's d*

$$d = \frac{\mu_A - \mu_B}{\hat{\sigma}}$$

Dabei entsprechen μ_a und μ_b dem jeweiligen Lageparameter der Gruppen/der Zeitpunkte, die jeweils durch die Angabe einer Standardabweichung $\hat{\sigma}$ dimensionslos standardisiert werden (Janczyk und Pfister 2013). Für die scheinbar gemeinsame Standardabweichung $\hat{\sigma}$ kommen zwei Berechnungsverfahren in Betracht, die hier aber nicht näher ausgeführt werden sollen (vgl. *gepoolte Standardabweichung*). Siehe auch Tabelle 10.2

Als Effektmaß bei nicht-parametrischen Test wie Wilcoxon oder Mann-Whitney-U wird das Effektmaß *r* genutzt. Dieses lässt sich aus dem z-Wert und der Stichprobengröße einfach bestimmen (Fritz, Morris und Richler 2012).

Tabelle 10.2.: Übersicht der genutzten Effektstärken (Bühner und Ziegler 2009; Janczyk und Pfister 2013)

Effektgröße	negativ	kein	klein	mittel	groß
d	bis 0	0	ab 0.2	ab 0.5	ab 0.8
r	bis 0	0	ab 0.1	ab 0.24	ab 0.37

Reliabilität: Cronbach's Alpha Der Koeffizient Cronbach's Alpha ist ein Maß für die interne Konsistenz einer Skala (Bortz und Döring 2006), d. h. im weitesten Sinne „die Ähnlichkeit, mit der die Aufgaben der Skala gelöst wurden“. Der Wert erhöht sich mit der Anzahl der Items pro Skala (ebd.), weshalb im allgemeinen ein Wert von $\alpha_{Cronbach} = 0.8$ also gut angesehen wird, bei Skalen mit wenigen Fragen ≤ 6 gelten auch Werte um $\alpha_{Cronbach} = 0.6$ als nutzbar (Riese und Reinhold 2014).

Rasch-Analyse Zur Raschanalyse als grundsätzlichem Auswerteverfahren ist der Übersicht halber das folgende Kapitel gewidmet.

Die klassische Testtheorie geht davon aus, Leistungen, Fähigkeiten und Einstellungen durch geeignete Items messen zu können und den so ermittelten, wahren Wert mit dazugehöriger Messabweichung angeben zu können (Bortz und Döring 2006). Tatsache ist aber ebenso, dass sich

„die überwiegende Zahl der Messgrößen in der naturwissenschaftsdidaktischen Forschung einer direkten Messung“ entzieht (Neumann 2014). Im Rahmen der Item-Response-Theorie wird daher davon ausgegangen, dass sich mit Tests oder Fragebögen allenfalls latente⁴ Konstrukte erfassen lassen (Bortz und Döring 2006; Neumann 2014; Boone, Staver und Yale 2014). Ein solches Konstrukt kann z. B. die Mechanik-Kompetenz wie auch das Fachinteresse an Physik sein. Nach Neumann (2014) ist weiterhin die reine Verknüpfung der Anzahl oder des Anteils gelöster Aufgaben in einem Fachtest keine geeignete Vergleichsgröße für einen Test. Dies wäre nur sinnfällig denkbar, wenn alle Aufgaben exakt „gleich schwierig“ wären, was unrealistisch scheint. Daher ist dieser Vergleich „ohne weitere Berücksichtigung der Schwierigkeiten der jeweiligen Aufgaben problematisch“ (ebd.). Geht man nun davon aus, dass die richtige Lösung einer Aufgabe (dichotome Charakteristik der Lösung angenommen) gewissen Schwankungen unterlegen ist, lässt sich ein Zusammenhang zwischen der Ausprägung des latenten Konstrukts (im Folgenden sei hier die Mechanik-Kompetenz gemeint und o. B. d. A. als Personenfähigkeit Θ bezeichnet) und der Wahrscheinlichkeit, eine Aufgabe zu lösen, unterstellen. Im Folgenden soll das Rasch-Modell als Anwendung der Item-Response-Theorie kurz dargestellt werden. Dieses stellt einen praktikablen Zusammenhang zwischen der Personenfähigkeit und der Aufgabenschwierigkeit her.

Den Zusammenhang zwischen Personenfähigkeit und Lösungswahrscheinlichkeit für ein dichotomes Item bezeichnet man als Itemcharakteristik oder Itemfunktion (Rost 2004). Als Verlauf kommen zumeist monoton steigende Funktionen in Frage (ebd.), da davon auszugehen ist, dass die Lösungswahrscheinlichkeit mit der Personenfähigkeit zunimmt oder zumindest nicht abnimmt. Beispielhafte Item Characteristic Curves (ICCs) zeigt Abb. 10.1.

⁴vgl. Duden: „nicht unmittelbar sichtbar oder zu erfassen“

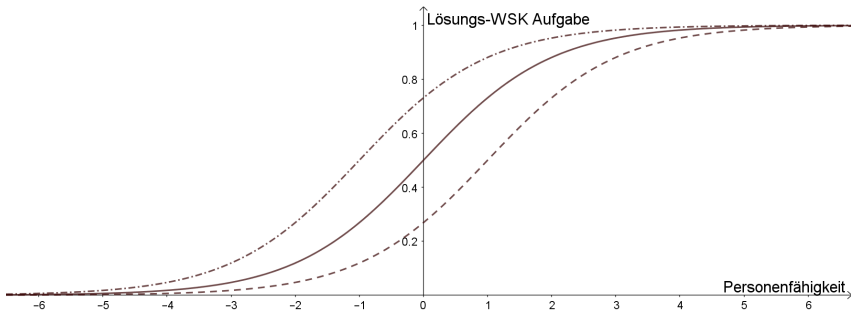


Abbildung 10.1.: Schematische Darstellung der Item Characteristic Curves dreier Aufgaben unterschiedlicher Schwierigkeit entsprechend des Rasch-Modells

Rost (2004) schlägt als Kriterien für mögliche Funktionen vor, dass diese möglichst einfach, plausibel und statistisch vorteilhaft sein sollen. Lineare Funktionen kommen nicht in Frage, da diese keine unbeschränkte Variable (latente Variable) mit einer beschränkten (Lösungswahrscheinlichkeit) verknüpfen können, ohne Bereiche der latenten Variable auszuschließen. Letzteres lässt u. U. eine Zuordnung bestimmter Personen zu einer Lösungswahrscheinlichkeit nicht zu (ebd.). Aufgrund derselben Konsequenz folgt ebenso, dass unterschiedliche Itemschwierigkeiten nicht sinnvoll modelliert werden können.

Dieser Problematik kann man nach Rost (ebd.) dadurch entgegen, nur einen bestimmten Teil der Funktion als linear anzunehmen und die Bereiche sehr hoher und niedriger Lösungswahrscheinlichkeit asymptotisch 0 bzw. 1 anzunähern.

Dazu wird der Quotient aus der Lösungswahrscheinlichkeit und Nicht-Lösungswahrscheinlichkeit (also der Gegenwahrscheinlichkeit) gebildet, welcher als „Wettquotient“ aufgefasst werden kann, siehe Gleichung (10.1).

$$\frac{P(X_{vi} = 1)}{P(X_{vi} = 0)} \tag{10.1}$$

Die so erhaltenen Werte bilden nicht nur den Bereich zwischen 0 und 1 (den Bereich der Lösungswahrscheinlichkeit), sondern den Bereich von 0 bis $+\infty$ ab. „Um diese Asymmetrie wieder zu beseitigen, und die gesamte Zahlengerade, also auch den negativen Wertebereich, mit einzubeziehen wird dieser Wettquotient logarithmiert, was man dann als den *Logit* der Wahrscheinlichkeit bezeichnet“ (ebd.). Die Wahrscheinlichkeit wird so symmetrisch auf ein Intervall von $-\infty$ bis $+\infty$ projiziert. Dabei kann

angenommen werden, dass die Transformation linear abhängig von der Personenfähigkeit ist, siehe Gleichung (10.2)

$$\log \frac{P(X_{vi} = 1)}{P(X_{vi} = 0)} = \Theta_v - \beta_j \tag{10.2}$$

Ist die Differenz aus Aufgabenschwierigkeit β_j und Personenfähigkeit nun negativ, ist die Lösungswahrscheinlichkeit kleiner als 50%, ist sie positiv, größer als 50%. Ist die Aufgabenschwierigkeit gleich der Personenfähigkeit beträgt die Lösungswahrscheinlichkeit 50%. Für die vollständige Herleitung siehe Rost (ebd.). Die Itemfunktion ergibt sich durch Umstellen zu

$$P(U_{ij} = 1 | \Theta_i, \beta_j) = \frac{e^{\Theta_i - \beta_j}}{1 + e^{\Theta_i - \beta_j}},$$

dabei steht β_j für die Schwierigkeit von Aufgabe i und Θ_i für die Fähigkeit von Person i.

Eine kompakte Darstellung von Schwierigkeiten und Fähigkeiten liefern *Person-Item-Maps* oder *Wright-Maps*. Diese zeigen übersichtlich die histogrammartige Verteilung der Personenfähigkeit und der Aufgabenschwierigkeit an der selben Skala ausgerichtet (Linacre 2019). Abbildung 10.2 zeigt eine solche Person-Item-Map, die Mittellinie kennzeichnet dabei einer gedachte Skala in logits, jedes A steht für eine Aufgabe und jedes X für eine Anzahl an Personen. Eine beispielhafte Interpretation dieser Map wäre, dass die Passung aus Aufgabenschwierigkeiten und den Fähigkeiten der Personen nicht ideal ist, da viele Aufgaben zu schwierig sind und es zu wenig leichte Aufgaben gibt. Dies verringert das Auflösungsvermögen dieses virtuellen Tests.

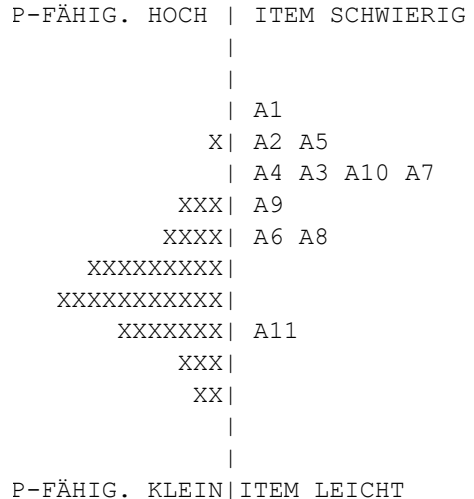


Abbildung 10.2.: Beispielhafte Person-Item-Map

Das Ergebnis der Rasch-Analyse, welches von primären Interesse ist, liefert in erster Linie für jede Aufgabe die entsprechende Aufgabenschwierigkeit, also den Parameter β_j , und die zu jeder Person gehörige

Personenfähigkeit. Die Werte werden für jede Kombination von Personen und Aufgaben neu berechnet. So sind die Ergebnisse für Personenfähigkeiten oder Aufgabenschwierigkeiten über eine bestimmte Analyse hinaus zunächst *nicht* miteinander vergleichbar.

Anforderungen Zur Anwendung des Rasch-Modells sollten einige Annahmen über den zugrunde liegenden Test erfüllt sein. Im Folgenden werden diese kurz skizziert:

Eine Voraussetzung ist die *Lokale stochastische Unabhängigkeit* (Neumann 2014; Strobl 2015). So ist gefordert, dass die Wahrscheinlichkeit eine bestimmte Aufgabe zu lösen, nicht davon abhängen darf, ob eine andere Aufgabe gelöst wurde oder nicht. Als Beispiel führt Strobl (2015) an, dass Ergebnisse aus vorangegangenen Aufgaben nicht insofern Basis einer neuen Aufgabe sein dürfen, als dass diese notwendig für deren Lösung sind. Gleichmaßen darf aus einer Aufgabe nicht „nicht automatisch die Lösung der anderen Aufgaben folg[en]“ (ebd.). Ebenso würden sog. „Testlets“ diese Anforderung verletzen, bei denen sich mehrere Fragen auf z. B. den selben Text bzw. den selben Aufgabenstamm beziehen. Diese Voraussetzung lässt sich bei der Konstruktion des Tests relativ einfach berücksichtigen.

Die nächste Voraussetzung ist die *Spezifische Objektivität*. Die Objektivität bezieht sich hier darauf, dass es für den Vergleich zweier Personen unerheblich sein muss, welche einzelne Aufgabe man für diesen Vergleich heranzieht (vgl. ebd.). So muss die Wahrscheinlichkeit eine beliebige Aufgabe zu lösen für eine Person größerer Fähigkeit stets größer sein als für eine Person mit geringerer Fähigkeit. Gleichbedeutend damit ist, dass die ICCs für alle Aufgaben keine Schnittpunkte aufweisen. Nach Neumann (2014) lässt sich die Spezifische Objektivität testen, indem zwei Zufallsstichproben darauf getestet werden, inwiefern sie z. B. gleiche Itemparameter liefern.

Weiterhin wird die *Eindimensionalität* des zu erhebenden Konstrukts für die hier durchgeführte Rasch-Analyse vorausgesetzt⁵, also dass „die Personen- und Aufgabenparameter auf einer gemeinsamen latenten Dimension liegen“ (Strobl 2015). Für das hier unterstellte Konstrukt der Mechanik-Kompetenz bedeutet dies beispielsweise, dass Aufgaben nicht gleichermaßen die Lesekompetenz testen dürfen⁶. Diese Annahme lässt sich auch argumentativ untermauern, eine bevorzugte Methode stellt aber die zur Analyse verwendete Software „Winsteps (Version 4.0.1)“ dar. Diese

⁵Wobei auch mehrdimensionale Raschmodelle existieren

⁶Natürlich kann nach Auffassung des Autors dieser Arbeit nicht verlangt werden, dass die Lesekompetenz keinen Einfluss auf die Lösungswahrscheinlichkeit der gestellten Aufgaben hat. Praxisnah ist aber das Vermeiden von unnötig komplexen Sätzen und umständlichen Formulierungen.

ermöglicht es, mittels einer Hauptkomponentenanalyse (Principal Component Analysis, PCA) multidimensionale Konstrukte in den Daten zu ermitteln. Werden zusätzliche Konstrukte in den Daten freigelegt, kann dies auch durch Zufälle bedingt sein (Linacre 2011). Erklären diese jedoch eine Varianz in der Stärke von mindestens zwei Items („Secondary dimension in the data must explain at least 2 items’ worth of variance“ (ebd.)), ist zunächst von mehrdimensionalen Daten auszugehen. Diese Grenze konnte durch Simulationen gezeigt werden. *Winsteps* bietet jedoch auch die Möglichkeit, Daten anhand der aktuellen Analyse so zu simulieren (ebd.), dass sich diese zwar an den Personenfähigkeiten und Itemschwierigkeiten orientieren, jedoch perfekt auf das Model passen. So erhaltene, „perfekte“ Werte für die Hauptkomponenten lassen eine Abschätzung der „realen“ Werte hinsichtlich deren Einfluss zu.

Vergleiche unterschiedlicher Tests Jede Raschanalyse berechnet aus der eingegebenen Datenmatrix die entsprechenden Personenfähigkeiten und Aufgabenschwierigkeiten. So sind diese Maße für jeden Test individuell und über unterschiedliche Tests, wie z. B. Pre- und Posttests, hinweg nicht vergleichbar. Dazu ist es also erforderlich, beide Tests an *derselben* Skala auszurichten. Zu diesem Zweck werden die Aufgabenschwierigkeiten eines Tests händisch als Vorgabe für die Aufgabenschwierigkeiten für den zweiten Test genutzt. Dieses *Anchoring* genannte Verfahren ist „the process of using anchor values to insure that different analyses produce directly comparable results“ (Linacre 2019). So werden die Personenfähigkeiten über beide Tests miteinander vergleichbar (Linacre (2004) dazu: „During an anchored analysis, the person measures are computed from the anchored item values“ (Worauf noch weitere Berechnungen folgen.))⁷.

Damit die Aufgabenschwierigkeiten übertragen werden können, ist es selbstverständlich erforderlich, dass zumindest ein Teil des Tests aus gleichen Items besteht (diese werden im Folgenden Anker-Items genannt). Als weitere Anforderung wird vorausgesetzt, dass die Ankeritems – bei getrennter Analyse von Vor- und Nachtest – ihre Lage zueinander nur insofern verändern, dass bei einem x-y-Plot der jeweiligen Schwierigkeiten eine verschobene Gerade erscheint. Zu weit von der Geraden entfernte Items sind dabei nicht als Ankeritems geeignet. Die verwendete Software *Winsteps* gibt weiterhin den Wert der Verschiebung als Displacement direkt in der Analyse aus. Mit Rückgriff auf Wright und Douglas (1976) bezeichnen O’Neill u. a. (2013) ein Displacement von etwa 0.5 logits als problemlos (wobei auch 0.6 akzeptiert werden).

⁷Alternativ zum Anchoring bietet sich das Stacking an, das Berechnen *aller* Aufgaben in einer Analyse.

Qualitätssicherung Zur Qualitätssicherung bedient man sich hinsichtlich Aufgabenschwierigkeiten und Personenfähigkeiten den sogenannten Fit-Maßen. Nach Boone, Staver und Yale (2014) geben die Maße des Fits an „how well data [is] conform to the Rasch model“. Beschrieben wird also die Güte der Modellpassung. Dabei wird zwischen den Infit-MNSQ- und den Outfit-MNSQ⁸-Werten unterschieden. Beide beschreiben in gewisser Weise den Unterschied zwischen erwarteter und tatsächlichem Verhalten der Person/des Items. Der Outfit, der etwas empfindlicher auf große Unterschiede reagiert (Neumann 2014) und in dieser Arbeit verwendet wird, berechnet sich als Summe der quadrierten Standard-Residuen geteilt durch die Anzahl der Personen/Aufgaben (Linacre und Wright 1994a). Für Infit und Outfit gilt, dass Werte von 1 eine ideale Modellpassung angeben. Weichen die Werte nach oben hin ab, zeigt sich eine schlechtere Passung, Werte unterhalb von 1 zeigen entsprechend eine zu gute Passung. Für „gewöhnliche“ (übersetzt nach Boone, Staver und Yale 2014) Multiple-Choice-Fragebögen sehen Boone, Staver und Yale (ebd.) (nach Linacre und Wright (1994b)) Outfit-Werte zwischen 0,7 – 1,3 als vernünftig⁹ an. Generell, und somit auch für Personen, geben Boone, Staver und Yale (2014) einen vernünftigen Bereich des Outfit-MNSQ von 0,5 – 1,5 an. Werte zwischen 1,5 und 2 verschlechtern die Messung nicht, tragen aber auch nicht produktiv zur Messung bei. Erst größere Werte als 2 werden als problematisch angesehen. Kleinere Werte als 0,5 werden ebenfalls als eher unproblematisch gesehen, können aber fälschlicherweise gute Reliabilitäten erzeugen (ebd.). Personen oder Items, die außerhalb dieser Werte liegen, können aus der Analyse ausgeschlossen werden.

Nach Neumann (2014) können jedoch „[z]ur Einschätzung, inwieweit Infit und Outfit signifikant von 1,0 abweichen, [...] noch entsprechende z-standardisierte Werte (ZSTD) oder T-Werte angegeben [werden]“. Diese geben eine Wahrscheinlichkeit dafür an, ob die Outfit-MNSQ -Werte zufällig zustande gekommen sein können. Sollte dieser Wert den Betrag 2 übersteigen, gilt dies als nicht zutreffend. Nach Linacre (2019) sollen die ZSTD-Werte solange vernachlässigt werden, wie die Fit-Werte in einem akzeptablen Bereich liegen. Das heißt am Beispiel der Personenfähigkeiten, dass erst bei Personen mit einem Outfit außerhalb der Grenzen die ZSTD-Werte betrachtet werden sollen und erst bei deren Übersteigen der Grenzen Konsequenzen gezogen werden sollten.

Es wird von vielerlei Seite davor gewarnt, Personen oder Items vor schnell aus der Analyse auszuschließen (Neumann 2014; Boone, Staver und Yale 2014; Linacre 2019). Zur Frage, ob alle nicht-perfekten Personen

⁸Meansquare

⁹Die Formulierung „vernünftig“ legt schon nahe, dass ein Item mit einem Outfit-MNSQ von bspw. 1,38 nicht unbedingt aus der Analyse ausgeschlossen werden muss.

und Items aus der Analyse ausgeschlossen werden müssen, äußern sich Boone, Staver und Yale (2014) fast schon poetisch:

„This conundrum is like improving the precision with which a building brick is made. For the construction of some buildings, it will not matter if the brick matches other bricks to a nanometer. Much depends upon what is measured.“

(ebd.)

Zusätzlich schlägt Linacre und Wright (2010) vor, *misfitting* Items oder Personen testhalber aus der Analyse auszuschließen und die nun erhaltenen Werte der Personenfähigkeit/Aufgabenschwierigkeit gegen die zuerst erhaltenen zu plotten. Zeigen sich so keine gravierenden Abweichungen, können die entsprechenden Personen in der Analyse verbleiben.

Letztlich ist nach Ansicht des Autors dieser Arbeit ein Ausschluss von Items oder Personen also in gewisser Weise auch eine Entscheidung die sowohl mit argumentativen Leitplanken (Outfits, ZSTD-Werte, Plots) als auch mit inhaltlichem Fingerspitzengefühl (z. B. „Weshalb soll die Aufgabe in der Analyse verbleiben?“) getroffen werden muss.

Teil V.

Ergebnisse

11. Ergebnisse der Vor- und Vergleichsstudien

11.1. Exploration

Während der Entwicklungsphase wurde eine Schulklasse eingeladen (N=26), die Unterrichtsidee der vorgestellten Lernumgebung zu erproben. Ziel dieser Veranstaltung aus *Perspektive der Materialentwicklung* war, das grundsätzliche Funktionieren der Lernumgebung in technischer Hinsicht festzustellen. Weiterhin sollte die grundsätzliche Akzeptanz des Lernprogramms evaluiert werden. Ebenso sollte das Arbeitstempo abgeschätzt werden, um für die spätere Durchführung der Studie Anhaltspunkte zu gewinnen. Aus *Perspektive der Schüler und Schülerinnen* war das Ziel, mit einer neuartigen Lernmethode – an Tablet-Computern – zu lernen. Auf eine formalisierte Überprüfung des Lernzuwachses wurde zugunsten des Explorationscharakters fast gänzlich verzichtet.

Die *Lernumgebung* war zum fraglichen Zeitpunkt noch nicht vollständig und eher prototypisch entwickelt, daher wurde nur das 2D-Konzept mit einer Klasse erprobt. Als Feedbackvariante wurde mangels Animationsfeedback das Textfeedback genutzt. In den zugehörigen Multiple-Choice-Fragen wurden vier Antwortoptionen dargeboten.

Der Ablauf des Vormittags gestaltete sich wie folgt: Nach einer kurzen Begrüßung und einem Überblick über die Elemente und Struktur der Lernumgebung begann für die Schüler und Schülerinnen (Jahrgang 7) das eigenständige Arbeiten am Tablet. Dafür waren zunächst 3 Zeitstunden eingeplant, unterbrochen von einer gemeinsamen Pause.

11.1.1. Evaluation

In diesem Abschnitt soll auf die Erkenntnisse eingegangen werden, die diese erste Exploration für die Weiterentwicklung ergeben hat.

Grundsätzlicher Eindruck

Im Grunde hatten die Schülerinnen und Schüler auf technischer Ebene nur wenig Probleme mit der Bedienung. So gelang das Öffnen der Simu-

lationen aus dem externen Ordner nach anfänglicher Unsicherheit und Eingewöhnung augenscheinlich gut. Das Schließen hingegen wurde von vielen Lernenden vernachlässigt, sodass mit fortschreitender Zeit zahlreiche Simulationen geöffnet waren, was durchaus zu Leistungseinbußen bei den Tablets führte.

Unklar war den Schülern mitunter, wann sie „das Ziel“ einer Simulation erreicht haben. Diese sind nämlich – auch nach wie vor – auf das Abarbeiten der Forschungsaufträge fokussiert. Dies geht nicht immer mit dem Erreichen eines Ziels einher. So erscheint in der Regel keine Meldung wie „Du hast es geschafft!“ – die Simulationen lassen sich also beliebig lange wiederholen. Diese Problematik im Kontext mit dem Forschungsauftrag führte vermeintlich auch zu den im Folgenden beschriebenen Problematiken. Bei bestimmten Simulationen wurde durchaus ein nicht zielgerichtetes Experimentieren beobachtet. Dies betraf z. B. die Simulation „Scooter“, bei der aus technischer Sicht frei mit den Scootern gefahren werden kann, oder die Simulation „Zielfoto“, bei der verschiedene Parameter der Bremsstärke scheinbar reizvoll verändert werden können.

Akzeptanz und Fachwissen

Im Folgenden sind die Fragen der abschließenden Evaluation aufgeführt. Die evaluierten Aspekte gliedern sich¹ grob in die Kategorien Verständlichkeit [Q1+7], Bedienung [Q3+4], Simulationen [Q5+6], Feedback [Q8+9+10+14], und Sonstiges [Q2+11+12+13+15] (getrennt durch waagerechte Linien).

- Die Texte waren gut verständlich. [Q1]
- Die Fragen waren verständlich formuliert. [Q7]

- Die Bedienung des Lernprogramms (vor, zurück etc.) hat gut funktioniert. [Q3]
- Die Bedienung der Simulationen hat gut funktioniert. [Q4]

- ²*Es war schwierig jede Simulation genau zu beobachten.[Q5]
- *Mir war manchmal unklar, was ich mit den Simulationen tun sollte. [Q6]

- Die Tipps und Hinweise haben mir weitergeholfen. [Q8]
- Ich habe mir die Videos angesehen. [Q9]
- Die kurzen Videos haben mir geholfen. [Q10]

¹Zur besseren Übersicht, ohne statistische Prüfung

²*=Invertiertes Item

- Ich habe die Hinweistexte nach den Fragen gelesen. [Q14]
- Die Texte waren nicht zu lang. [Q2]
- Mir ist bewusst, was ich mit dem Programm an den Tablets gelernt habe. [Q11]
- *Ich habe mich manchmal gelangweilt. [Q12]
- Das Design des Lernprogramms hat mir gefallen. [Q13]
- Mir hat das Lernprogramm gefallen. [Q15]

Zusätzlich wurden kompakte Fachwissen-Fragen in Form von Multiple-Choice (bzw. Single-Choice) Items gestellt, wie in Abb. 11.1 dargestellt.

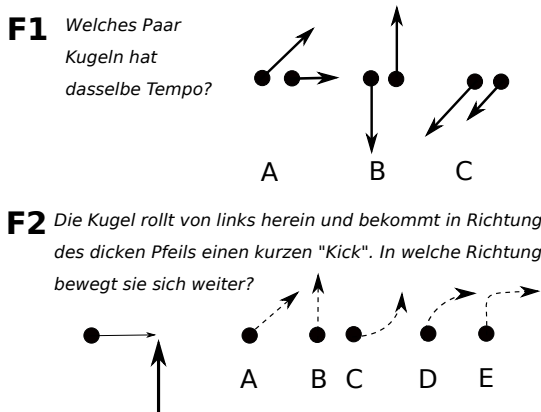


Abbildung 11.1.: Die Multiple-Choice-Aufgaben F1 und F2

Ergebnisse Akzeptanz 84,6% der befragten Schüler gaben an, die Texte der Lernumgebung gut oder eher gut verständlich zu finden. 7,7% lehnen die Aussage hingegen komplett ab. Mehr Varianz zeigt sich bei der Verständlichkeit der Formulierungen zu den MC-Fragen. Dort sind etwa 65% der Befragten der Auffassung, dass diese gut oder eher gut verständlich formuliert sind. 18% lehnen diese Aussage ab bzw. eher ab. Näheres dazu in Abbildung 11.2.

Auch wenn die Schülerinnen und Schüler aller Wahrscheinlichkeit nach eher erfahrene Nutzer von Tablets und anderen touchfähigen Geräten sind, sollte erhoben werden, wie gut diese mit der Bedienung des Lernprogramms [Q3] und der Simulationen [Q4] zurecht gekommen sind, um etwaigen Problemen vorbeugen zu können. Erfreulicherweise geben 82,3% der Befragten an, dass das Bedienen der Simulationen gut (38,5%) oder

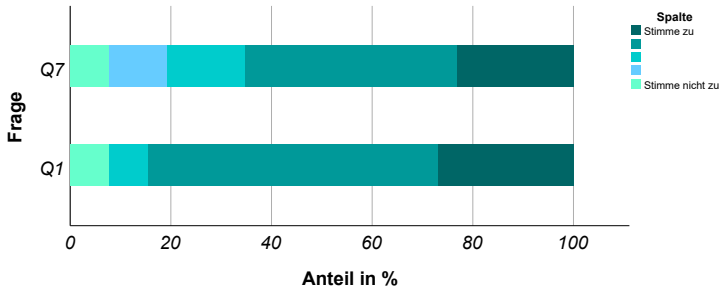


Abbildung 11.2.: Ergebnisse bzgl. Verständlichkeit

eher gut funktioniert hat. Kein Schüler oder Schülerin stimmte dieser Aussage nicht oder eher nicht zu. Die Bewertung der Bedienung des Lernprogramms erfolgt in augenscheinlich ähnlicher Art. 50,0% attestieren hier gute Bedienungsmöglichkeiten, 26,9% eher gut. Auch hier lehnt niemand diese Aussagen ab oder eher ab (vgl. Abbildung 11.3).

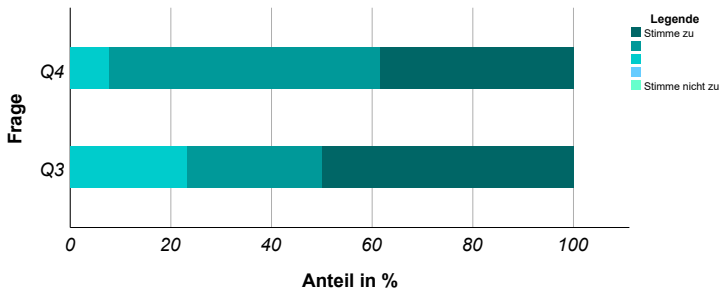


Abbildung 11.3.: Ergebnisse bzgl. Bedienung

Zentral für die spätere Studie ist nicht nur das „Bedienen-Können“ der Simulationen, sondern auch der inhaltliche Umgang mit ihnen. Daher wurde die Klarheit der Forschungsaufträge erfragt [Q6], ebenso wie die Einfachheit der Beobachtung [Q5].

Etwa 15% der Befragten gaben an, dass sie *manchmal* nicht wussten, was sie mit den Simulationen tun sollen. Bei weit über der Hälfte der Schüler war dies hingegen (eher) nie der Fall. Das (sachgemäße), genaue Beobachten der Simulationen fiel keinem Befragten (eher) schwer. 88,4% geben an, diesbezüglich niemals Probleme gehabt zu haben (vgl. Abb. 11.4).

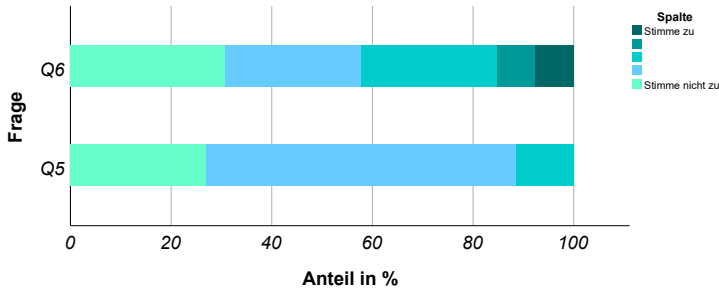


Abbildung 11.4.: Ergebnisse bzgl. Umgang mit Simulationen

Der Umgang mit dem Feedback war Gegenstand der folgenden Fragen. Die Hilfe durch die Hinweistexte wurde von den Schülern unterschiedlich aufgefasst [Q8]. Über 42% der Befragten empfanden die Hinweistexte nach falschen MC-Fragen als (eher) hilfreich, etwa ein Drittel stimmten dazu neutral ab. 25% der Befragten hingegen empfanden die Hinweistexte nicht immer als hilfreich. Gelesen wurden die Hinweistexte hingegen von der überwiegenden Mehrheit (77,3%) der Schüler [Q14]. Nur 13,6% gaben an, diese (fast) nie gelesen zu haben (vgl. Abb. 11.5). Die Konsistenz und damit die Aussagekraft der Daten ist an dieser Stelle im Kontext beider Fragen eher kritisch zu sehen, da auch Schüler/innen, die angaben die Hinweise fast nie gelesen zu haben, deren Qualität beurteilten.

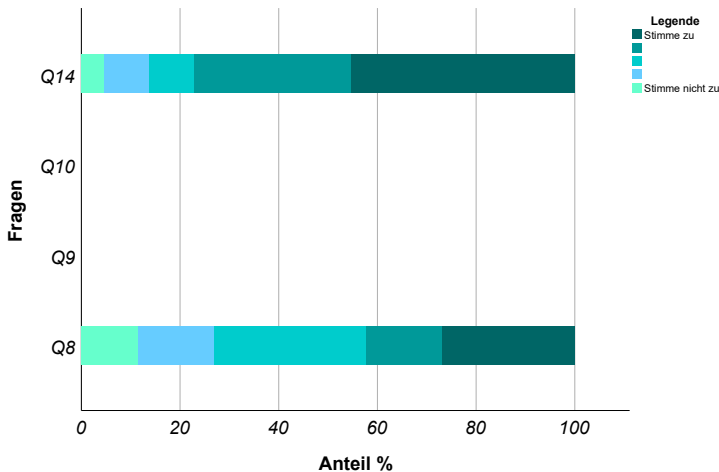


Abbildung 11.5.: Ergebnisse bzgl. Umgang mit Feedback

Die Fragen [Q9] und [Q10] beziehen sich auf das Animationsfeedback, das zum Zeitpunkt der Durchführung aber noch nicht eingearbeitet war. Daher wurden diese Fragen auch nicht ausgewertet.

Der letzte Teil umfasst Fragen zu sonstigen Aspekten. Zu diesen gehört etwa die Frage nach angemessener Textlänge [Q2]. Die Hälfte der Schüler zeigt sich diesbezüglich neutral, etwa 34% stimmten der Aussage voll oder eher zu. Das Bewusstsein über die durch das Lernprogramm gelernte Inhalte ist relativ ausgeprägt. 88,5% stimmen der Aussage über dieses Bewusstsein zu verfügen voll oder eher zu - während sie von keinem Schüler abgelehnt oder eher abgelehnt wird. Die Aussage sich manchmal gelangweilt zu haben [Q12] wird von 65,3% der Befragten abgelehnt bzw. eher abgelehnt. Nur 15% und damit 4 Schüler/innen stimmten der Aussage komplett zu.

Das Lernprogramm als solches wie auch dessen Design wurden überwiegend positiv bewertet. Alle befragten Personen gaben an, dass ihnen das Lernprogramm grundsätzlich gefallen habe (73,1% davon voll, 26,9% eher). 69,2% hielten das Design der Lernumgebung für gelungen bzw. eher gelungen, während es von niemandem komplett, jedoch von 11,5% teilweise abgelehnt wurde (vgl. Abb 11.6).

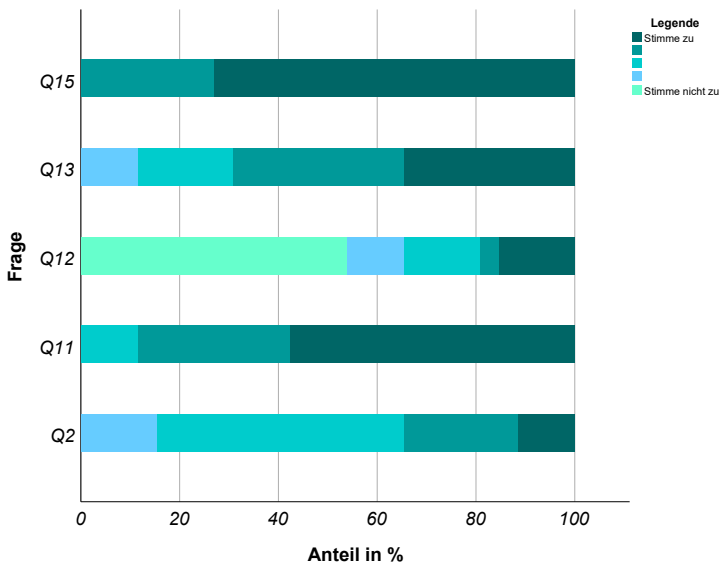


Abbildung 11.6.: Ergebnisse zu weiteren Aspekten

Ergebnisse Fachlich Die Ergebnisse dieser eher „blitzlichtartigen“ Abfrage des gewonnenen Fachwissens stellen sich recht zufriedenstellend dar. 96,2% der Kinder wählten in der Aufgabe F1 das korrekte Kugelpaar B aus. Die grundsätzliche Bedeutung der Pfeile scheint also in gewissem Rahmen in der kurzen Explorations-Zeit zu transportieren gewesen zu sein. Auch zu Aufgabe F2 wählten 73,1% der Schüler die korrekte Antwort A. Etwa ein Viertel hingegen hielt auch nach dem Lehrgang an der Alltagsvorstellung „Einwirkungsrichtung = Bewegungsrichtung“ fest (vgl. Kapitel 3.3).

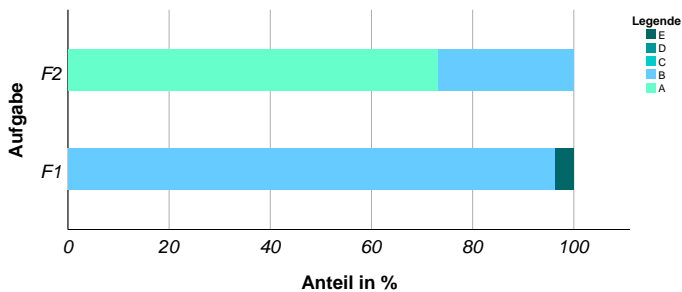


Abbildung 11.7.: Ergebnisse der Fachwissensaufgaben F1 und F2

MC-Items

Im Folgenden wird knapp das Antwortverhalten beim Bearbeiten der Multiple-Choice-Fragen, welche das Feedback generieren, skizziert. Ziel ist es hier an vier ausgewählten Beispielen das grundsätzliche Vorgehen der SuS darzustellen – und nicht, die Lösungshäufigkeiten der zahlreichen Antwortmöglichkeiten im Detail darzustellen. Dies wäre auch kaum zielführend, da wie bereits erwähnt eine prototypische Version der Lernumgebung exploriert wurde, welche später noch diverse (inhaltliche) Entwicklungsschritte durchlief.

Zum einen ist bei der Interpretation der Diagramme (Abb. 11.8 und 11.9) zu berücksichtigen, dass jeder SuS letztlich einmal die richtige Antwort auswählen muss, um in der Lernumgebung fortzuschreiten. Zum anderen ist zu berücksichtigen, dass die *kumulierte* Antworthäufigkeit dargestellt ist, also die gesamte Häufigkeit mit der die Antwortoptionen gewählt wurden. Es ist *nicht* dargestellt, welche Option erstmalig gewählt wurde. Als letzter Hinweis sei gesagt, dass auch Häufigkeiten auftreten können, die größer als die Anzahl der Versuchspersonen ist, was zunächst unlogisch scheint. Dies ist nach Abklärung der Logfiles durch mehrmaliges

Beantworten der Frage durch Vor- und Zurückspringen innerhalb der Lernumgebung zu begründen.

Im Folgenden nun die kurze Darstellung der Fallbeispiele:

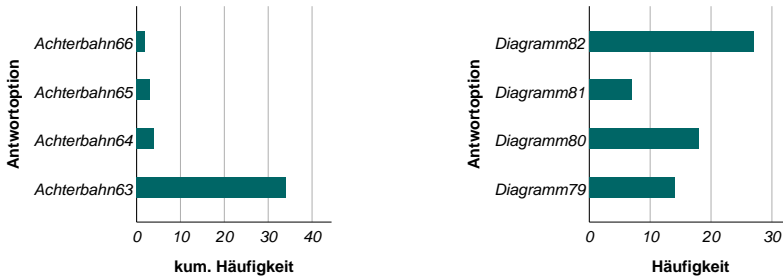


Abbildung 11.8.: Kumulierte Häufigkeiten der gewählten Antworten zu zwei Beispiel MC-Aufgaben. Korrektes Item jeweils am häufigsten gewählt.

Das erste Beispiel (Abb. 11.8, links) zeigt das Antwortverhalten eines augenscheinlich einfachen MC-Items. Distraktoren wurden jeweils höchstens 2-4 mal gewählt. Dieses Muster tritt vermehrt zu Beginn der Lernumgebung auf, wo die Simulationen noch relativ schlicht gehalten waren. Das rechte Beispiel zeigt das Antwortverhalten, bei dem eine Antwortoption eher vernachlässigt ausgewählt wird – ein Muster das in der Exploration nicht systematisch auftrat.

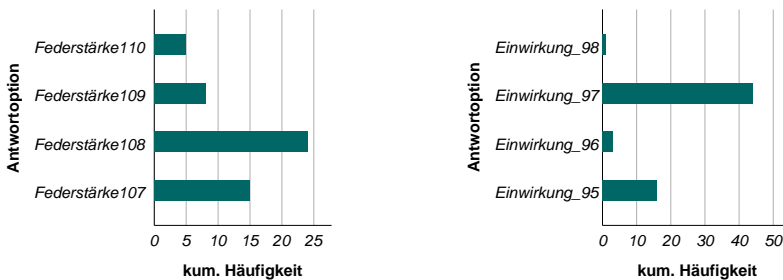


Abbildung 11.9.: Kumulierte Häufigkeiten der gewählten Antworten zu zwei Beispiel MC-Aufgaben. Korrektes Item jeweils am häufigsten gewählt.

Das linke Beispiel in Abb. 11.9 zeigt ein relativ ausgewogenes Antwortverhalten. Auch wenn ein Distraktor deutlich überwiegt, werden die übrigen relativ häufig ausgewählt. Dieses Muster zeigt sich relativ aufgabenspezifisch, vermehrt jedoch zu Simulationen im Bereich *Kraft*. Hingegen ist im rechten Beispiel ein Muster zu erkennen, bei dem nur ein Distraktor für die SuS tatsächlich plausibel scheint. Auch dieses Muster zeigt sich aufgabenspezifisch.

Zu bemerken ist also, dass sich unterschiedliche Muster bei der Beantwortung der MC-Fragen zeigen, die verschieden häufig auftreten. Die Qualität der Distraktoren scheint dabei ebenfalls unterschiedlich zu sein.

11.1.2. Diskussion

Zusammenfassend war für die weitere Entwicklung der Lernumgebung erfreulicherweise zu konstatieren, dass die Schülerinnen und Schüler mit der Lernumgebung auf technischer Ebene gut umgehen konnten. Dies bestätigen die Befragten in der Evaluation selbst, auch wenn durch den Studienleiter mitunter größere Probleme beim konsequenten Schließen von Simulationen beobachtet wurden. Darauf ist bei Folgestudien verstärkt zu achten. Ebenfalls zu berücksichtigen ist die von den SuS erwartete *Gamification* der Simulationen, d. h. dass entweder ein Ziel in den Simulationen erwartet oder mit den Simulationen unverhältnismäßig viel „gespielt“ wurde. Beides ist bei Folgeuntersuchungen durch entsprechende Instruktion bzw. geschicktere Forschungsaufträge zu minimieren. Eine grundlegende Umgestaltung der Simulation hin zu einer *Gamification* ließ sich zum Zeitpunkt der Exploration schon nicht mehr ökonomisch durchführen. Trotz der Problematik des vereinzelt auftretenden „Spielens“ herrschte seitens der Schüler eine relativ große Klarheit darüber, was mit den Simulationen zu tun sei. Die Gestaltung der Lernumgebung sowie die Länge und Verständlichkeit der Texte wurden ebenfalls positiv bewertet.

Die stichprobenartige Überprüfung des Lernzuwachses durch zwei Aufgaben zeigte erfreuliche Ergebnisse, sodass eine gewisse Lernwirksamkeit in den Kernaspekten unterstellt werden kann. Ebenso ist sich der Großteil der Befragten des eigenen Lernzuwachses bewusst.

Die Beantwortung der MC-Fragen bzw. die Auswahl der Distraktoren folgte unterschiedlichen Mustern. So schienen nicht immer alle Distraktoren ähnlich attraktiv. Zwar empfanden diese auch hier über die Hälfte der Befragten als verständlich, jedoch scheint es eine sinnvolle Option zu sein, die Anzahl der Antwortoptionen auf drei zu reduzieren um plausible Distraktoren zu erstellen sowie den Leseaufwand zu verringern (vgl. Kapitel 4.3).

Abgesehen von den erwähnten Aspekten zeigte sich, dass die Lernumgebung für eine Dauer von 2 – 3 Doppelstunden in der angestrebten Jahrgangsstufe eingesetzt werden kann.

11.2. Vorstudie

Die Vorstudie (siehe 9) wurde in zwei zehnten Klassen vorgenommen (N=45), davon wurden N=23 SuS nach dem 2D-Konzept unterrichtet, die weiteren nach dem 1D-Konzept. Angesichts des Vorwissens und vergleichsweise fortgeschrittenen Alters der SuS wurde die Lernumgebung an manchen Stellen leicht gekürzt, nicht mehr altersgerechte Passagen wurden geändert und im Abschluss wurde ein Ausblick auf den Beschleunigungsbegriff gegeben. Ziel der Vorstudie sollte sein, neben allgemeinen Beobachtungen zum Ablauf, auch die Fragebögen und Tests mit einer Reihe von Items auf Konsistenz zu überprüfen. Aus diesem Grund erfolgt auch keine ebenso ausführliche Darstellung der Ergebnisse, wie im entsprechenden Kapitel zur Hauptstudie.

11.2.1. Allgemeines

Während der Untersuchung zeigte sich durch Beobachtungen der Hilfskräfte, dass das Feedback nach den MC-Fragen scheinbar nur oberflächlich wahrgenommen wurde und die Tipps und Hinweise nicht zwingend gelesen wurden. Dies könnte auf die markanten roten und grünen Symbole zurückzuführen sein, die eine richtige oder falsche Antwortoption sofort als solche kennzeichneten (vgl. auch Kapitel 4.3).

Die übrigen Beobachtungen decken sich in etwa, mit dem in der Exploration (vgl. Kapitel 11.1). Vor allem der Effekt des vielfachen Öffnens der Simulationen trat trotz expliziter Hinweise wieder zu Tage. Des Weiteren wurden auftretende technische Probleme erfasst, gelöst oder zumindest beherrscht.

Letztlich wurde aufgrund der Beobachtung zum Feedback im später verwendeten Lernprogramm davon Abstand genommen, ein *General Feedback* anzugeben. Die Richtigkeit der Antwort lässt sich mit der verwendeten Methodik nun prinzipiell erst durch tatsächliches Lesen des jeweiligen Hinweises feststellen.

11.2.2. Fragebogen

In diesem Abschnitt werden die Daten der Fragebögen zunächst dargestellt und anschließend diskutiert. Dies erfolgt durchweg vor allem hinsichtlich der *Qualität* der sich anschließenden Hauptstudie. So können hier beispielsweise unpassende Items identifiziert werden und für die spätere Verwendung gestrichen oder modifiziert werden. Die Fragen der Hauptstudie entsprechen folglich im Detail nicht durchweg denen dieser Vorstudie, da sich Änderungen hin zur Hauptstudie ergeben haben. Die

Fragen befinden sich der besseren Übersicht halber im Anhang (siehe Kapitel A). Daher erfolgt die *inhaltliche* Darstellung der Ergebnisse auch nicht im gleichen Detailgrad wie in der Hauptstudie.

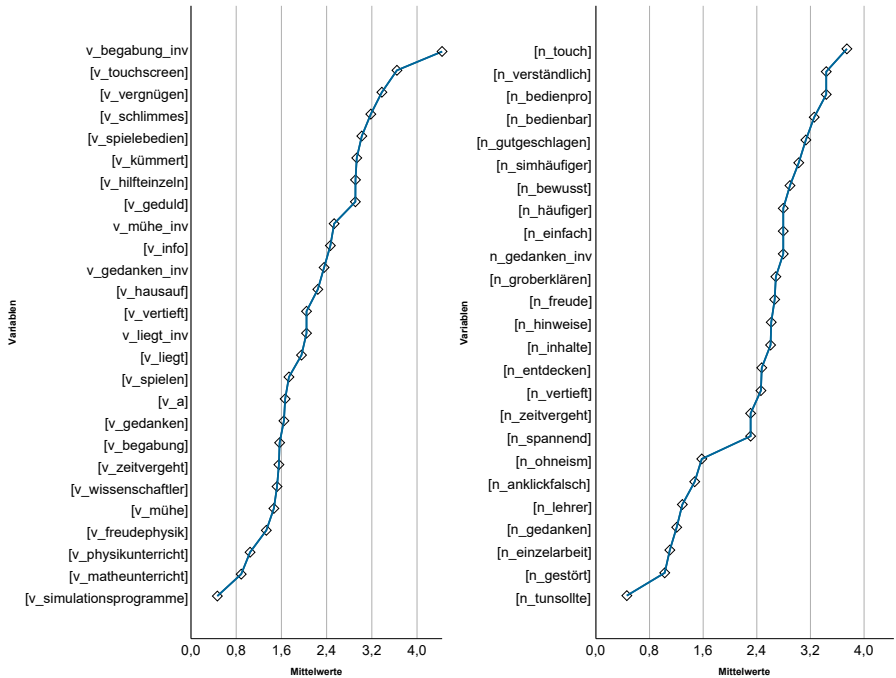
Qualitätssicherung

Angaben zur Reliabilität einer Skala, der Trennschärfe der Items, der Fallzahl sowie dem Anteil fehlender Werte finden sich an zugehöriger Stelle der Ergebnisdarstellung. Im Folgenden soll die Schwierigkeit der Items dargestellt werden.

Die Schwierigkeit von Items lässt sich auch für Likert-Items angeben (Bortz und Döring 2006). Diese sollte „wünschenswerterweise“ auf einer Skala von 0 bis 1 zwischen 0,2 und 0,8 liegen. Für die Schwierigkeit der vorliegenden Items bedeutet dies, dass ein Bereich zwischen 0,8 und 3,2 als optimal gilt (vgl. ebd.). Darüber oder darunterliegende Werte sollten als potentiell problematisch angesehen werden. Abb. 11.10 gibt eine Übersicht über alle Items der Fragebögen aus Vor- und Nachuntersuchung.

Erfreulicherweise liegt die überwiegende Mehrheit der Items innerhalb der beschriebenen Grenzen. Im Vorfragebogen unterschreitet lediglich die Frage nach der Häufigkeit des Simulationseinsatzes und im Nachfragebogen die Frage nach der Klarheit der Instruktion die 0,8er-Grenze. Zu leicht fallen im Vorfragebogen lediglich die Items nach der empfundenen Kompetenz der Touchscreenbedienung und der Häufigkeit der Internetnutzung zum Vergnügen aus. Im Nachfragebogen betrifft dies ebenfalls die Kompetenz der Bedienung des Touchscreens und des Lernprogramms, wie auch die Verständlichkeit der Texte, die in der Lernumgebung vorkommen. Mit Hinblick auf die Hauptstudie werden die Items so belassen. Die Klarheit der Instruktion und der Bedienung von Programm und Touchscreen könnte nach Ansicht des Autors durchaus Alterseffekten unterliegen, sodass evtl. geringere bzw. höhere Zustimmungswerte zu erwarten sein könnten. Die Frage zur Häufigkeit des Simulationseinsatzes wird belassen, da die niedrige Zustimmung zugunsten der so gewonnenen Information in Kauf genommen wird (und gewissermaßen erwartbar ist).

Weitere Aspekte, die die Auswahl von Items oder Konstrukten für die Hauptstudie betreffen finden sich im folgenden Abschnitt, in dem auch inhaltliche Ergebnisse dargestellt werden.



(a) Mittelwerte des Vorfragebogens der Pilotierung (b) Mittelwerte des Nachfragebogens der Pilotierung

Abbildung 11.10.: Mittelwerte aller Items aus beiden Fragebögen der Pilotierung

Ergebnisse

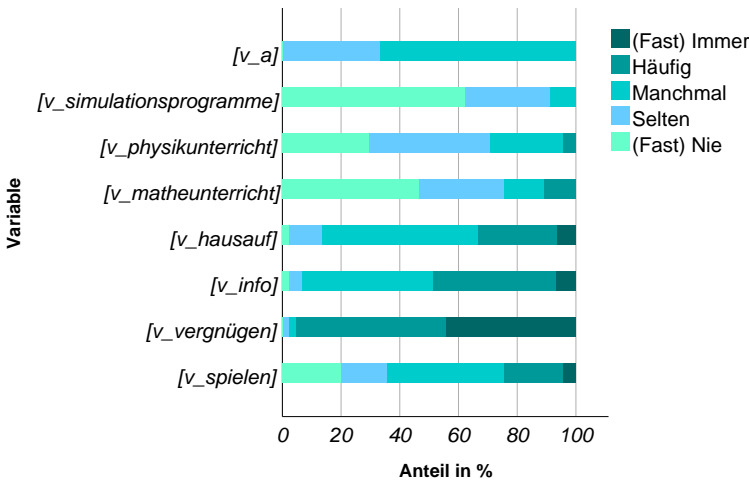


Abbildung 11.11.: Nutzung digitaler Medien.

Nutzung von Medien Die Ergebnisse zur Nutzung von Medien im privaten und schulischen Bereich sollen hier nur entsprechend knapp dargestellt werden.

Im privaten Bereich nutzen 91,1% der SuS digitale Medien zum Vergnügen (z. B. Facebook, Youtube) (siehe Abb. 11.11, v_vergnügen). Zur Hausaufgabenenerledigung und Informationsbeschaffung zeigen sich ähnliche Nutzungsanteile, wobei die beiden letzten Kategorien (Nie/selten) nur von 8 – 13% angekreuzt wurden. Im Bereich der spielerischen Nutzung geben immerhin 20% an dies nie zu tun, jedoch zeigt sich hier ein signifikanter und großer Geschlechtereffekt ($z(45) = -3,061; p = .002, r = .45$), wobei die Jungen häufiger spielen als die Mädchen.

Im schulischen Bereich werden für 70% der Schüler neue Medien (fast) nie oder selten eingesetzt, speziell Simulationen betreffend sind dies sogar 91,1%. Beide Klassen der Vorstudie unterscheiden sich hinsichtlich des Einsatzes von Simulationen und Neuen Medien im Physikunterricht jedoch signifikant ($z(45) = -2,782; p = .005, z(44) = -2,950; p = .010$) und mit jeweils großem Effekt ($r = .41, r = .44$). Betrachtet man die Mittelwerte, sind diese jedoch jeweils beide im Bereich zwischen selten und nie angesiedelt. Im Zuge der Hauptstudie ist also durchaus ein Test auf Klasseneffekte sinnvoll.

Zu beachten ist, dass die Frage v_a nach anderweitiger Nutzung nur von 3 Schülern beantwortet wurde.

Tabelle 11.1.: Übersicht zu den Skalen der Vorstudie. Hinweis: Beim Anteil fehlender Werte werden nicht-anwesende SuS nicht berücksichtigt (siehe 10.1.1).

Frage	$\alpha_{Cronbach}$	Trennschärfe	N _{gültig}	Fehlend _{Anteil}
Pre Selbstkonzept	.908	.811 – .827	44	2,2%
Post Selbstkonzept	.058	.030 – .030	38	2,5%
Pre Interesse	.772	.637 – .637	44	2,2%
Post Interesse	.762	.550 – .667	38	2,5%
Pre Flow	.583	.334 – .434	45	0%
Post Flow	.579	.320 – .507	39	0%
Pre Fehlerkultur	.825	.635 – .680	45	4,4%
Post Fehlerkultur	.176	.320 – .507	39	0%

Selbstkonzept Die Kennwerte zur Skala im **Vorfragebogen** liegen im sehr guten Bereich (siehe Tabelle 11.1). Das Selbstkonzept erreicht für alle SuS im Vorfragebogen einen Mittelwert von 2,35. Signifikante Unterschiede zwischen den einzelnen Zugängen respektive Lerngruppen (in der Pilotierung identisch) lassen sich mittels MWU-Test nicht identifizieren. Es zeigt sich allerdings ein signifikanter Unterschied zwischen beiden Geschlechtern $z(44) = -3,595; p < .001$, dem auch ein starker Effekt zugrunde liegt ($r = .54$). Dieser fällt zugunsten der Jungen aus (vgl. auch Abb. 11.12).

Im **Nachfragebogen** hingegen sind weder Reliabilität noch Trennschärfe in irgendeiner Weise akzeptabel (siehe auch Diskussion). Wertet man diesen dennoch aus ergibt sich ein Skalenmittelwert von 2,91. Auch hier lässt sich kein signifikanter Unterschied zwischen Klassen oder Zugang feststellen. Der Unterschied innerhalb der Geschlechter lässt sich jedoch noch nachweisen $z(38) = -3,8; p < .001$.

Im **Vergleich**³ zwischen Vor- und Nachfragebogen ist ein signifikanter Unterschied bezüglich des Skalenmittelwertes festzustellen $z(74) = -2,968; p = .003$. Dies entspricht noch einem mittleren Effekt bzw. knapp keinem starken ($r = .35$). In den Untergruppen Zugang, Feedback und

³Der Einfachheit halber wird in diesem Kapitel von Vor- und Nach- bzw. Pre- und Post-Konstrukten geschrieben. Korrekterweise ist natürlich das Konstrukt im Bezug auf den vorigen Unterricht (Pre, Vor) im Vergleich zum Konstrukt im Bezug auf die Intervention (Post, Nach) gemeint!

Geschlecht zeigen sich durchweg mittlere bis starke Effekte (den größten Effekt und Zuwachs verzeichnen dabei die Mädchen) (vgl. jeweils Abb. 11.12). Die Zuwächse zwischen den Untergruppen unterscheiden sich jedoch rechnerisch tatsächlich *nicht* signifikant, sodass beispielsweise keine signifikanten Unterschiede zwischen Mädchen und Jungen bestehen. Auf eine detaillierte rechnerische Betrachtung von Interaktion zwischen Geschlecht oder Feedback und Zugang wurde bewusst verzichtet, um keine zu kleine Population zu betrachten. Der Vollständigkeit halber sind diese aber in den jeweiligen Grafiken mit aufgeführt, siehe hierzu Abb. 11.12.

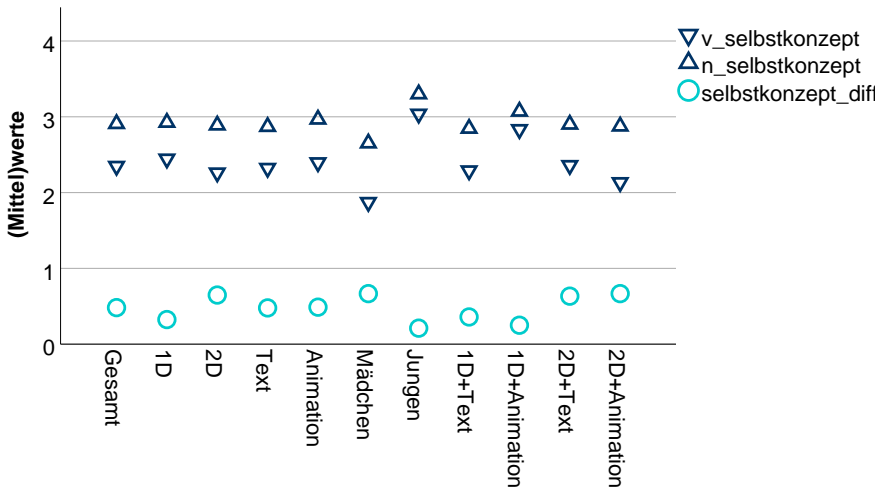


Abbildung 11.12.: Vergleich der Skalen(mittel)werte zum Selbstkonzept

Interesse Die Kennwerte zur Skala im **Vorfragebogen** hinsichtlich Trennschärfe und Reliabilität liegen im akzeptablen bis guten Bereich. Das Interesse erreicht für alle SuS im Vorfragebogen einen Mittelwert von 1,42. Signifikante Unterschiede zwischen den einzelnen Zugängen respektive Lerngruppen (in der Vorstudie identisch) lassen sich mittels MWU-Test nicht identifizieren. Zwischen beiden Geschlechtern hingegen bestehen signifikante Unterschiede $z(44) = -2,238; p = .025$, was einem mittleren Effekt entspricht $r = .34$ (vgl. auch Abb. 11.13). Die Jungen zeigen dabei ein größeres Interesse als die Mädchen (1,85 ggü. 1,15). *Unbedingt* zu berücksichtigen ist, dass die Skala im Vortest hier versehentlich nur aus zwei Items gebildet wurde. In der Hauptstudie wird das unterschlagene Item selbstverständlich wieder mit aufgenommen.

Die Kennwerte des Interessenkonstrukts im **Nachfragebogen** liegen durchweg im guten Bereich. Es wird ein Skalenmittelwert von 2,59 erreicht. Wie schon im Vorfragebogen lassen sich auch hier keine signifikanten Unterschiede zwischen Klassen oder Zugang identifizieren. Der signifikante Unterschied innerhalb der beiden Geschlechter, wie er noch im Vorfragebogen auftrat, lässt sich hier nicht mehr nachweisen. Hinsichtlich einer noch besseren Vergleichbarkeit des Interesses in Pre- und Postfragebogen wird für die Hauptstudie eine bessere Angleichung des Item-Textes im Nachfragebogen vorgenommen.

Im **Vergleich** zwischen Vor- und Nachfragebogen ist ein signifikanter Unterschied bezüglich des Skalenmittelwertes festzustellen ($z(74) = 4,619; p < .001$). Dies entspricht einem großen Effekt ($r = .54$). In den Untergruppen Geschlecht, Zugang und Feedback zeigen sich durchweg eher große Effekte im Vergleich zwischen Pre- und Postfragen, die in den meisten Untergruppen ähnlich gelagert sind (vgl. jeweils Abb. 11.13). Ein deutlicher Unterschied zeigt sich zwischen Jungen und Mädchen. Bei ersteren ist der Effekt deutlich schwächer ausgeprägt (wenn auch an sich nach wie vor groß). Die Mädchen zeigen damit die deutlichste Differenz/Zuwachs aller Untergruppen, der aber im Vergleich zu dem der Jungen hier nicht signifikant ist ($p=.057$).

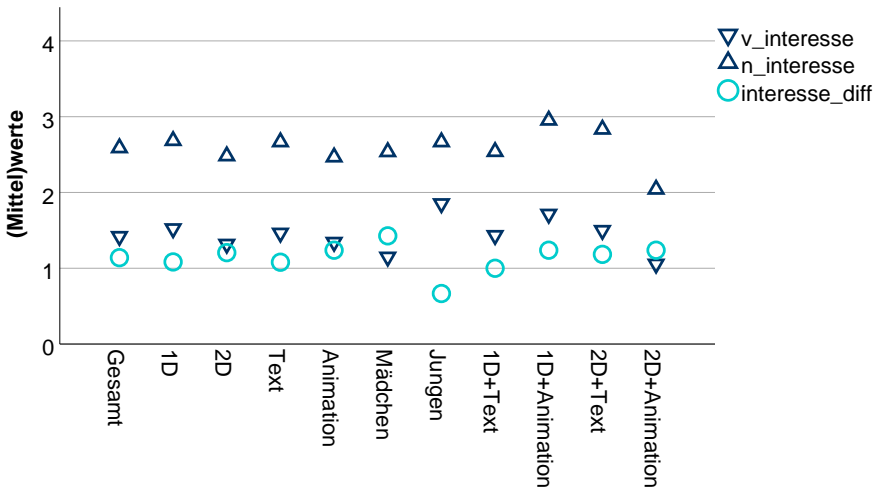


Abbildung 11.13.: Vergleich der Skalen(mittel)werte zum Interesse

Flow Die Kennwerte zur Skala im **Vorfragebogen** liegen im noch akzeptablen Bereich, wenn auch die Werte der inneren Konsistenz der Skala

eher (!) schlecht sind. Der Flow erreicht für alle SuS im Vorfragebogen einen Mittelwert von 1,99. Signifikante Unterschiede zwischen den einzelnen Zugängen respektive Lerngruppen (in der Pilotierung identisch) lassen sich mittels MWU-Test nicht identifizieren. Auch zwischen beiden Geschlechtern bestehen keine signifikanten Unterschiede (vgl. auch Abb. 11.14).

Der Flow erreicht für alle SuS im **Nachtest** einen Wert von 2,52. Wie schon im Vorfragebogen lassen sich auch hier keine signifikanten Unterschiede zwischen Klassen oder Zugang feststellen. Auch der im Vorfeld bestehende Unterschied innerhalb der Geschlechter lässt sich nicht mehr nachweisen.

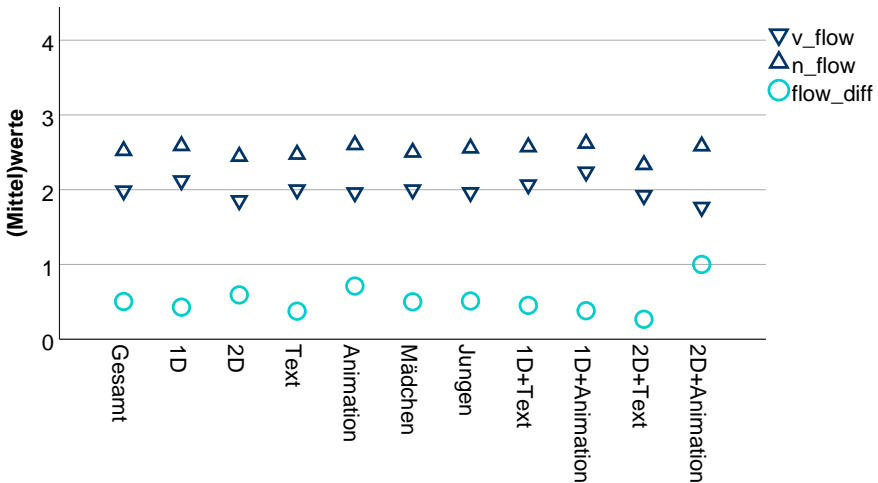


Abbildung 11.14.: Vergleich der Skalen(mittel)werte zum Flow

Im **Vergleich** zwischen Vor- und Nachfragebogen ist ein signifikanter Unterschied bezüglich des Skalenmittelwertes festzustellen $z(78) = -3,618; p < .001$. Dies entspricht einem großen Effekt ($r = .41$). Innerhalb der Untergruppen zeigen sich durchweg mittlere Effekte, die allen Untergruppen ähnlich gelagert sind (vgl. jeweils Abb. 11.12). Zwischen den einfachen Untergruppen bestehen keine signifikanten Unterschiede im Zuwachs/Differenz.

Fehlerkultur Die Kennwerte zur Skala der Fehlerkultur liegen im **Vorfragebogen** im sehr guten Bereich. Es wird ein Skalenmittelwert von 3,00 erreicht. Untergruppenunterschiede bestehen lediglich zwischen den

Geschlechtern ($z(43) = -2,005; p = .045$), wobei hier die Mädchen die Unterstützung der Lehrkraft im Fehlerfalle höher einschätzen (MW: 3,22) als die Jungen (MW: 2,69). Dies entspricht einem mittleren Effekt ($r = .31$).

Die Kennwerte zur Skala im **Nachfragebogen** liegen im schlechten Bereich. Besonders die interne Konsistenz ist auch nach Ausschluss einzelner Items in keinsten Weise akzeptabel. Der Vollständigkeit halber soll dennoch der Skalenmittelwert von 1,25 (bzw. 2,75 umgepolt) erwähnt werden.

Ein **Vergleich** zwischen Vor- und Nachtest wie auch der Zuwächse findet aus nachvollziehbaren Gründen nicht statt.

Weitere Items Abbildung 11.15 gibt einen Überblick über die Items des **Vorfragebogens**, die keiner Skala zugeordnet sind. Hier zeigt sich deutlich, dass sich die Schüler bei der Bedienung von Touchscreens als sehr kompetent einschätzen, da 88,9% der entsprechenden Frage voll oder eher zustimmen. Immerhin noch 75% der SuS geben an, auch in Computerspielen einfach die interaktiven und bedienbaren Elemente identifizieren zu können, wobei Mädchen hier signifikant größere Probleme haben ($z(44) = -3,030; p = .002; r = .46$).

Abbildung 11.16 gibt einen Überblick über die Items des **Nachfragebogens**, die keiner Skala zugeordnet sind. Dazu gehören beispielsweise Items zur Bedienung und Akzeptanz der Lernumgebung. Relevante Unterschiede in den Untergruppen werden nur erwähnt, wenn diese bezogen auf einen Vergleich mittels MWU-Test signifikant erscheinen.

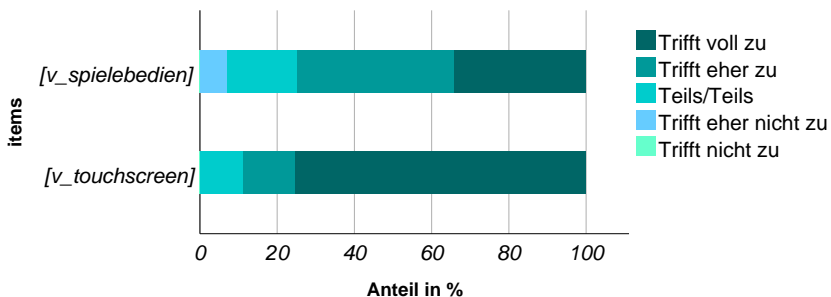


Abbildung 11.15.: Überblick über die weiteren Items im Vorfragebogen

Die Texte der Lernumgebung sind für die überwiegende Mehrheit (92,3%) der Befragten (eher) verständlich⁴, nur 7,7% zeigen sich unentschlossen. Die Inhalte der Lernumgebung werden von nur sehr wenigen

⁴Mit (eher) ist die Zusammenfassung beider Items höchster Zustimmung gemeint

V. Ergebnisse

Schülerinnen und Schülern als voll „spannend“ beschrieben. Jeweils 43,6% bzw. 41,0% stimmen dieser Aussage aber eher zu bzw. stimmten neutral ab. 12,8% lehnen die Aussage (eher) ab.

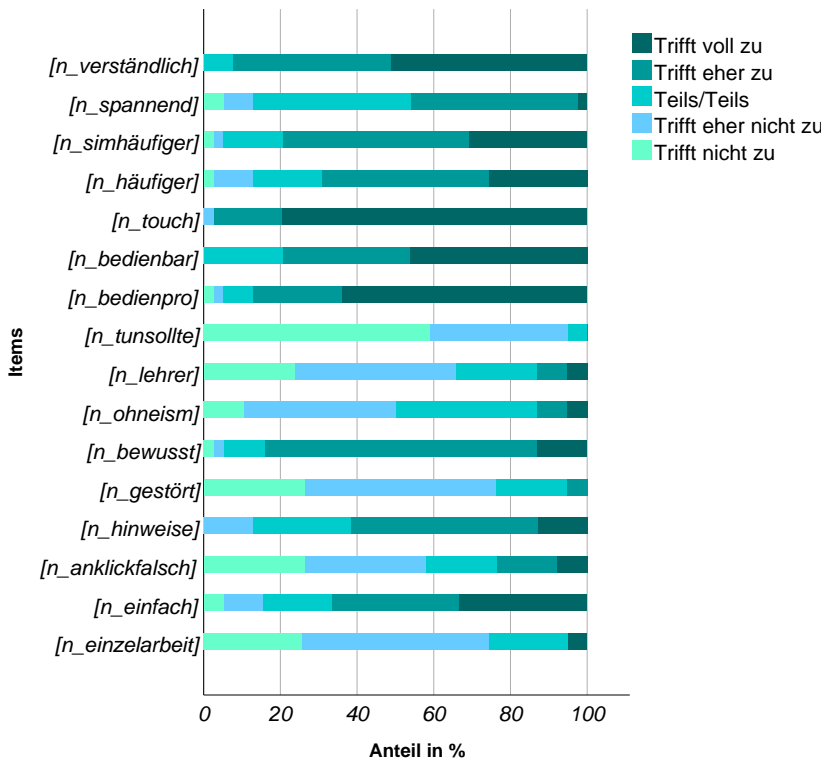


Abbildung 11.16.: Überblick über die weiteren Items im Nachfragebogen.

79,5% der SuS geben an, dass sie Simulationen wie die genutzten (eher) häufiger im Unterricht einsetzen möchten. Lediglich 5,2% sprechen sich (eher) dagegen aus. Ähnlich verhält es sich bezüglich des Lernprogramms, welches von 69,2% der Befragten (eher) in ähnlicher Form wieder gerne genutzt werden würde.

Bezüglich der Touchscreenbedienung von Lernprogramm und Simulationen geben 79,5% an, diese problemlos zu beherrschen, weitere 17,9% stimmen dem (eher) zu (Deckeneffekt, siehe Qualitätssicherung). Ähnlich kompetent bewerten alle SuS ihr Vermögen, bei den Simulationen die interaktiven Elemente zu identifizieren. Bei diesem Item liegt, anders als im Vorfragebogen, kein signifikanter Unterschied zwischen den Geschlechtern vor. 79,5% geben an, dass dies grundsätzlich (auch eher) problemlos

gelänge. Die Übrigen schätzen sich diesbezüglich neutral ein. Auch die Bedienung des Lernprogramm an sich – vor, zurück, speichern etc. – bereite 87,2% der SuS (eher) keine Schwierigkeiten. Bezüglich des zielgerichteten Umgangs mit den Simulationen wurde erfragt, inwiefern Klarheit für den Umgang mit den Simulationen herrscht. Die überwiegende Mehrheit (94,9%) stimmt dahingehend der Aussage „nicht zu wissen, was sie tun sollten“ (eher) *nicht* zu. Auch bei diesem Item liegt knapp ein signifikanter Unterschied $z(41) = -2,279$; $p = .049$ mit einem mittleren Effekt ($r = .35$) vor; hier zwischen den Klassen bzw. Zugängen. Dennoch liegen beide Gruppen bzgl. des Skalenmittelwerts in einem sehr niedrigen Bereich (1D: 0,67; 2D: 0,22).

Im Bezug auf das Lernen wurde erfragt, inwiefern den SuS das Lernen auch nur mit Lernprogramm – ohne Simulationen – gelungen wäre. Der Aussage steht eine grundsätzlich eher ablehnende Haltung der SuS gegenüber. 50% der Befragten geben diese (eher) zum Lernen ohne Simulationen an. Ein Bewusstsein für den Lernerfolg wurde bei 84,3% der SuS (eher) erreicht. Nur 5,2% der SuS ist (eher) unklar, was sie mithilfe des Lernprogramms gelernt haben.

Da die Skala zur Fehlerkultur im Nachfragebogen nicht in diesem Sinne ausgewertet werden konnte, sollen hier die Items knapp einzeln dargestellt werden. Im Bezug auf das Feedback geben 76,3% der Befragten an, durch die Hinweise oder Animationen nicht oder eher nicht abgelenkt/gestört worden zu sein. Niemand empfand sie tatsächlich als störend, nur von knapp 5,3% wurden sie *eher* als störend empfunden. Damit harmonisierend geben 61,5% der SuS an, die Hinweise oder Animationen (eher) als hilfreich empfunden zu haben. Etwa ein Viertel steht dem unentschlossen gegenüber, während 12,8% die Angebote eher nicht als Hilfe empfanden. Wurden MC-Fragen – auf die sich das Feedback letztlich bezieht – nicht korrekt beantwortet, war die Mehrheit von 57,9% der SuS *nicht* (eher) enttäuscht. Immerhin 23,7% der SuS hingegen zeigten sich in diesem Falle aber (eher) sehr enttäuscht.

Bezüglich der Arbeitsform *Einzelarbeit* überwiegen mit 74,3% Meinungen der SuS, die Einzelarbeit (eher) nicht als „anstrengend“ empfanden. Nur 5,1% hingegen stimmen dem voll zu (Man beachte, dass es sich dabei lediglich um 2 Schüler/innen handelt!), während die übrigen Schüler neutral abstimmten. Letztlich wurde das Lernprogramm von einem Drittel der Schüler als zu einfach empfunden, ein weiteres Drittel stimmt dieser Aussage (eher) zu.

Offene Frage Insgesamt wurden in der Pilotstudie 14 Kommentare durch die SuS abgegeben. An dieser Stelle scheint es – auch aufgrund der geringen Anzahl der Kommentare – zweckmäßig, direkt eine Kate-

gorisierung vorzunehmen. Auch das Ziel der Vorstudie, dem Schaffen eines Überblicks spricht für diese eher pragmatische Vorgehensweise. Aus Gründen der Transparenz sind die Originalantworten in Tabelle 11.2 mit ihrer jeweiligen Zuordnung aufgeführt. Kommentare, die in mehrere Kategorien eingeordnet werden könnten, wurden entsprechend gesplittet oder vollständig in beide Kategorien eingeordnet.

Tabelle 11.2.: Übersicht der freien Schülerkommentare der Vorstudie

Kommentar	Kategorie
Gut und nachvollziehbare Simulationen, die passend zum Thema sind Sehr ausführlich vllt zu ausführlich erklärt, sodass man weiter klickt; strukturiert, übersichtlich Ich habe Spaß gehabt, die Aufgaben zu bearbeiten und durch die guten Veranschaulichungen habe ich auch viel gelernt. Ich möchte ein Lernprogramm für meine Schule, weil man mehr selber lernt. Es ist interessanter als Schule Sie können gerne wiederkommen.	Spezifische positive Anmerkungen
Alles andere, Daumen hoch! (Anm: Manches aber nicht spannend) Hat Spaß gemacht. Gruß an Harald Lesch war besser als die vom letzten Jahr	Unspezifisch positive Kommentare
Man sollte besser nebendran ein Audio haben, welches die Texte vorliest (sic!), da die Konzentration schnell abnimmt. Man könnte vielleicht noch Musik einbauen. Einige Dinge könnten noch ein wenig spannender bzw. umfangreicher gestaltet werden.	Änderungsvorschläge
Viele der Inhalte waren nicht neu. Gefühl habe ich nichts neues gelernt. Von Kreisbewegung gab es im ersten Durchlauf nichts. Es war sehr kindlich dargestellt, trotzdem hat es Spaß gemacht die Aufgaben mit den Simulationen zu lösen. Zu viele Rechtschreibfehler	Spezifische negative Anmerkungen
Die Bedienung der Simulationen war nicht immer ohne Probleme. Anstrengend die ganze Zeit auf den Bildschirm zu achten.	Technische Aspekte

Die Aussagen lassen sich den Kategorien *Spezifische positive Anmerkungen*, *Unspezifische positive Anmerkungen*, *spezifische negative Anmerkungen*, *Änderungsvorschläge* sowie *Technische Aspekte* zuordnen. Angesichts der kleinen Stichprobe können die Antworten nur als Indikatoren für die (Nicht)akzeptanz des Lernprogramms sein. In diesem Rahmen sind keine Kommentare zu verzeichnen, die schwerwiegende Mängel vermuten lassen. Einzelne Nennungen betreffen eine nicht immer problemlose Simu-

lationsbedienung, wie die eher „kindliche Darstellung“, Rechtschreibfehler oder die Themenauswahl (viel sei schon bekannt). In Übereinstimmung mit den Items der Fragebögen besteht aber auch nach der Auswertung der offenen Frage neben redaktionellen Änderungen kein akuter Handlungsbedarf für die Hauptstudie.

Diskussion & Zusammenfassung

Die Vorstudie in zwei 10. Klassen bestätigte, wie schon die Explorationsstudie, die grundsätzliche Tauglichkeit der Lernumgebung im Rahmen des Forschungsvorhabens.

Wesentliche Änderungen zur Hauptstudie: Wie bereits erwähnt wurde in der Vorstudie aus Gründen der Passung zu Alter und Vorwissen eine leicht veränderte Version der Lernumgebung eingesetzt. In der Hauptstudie wurde daher die ursprüngliche Variante eingesetzt. Als bedeutende Änderung auf struktureller Ebene wurde zur Hauptstudie die Gestaltung des Feedbacks geändert. Da beobachtet wurde, dass Schüler bisweilen lediglich die „Richtig-Falsch-Information“ betrachteten, ohne auf das elaborierte Feedback einzugehen, schien diese Änderung nötig. Letztlich wurde aufgrund dieser Beobachtung im später verwendeten Lernprogramm davon Abstand genommen, solch ein General Feedback anzugeben. Die Richtigkeit der Antwort lässt sich mit der verwendeten Anordnung der Elemente und dem Verzicht auf farbliche Markierungen nun prinzipiell erst durch tatsächliches Lesen des jeweiligen Hinweises feststellen.

Die Fragebögen zeigten eine erfreuliche Schwierigkeit, wobei nur vereinzelte Items als zu leicht oder zu schwer eingestuft werden mussten. Da jeweils Argumente für den Verbleib der Items sprachen, wurden diese im Falle des Vorfragebogens übernommen. Im Nachfragebogen ergeben sich neben kleineren Anpassungen jedoch deutliche Änderungen. Vor allem betrifft dies die Items zu den Skalen *Selbstkonzept* und *Fehlerkultur*. Im ersten Fall werden die Items aufgrund schlechter Skalenkennwerte durch neue Items, die stärker denen des Pretests entsprechen, ersetzt. Im zweiten Fall ist diese Art der Itemformulierung nicht ohne weiteres möglich gewesen, weshalb für den Nachtest neue Items geschaffen wurden. Aufgrund schlechter Vergleichbarkeit wird daher in der Hauptstudie keine Auswertung der Skala Fehlerkultur vorgenommen, sondern eine Auswertung auf Itemebene. Die übrigen Fragen des Nachtests konnten das Erkenntnisinteresse befriedigen.

Inhaltliche Betrachtung: Bei Auswertung der Konstrukte Interesse, Selbstkonzept und Flow ist zu beachten, dass sich die Konstrukte auf unterschiedliche Zeiträume beziehen – d. h. streng genommen ist jeweils beispielsweise das Interesse bezogen auf den Vorunterricht (Vorfragebogen) und das Interesse bezogen auf die Intervention (Nachfragebogen). In

diesem Zusammenhang ist ebenso zu beachten, dass es sich speziell bei Interesse und Selbstkonzept um stabile Konstrukte handelt, die durch eine recht kurze Intervention wie diese nicht dauerhaft verändert werden können. Umso mehr ist zu beachten, dass lediglich die Ausprägung eines Konstrukts im Bezug auf den Vorunterricht und auf die Intervention erhoben wird.

Im Vergleich zum vorigen Unterricht ist bezogen auf die Intervention ein deutlicher Anstieg des Interesses zu verzeichnen. Bezogen auf alle Schüler wie auch innerhalb der Untergruppen Geschlecht, Feedback und Zugang sind dabei ähnliche Zuwächse zu verzeichnen. Auffällig ist jedoch, dass die im Vorfragebogen gemessenen Unterschiede zwischen Mädchen und Jungen bezogen auf die Intervention nicht mehr festzustellen waren. Aufgrund des relativ seltenen Einsatzes digitaler Medien in den befragten Klassen können diese Zuwächse mitunter mit Neuigkeitseffekten (siehe Kapitel 4.2.1) in Bezug auf Nutzung von Tablets in einem ungewohnten Unterrichtssetting kofundieren.

Hinsichtlich des Selbstkonzepts sind im Mittel ebenfalls Zuwächse zu verzeichnen, wenn auch in geringerem Maße als noch beim Interesse⁵. Die Schülerinnen und Schüler zeigen keine Unterschiede hinsichtlich des Zugangs oder des Feedbacks. Auffällig ist jedoch, dass die Mädchen einen erheblichen Zuwachs des interventionsbezogenen Selbstkonzepts erfahren, die Jungen dagegen nur einen relativ kleinen Zuwachs. Dennoch erreicht das Niveau im *Postfragebogen* der Mädchen nicht das Niveau der Jungen im *Prefragebogen*.

Das Flowerleben zeigt im Vergleich zum Vorinteresse wieder in allen Untergruppen moderate und signifikante Zuwächse, die sich aber zwischen den Gruppen kaum unterscheiden. Speziell hier sind auch keine Unterschiede zwischen Mädchen und Jungen zu erwarten.

Das Feedback durch Text oder Animation wurde – durch vier einzelne Items erhoben – von über der Hälfte der Schüler als hilfreich bzw. eher hilfreich empfunden. Jedoch lässt sich kein Unterschied zwischen den Feedbackvarianten feststellen. Dies ist erstmal hinzunehmen, zumal auch keine Hypothese hinsichtlich eines Unterschieds besteht. Evtl. zeigt sich im Rahmen der Hauptstudie und größerem Stichprobenumfang ein Unterschied in der Nutzung. Zusätzlich sind in der Hauptstudie Interviews geplant, die weiteren Aufschluss über das Nutzen des Feedbacks geben können. Evtl. könnte sich die Einstellung zum Feedback auch dadurch ändern, dass die Verarbeitungstiefe des Feedbacks durch bereits geschilderte Maßnahmen erhöht werden soll. 40% der Schülerinnen und Schüler zeigen sich durch „falsche“ Antworten teilweise, eher oder voll

⁵Achtung! Dies kann nur ein Anhaltspunkt sein, da die Skalen prinzipiell *nicht* vergleichbar sind!!

enttäuscht. Das ist vielleicht nachvollziehbar, entspricht aber nicht dem formativen Charakter des Feedbacks innerhalb einer *Lernsituation*. Besonderes Augenmerk ist in der Hauptstudie daher bei der anfänglichen Instruktion darauf zu achten, den helfenden (und nicht den vermeintlich bewertenden!) Charakter der Multiple-Choice-Fragen herauszustellen!

Die Bedienung und Steuerung der Lernumgebung und Simulationen bereitete den Schülern kaum Probleme. Prinzipiell kann man dies durch informelle Beobachtungen aus dem Unterricht stützen. Jedoch ist zu berücksichtigen, dass stets Hilfskräfte anwesend waren, die bei Problemen weiterhelfen konnten. Andererseits geben die Schüler eine – erwartungsgemäß – große Kompetenz in der Bedienung von Touchscreens an. Dazu ist wiederum anzumerken, dass nicht alle Schülerinnen und Schüler die Bedienung per Touch nutzten. Dringende Änderungen hin zur Hauptstudie scheinen aber nicht angebracht. Gleiches gilt für die Verständlichkeit von Texten und Forschungsaufträgen zu Simulationen, die jeweils als hoch eingeschätzt wurden. Hier ist jedoch zu beachten, dass die Schülerinnen und Schüler der Hauptstudie etwa 2 – 3 Jahre jünger sein werden, als die der Vorstudie.

Aus den Antworten zum offenen Aufgabenformat ergeben sich keine Hinweise auf dringende Änderungen zur Hauptstudie hin.

11.2.3. Fachwissenstest

Allgemeine Anmerkungen zu den Items Die Darstellung des Fachwissenstest erfolgt relativ kompakt mit Blick auf die Hauptstudie. Der Fachwissenstest umfasste 24 Items⁶. Die relativen Lösungshäufigkeiten der Items in Pre- und Posttest weisen darauf hin, dass die Items unterschiedlich schwierig sind. Die Bandbreite erstreckt sich dabei von Items, die nahezu immer korrekt bis hin zu Items, die tatsächlich (fast) nie korrekt beantwortet wurden. Dies ist für die Auswahl eines künftigen Tests, der zwischen kompetenten und weniger kompetenten SuS unterscheiden soll, von Vorteil.

Die schwierigsten Items in Vor- und Nachtest (rel. Lösungshäufigkeit kleiner 0,2) betreffen inhaltlich die richtungsändernde Eigenschaft einer Kraftwirkung wie auch die Kreisbewegung (welche in den betroffenen Klassen zum fraglichen Zeitpunkt auch noch nicht unterrichtet wurde). Die einfachsten Items (rel. Lösungshäufigkeit deutlich größer 0,8) betreffen inhaltlich z. B. die Interpretation von einfachen t-x-Diagrammen wie

⁶Drei weitere Items wurden nicht ausgewertet, da diese quasi nicht bearbeitet wurden bzw. sich der Auswerteaufwand der offenen Fragen als kaum für eine Vorstudie ökonomisch erwies.

auch den geschwindigkeitsbetragändernden Aspekt der Kraftwirkung (vgl. Abb. 11.17).

Auffällig ist jedoch, dass sich die Lösungswahrscheinlichkeiten der einzelnen Aufgaben zwischen Pre- und Posttest kaum deutlich unterscheiden. Präzisere Untersuchungen bezüglich des Lernzuwachses werden mittels der Raschanalyse über die Personenfähigkeit hergestellt.

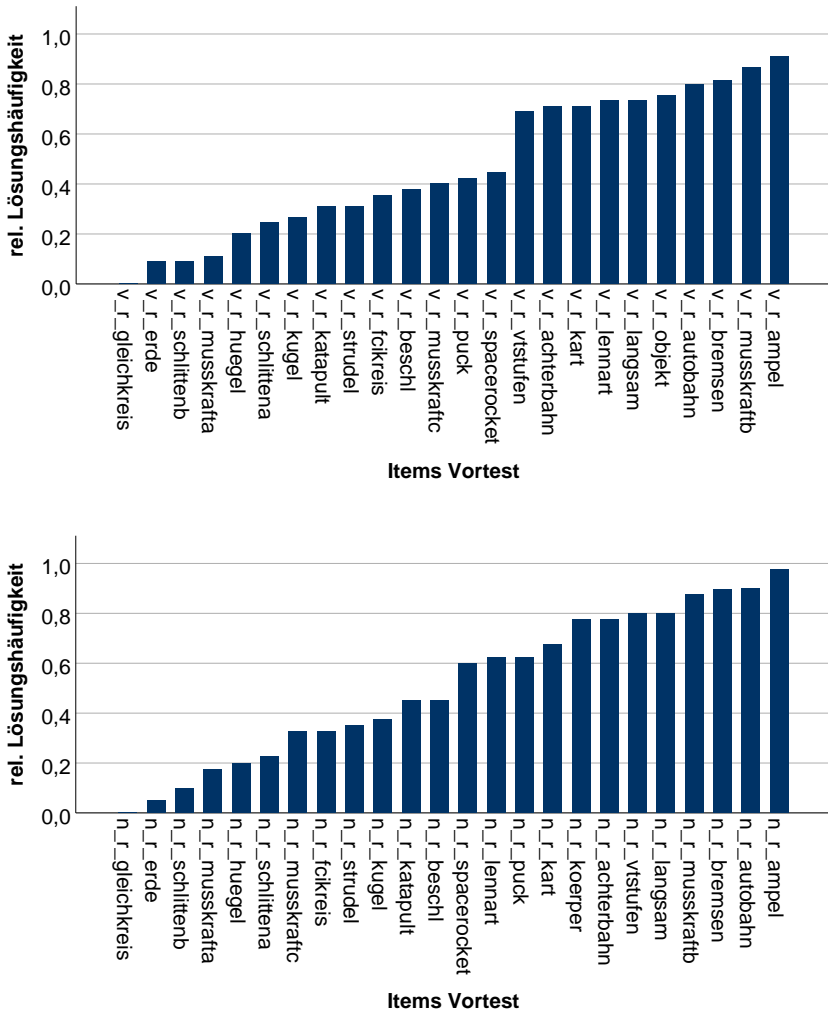


Abbildung 11.17.: Überblick über die relativen Lösungshäufigkeiten im Vor- und Nachttest (oben bzw. unten)

Raschanalyse Im Bereich der Items weist der **Pretest** ein $\alpha_{Cronbach}$ von 0,55 auf. Dies scheint im Falle eines Vortests noch ebenso akzeptabel. Eine angemessene Eindimensionalität scheint nach Simulation von Daten und anschließendem Vergleich gegeben (vgl. Kapitel 10.3). Eine Analyse der Items hinsichtlich der Infitwerte (vgl. Kapitel 10.1) erfolgte ebenso. Außerhalb des allgemein anerkannten Bereichs⁷ von 0.8 bis 1.2 befinden sich keine Items. *Hinweis: Das Item v_gleichkreis wurde aus der Analyse entfernt, da es von keinem Schüler in Vor- und Nachtest korrekt gelöst wurde und somit keine nützlichen Informationen enthält.*

Im **Posttest** zeigt eine deutlich verbesserte Reliabilität, ein verbessertes $\alpha_{Cronbach}$ von 0,67. Kritisch zu betrachten ist hier die Eindimensionalität des Konstrukt, da die PCA im Vergleich zum Vortest Indizien für eine weitere Komponente liefert. Die Analyse wird dennoch durchgeführt, da die Verletzung zum einen nicht sehr schwerwiegend ist und zum anderen die geringe Anzahl der Teilnehmer der Vorstudie die Aussagekraft ohnehin einschränkt. Die Analyse der Items hinsichtlich der Infits zeigt sich unkritisch. Im Bereich der Personen zeigen sich nach Abklärung aller Kennwerte Outfit-Werte (für die im Allgemeinen ein Bereich von 0.5 – 1.5/2 angenommen wird, vgl. Kapitel 10.3) nur 2 SuS, die diese Grenze knapp unter- bzw. überschreiten. Da jedoch die ZSTD-Werte im angemessenen Rahmen sind und auch Crossplots der Itemschwierigkeit und Personenfähigkeit nach Ausschluss von Personen/Items keinen Einfluss auf die Messung zeigte, werden die Personen in der Analyse belassen. Im Nachtest können nach gleichem Muster ebenso alle Personen in der Analyse verbleiben (vgl. Kapitel 10.3).

Als Ankeritems konnten nach grafischer Analyse wie Betrachtung der Displacementwerte alle Items außer n_lennart, n_erde und n_ampel verwendet werden.

⁷Ohne dass dieser eine unumstößliche Gesetzmäßigkeit wäre! (vgl. Neumann (2014))

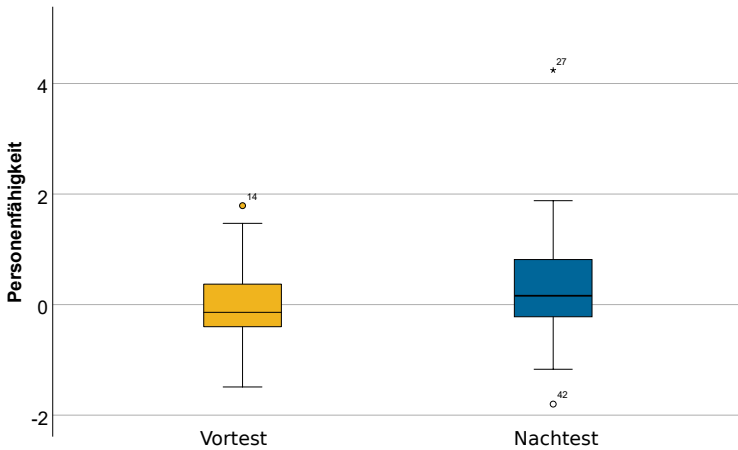


Abbildung 11.18.: Überblick der Personenfähigkeiten aller SuS

Abbildung 11.18 gibt einen Überblick über die Personenfähigkeiten aller Schüler im Vergleich von Pre- zu Posttest. So ist der Unterschied beider Testzeitpunkte signifikant $z(40) = -3,154$; $p = .002$. Dies entspricht einem mittleren Effekt von $r = .35$. Eine Betrachtung unterschiedlicher Untergruppen im nächsten Kapitel kann womöglich näheren Aufschluss geben.

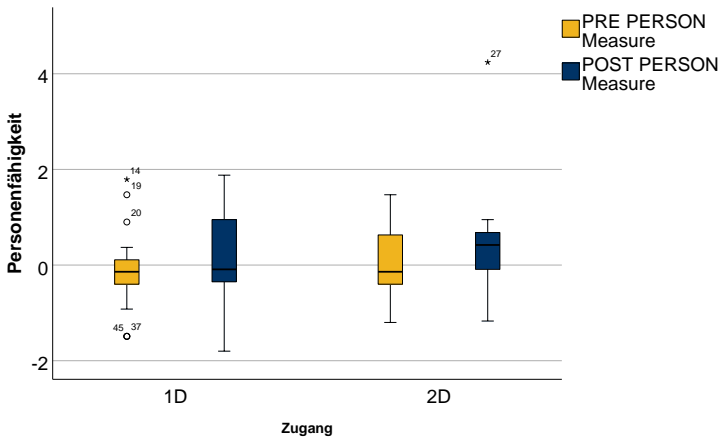


Abbildung 11.19.: Darstellung der Personenfähigkeiten in der 1D- und 2D-Gruppe (links bzw. rechts)

Formal zeigt Abbildung 11.19 die Personenfähigkeit nach Zugang getrennt. Faktisch wird es kaum möglich sein, aufgrund *dieses* Merkmals Rückschlüsse zu ziehen. Da je einer Klasse ein Zugang zugeteilt wurde, entsprechen die in der Abbildung dargestellten Werte auch gleichzeitig einem Klasseneffekt. Letzteres ließe sich lediglich durch größere Stichproben abmildern. Allerdings lassen sich auch hier keine signifikanten Unterschiede zwischen den Gruppen innerhalb des Pre- und Posttests feststellen. Sowohl das Vorwissen als auch die Ergebnisse des Nachtests unterscheiden sich nicht deutlich. Innerhalb der 1D-Gruppe ist jedoch ein signifikanter Zuwachs von Pre- zum Posttest zu verzeichnen ($z(42) = -3,012; p = .003, r = .46$). In der 2D-Gruppe ist solch ein Zuwachs nicht nachzuweisen.

Abbildung 11.20 zeigt die Personenfähigkeit nach Aufteilung der beiden Geschlechter. Wie die Abbildung schon nahelegt, zeigen sich sowohl zwischen den Vortests als auch zwischen den Nachtests signifikante Unterschiede (Pretest: $z(90) = -2,937; p = .003$; Posttest: $z(80) = -2,799; p = .005$). Dabei liegt sowohl im Vortest als auch im Nachtest mit $r = .31$ ein mittlerer Effekt vor. Der Zuwachs der Personenfähigkeit ist innerhalb beider Gruppen signifikant. Dabei liegt bei den Mädchen ein mittlerer Effekt vor ($z(50) = -2,412; p = .016, r = .34$). Der Zuwachs vom Pre- zum Posttest ist bei den Jungen knapp nicht signifikant ($z(30) = -1,931; p = .053$). Zu beachten ist natürlich, dass die Stichprobengrößen (15 vollständige Datensätze in Vor- und Nachtest) relativ klein ist.

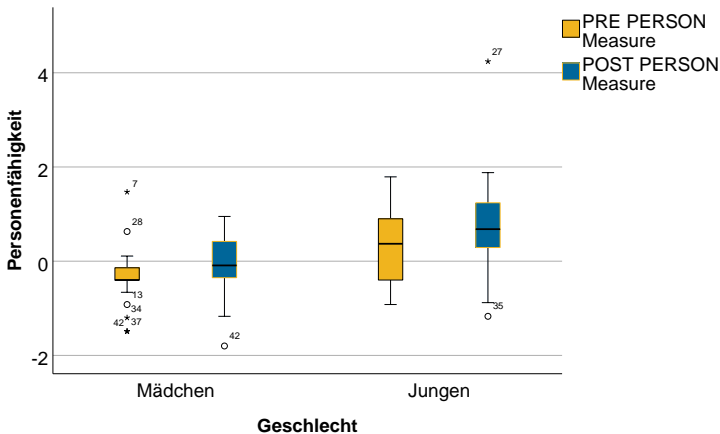


Abbildung 11.20.: Darstellung der Personenfähigkeiten von Mädchen und Jungen (links bzw. rechts)

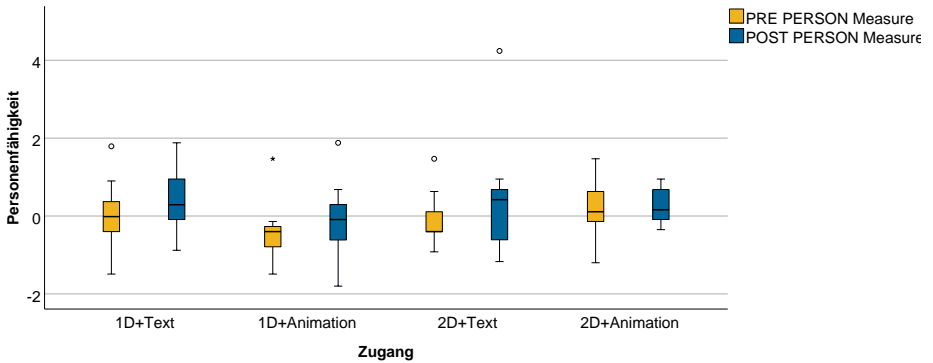


Abbildung 11.21.: Darstellung der Personenfähigkeiten je nach Versuchsbedingung

Abbildung 11.21 zeigt die Personenfähigkeiten nach Versuchsbedingung aufgeschlüsselt. Dabei zeigt sich, dass einzig in der Gruppe 1D+Text-feedback eine signifikante Änderung der Personenfähigkeit nachzuweisen ist ($z(28) = -2,608; p = .009$). Man beachte aber auch hier wiederum die geringe Größe der Stichprobe. Von einer näheren Betrachtung wird hier daher abgesehen.

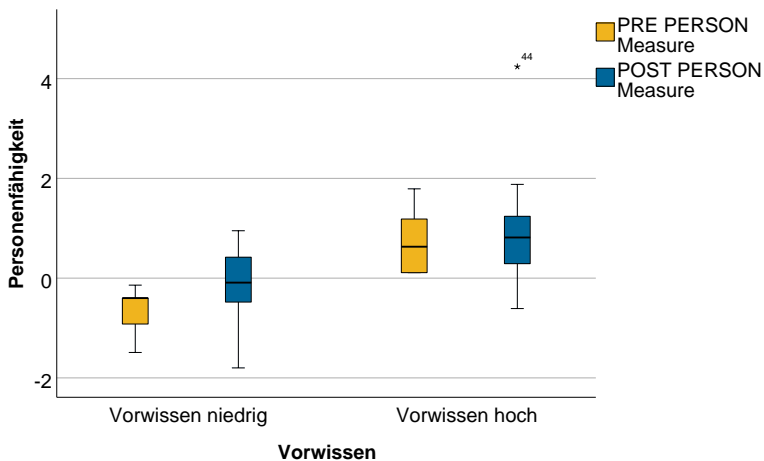


Abbildung 11.22.: Darstellung der Personenfähigkeiten von Schülern unterschiedlichen Vorwissens

Zuletzt sind in Abbildung 11.22 die Personenfähigkeiten von Schülern mit (eher) großem Vorwissen und (eher) niedrigem Vorwissen dargestellt. Zur Unterteilung wurde ein Mediansplit anhand der im Vortest ermittelten Personenfähigkeit durchgeführt.

Sowohl im Vortest als auch im Nachtest unterscheiden sich SuS unterschiedlichen Vorwissens signifikant (Vortest: $z(45) = -5,792; p < .001$; Nachtest: $z(40) = -3,027; p = .002$). SuS hohen Vorwissens können jedoch im Vergleich zu denen niedrigen Vorwissens keinen signifikanten Lernzuwachs verzeichnen (SuS Vorwissen niedrig: $z(38) = -2,980; p = .003$, $r = .48$).

Bearbeitung der Tempo-Aufgabe

Die Tempo-Aufgabe (siehe A.2.1) wurde außerhalb des raschskalierten Tests ausgewertet, da die Aufgabe für die Teilnehmer je nach Zugang unterschiedlich zu bearbeiten war bzw. unterschiedlich als korrekt gewertet wurde. Antwortoption *c* bezog sich dabei auf den korrekten Geschwindigkeitsbegriff in einer Dimension, die Optionen *a* und *d* auf die korrekte Interpretation nach 2D⁸.

Im Falle der 1D-Klasse ändert sich der Anteil von Antwort *c* kaum von 77,3% auf 71,4%, was einer Nennung weniger entspricht. Die übrigen Antworten verteilen sich auf die restlichen Optionen. Im Falle der 2D-Klasse beträgt der Anteil von Antwort *c* zunächst 87%, im Postfragebogen 36,8%. Die korrekte Antwortoptionen *a* & *d* ist mit 26,3% vertreten und Teilantwort *a* mit 15,8%.

Diskussion

Bei der Diskussion der Ergebnisse der Fachtests ist zunächst zu berücksichtigen, dass die Schüler in dieser Vorstudie selbstverständlich über schulisches Vorwissen zu den Inhalten verfügten. Darüber hinaus ist zu beachten, dass Vorstudie und Hauptstudie in unterschiedlichen Jahrgängen durchgeführt werden und daher besonders hinsichtlich des Fachtests (beispielsweise im Gegensatz zur Usability), nicht ohne weiteres vergleichbar sein werden. Deshalb sind folgende Ausführungen mit einer gewissen Vorsicht zu betrachten. Vernünftigerweise soll aber angenommen werden, dass grobe Probleme mit den Tests in der Vorstudie auch im Rahmen der Hauptstudie in jüngeren Jahrgängen auftreten würden.

⁸ „(c) Tempo und Geschwindigkeit sind das Gleiche.“ bzw. „(a) Zwei Körper können sich mit gleichem Tempo, aber unterschiedlicher Geschwindigkeit bewegen.“ und „(d) Bewegen sich zwei Körper mit gleicher Geschwindigkeit, so bewegen sie sich automatisch auch in die gleiche Richtung.“

In Bezug auf die Tauglichkeit des Tests als Testinstrument ist zunächst festzuhalten, dass Aufgaben unterschiedlicher Schwierigkeit im Aufgabenpool enthalten sind (Abb. 11.17). Aufgaben, die hier im Vortest schon mit einer geringen Lösungswahrscheinlichkeit auftraten (z. B. v_schlittena, v_schlittenb, v_huegel), sollten in der Hauptstudie in den Nachtest übernommen werden. Aufgaben aus dem Themenbereich der Kreisbewegung werden für die Hauptstudie komplett aus dem Test entfernt. Dies ist zum einen derer relativ großen Schwierigkeit geschuldet, zum anderen liegen – damit verknüpft – inhaltliche Gründe vor. Die Kreisbewegung wird inhaltlich in der Lernumgebung nicht thematisiert, daher scheint die Transferleistung bei der Beantwortung der Fragen zu groß zu sein.

Die Qualität der Raschanalyse scheint bislang mäßig aber noch akzeptabel zu sein. Dies betrifft die Reliabilität wie die Eindimensionalität des Konstrukts. Die qualitätssichernden Kennwerte für Items und Personen sind hingegen durchweg gut. Zu beachten ist hier, dass die Anzahl der Datensätze (40 pro Analyse) grenzwertig ist (vgl. Neumann 2014). Im Rahmen der Hauptstudie treten deutlich größere Fallzahlen auf, was den Modellannahmen zuträglich ist. Daher wird das Auswerteverfahren auch für die Hauptstudie genutzt.

Aus inhaltlicher Perspektive sind Lernzuwächse trotz schulischen Vorwissens erkennbar, sodass unterstellt werden kann, dass diese auch in der Hauptstudie auftreten werden. Es konnte jedoch aufgrund der Ununterscheidbarkeit von Effekten der Klasse und des Zugangs keine differenzierte Unterscheidung der Wirkung der Zugänge vorgenommen werden. Im fachlichen Test zeigen sich ergänzend deutliche Gendereffekte zu Ungunsten der Mädchen. Es existieren keine generellen Unterschiede zwischen Versuchsbedingungen. Das Feedback wurde nur marginal dargestellt, da an der Präsentation des Feedbacks zur Hauptstudie Veränderungen vorgenommen wurden (siehe auch Kapitel 11.2.2).

Deutliche Unterschiede hinsichtlich des Zuwachses der Personenfähigkeit zeigen sich bei Betrachtung des Vorwissens. Dort zeigen Schüler mit niedrigem Vorwissen signifikante Zuwächse, Schüler mit hohem Vorwissen nicht. Ein solcher Sättigungseffekt legt die Vermutung nahe, dass die behandelten Inhalte der Lernumgebung auch schulisch bereits bekannt waren. Daher ist ein solcher Effekt für die Hauptstudie bzw. Anfangsunterricht nicht zu befürchten, soll aber dennoch untersucht werden – gerade hinsichtlich der Angleichung des absoluten Lernerfolgs von Schülern hohen und niedrigen Vorwissens.

12. Ergebnisse Hauptstudie

12.1. Qualitative Daten

12.1.1. Ergebnisse der Befragung der Lehrkräfte

Vorbemerkungen

Abgesehen von einer Lehrkraft wurden die Fragebögen (siehe A.1.1) von allen zurückgegeben. Die Umfänge an geschriebenen Worten sind bei papierbasierter und elektronischer Form in etwa ähnlich, weshalb nicht angenommen werden sollte, dass eine Version zu Gunsten der anderen abweicht. Die Antworten lassen sich nicht immer einzelnen Klassen zuordnen, da bei mehreren von einer Person unterrichteten Klassen mitunter für diese bei ähnlichen Beobachtungen nur ein Bogen abgegeben wurde.

Die Auswertung wurde nicht stringent nach der qualitativen Inhaltsanalyse nach Mayring durchgeführt, da die Antworten teils sehr knapp gehalten waren oder kaum inhaltliche Deckung aufwiesen. Daher erwies sich der potentielle Mehrwert als gering. Daher werden die Antworten hier zum Verschaffen eines Überblicks dargestellt.

Ergebnisse

(Auffällig) Als Auffälligkeiten wurde in einer Klasse genannt, dass die SuS Schwierigkeiten hatten konsequent alleine zu arbeiten und mit den Simulationen herumspielen bzw. sich sehr lange aufhalten. Mitunter wurde an anderer Stelle eine reflektierte Auswertung, konzentrierte Bearbeitung, regelmäßiges Notizenmachen, kein Herumspielen gelobt. Weiterhin wurde in einer weiteren Lerngruppe auf den kritischen Aspekt „Spielen vs. Lernen“ sowie auf das vernachlässigte Anfertigen von Notizen hingewiesen. In einer anderen Schule habe eine eher unruhige Lerngruppe motiviert und konzentriert mitgearbeitet. An anderer Stelle waren SuS sehr unkonzentriert und unmotiviert und fertigten unzureichende Aufzeichnungen an. Weiterhin wurde eine sehr kommunikative Klasse beschrieben, in der Mitschriften unterschiedlicher Qualität angefertigt wurden.

(Schwierig) In einigen Klassen fiel die Unterscheidung von relevanten und nicht relevanten Inhalten schwer, so dass nicht durchgängig klar war,

worauf es genau ankommt. In einer Klasse trat zu Beginn vereinzelt das „Trial and Error“-Verfahren auf, wobei aber schnell realisiert wurde „worauf es ankommt bzw. was sie beobachten sollen.“ In einer weiteren Gruppe wurde seitens der Lehrkraft die stringente Auslegung der Einzelarbeit für die an sich diskussionsfreudige Klasse kritisiert.

(KritSim) Den Schülern soll ermöglicht werden, im eigenen Tempo zu arbeiten und Größen zu variieren. Dabei soll die Simulation realistisch¹ sein. Weiterhin sollte sie bezüglich der Bedienung selbsterklärend sein. Dieser Aspekt wird von einer weiteren Lehrkraft betont – im gleichen Zuge, wie sie die Passgenauigkeit der Simulation zum Lerngegenstand fordert. Ebenso sei eine „Schrittfunktion“ wünschenswert, sodass Schüler die zeitliche Abfolge des Sachverhaltes selbst beeinflussen können. Eine weitere Lehrkraft betont die Wichtigkeit der Einbettung der Simulation in den Unterricht, ganz analog zur Wichtigkeit der Einbettung von Realexperimenten. Entsprechend empfindet sie Simulationen als gewinnbringend, „wenn ihr Einsatz gezielt erfolgt, gut vorbereitet ist und entsprechend nachbereitet wird.“

Eine Lehrkraft bezog die Frage auf die in der Lernumgebung genutzten Simulationen. In diesem Zusammenhang hätte sie die Arbeit mit den Simulationen fruchtbarer empfunden, wenn die Schüler nicht komplett eigenständig, sondern durch die Lehrkraft angeleitet, am Programm gearbeitet hätten.

(2D) (Frage nicht von allen Lehrkräften beantwortet) Im Bezug auf die Lernumgebung wird kritisiert, dass die Einführung von physikalischen Aspekten wie Kraft, Ort und Geschwindigkeit parallel erfolgt und nicht wie im KC vorgesehen getrennt voneinander². Eine weitere Lehrkraft betont die Vorteile eines zweidimensionalen Zugangs im Bezug auf die Anschlussfähigkeit hinsichtlich Kraft, Kreisbewegung, waagerechtem Wurf und Ablenkung geladener Teilchen in elektrischen und magnetischen Feldern.

(1D) (Frage wurde nur von einer Lehrkraft beantwortet) Eine reine eindimensionale Kinematik wurde durch die beantwortende Lehrkraft

¹Anm. d. A.: Im Bezug auf das dargestellte Szenario alltagsbezogen oder zumindest vorzustellen sein. Unrealistische Aspekte wie Vernachlässigung von Störfaktoren sind hier vermutlich nicht gemeint

²Anm. d. A.: Das Kerncurriculum gibt bezüglich der Reihenfolge dort keine Vorgaben außer der Zuordnungen zu Doppeljahrgängen. Angaben über den daran orientierten schulinternen Lehrplan können hier u. a. mangels Kenntnis nicht getätigt werden.

auch im Anfangsunterricht nicht unterrichtet. So würde sie gleich die Geschwindigkeit als Größe mit Richtung und Betrag einführen, um spätere Lernhemmnisse abzubauen.

(Sonst) Vier Kommentare zur Frage nach sonstigen Aspekten beziehen sich auf die Lerngruppen³. Dabei wird teilweise das konzentrierte Arbeiten, die Anschlussfähigkeit des erworbenen Wissens, wie die gute Qualität der Notizen gelobt. Ein Kommentar zu einer der Klassen ist dem entgegengesetzt, so sei vielen SuS unklar, was mit dem Programm gelernt werden könne. Eine weitere Lehrkraft betont die Qualität des Lernprogramms/der Simulationen und äußert absolutes Unverständnis über das desinteressierte Verhalten der Klasse.

Eine Lehrkraft kritisiert die eher spielerische Einbindung der Simulationen. Dies könne dazu führen, dass sich Lernen zunehmend mit Spaß verknüpft und der Aspekt „Ich will Lernen, um besser zu werden, um Ziele zu erreichen“ in den Hintergrund rückt.

Weiterhin wird von anderer Seite auf die vermeintlich unterschiedliche Qualität der Simulationen hingewiesen. Diese sei im Bereich *Kraft* deutlich höher als im Bereich *Geschwindigkeit* oder *Diagramme* – was sich für den späteren Unterricht in unterschiedlichem Nachbereitungsaufwand widerspiegelt.

Diskussion

In einer kritischen Nachbetrachtung fällt auf, dass die einzelnen Fragen des Bogens eine negative Konnotation aufweisen (Auffälligkeiten, Schwierigkeiten, Probleme). Künftig sollte ausdrücklich nach gelungenen Aspekten gefragt werden.

Insgesamt zeigt sich ein heterogenes Bild in den Beobachtungen der Lehrkräfte zwischen den Klassen. Die Spannweite reicht dabei von Klassen, die „gut mitgearbeitet“ haben, bis hin zu Lerngruppen, die das Lernprogramm nicht als Lerngelegenheit genutzt haben – ohne hier über Gründe spekulieren zu wollen. Eventuell ist eine entsprechend getrennte Auswertung nach Lerngruppen sinnvoll!

Zu den Vorteilen eines 2D-Lehrgangs herrscht hingegen weitestgehend Einigkeit zu dessen Gunsten.

Als Kriterien für den Einsatz von Simulationen werden selbsterklärende Bedienung, Passgenauigkeit zum Unterrichtsgegenstand wie geeignete Möglichkeiten zur Beeinflussung gefordert.

³Frage wurde ähnlich wie zu Auffälligkeiten beantwortet

12.1.2. Ergebnisse der freien Schülerkommentare (n_kommentar)

Kategorisierung

N=111 Schüler haben die offene Frage der Nachfragebögen beantwortet. Ziel der Auswertung der Kommentare ist es, diese zu kategorisieren um so typische Antworten erkennen zu können. In einem ersten Schritt wurden dazu in zwei Stufen Reduktionen vorgenommen (vgl. Mayring 2015). Zunächst wurde die Aussage auf ein abstrakteres Niveau gehoben. Auf Basis dieser Stufe wurden induktiv 9 Kategorien gebildet. Dazu gehören auch 4 Paare von Kategorien, die eine *positive* und eine *negative* Ausprägung haben. Die Reduktionen wurden dann den Kategorien zugeordnet. Bei mehrteiligen Äußerungen war dabei auch eine mehrfache Zuordnung möglich. Nach Überarbeitung der Kategorien wurde ein abschließendes Rating der Schülerantworten von zwei Personen durchgeführt. Es ergab sich eine Interrater-Reliabilität von $\kappa = .643$, was einer guten Übereinstimmung entspricht⁴ (vgl. z. B. Bortz und Döring 2006).

Ergebnisse

Die Zuordnungen *Reduktion* → *Kategorie* wurden induktiv auf der Basis der Reduktionen durch einen Rater vorgenommen. Nach Ausformulierung der so erhaltenen Kategorien wurde auf dieser Grundlage von einem zweitem Rater eine Zuordnung vorgenommen. Nach Überarbeitung der Kategorien und eines erneuten Rating-Durchgangs verbesserte sich die Reliabilität deutlich. Die entsprechenden Kategoriebeschreibungen finden sich im Anhang (siehe Kapitel A.4). Es ist selbstredend möglich, dass der Kommentar eines Schülers in mehrere Kategorien eingeordnet werden kann, wenn dieser unterschiedliche Bedingungen erfüllt.

Im Folgenden werden charakteristische Beispiele zu den Kategorien dargestellt. Insgesamt wurden induktiv 9 Kategorien (A – I) ermittelt. In vier bzw. acht Fällen treten Kategorien doppelt auf, jeweils in positiver bzw. negativer Ausprägung (A – H). Die Häufigkeiten einer jeden Kategorie sind in Abbildung 12.1 aufgeführt.

⁴Maß für die zufallskorrigierte Übereinstimmung von hier zwei Ratern.

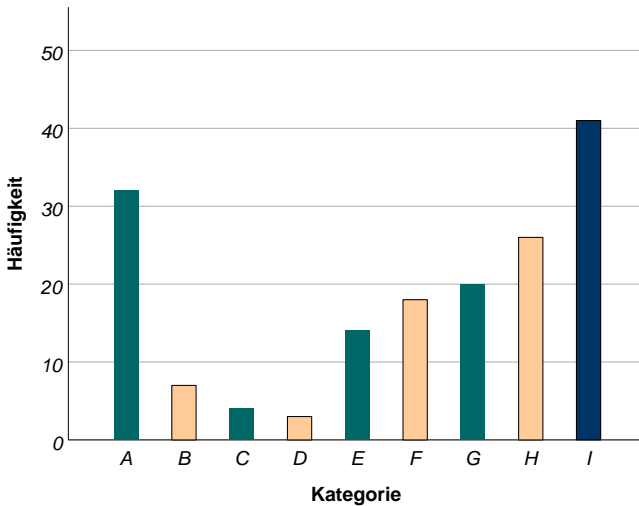


Abbildung 12.1.: Häufigkeiten der Zuordnungen pro Kategorie

In die Kategorie **(A) allgemein positive Anmerkungen** fallen die zweithäufigsten Nennungen, insgesamt 32 Stück. Entsprechend der Kategorie fallen diese eher unspezifisch aus („Ich fand es abwechslungsreich.“ „Sehr schön.“ oder „Gibt von mir nen Daumen nach oben! :)“). Mit nur 8 gibt es in dieser Kategorie deutlich weniger Bemerkungen zur zugehörigen Kategorie **(B) allgemein negative Anmerkungen**. Mehrfach wurde dabei der Faktor Ermüdung genannt („Wirkt ermüdend.“, „Bin zum Schluss immer müde geworden“) bzw. damit verwandt („Ich fand es langweilig und anstrengend.“). Bezüglich der betreuenden Personen (nicht der Lehrkräfte!) wurden nur in 5 bzw. 4 Fällen Angaben gemacht. Da alle 4 negativen Äußerungen in der Kategorie **(D) negativer Eindruck Personen** („Die Helfer waren teilweise etwas unhöflich“) in einer einzigen Klassen auftraten, ist hier von einem wie auch immer gearteten Vorfall auszugehen, der zu den Äußerungen verleitete. Besonders ist dabei zu bemerken, dass fast alle Kommentare der Kategorie **(C) positiver Eindruck Personen** („nette Menschen“) ebenfalls aus dieser Klasse stammen. Eine generalisierbare Aussagekraft dieser Kategorie ist also zweifelhaft.

14 bzw. 18 Antworten entfallen auf **(E) positiven Eindruck von allgemeiner (Tablet)arbeitsform inkl. technischer Aspekte** bzw. **(F) negativen Eindruck von allgemeiner (Tablet)arbeitsform inkl. technischer Aspekte**. Als Ankerbeispiel der positiven Kategorie kann stellvertretend „Ich finde es gut mal mit Tablets zu arbeiten und nicht immer mit Buch und Zettel.“ dienen, da viele Kommentare das scheinbar im

Unterricht unübliche Tablet loben. Kritische Anmerkungen beziehen sich in der Hauptsache auf problematische oder fehlerhafte Bedienung sowie Funktion („Es hat mich gestört, dass das Programm nicht richtig gespeichert hat.“; „Es ist ziemlich umständlich vom Lernprogramm zum Ordner und wieder zurück“). Ebenfalls wird auch die Arbeitsform kritisiert, die teilweise – speziell in Einzelarbeit – als langweilig empfunden werde (vgl. „Auf Dauer langweilig, weil man nur mit sich selbst beschäftigt ist.“).

Die 20 Aspekte, die den positiv empfundenen Umgang mit dem Lernprogramm (inkl. der Simulationen) darstellen, werden in Kategorie **(G) positiver Eindruck des Lernprogramms** zusammengefasst. Ohne eine geratete Unterteilung vorgenommen zu haben, zielen die Kommentare häufig auf den Lernerfolg ab (vgl. „Das Lernprogramm ist hilfreich und sollte im Unterricht eingeführt werden.“, „Ich habe viel Neues erfahren.“). Auch die empfundene Freude am Programm (vgl. „Ich fand die Simulationen echt cool zum ausprobieren und es hat mir Spaß gemacht“) wurde kundgetan, der Rest gliedert sich augenscheinlich unspezifisch.

Zu den negativ empfundenen Positionen (26) der Kategorie **(H) negativer Eindruck des Lernprogramms** gehören häufig die scheinbare Eintönigkeit/Langeweile des Programms (vgl. „Simulationen waren nicht sehr spannend“, „Die Simulationen waren ein bisschen langweilig und die Fragen zu einfach.“), sowie Unklarheiten in Texten, Fragen und Instruktionen („jedoch waren die Fragen zum Teil unverständlich!“, „Anfangs wusste man überhaupt nicht, was man machen muss.“).

Die mit Abstand häufigste Zuordnung (41) erfährt die Kategorie **(I) Wünsche / konkrete Verbesserungsvorschläge**. Dort sind mehr oder weniger – auch eher implizite – Wünsche zugeordnet. Die Bandbreite an Wünschen ist dabei breit gefächert. Die Nennungen betreffen den Wunsch nach Wiederholung (vgl. „Sowas sollte man öfter machen.“). Jeweils 1 – 3 Nennungen entfallen auf die Simulationen (vgl. „Bei manchen Simulationen würde ich mir wünschen, dass es ein eindeutiges und nicht so einfach zu erreichendes Ziel gibt.“), die Texte (vgl. „Nicht soviel Text.“) oder auch die Konzeption („Ich würde so etwas auch gerne öfter machen, aber wenn, dann nur in Abwechslung von Unterricht (normal)/ Beantwortung der Fragen durch Lehrer.“) sowie Sonstiges.

12.1.3. Auswertung der Schülerinterviews

Zu den Hintergründen des Interviews siehe Kapitel 8.4.

Top 3

Angelehnt an Tobias (2010) wurde nach den *Top 3*-Dingen gefragt, über die die interviewten Personen etwas gelernt haben. Die Auswertung er-

folgt nicht über das Mayring'sche Schema, sondern zweckmäßigerweise quantitativ. Zu beachten ist, dass sowohl nicht immer eine Reihenfolge benannt wurde wie auch, dass nicht immer drei Begriffe genannt wurden. Aus diesen Gründen wird auf eine Betrachtung der Rangfolge verzichtet. Ein Überblick findet sich in Tabelle 12.1.

Tabelle 12.1.: Top 3 Begriffe

Begriffe	Anzahl
(Arten von) Bewegungen	3
Kräfte	2
t-v/t-x-Diagramme	5
Geschwindigkeit	6
►davon Tempo und Richtung	2
Zusatzgeschwindigkeit	1
Einwirkung	1
Energie(spannung)	1
Simulationsbezogen	2

Die häufigsten Nennungen betreffen dabei den Geschwindigkeitsbegriff (in allen Interviews 6 Nennungen). Von den insgesamt 5 Befragten, die nach dem 2D-Konzept unterrichtet wurden, zählten 2 Aspekte zur Unterscheidung von Tempo, Richtung und Geschwindigkeit zu den Top-3-Begriffen. Diagramme wie das t-v-, oder t-x-Diagramm zählen mit 5 Nennungen zur zweithäufigsten Nennung, wovon nur eine Nennung von 2D-unterrichteten Schülern stammt. Bewegungen und Kräfte wurden mit 2 bzw. 3 Nennungen sporadisch erwähnt. Zusatzgeschwindigkeit und Einwirkung wurden als letzte thematisch passende Aspekte aufgezählt. Der Begriff *Energiespannung* bzw. *Energie* wird zwar genannt, ist aber zum einen falsch und zum anderen wurde der Energieaspekt im Lernprogramm konsequent vermieden. Weiterhin wurden Aspekte genannt, die sich nicht auf physikalische Inhalte beziehen, sondern auf Spezifika der Simulationen (z. B. „Ich fand sehr spannend das mit den Zielfotos, wie man diese Bremsstärken und so einstellen musste“).

Das häufige Auftreten von Geschwindigkeit und Aspekten zu Diagrammen zur Kinematik lässt sich evtl. dadurch erklären, dass es sich um vergleichsweise (scheinbar) vertraute Begriffe oder Darstellungen handelt. Diese könnten/sollten den Schüler bisweilen aus dem Mathematikunterricht bekannt sein – ob nun physikalisch korrekt oder nicht.

Überraschend ist das seltene Erwähnen der Zusatzgeschwindigkeit unter den 2D-Schülern, wo doch diese im letzten Drittel der Lernumgebung erheblichen Raum einnimmt. Gleiches gilt für den Kraftbegriff.

Realexperimente

Die Arbeit mit Simulationen brachte im vorliegenden Fall den kompletten Verzicht auf Realexperimente mit sich, was u. U. für den Physikunterricht ungewöhnlich sein könnte. Das Fehlen dieser Realexperimente bzw. der Wunsch nach solchen wurde von den befragten Personen wie folgt beurteilt.

Die durch die Reduktion erhaltenen Bündelungen werden hier als Kategorien vorgestellt und an Ankerbeispielen verdeutlicht. Wörtliche Zitate sind in den folgenden Ausführungen zu den Interviews Schüleraussagen)

■ R1: Realexperimente hier unnötig

► Simulation ausreichend ► Praktische Gründe (Realisierbarkeit von Realversuchen) ► Langfristig Notwendigkeit gegeben

Realexperimente werden aus diversen Gründen im Kontext der Studie nicht zwingend als notwendig angesehen. Die Gründe dafür sind vielfältiger Natur. Die Simulationen wurden beispielsweise als für den Unterrichtsgegenstand ausreichend empfunden (vgl. „ich fand man hätte keine zusätzlichen Experimente gebraucht, die Simulationen haben das eigentlich schon ganz gut erklärt [...]“). Der langfristige Bedarf an Realexperimenten wurde aber ebenfalls erkannt, wie die Aussage „Experimente würden mir glaube ich in längerer Zeit fehlen. Aber in diesen vier Stunden haben sie mir nicht gefehlt.“ zeigt. Mitunter spielten praktische Erwägungen ebenfalls eine Rolle. Zum einen wurden zum Beispiel Probleme das unterschiedliche Arbeitstempo betreffend angesprochen. Zum anderen werden äquivalente bzw. gleiche Realexperimente wie in den Simulationen als zu aufwändig oder zu wenig anschaulich charakterisiert.

■ R2: Einsatz von Realexperimenten nicht relevant

Zwei Befragte gaben darüber hinaus an, sich über das Fehlen oder Wünschen von „echten“ Versuchen während der Arbeit am Tablet keine Gedanken gemacht zu haben.

■ R3: Realexperimente hier (eher) nötig

► Langfristig nötig ► Freude am Experimentieren

Realexperimente werden aus unterschiedlichen Gründen als durchaus relevant angesehen. Zwei mal wird dabei die langfristige Perspektive erwähnt, die bereits in der obigen Kategorie schon genannt wurde. Entsprechend dieser seien Realexperimente für den Phy-

sikunterricht wichtig, für einen kurzen Zeitraum aber durchaus auch durch Simulationen zu ersetzen (vgl. R1 und z. B. „Das mit den Tablets finde ich gut aber halt auf Dauer brauch man noch Experimente.“). Weiterhin könnten Realexperimente mehr Freude bereiten (vgl. „Mir macht das an sich schon mehr Freude selber diese Versuche durchzuführen oder so reale Versuche zu sehen.“).

■ R4: Lernerfolg mit Simulationen und Realexperimenten

►keine Unterscheidung möglich ►ähnlicher Lernerfolg ►Lernerfolg mit Realversuchen größer

Die Befragten wurden dazu angeregt sich zum erwarteten oder vermuteten Lernerfolg zu äußern. In zwei Fällen war die Unterscheidung von Lernerfolgen mit Simulation oder Realexperiment nicht möglich (vgl. „Das kann ich nicht einschätzen“ bzw. „kann ich nicht sagen, weil bei echten Versuchen da kann ja auch mal einiges schief gehen.“). In einem Fall wurde der Lernerfolg als ähnlich eingeschätzt (vgl. „IN: Glaubst du denn das du mit echten Versuchen auch besser gelernt hättest als mit den Simulationen? S: Nö IN: Oder mit Simulationen besser als mit echten? S: Ne nicht wirklich“).

Ein größerer Lernerfolg wurde aber Realversuchen zugesprochen, weil diese besser zur Schülerpersönlichkeit passen würde, ohne, dass dies näher erläutert wurde („Ähm, ja, da denke ich schon, weil es halt für mich als Person besser ist mit Versuchen.“). Mitunter wurden aber von einem weiteren Befragten zum gleichen Aspekt weitere Gründe für größeren Lernerfolg gegeben („ich finde echte Experimente einfach irgendwie besser zu verstehen, wenn man das selber ausprobieren kann und sowas“). In Abgrenzung zu Simulationen „Ja, aber man ist nicht so selbst dabei so, also wie das Ganze passiert und so, sondern man kann das nur sehen, allerdings weiß man nicht so genau, was da vorgeht und so“. In einer ähnlichen Richtung argumentiert eine weitere befragte Person, die nach den Gründen für größeres Verständnis gefragt, darlegt, dass dies darin begründet sein könnte „weil man es glaub ich direkt so vor sich hat. Ja. Vielleicht. Man sieht es dann ja auch, was da passiert“. In einem Fall wurden die Simulationen aber dennoch als gleichwertig betrachtet („Man hats dann vielleicht auch früher drauf. Aber ich fand's so ganz gut“).

Zukunftsperspektiven

Jede/r Befragte äußerte sich auf konkrete Nachfrage mit einer Äußerung zur Frage, ob er/sie „wieder so lernen“ wolle.

■ Z1: Lernmethode in Zukunft vorstellbar

Zwei Personen können sich das Lernen mit Tablets wie im Rahmen der Studie vorstellen, ohne dabei wie auch immer geartete Einschränkungen zu nennen (vgl. „Ja, kann ich mir sehr gut vorstellen sogar.“). Die zweite Person begründete die Vorzüge der eingesetzten Lernmethode mit Lernprogramm und Simulationen mit einer besseren Fokussierung im Vergleich zum Einsatz von „normalen Experimenten“.

■ Z2: Lernmethode in Zukunft mit (leichten) Einschränkungen vorstellbar

- ➡ z. T. fehlende Interaktion (Lehrererklärungen und Tafelarbeit fehlen)
- ➡ mangelnde Stillarbeitsphasen

Die meisten (6) Antworten entfallen in diese Kategorie. Dabei werden seitens der Lernenden mehr oder weniger gravierende Einschränkungen oder Ergänzungen zur Lernmethode erwähnt, wobei die grundsätzliche Offenheit der Methode gegenüber aber gewahrt blieb. Etwa die Hälfte der Einschränkungen bzw. Wünsche entfällt auf die Interaktion mit den Lehrkräften, die zum streckenweisen Erklären als wünschenswert erachtet wird, wie folgende zwei Kommentare zeigen: „Ich könnte es mir vorstellen, ja, aber ich find’s teils besser von nem Lehrer also von einer Person Sachen erklärt zu bekommen und teils find ich sowas besser an nem PC.“; „Finde manche Dinge sollten Lehrer auch erklären weil * manchmal ist es * bei manchen Dingen ist es verständlicher wenn es erklärt wird sozusagen.“ Ein weiterer Aspekt fehlender Interaktion wird das fehlende Abschreiben von der Tafel genannt (vgl. „Genau so glaube ich nicht, weil ich find ein bisschen ist dieser praktische Teil an der Tafel wichtig, dass man selber mitschreibt und nicht alles auf dem Tablet hat.“). Bezüglich der Sozialform wurde kritisiert, dass die „Stillarbeitsphasen“ nicht ausreichend durchgesetzt wurden (vgl. „aber äähm das Problem war das es also das ich eigentlich niemand an die Stillarbeitsphase gehalten hat“).

■ Z3: Lernmethode in Zukunft nicht vorstellbar

Eine befragte Person lehnt das Lernen nach dem Konzept der Studie ab, indem die fehlende Interaktion mit Lehrkräften bemängelt wird (vgl. „Ich da immer das Gefühl man konnte besser diese Fragen an die Lehrer stellen oder [sic!]“).

Einzelarbeit

■ E1: Einzelarbeit gut

Das auf sich allein gestellte Arbeiten am Tablet wurde von den befragten Personen überwiegend positiv aufgenommen. Die jeweiligen Begründungen fallen unterschiedlich aus. Faktoren sind dabei die Konzentration („Ich fand’s besser, weil man sich besser konzentrieren konnte als in Partnerarbeit oder so“), das höhere Arbeitstempo („Es ging relativ schnell dadurch dass man Einzelarbeit halt hatte und nicht von anderen Personen auch abgelenkt wurde“). Weiterhin wurde die Einzelarbeit als selbstverständlich oder gut hingenommen (vgl. „Machen wir ja sonst in der Schule auch und äh also da hab ich überhaupt kein Problem mit“).

■ E2: negative Einschränkungen der Einzelarbeit

Auch unter dieser Kategorie subsumieren sich Anmerkungen der Befragten, die der Einzelarbeit eher bzw. neutral positiv gegenüberstehen, jedoch auch kritische Nuancen enthalten. Im Einzelnen wurde z. B. der Austausch mit dem Partner gewünscht (vgl. „Das wäre vielleicht auch noch ganz cool, wenn man so mit anderen sich austauschen könnte während der Arbeit.“), wobei dieser mitunter auch schon statt gefunden hat (vgl. „Man hat dann schon mal so dem Sitznachbarn gezeigt, wenn er was nich verstanden hat“, vgl. auch Kapitel 9.1.1). Bemängelt wurde weiterhin die Lautstärke während der Arbeitsphase, die phasenweise als zu laut empfunden wurde, um konzentriert zu arbeiten (vgl. „Also bei mir war das zumindest so, dass wenn zum Beispiel die anderen dann so geredet haben oder so dann konnte man sich nicht ganz hundertprozentig so auf den Text konzentrieren“).

Allgemeine negative Kritik und Änderungswünsche

■ NKÄ1: Änderungswünsche nach Kritik an Lernumgebung

► technischer Natur ► konzeptioneller Natur ► inhaltlicher Natur

Hier sind die Änderungswünsche zusammengefasst, die auf einer konkret formulierten Kritik aufbauen. Bezüglich konzeptioneller Probleme wurde bemängelt, dass man die korrekte Antwortoption zu den MC-Fragen durch mehrfaches Raten erreichen konnte, ohne parallel das Verständnis für diese Antwort zu entwickeln („Aber dann, wenn man halt das falsch angeklickt hat, dass man dann einfach zurückgehen konnte, einfach das nächste anklicken konnte und wenn das dann auch wieder falsch war, äh, dass man dann wieder zurückgehen konnte, bis man halt das richtige hatte“). Daraufhin wurde das tiefergehende Erklären der gemachten Fehler inkl. Nennung der richtigen Antwort gewünscht. Weiterhin wurden in einem Fall kürzere Texte gewünscht („Also das hätte man vielleicht so, mmhh, so kürzer zusammenfassen können.“)

Die Textmenge betreffend, wurde von einer weiteren befragten Person ein ausgeglicheneres Verhältnis von Textmenge zu zugehörigen Simulationen gewünscht, da diese stellenweise zu groß oder zu klein erschienen. (vgl. „es war find fand ich dann zu viele Simulationen ähm für zu wenig Text.“). Speziell bezog sie sich dabei auf die Rafting-Simulation, die in zweifacher Ausführung vorhanden war.

Weiterhin wurde bemängelt, dass die Simulationen relativ umständlich „per Hand“ zu öffnen waren. Daher wurde vorgeschlagen, die Simulationen zu verlinken und den Workflow so deutlich zu verbessern („vielleicht könnte man dann direkt auf die Simulation zugreifen auf den Seiten das man nicht immer wieder zurück hin und her müsste“).

■ NKÄ2: Änderungswünsche nach Kritik an Umsetzung der Lernmethode

In dieser Kategorie werden Bemerkungen zusammengefasst, die nach ausdrücklich formulierter Kritik Änderungswünsche an der Lernmethode nahelegen. Einziger Kommentar in dieser Sektion bezieht sich auf die Lautstärke während der Bearbeitung der Lernumgebung („man sollte auch häufiger auf die Lautstärke achten, weil wenn die Lautstärke ja so hoch, also weil wenn es ne hohe Lautstärke ist, kann man Texte auch nicht so gut verstehen“).

■ NKÄ3: Änderungswünsche an Lernumgebung

Hier werden Änderungswünsche der Befragten aufgenommen, die ohne explizit formulierte Kritik einhergehen.

Beispielsweise äußert ein Befragter den Wunsch nach Beispielbildern (auch wenn er/sie die nur auf ein Beispiel bezieht) („Hmm *2* ahh vielleicht so* so ein so ein kleines Beispielfoto vielleicht zum Beispiel, einfach so zum Beispiel das Beispiel eben von der Strecke das man hier diese Strecke einfach so eine eine kurze Strecke[. . .]“). Ein Befragter äußerte den Wunsch nach mehr Videos zuungunsten von Simulationen (vgl. „Vielleicht auch mal, dass man so ein Video hat, dass man nicht immer alles selbst also das man immer alles ausprobieren sondern auch mal gucken kann, wie andere das ausprobieren“). Auch zwei weitere Bemerkungen beziehen sich auf den Einsatz der Medien auf Ebene der Lernumgebung. Gewünscht wurde beispielsweise in einem Fall der Verzicht auf Simulationen und mithin der alleinige Einsatz von Texten und Bildern (vgl. „Man könnte es theoretisch auch ähm mit Texten oder äh Videodateien machen. Das fände ich besser anstatt ääh anstatt des ääh anstatt der Simulationen.“) Im Bereich der MC-Fragen wurde von einer weiteren Person das häufigere Zurverfügungstellen von Videos gewünscht („Genau also mehr Bilder im Sinne von diesen Videos unter den äähm Fragen wenn man da was falsch hatte“).

Eine weitere Person wünscht ihrerseits Zusammenfassungen am Ende der Lernumgebung („jetzt am Ende nochmal ne kleine Wiederholung der ganzen Lerninhalte, wo wir das Ganze nochmal sozusagen Revue passieren lassen“).

■ NKÄ4: Änderungswünsche an Lernmethode

Teils wurden zusätzlich zu den Simulationen Erklärungen von der Lehrkraft gewünscht. In einer weiteren Nennung wurde mehr Zeit gewünscht, um die Materialien ohne Zeitdruck und ungestört durcharbeiten und verstehen zu können („mehr Zeit hätte ich gut gefunden damit man sich das langsam ja noch mal erklä- äh durchlesen kann“). Gleiche befragte Person wünscht sich darüber reale Experimente („nicht nur diese Simulationen sondern auch nebenbei noch andere praktische Versuche die man selbst ausprobieren kann“) (siehe dazu auch R1 – 3 in diesem Kapitel).

■ NKÄ5: Kritik an Lernumgebung

➡ technischer Natur ➡ inhaltlicher Natur

Kritik an der Lernumgebung war in zwei Fällen auf (scheinbare) technische Unzulänglichkeiten wie fehlerhaftes Abspeichern oder inkohärente Benennung von Simulationen bezogen (vgl. „Wenn man das Programm nicht richtig gespeichert hatte und von vorne anfangen musste“).

Einem Befragten waren mitunter Begriffe unklar, da diese seiner/ihrer Auffassung nach nicht ausreichend erklärt wurden („einzelne Texte, wo äh Begriffe genannt wurden die gar nicht erklärt wurden, aber das war auch nur sehr selten“).

Kritisiert wurde von einer Stelle weiterhin, dass die Simulationen willkürlich zu öffnen sind, was sich der Kontrolle der Lehrenden entzieht („Dass man da immer das öffnet was man will, ich glaube zu Zweit können Sie das ja auch bei 30 Schülern * nicht gerade häufig und gut kontrollieren“).

■ NKÄ6: Kritik an Lernmethode

► Langeweile

An der Lernmethode wurde – ohne konkrete Verbesserungsvorschläge oder Gründe vorzubringen – einzig die aufkommende Langeweile kritisiert, die sich gegen Ende der jeweiligen Unterrichtsstunden bei einer befragten Person einstellte („Es war interessant, aber irgendwann wurde es gewissermaßen auch eine Art von Langweilung (sic!)“)

Allgemeine positive Kritik

■ PK1: Abwechslung/Kontrast zum konventionellen Unterricht

Ein mehrfach genanntes Kriterium für die positive Bewertung war die Abwechslung im Vergleich zu dem Physikunterricht, den die Befragten gewohnt sind („eine kleine Abwechslung und nicht so stumpfer Physikunterricht, das fand ich ganz cool“; „auch ein bisschen mehr Spaß gemacht als im normalen Physikunterricht“).

■ PK2: Gut (ohne Spezifikation)

In dieser Kategorie werden alle Kommentare, die sich auf die Qualität der Lernumgebung/Methode beziehen subsumiert, die positiv ausgelegt werden können – aber nicht näher begründet werden. Dazu gehören Bemerkungen wie „Ich hab nichts zu sagen wirklich außer, dass es eine gute Sache war, das mal zu testen so im Unterricht.“, „Jooa, ich fand das ganz cool.“ oder „ich fand es war eine coole Erfahrung so was neues auszuprobieren“.

■ PK3: Gut (mit Spezifikation)

► Lernmethode ► Inhaltlich

Spezifizierte positive Äußerungen beziehen sich zunächst auf die methodische Aufbereitung der Lernumgebung. Dies umfasst in einem Beispiel die Anschaulichkeit und die Möglichkeit, Dinge selbst zu erkunden. („Gute Idee, dass man auch einmal irgendwie etwas zum Anschauen hatte oder zum Ausprobieren und nicht immer nur in den Büchern liest und sich diese Infokästen oder sowas durchliest“)

In einem weiteren Interview wurde weiterhin das eigenständig festzulegende Arbeitstempo gelobt, sodass es auch möglich war, sich nochmals intensiv mit Inhalten auseinanderzusetzen („ganz lehrreich wenn man dann so einen Text gelesen hat und dann kann man sich das nochmal durchlesen und nochmal durchlesen wenn man das vielleicht einmal nicht verstanden hat“)

Bezüglich inhaltlicher Aspekte verständlich und einfach („alles super verständlich, auch recht einfach“).

Nutzung des Feedbacks

■ F1: Das Feedback wurde bei falschen Antworten zielgerichtet genutzt, bisweilen intensiver als bei korrekten.

In diese Kategorie fallen Äußerungen, die nahelegen, dass das Feedback vor allem bei falschen Antworten genutzt wurde. Besonders deutlich wird dies anhand von Äußerungen wie „Bei falschen Antworten habe ich mir immer noch die Simulation danach angeguckt und geguckt, warum äh, wie ich jetzt auf mein falsches Ergebnis gekommen bin“.

■ F2: Das Feedback wurde bei richtigen Antworten (auch eher knapp) gelesen und mitunter auch genutzt.

► Richtig-Falsch-Information ► „Überfliegen“, „Intensivere Auseinandersetzung“

Bei korrekten Antworten zu den MC-Fragen wurde das Feedback von den befragten Schülern auf eine andere Art und Weise genutzt als bei falschen Antworten. Eine wiederholte Auseinandersetzung mit den Simulationen findet dabei nicht zwingend statt. Die Bandbreite reicht dabei von relativ intensiver Auseinandersetzung bis hin zum Übergehen des Feedbacks, wie folgende Beispiele zeigen: „Genau nach richtigen Antworten und ich habe mir die dann halt noch einmal durchgelesen und diese. Ich habe mir meistens erst bei diesen

Hinweisen nach den Fragen meine Beobachtung auf meinem Zettel aufgeschrieben.“ bzw. „Also ich bin immer drauf eingegangen, ja.“. Zwei weitere Schüler äußerten sich entsprechend entgegengesetzt. So wurden die Antworten eher nur „überflogen“ bzw. nur auf die Richtig-Falsch-Information überprüft. Das bestätigen die Aussagen I: Aber wenn es richtig war hast du es dann so überflogen? S1: Ja mehr wie auch „Also ich hab es äh bei einigen hab ich die Inhab ich die hab ich nur die ääh Information, ja, richtig oder falsch genutzt“. Aus den Antworten ist jedoch nicht ersichtlich, inwiefern das Überfliegen der Antworten nicht auf bloß dem Erhalten der Falsch-Richtig-Information diene.

- F3: Das Feedback wurde, vor allem bei richtigen Antworten, nur knapp, mitunter auch garnicht, gelesen/genutzt.

Als Ankerbeispiel für die Kategorie dient folgende Äußerung eines Schülers: „Den hab ich dann nicht weiter verfolgt, da bin ich dann gleich weiter gegangen.“.

- F4: (Eigenständige) Nutzung der Simulation zur Überprüfung.

Vier der befragten Personen äußerten sich konkret zum Nutzen der Simulationen nach Anregung durch das Feedback. Dabei lassen sich elaborierte Strategien erkennen, wie z. B. das Aufnehmen des Feedbacks und daraufhin das Nachvollziehen innerhalb der Simulation, wobei die Verarbeitungstiefe dabei variiert. Eine ausführliche Beschäftigung mit der jeweiligen Simulation beschreibt folgender Schüler: „manchmal bin ich dann auch nochmal in die Simulation gegangen und hab geguckt, was ich falsch gemacht habe. [...] ich hab auch den Tipp ausprobiert aber ich hab auch andere Sachen nochmal angeguckt [...] . Also ich hab das dann nochmal gründlicher erforscht diese Simulation.“. Ebenfalls eine tiefgehende Beschäftigung mit der Simulation, allerdings ganz bewusst *ohne* die Feedback-Hinweise dabei zu beachten, beschreibt eine weitere Person: „also ich fands eigentlich so wenn man selber macht, find ich einfacher. Wenn man sich da selbst Gedanken drüber gemacht hat.“

Eine abgeschwächte Auseinandersetzung mit den Simulationen zeigen zwei weitere Befragte. Zumindest die Formulierung legt zwar eine Auseinandersetzung mit den Simulationen nah, aber eher auf der Ebene des *Nachvollziehens*: „Ich hab ich hab mir die Simulation dann noch mal angeguckt * und versucht das nachzuvollziehen.“ bzw. „und dann bin ich das noch mal durchgegangen um es zu verbessern und es mir noch mal an zu gucken“. Von anderen Befragten

wurden keine ausdrücklichen Hinweise zur weiteren Nutzung der Simulationen getätigt.

■ F5: Hinweise sind leicht zu merken und bisweilen auch umzusetzen.

Im Allgemeinen seien nach Einschätzung der vier dazu Befragten die durch Text-Feedback erhaltenen Hinweise leicht zu merken und bisweilen auch leicht umzusetzen. Dies soll durch die folgenden Zitate illustriert werden: „IN: waren diese Hinweise leicht oder schwierig für dich zu merken du umzusetzen? S: Leicht. Also es war relativ leicht die umzusetzen.“, „merken waren sie äh leicht“, „Also ich konnte sie leicht umsetzen. Ich glaube also mir ist es nicht schwer gefallen die umzusetzen.“. Eng mit dieser Kategorie verknüpft ist allerdings Kategorie F7 (siehe unten).

■ F6: Animationen waren teils überflüssig, teils potentiell hilfreich.

Konkret zu den Animationen äußerten sich zwei Schüler/innen. Eine Äußerung schreibt diesen dabei eine sinnvolle Unterstützungsfunktion zu, obwohl die betreffende Person selbst mit Text-Feedback gearbeitet hat. (vgl. „Ich würde es gut finden wenn das dann bei jedem da so war damit man sich das noch mal angucken könnte und dann noch mal die Fragen beantworten kann, damit äh man direkt weiß was da gem- was man da hätte tun sollen.“). Es handelt sich bei der Basis dieser Kategorie also nicht um eine direkte persönliche Erfahrung, sondern eher um eine Einschätzung.

Die Äußerung einer weiteren Person legt nahe, dass Animationen nicht grundsätzlich als hilfreich empfunden werden. Die – wenn überhaupt – gemachten Fehler schienen derart einsichtig, dass ein Betrachten der Animation als nicht zwingend notwendig empfunden wurde („war es halt so, dass man die Animation eigentlich nicht mehr angucken brauchte, weil man schon gesehen hat, dass man die wo der Fehler lag. *2* D weil ich fand das jetzt nicht SO schwer“).

■ F7: Feedbacknutzung abhängig von Verständlichkeit der jeweiligen Instruktion.

Eine befragte Person gab an, die Simulationen nach entsprechenden Hinweisen mittels Feedback nur genutzt zu haben, wenn die Instruktion des Feedbacks klar und verständlich war (vgl. „Wenn du’s nicht gemacht hast, warum nicht? S: Ääh, weil ich diese Fragen einfach unverständlich fand oder ich jetzt nicht wissen äh wusste was

ich genau da zu tun hatte.“). Diese Entscheidung des Schülers/der Schülerin scheint unmittelbar einsichtig und es ist nicht unwahrscheinlich, dass auch viele andere Schüler so verfahren – jedoch erfolgte hier die einzige explizite Nennung.

■ F8: Das Feedback wird als automatisches System geschätzt.

Die befragte Person bemängelte, dass es in seiner/ihrer Lerngruppe häufig zur Einmischung anderer Personen in das eigene Arbeiten käme. Daher wurde seitens dieser Person das gegenwärtige Feedbacksystem geschätzt, sodass jeder für sich alleine arbeiten kann (vgl. „um Beispiel so ein Junge hat aus unserer Klasse der der verbessert immer dauernd alle und das ist, würde mir dann schon ein wenig auf die Nerven gehen. Also fand ich das dann schon so ganz gut.“).

Die beiden letzten Kategorien entfallen jeweils nur auf eine befragte Person und ließen sich nicht ohne Weiteres einer anderen Kategorie zuordnen.

Fachliche Aspekte

In diesem Abschnitt wird ergänzend zu Aspekten der Lernumgebung die fachliche Komponente der Interviews beleuchtet. Die einzelnen Codiereinheiten wurden den Codes aus Abschnitt D.1 aus dem Anhang zugeordnet. Die extrahierten Codiereinheiten wurden von zwei Ratern in zwei Durchgängen geratet. Letztlich ergab sich eine Interrater-Reliabilität von $\kappa = .82$, was als (sehr) gut zu bezeichnen ist. Auf vergleichende/hypothesentestende Verfahren wird hier aufgrund der Fallzahlen verzichtet. Zu beachten ist, dass mehrere Codes zu einer Versuchsperson gehören können – also durchaus falsche Aspekte wie auch korrekte Antworten zu einer Person gehören können.

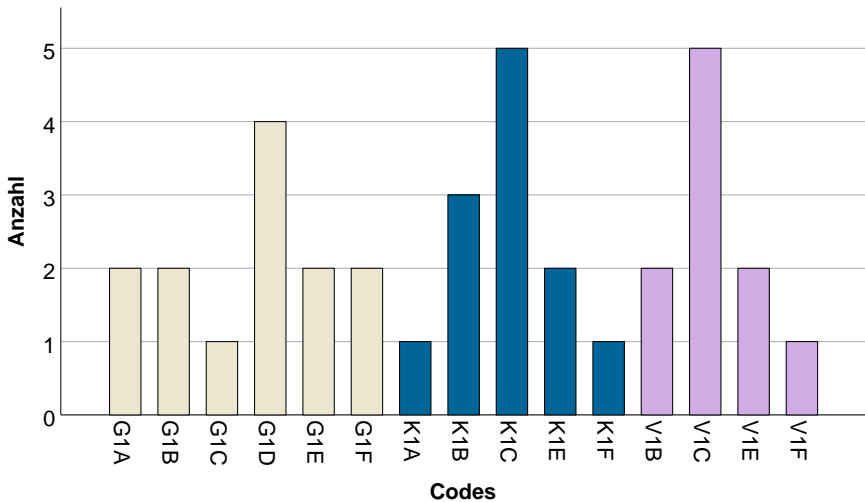


Abbildung 12.2.: Häufigkeit der Zuordnungen pro Kategorie 1D

In Abbildung 12.2 sind die Häufigkeiten der Codierung für Interviews mit 1D-Schülern abzulesen.

WIE BESCHREIBT EIN PHYSIKER GESCHWINDIGKEIT? 1D

Der häufigste Code (4) im Bereich der Geschwindigkeit ist G1D, d. h. Geschwindigkeit als Strecke pro Zeit („mit der Geschwindigkeit ist gemeint, wie viel Strecke ein Körper in einer bestimmten Zeit schafft.“). Zu beachten ist dabei, dass in allen Fällen (teils deutliche) Hilfestellung seitens des Interviewers gegeben wurde (vgl. „Und wenn man in dieser Erklärung jetzt noch Begriffe wie „zurückgelegte Strecke“ und „Zeit“ einfließen lassen würde?“). Eher oberflächliche Anmerkungen zu Geschwindigkeit als

Schnellsein oder Nennung von Begriffen kommen ebenfalls vor. Unspezifische oder falsche Angaben wurden in der Summe ebenfalls vier mal codiert (vgl. „Und unter Geschwindigkeit verstehe ich die Bewegungskraft“). Ein Definitionsversuch mittels Beispielrechnung wurde einmal unternommen. Der Aspekt des Vorzeichens und Bewegungsrichtung wird von keinem der Interviewten thematisiert. Es zeigt sich also eine erschöpfende Nutzung der Kategorien.

WORAN ERKENNT MAN, DASS HIER EINE KRAFT WIRKT? 1D

Im Kontext der vorgelegten 2D-Simulation wurden zugehörige Äußerungen überwiegend als K1C – also Veränderung von Richtung oder Geschwindigkeit – codiert (vgl. „wenn der fährt und ich ihm hinten reinfahre, dann wirkt eine Kraft auf den, auf den Max und der wird dann schneller, weil ich den von hinten anstoße“). In weiteren Fällen äußerten sich die Schüler entsprechend der Codierung K1B nur implizit anhand von Beispielen zur Geschwindigkeits- oder Richtungsänderung (hier auch eine deutliche Hilfestellung). Ausdrücklich falsche Nennungen wurden mit K1A und K1F zweimal codiert (vgl. „Also in der Simulation hat man eine Kraft drin erkannt, dass sich etwas bewegt hat“). Allgemeine Formulierungen unter Zuhilfenahme des Begriffs *Bewegungszustand* wurden der Codierung nach nicht gegeben.

VARIABLENKONTROLLE 1D

Kein Interviewter erkennt den dargestellten Fehler in der präsentierten Simulation *nicht*. Die häufigste Codierung V1C beschreibt Fälle, bei denen die mangelnde Variablenkontrolle zwar erkannt und auch am Beispiel erläutert wird, jedoch nicht mehr oder weniger explizit benannt wird („wenn er mehr Stärke da reinbringt. Aber er hat halt auch die Masse auf 3 erhöht und da ist diese Tendenz gleich geblieben, wie wenn man eins eins macht und deswegen ist die Geschwindigkeit von 4,5 km/h auch gleich geblieben.“). Letzteres wurde niemals codiert. Unspezifische Äußerungen oder mäßig elaborierte Beispiele wurden zwei mal codiert.

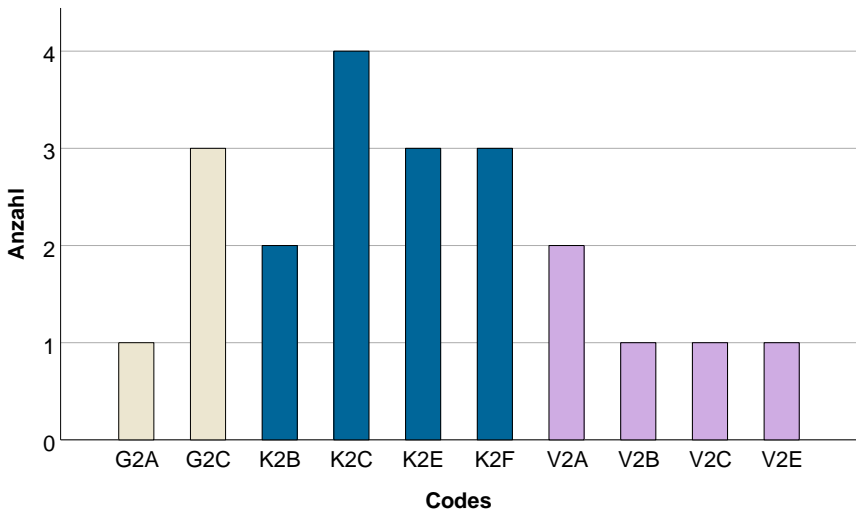


Abbildung 12.3.: Häufigkeiten der Zuordnungen pro Kategorie 2D

In Abbildung 12.3 sind die Häufigkeiten der Codierung für Interviews mit 2D-Schülern abzulesen.

WIE BESCHREIBT EIN PHYSIKER GESCHWINDIGKEIT? 2D

Die Unterschiedlichkeit in den Codierungen fällt bei interviewten Schülern des 2D-Zugangs deutlich geringer aus. Nur ein Code bezieht sich auf Geschwindigkeit als Schnellsein, die übrigen (G2C) zielen auf die Nennung von Richtung und Tempo als (auch vage) zur Geschwindigkeit gehörend („Die Zusammenfassung von Schnelligkeit, also von Tempo, und äh von Geschwindigkeit war das jetzt? Äh, ich glaube das war allgemein das Tempo und äh *6* (lacht) also entweder äh ich habs vergessen also das Tempo glaube ich und die enter die Richtung oder die oder die wie heißt es *5* ich habe das Wort vergessen * (lacht) IN: (lacht) Jetzt sind wir der Lösung schon ziemlich ziemlich nah.*“). Eine explizite Erläuterung der Geschwindigkeit als gerichtete Größe bei der ein Pfeil die Schnelligkeit und Richtung angeben kann wurde nicht codiert. Die Thematisierung von Geschwindigkeitspfeilen wurde nicht erfasst. Gleiches gilt für falsche oder unspezifische Äußerungen.

WORAN ERKENNT MAN KRAFTWIRKUNG? 2D

Bei Vorlage der 1D-Simulation wurde eine Argumentation über die Zusatzgeschwindigkeit nicht codiert. Abgesehen von der größeren Anzahl von unspezifischen oder falsch codierten Äußerungen, ähneln die Begründungen der Schüler daher denen der 1D-Schüler. So dominiert mit K2C auch hier

die Argumentation über Geschwindigkeitsänderungen (spez. schneller oder langsamer werden) oder implizit über Beispiele (K2B).

VARIABLENKONTROLLE

Die Fehler im gezeigten Video wird 2 Codierungen nach nicht erkannt. Nur in einer Äußerung spiegelt sich anhand beispielhafter Äußerungen des Schülers ein Verständnis für die mangelnde Variablenkontrolle wider. Unspezifische Äußerungen wie ein mäßig elaboriertes Beispiel entfallen auf jeweils eine Codierung.

12.2. Quantitative Daten

12.2.1. Fragebögen

Einführung In diesem Kapitel wird die Auswertung der Vor- und Nachfragebögen der Hauptstudie beschrieben. Zunächst erfolgt dabei eine eher deskriptive Betrachtung der Konstrukte Interesse, Selbstkonzept, Flow und Fehlerkultur sowie abschließend eine solche Betrachtung separater Fragen. Eine inhaltliche Diskussion findet im Anschluss statt.

Qualitätssicherung Wie schon in der Darstellung der Ergebnisse der Vorstudie finden sich die Angaben zur Reliabilität der Skalen im jeweiligen Abschnitt.

Im Vorfragebogen (Abb. 12.4 (a)) überschreiten alle vier Items zur Fehlerkultur die Grenze von 3,2 um 0,07 bis 0,24 Punkte – sowie auch das Item zur Touchscreenbedienung um 0,35 Punkte. Am unteren Ende der Skala erreichen zwei Items zur Häufigkeit der Verwendung von Simulationen im Unterricht die Grenze von 0,8 nicht. Die Skala zur Fehlerkultur scheint also einem Deckeneffekt zu unterliegen, was die Aussagekraft bisweilen einschränken wird. Alle weiteren, einzelnen Items werden mit dem Wissen um Decken- und Bodeneffekte beibehalten, weil die enthaltenen Informationen wertvoller scheinen als ein Verwerfen zuträglich wäre. Weiterhin reicht das Spektrum der Skala bei manchen Items von *nie* bis *immer* und bietet mit Hinblick auf abgefragte Häufigkeiten zur Simulationsnutzung kaum Möglichkeiten zur Vermeidung von Bodeneffekten, sofern abzusehen ist, dass sich Antworten eher im „Selten“-Bereich sammeln würden. Es wäre aber keine Alternative die Option „Immer“ nicht anzubieten!

Im Nachfragebogen über- bzw. unterschreiten drei Items zur Bedienung des Lernprogramms den optimalen Bereich um 0,01 bis 0,32 Punkte sowie ein Item der Skala *Selbstkonzept* um 0,05 Punkte. Letzteres wird zu Gunsten der Skala aus Gründen der Verhältnismäßigkeit und der

Geringfügigkeit ignoriert. Die erstgenannten, einzelnen Items werden mit Berücksichtigung des Decken- bzw. Bodeneffekts ausgewertet und kritisch betrachtet.

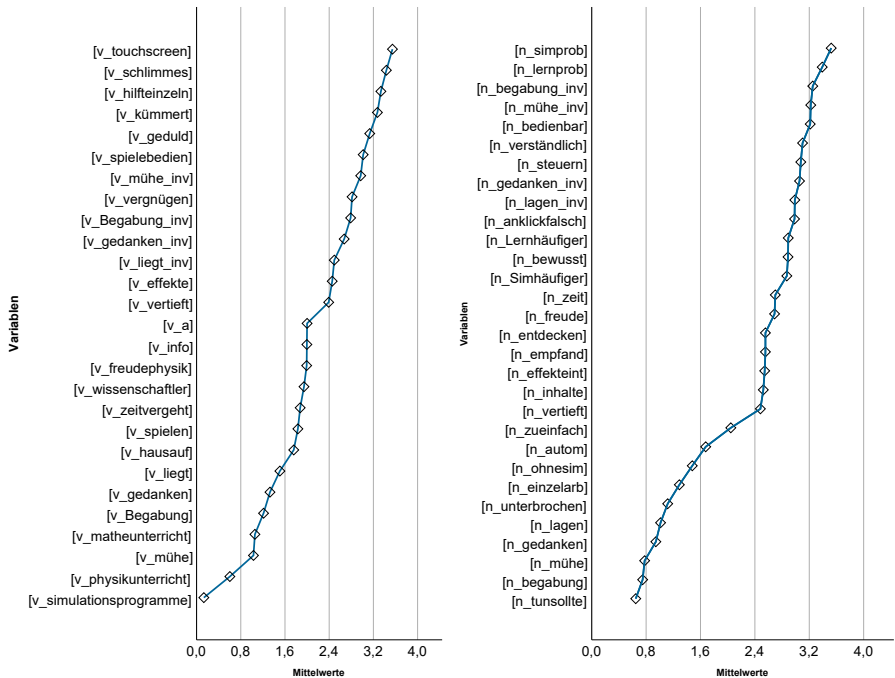


Abbildung 12.4.: Mittelwerte aller Items aus beiden Fragebögen

Nicht vollständige Fälle wurden listenweise ausgeschlossen. So kommen unterschiedliche Fallzahlen für verschiedene Skalen oder Items zustande. Die durchweg niedrigeren Fallzahlen im Nachfragebogen beruhen auf der Tatsache, dass nicht alle Schüler die Lernumgebung abgeschlossen haben oder in der letzten Stunde der Intervention nicht anwesend waren (vgl. Kapitel 9.1.1).

Leistungen Für die bisherigen Leistungen der Schülerinnen und Schüler sollten über die Abfrage der Mathematik-, Physik und Deutschnoten Indikatoren erfasst werden. Beim Ziehen von Rückschlüssen aufgrund der Note ist jedoch der relativ (!) hohe Anteil von fehlenden Werten speziell bei der Physiknote (7,1%) zu beachten.

- **Physik** Bei N=255 gültigen Nennungen ergibt sich eine durchschnittliche Physiknote von 2,42. Es gibt dabei keine signifikanten Unterschiede zwischen Jungen und Mädchen ($(t(250) = -0,870; p = .385)$) und ebenso keine signifikanten Unterschiede zwischen den Jahrgängen 7 und 8 ($t(253) = -0,295; p = .768$).
- **Mathematik** Bei N=260 gültigen Nennungen ergibt sich eine durchschnittliche Mathematiknote von 2,60. Es gibt dabei keine signifikanten Unterschiede zwischen Jungen und Mädchen ($(t(255) = -1,226; p = .221)$) und ebenso keine signifikanten Unterschiede zwischen den Jahrgängen 7 und 8 ($t(208) = 1,048; p = .296$).
- **Deutsch** Bei N=262 gültigen Nennungen ergibt sich eine durchschnittliche Deutschnote von 2,65. Es gibt dabei signifikante Unterschiede zwischen Jungen und Mädchen ($(t(257) = -4,296; p < .001)$), wobei hier mit $d = .27$ ein kleiner Effekt zu Grunde liegt. Mädchen haben mit 2,41 im Schnitt bessere Noten als die Jungen mit 2,86. Keine signifikanten Unterschiede bestehen hier zwischen den Jahrgängen 7 und 8 ($t(260) = 1,115; p = .266$).

Private Nutzung von digitalen Medien Die private Nutzung von Tablets und Computern wurde in den Bereichen Vergnügen, Spielen, Informationsbeschaffung und Hausaufgaben unterteilt. Abbildung 12.5 zeigt dabei die Ergebnisse über alle Befragten. Im Bereich der Informationsbeschaffung und Hausaufgabenenerledigung zeigt sich dabei ein ausgeglichenes Bild. Es überwiegen über alle SuS Nennungen von *Manchmal*, *Selten* bzw. *Häufig*. Extrempositionen werden eher selten vertreten. Etwa drei Viertel der Befragten nutzt die Geräte nahezu *immer* oder *häufig* zum Vergnügen, womit etwa das Schauen von Youtube©-Videos gemeint sein kann. Das verbleibende Viertel der Befragten gibt an, die Medien *manchmal*, *selten* oder nahezu *nie* zu nutzen. Der Bereich des Spielens zeigt wieder ein ausgeglichenes Bild über die meisten Antwortoptionen mit Ausnahme von (*Fast*) *immer*. Das Spielen ist jedoch die einzige der abgefragten Kategorien, die einen erheblichen Unterschied zwischen den Geschlechtern zeigt. Die Jungen (Mittelwert der Skala 2,36) zeigen dabei eine signifikant häufigere Nutzung als die Mädchen (1,28) ($t(262) = -8,091; p < .001$), was einem großen Effekt von $d = .99$ entspricht.

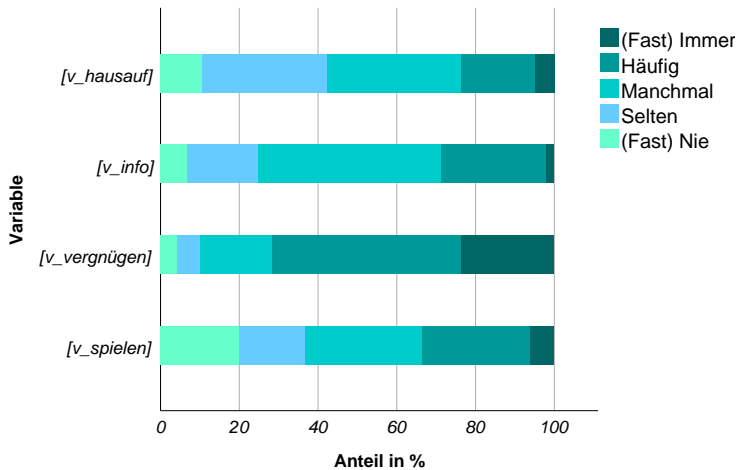


Abbildung 12.5.: Einsatz von Computer, Tablet oder Smartphone im privaten Bereich

Schulische Nutzung von digitalen Medien: (v_matheunterricht), (v_physikunterricht), (v_simulationsprogramme), (v_a) Computer, Tablets und Smartphones werden im Mathematik- und Physikunterricht der befragten SuS eher selten eingesetzt (siehe Abb. 12.6). 50,2% aller Befragten nutzen entsprechende Medien im Mathematikunterricht (fast) nie, im Physikunterricht sind es sogar 59,5%. Immerhin 15,2% (Mathematikunterricht) bzw. 23,8% (Physikunterricht) der Schüler nutzen diese Medien zumindest selten. Da diese Werte auf individueller Schülerebene kaum Unterschiede zeigen dürfte, dafür eher auf Klassenebene, werden diese im Folgenden untersucht.

Jedoch sind mittels ANOVA bezüglich der Nutzung von Simulationen und dem Einsatz von Tablets etc. im Physikunterricht für den Skalenmittelwert keine signifikanten Unterschiede zwischen den untersuchten Klassen feststellen. Einzig im Falle des Mathematikunterrichts zeigen sich signifikante Unterschiede ($F(1, 10) = 5,331, p < .001$). So existieren Klassen mit im Schnitt gelegentlicher Nutzung der Medien, wie auch Klassen, in denen diese fast keiner Nutzung unterliegen.

Weiterhin denkbar wäre eine unterschiedliche Verbreitung der Medienutzung in den unterschiedlichen Jahrgängen. Bei keinem der drei Items lässt sich aber entsprechender Unterschied per t-Test nachweisen.

Der Einsatz von Simulationen im Physikunterricht erfolgt nach 88,7% der Befragten (fast) nie. 9,8% geben jedoch einen seltenen Einsatz von Simulationen an. In den Gruppen *Zugang*, *Jahrgangsstufe* oder *Lerngruppe*

lassen sich per t-Test oder MWU-Test keine signifikanten Unterschiede zeigen.

Die Frage nach dem Einsatz anderer digitaler Medien [v_a] wurde nur von 7 SuS beantwortet und wird hier im Rahmen quantitativer Betrachtungen nicht weiter verfolgt.

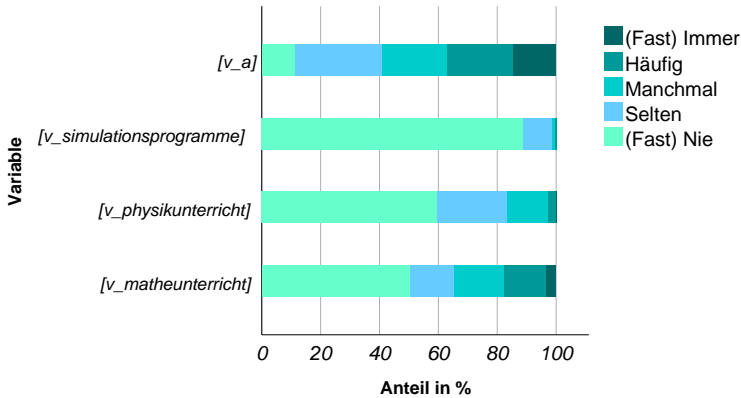


Abbildung 12.6.: Einsatz von Computer, Tablet oder Smartphone im jeweiligen Unterricht / von Simulationen im Physikunterricht

Interesse

Die zugehörigen Fragen zum Konstrukt finden sich in Tabelle A.1.

Tabelle 12.2.: Übersicht zur Skala Interesse

Frage	$\alpha_{Cronbach}$	Trennschärfe	$N_{gültig}$	Fehlend _{Anteil}
Pre Interesse	.751	.511 – .633	260	4,1%
Post Interesse	.774	.582 – .661	232	5,3%

Vorfragebogen Die Kennwerte zur Skala liegen im guten, akzeptablem Bereich (vgl. Tabelle 12.2). Im *Vorfragebogen* bestehen in den Untergruppen Geschlecht und Zugang dabei signifikante Unterschiede. Das Interesse von Jungen (2,40) unterscheidet sich im Vorfragebogen hochsignifikant von dem der Mädchen (1,84) mit $(t(252) = -5,471; p < .001)$ und einem mittleren Effekt ($d = .68$). Ebenso unterschiedlich ist das Interesse vor der Erhebung in den jeweiligen Zugängen, (1D: 2,30; 2D: 1,96), welches mit

($t(258) = 3,187; p = .002$) signifikant und einem kleinen Effekt ($d = 0,37$) zwischen 1D und 2D unterscheidet. Da die unterrichteten Zugänge klassenweise aufgeteilt wurden, legt dies den Verdacht nahe, dass Unterschiede im Interesse auf Klasseneffekten beruhen könnten. Eine ANOVA mit $F(1, 10) = 3,232, p = .001$ bestätigt diese Vermutung zunächst. Auch nach Ausschluss einer Klasse mit auffällig niedrigem Interesse bleibt dieses Ergebnis signifikant.

Nachfragebogen Im *Nachfragebogen*, der zu dieser Skala das Interesse am Gegenstand der Intervention abfragt, bestehen zwischen dem 1D- und 2D-Zugang kein signifikanter Unterschied mehr ($t(230) = 1,292; p = .198$). Ebenso sind die auf den Vorunterricht bezogenen signifikanten Unterschiede zwischen Jungen und Mädchen *nach* der Intervention nicht mehr feststellbar ($t(224) = -1,684; p = .094$). Ähnliche Werte ergeben sich für potentielle Unterschiede zwischen Feedback-Varianten und Jahrgangsstufen.

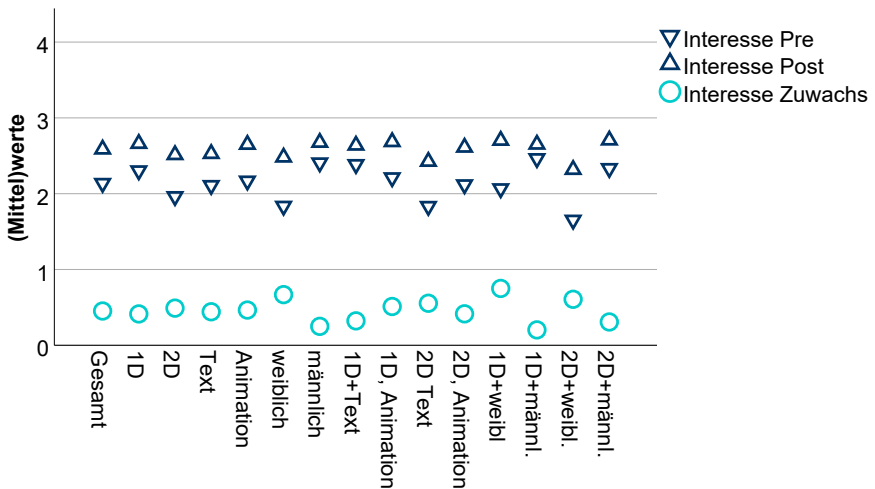


Abbildung 12.7.: Mittelwert und Zuwachs des Interesses

Vergleich Das Interesse im Pre-Fragebogen erreicht über alle Schüler einen Wert von 2,12. Das Interesse im Post-Fragebogen, bezogen auf die Intervention, 2,59 (N=245) (N=224 paarweise vorhandene Skalen). Dies entspricht einem höchst signifikanten Unterschied ($t(223) = -6,079; p < .001$) und einem mittleren Effekt ($d = .45$).

Wie Abbildung 12.7 schon vermuten lässt, zeigt sich in allen betrachteten Untergruppen ein signifikanter Unterschied zwischen Pre- und Postfragebogen bzw. Zuwachs (außer $\sigma + 1D$). In der Regel sind dabei mittlere Effekte zu verzeichnen, ohne dass sich „parallele“ Gruppen (1D/2D, Text/Animation. . .) deutlich unterscheiden. Einzig Jungen und Mädchen zeigen hinsichtlich des Effekts Unterschiede. So zeigt sich bei den Mädchen ein mittlerer Effekt ($d = .61$), bei den Jungen noch knapp ein kleiner Effekt ($d = .22$). Bei der Aufteilung nach Zugang und Geschlecht zeigen sich bei den Jungen keine signifikanten Unterschiede mehr. Bei den Mädchen sind solche hingegen festzustellen. Die zugehörigen Effekte sind im mittleren Bereich angesiedelt – im Zusammenhang mit dem 1D-Konzept etwas stärker als mit dem 2D-Konzept ($d = .42$ vs. $d = .37$). Eine ANOVA zeigt allerdings, dass die Unterschiede auch in diesen Untergruppen auf das Geschlecht und nicht auf den Zugang oder die Interaktion zurückzuführen sind.

Selbstkonzept

Die zugehörigen Fragen zum Konstrukt finden sich in Tabelle A.2.

Tabelle 12.3.: Übersicht zur Skala Selbstkonzept

Frage	$\alpha_{Cronbach}$	Trennschärfe	N _{gültig}	Fehlend _{Anteil}
Pre Selbstkonzept	.814	.640 – .680	264	2,6%
Post Selbstkonzept	.705	.472 – .578	232	4,5%

Die Kennzahlen der Skala liegen im Vorfragebogen im guten, im Nachfragebogen im akzeptablem Bereich.

Vorfragebogen Der durchschnittliche Kennwert der Skala bezüglich des Physikunterrichts liegt über alle Schüler und Schülerinnen bei 2,75. Hinsichtlich des Geschlechts gibt es signifikante Unterschiede in den Untergruppen *Zugang* und *Geschlecht*. Bei letzterem ist der Unterschied die Differenz zwischen 2,56 (Mädchen) und 2,94 (Jungen) ($t(256) = -3,418; p = .001$) und entspricht einem mittlerem Effekt ($d = .41$). Die Mittelwerte in den Zugängen unterscheiden sich von 2,57 (2D) zu 2,93 (1D), was ebenfalls einem mittlerem Effekt entspricht ($d = .40$) ($t(262) = 3,275; p = .001$). Dieser Unterschied bezüglich der vorgefundenen Bedingungen legt wie schon im Falle des Interesses Klasseneffekte nah. Eine daraufhin durchgeführte ANOVA auf Klassenebene zeigt eine entsprechende Signifikanz

($F(1, 10) = 2,323, p = .012$). Nach Ausschluss einer Klasse mit auffällig niedrigem Selbstkonzept sind mittels der ANOVA keine signifikanten Unterschiede zwischen den Gruppen nachzuweisen ($F(1, 9) = 1,140, p = .335$).

Nachfragebogen Der durchschnittliche Kennwert der Skala des Selbstkonzepts bezüglich der Intervention über alle Schüler beträgt im *Nachfragebogen* 3,17. In der Untergruppe *Zugang* ist dabei kein signifikanter Unterschied mehr festzustellen – unabhängig von der (Nicht-)Berücksichtigung der oben angesprochenen Lerngruppe. In der Gruppe *Geschlecht* besteht nach wie vor ein knapp signifikanter Unterschied ($t(226) = -1,980; p = .049$), der Effekt ist jedoch kleiner ($d = .30$).

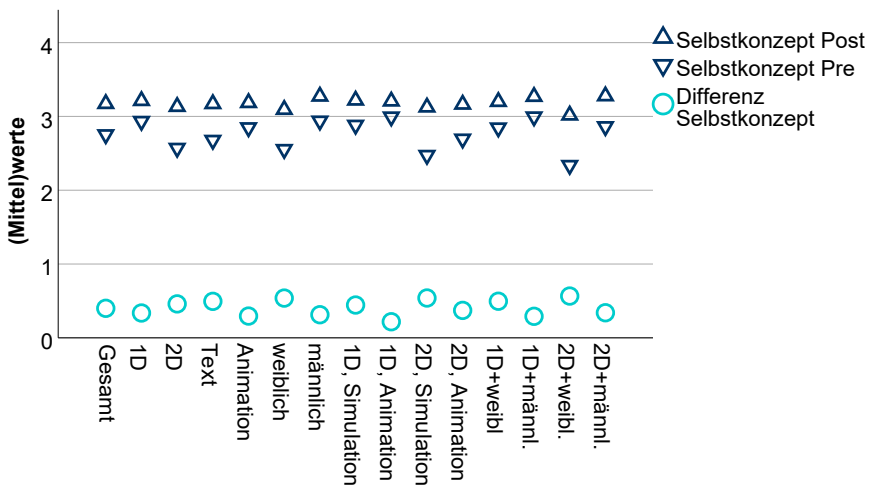


Abbildung 12.8.: Mittelwert und Zuwachs des Selbstkonzepts

Vergleich Im Vergleich zwischen Selbstkonzept im Vorunterricht und interventionsbezogenem Selbstkonzept zeigen sich in den Untergruppen *Zugang* (je $p < .001$), *Feedback* (Text: $p < .001$; Anim.: $p = .002$) und *Geschlecht* (♀ : $p < .001$; ♂ : $p = .001$) jeweils signifikante Unterschiede. Die Effekte sind dabei im 2D-Zugang relativ am größten ($d = .49$; 1D $d = .36$), ebenso bei den Mädchen ($d = .60$; Jungen $d = .30$). Entsprechend profitieren die 2D-unterrichteten Mädchen hinsichtlich des Selbstkonzepts am meisten (siehe Abbildung 12.8).

Flow

Die zugehörigen Fragen zum Konstrukt finden sich in Tabelle A.3.

Tabelle 12.4.: Übersicht zur Skala Flow

Frage	$\alpha_{Cronbach}$	Trennschärfe	N _{gültig}	Fehlend _{Anteil}
Pre Flow	.530	.306 – .380	257	5,2%
Post Flow	.618	.367 – .520	243	2,4%

Die Kennwerte der Skala (siehe Abb. 12.4) liegen im eher schlechten, aber noch akzeptablen Bereich – fallen im Nachfragebogen dabei aber günstiger aus. Dies könnte in der spezifischeren Formulierung der Items im Nachfragebogen begründet sein.

Vorfragebogen Der Kennwert für das Flow-Erleben im Vorfragebogen liegt bei 2,31. Auch hier lassen sich signifikante Unterschiede bezüglich des Zugangs feststellen ($t(255) = 2,418; p = .016$). Klasseneffekte liegen entsprechend einer ANOVA jedoch nicht vor. Entsprechend lässt sich diese Tatsache auch nicht durch das probeweise Ausschließen womöglich auffälliger Klassen abmildern. Des Weiteren ist ein signifikanter Unterschied zwischen Jungen und Mädchen festzustellen ($t(249) = -2,381; p = .018$) – der zugehörige Effekt ist jedoch klein ($d = .26$). Es gibt keine signifikanten Unterschiede in der Feedbackvariante.

Nachfragebogen Der Wert für das Flow-Erleben im Nachfragebogen liegt bei 2,75. Dabei zeigt sich kein signifikanter Unterschied zwischen den Zugängen mehr, gleiches gilt nach wie vor für die Unterscheidung nach Feedback-Variante. Ebenfalls ohne signifikante Unterschiede fällt der Vergleich innerhalb der Interaktion zwischen Feedback und Zugang aus. Auch eine Unterscheidung nach beiden Geschlechtern liefert sehr ähnliche Mittelwerte und mithin keine signifikanten Unterschiede.

Vergleich Der Unterschied zwischen Flow und interventionsbezogenem Flow ist in den Untergruppen durchweg signifikant (jeweils $p < .001$). Wie schon im Falle des Selbstkonzepts verzeichnen die Mädchen dabei einen größeren Effekt als die Jungen ($d = .52$ vs. $d = .34$). Im 2D-Zugang herrscht ein größerer Effekt vor als im 1D-Zugang ($d = .41$ vs. $d = .45$). Mithin zeigen auch die Untergruppen ♀+1D und ♀+2D größere Effekte als ihre männlichen Pendanten.

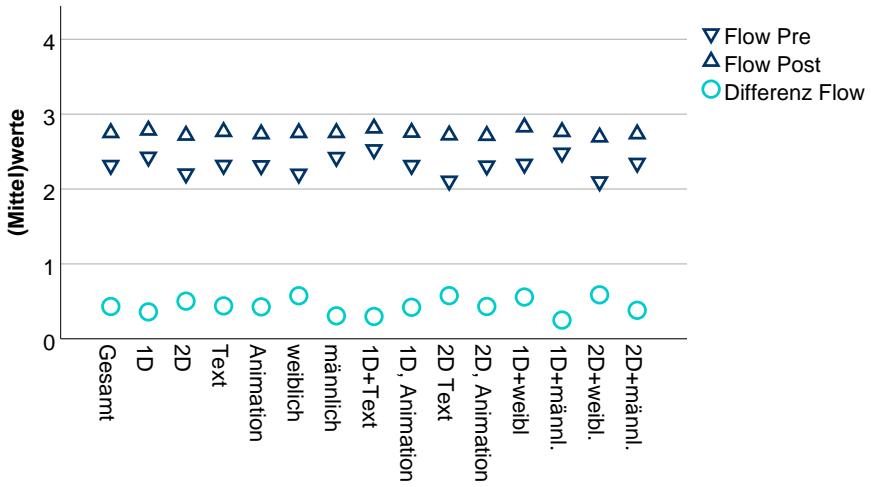


Abbildung 12.9.: Mittelwert und Zuwachs des Flows

Einzelne Fragen Vorforschungsbogen

(v_touchscreen), (v_spielebedien) Die Kompetenz Touchscreens zu bedienen sowie interaktive Elemente in Spielen zu identifizieren, wird von den Schülerinnen und Schülern hoch eingeschätzt (vgl. Abb 12.10). 90,6% aller Befragten stimmen der problemlosen Touchscreenbedienung voll oder eher zu, immerhin noch 71,7% der problemlosen Identifikation entsprechender Elemente in Spielen. In den Untergruppen unterscheiden sich Mädchen und Jungen signifikant ($t(251) = -3,036; p = .003$) hinsichtlich letztem Punkt. Dabei ist ein kleiner Effekt ($d = .38$) zu beobachten.

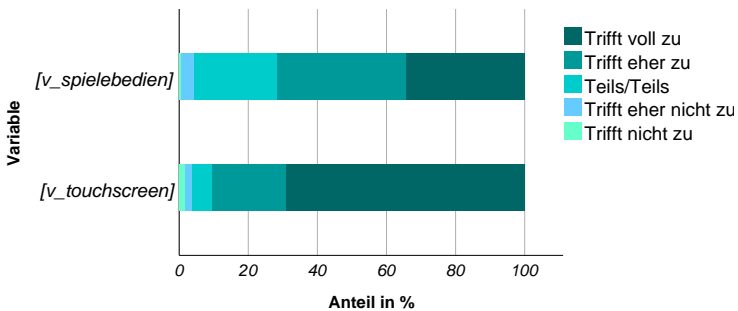


Abbildung 12.10.: Ergebnis der Fragen zur selbst eingeschätzten Bedienkompetenz von Touchscreens und Spielen.

(v_kümmert), (v_hilfteinzel), (v_schlimmes), (v_geduld) Die Fragen zur vorherrschenden Fehlerkultur im Unterricht der jeweiligen Lehrkraft⁵ zeigen einen deutlichen Deckeneffekt (siehe auch Abschnitt *Qualitätssicherung*). Über alle Schüler betrachtet stimmen 45,7% (v_geduld) bis 61,1% (v_schlimmes) jeweils der höchsten Ausprägung *Trifft voll zu* zu. Gemeinsam mit der zweithöchsten Kategorie *Trifft eher zu* ergeben sich Zustimmungen von 79,2% (v_geduld) bis 89,3% (v_schlimmes). Alle weiteren Kategorien erhalten maximal 15,7% der Zustimmung (siehe Abbildung 12.11). Signifikante Geschlechterunterschiede sind per t-Test nicht nachzuweisen.

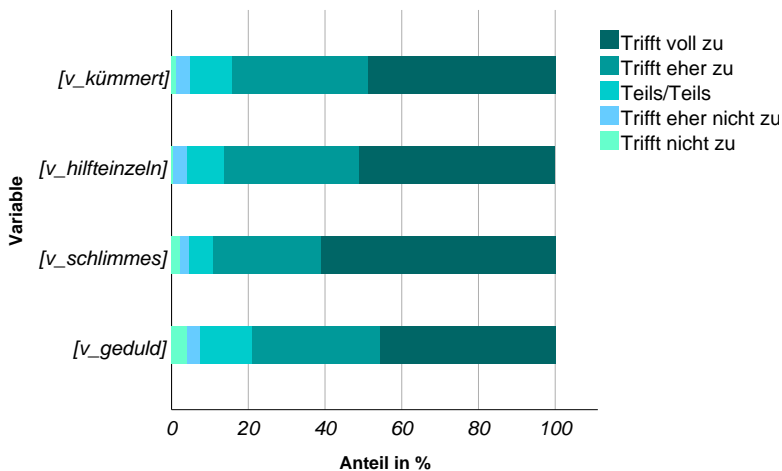


Abbildung 12.11.: Ergebnis der Fragen zur Fehlerkultur seitens der Lehrkraft.

Nutzen des Feedbacks: (n_empfand), (n_anklickfalsch), (n_unterbrochen)

Von 56,8% der Schüler wurde das jeweilige Feedback voll (oder eher) als hilfreich empfunden. 17,8% nehmen die gegenteilige Position ein, während sich ein Viertel unentschlossen zeigt (siehe Abbildung 12.12). Zu bemerken ist hierbei, dass sich keinerlei signifikante Unterschiede zwischen den Feedbackvarianten zeigen (ebenso wie zwischen den Zugängen wie auch zwischen den Geschlechtern). Im Sinne des formativen Charakters des Feedbacks in beiden Varianten zeigten sich 73,1% der Befragten (auch eher) nicht enttäuscht, so sie eine Frage der Multiple-Choice-Fragen falsch beantwortet haben. Nur 13,3% nehmen die gegenteilige Position ein (13,6%

⁵In der Hauptstudie als einzelne Fragen ausgewertet. Siehe Vorstudie.

neutral). Signifikante Unterschiede lassen sich zwischen den Zugängen, Feedbackvarianten oder Geschlechtern nicht feststellen (*n_anklickfalsch*). 13,4% der Befragten zeigten sich durch das Feedback in ihren Gedankengängen gestört, während 70% dem (auch eher) nicht zustimmen. Es sind keine signifikanten Unterschiede zwischen Untergruppen feststellbar (*n_unterbrochen*). Für 45,4% der Schüler ist das automatische Feedback dahingehend (auch eher) unkritisch, als dass sie sich *kein* ersatzweises Feedback von der Lehrkraft wünschen. Während ein Drittel dieser Aussage neutral gegenübersteht, zeigen aber 20% der Schüler (auch eher) Interesse an Rückmeldungen von der Lehrkraft statt automatisch vom Lernprogramm. Auch hier bestehen keine signifikanten Unterschiede zwischen den Feedbackvarianten.

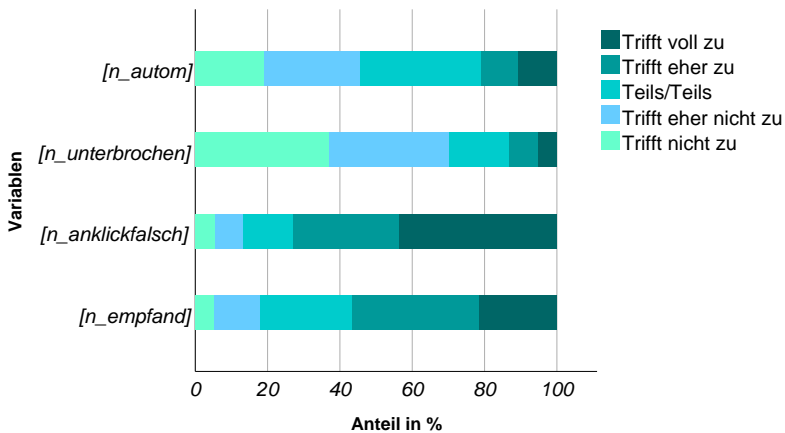


Abbildung 12.12.: Ergebnisse der feedbackbezogenen Fragen

Bedienung: (n_simprob), (n_lernprob), (n_bedienbar), (v_steuern) 86,7% aller Schüler und Schülerinnen beschreiben das Bedienen der Simulationen per Touchpad als voll oder eher unproblematisch. 84,1% geben dies auch für die Bedienung des Lernprogramms an. 72,8% der Schüler sind der Auffassung, dass das Steuern des Lernprogramms (also z. B. das Wechseln zwischen den einzelnen Seiten) (auch eher) einfach war. 79,4% der Befragten geben an, dass sie bedienbare Elemente der Simulationen (auch eher) gut identifizieren konnten (vgl. Abb. 12.13). Der Unterschied zwischen Mädchen und Jungen im Bezug auf die Identifikation interaktiver Elemente bleibt im Vergleich zum Vorforschfragebogen aber fast unverändert bestehen ($t(230) = -3,026; p = .003$). Dieses Item ist auch das einzige der vier, in welchem sich signifikante Unterschiede nachweisen lassen.

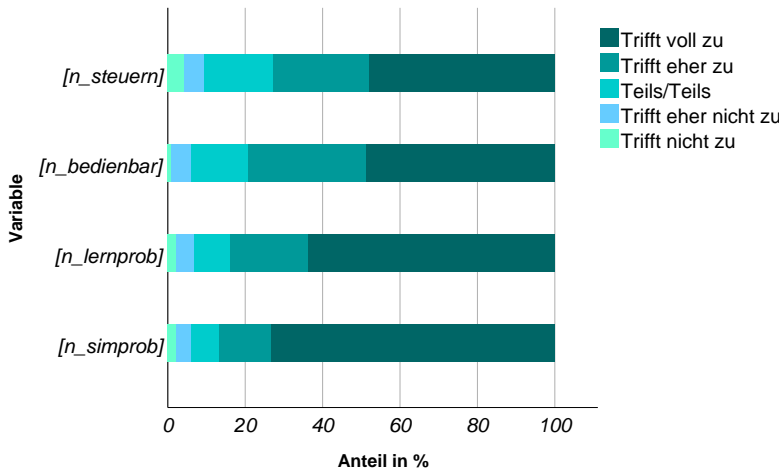


Abbildung 12.13.: Ergebnis der Fragen zur selbst eingeschätzten Bedienkompetenz von Simulationen und Lernprogramm.

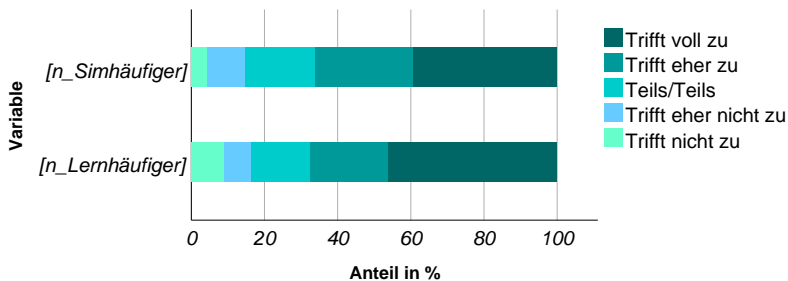


Abbildung 12.14.: Ergebnis der Fragen nach dem Wunsch zum häufigeren Einsatz von Lernprogramm oder Simulationen

Zukunftsperspektive: (n_simhäufiger), (n_lernhäufiger) 67,8% aller Schülerinnen und Schüler möchten ähnliche Lernprogramm häufiger im Unterricht einsetzen, bzw. stimmen dieser Aussage voll sowie eher zu. 17,2% stimmen der Aussage nicht oder eher nicht zu. Ähnlich gelagert ist die Zustimmung zur Frage, ob ähnliche Simulationen – ohne ein Lernprogramm – häufiger eingesetzt werden sollen. Hier liegt die volle bzw. eingeschränkte Zustimmung bei 66,4%. 14,7% stimmen der Aussage nicht oder nur eingeschränkt zu (vgl. Abb. 12.14).

Während in den Untergruppen *Zugang* und *Feedback* keine signifikanten Unterschiede zu verzeichnen sind unterscheiden sich Jungen und

Mädchen deutlich hinsichtlich des Wunsches nach häufigerem Simulationseinsatz ($t(230) = -2,037; p = .043$). Hier zeigt sich auch ein kleiner Effekt ($d = .29$) zugunsten der Jungen. Bezüglich des Einsatzes von Lernprogrammen gibt es hier keine signifikanten Unterschiede.

Lernerfolg und Komplexität: 72,5% aller Schülerinnen und Schüler sind voll oder eher der Auffassung, dass sie sich bewusst darüber sind, was sie mithilfe der Lernumgebung und der Simulationen gelernt haben (vgl. Abb. 12.15). 9,6% stimmen dieser Aussage hingegen nicht oder eher nicht zu. Signifikante Unterschiede wurden in den Gruppen *Zugang*, *Feedback* oder *Geschlecht* nicht festgestellt. Auch Varianzanalysen über die Versuchsbedingungen oder über *Zugang* und *Geschlecht* zeigten keinerlei auffällige Unterschiede.

Ob die Lernumgebung „zu einfach“ gewesen sei bejahen voll oder eher 32,6% aller befragten. Allerdings lehnen auch ca. 28,9% der Schüler diese Aussage ab – was eine etwas größere neutrale bzw. „teils/teils“-Gruppe zur Folge hat. Diese Gesamtlage der Daten spricht für eine durchaus ausgewogene Schwierigkeit der Lernumgebung. Aufschlussreich ist hier aber eine Aufschlüsselung der Antwort nach beiden Geschlechtern getrennt. Das Empfinden einer „zu einfachen“ Lernumgebung der Jungen unterscheidet sich signifikant von dem der Mädchen ($t(231) = -3,793; p < .001$), wobei die Jungen sich eher unterfordert fühlten als die Mädchen. Dabei ist auch ein mittlerer Effekt zu verzeichnen ($d = .50$; MW ♂: 2,31; MW ♀: 1,79). Andere Untergruppen wie *Zugang*, *Feedback* oder *Zugang vs. Feedback* zeigen keine weiteren signifikanten Unterschiede.

Mit 57,9% sind die meisten befragten Personen voll oder eher der Auffassung, dass sie ohne die Simulationen – und nur mit den Lehrtexten – *nicht* „so gut“ gelernt hätten wie mit den Simulationen. Jedoch beziehen auch 18,3% der Schüler die gegenteilige Position während sich etwa ein Viertel unentschlossen zeigt. Auch hier sind keine signifikanten Unterschiede in den bisher betrachteten Untergruppen feststellbar.

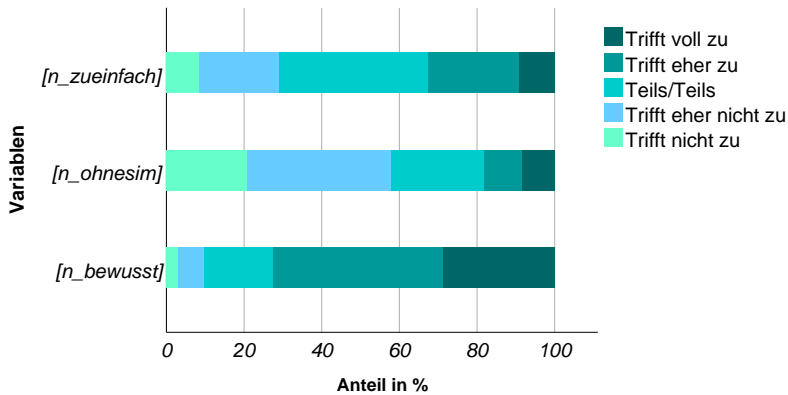


Abbildung 12.15.: Ergebnis der Fragen nach dem empfundenen Lernerfolg

Klarheit: (n_tunsollte), (n_verständlich) Nur 4,1% aller befragten Personen sind voll oder eher der Auffassung, dass Unklarheit über die Tätigkeiten mit den Simulationen herrschte. Die überwiegende Mehrheit (85,6%) hingegen lehnt die Aussage voll oder eher ab (vgl. Abb. 12.16). In keiner der üblicherweise betrachteten Untergruppen lassen sich dabei signifikante Unterschiede feststellen.

Ähnlich hoch sind die Zahlen im Falle der Textverständlichkeit gelagert. 78% der Befragten bezeichnen die Texte innerhalb der Lernumgebung als voll oder eher verständlich. Hingegen sind nur 4,2% aller Schülerinnen und Schüler gegenteiliger Ansicht. Wie schon bei vorangegangener Frage lassen sich auch hier keine Unterschiede in den Untergruppen feststellen (vgl. Abb. 12.16). Eine Korrelation zwischen der Deutschnote als vagem Indikator für Textverständnis und der empfundenen Verständlichkeit der Texte der Lernumgebung lässt sich nicht nachweisen.

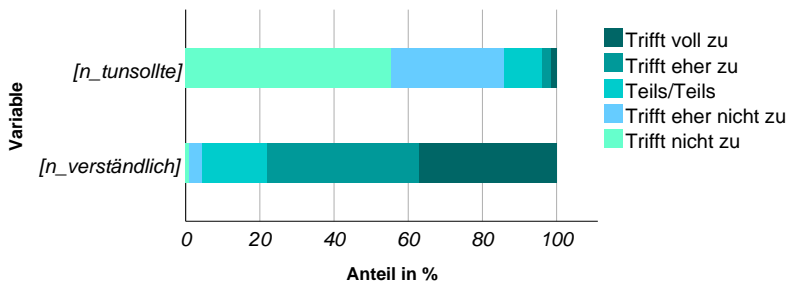


Abbildung 12.16.: Ergebnis der Fragen nach der empfundenen Klarheit

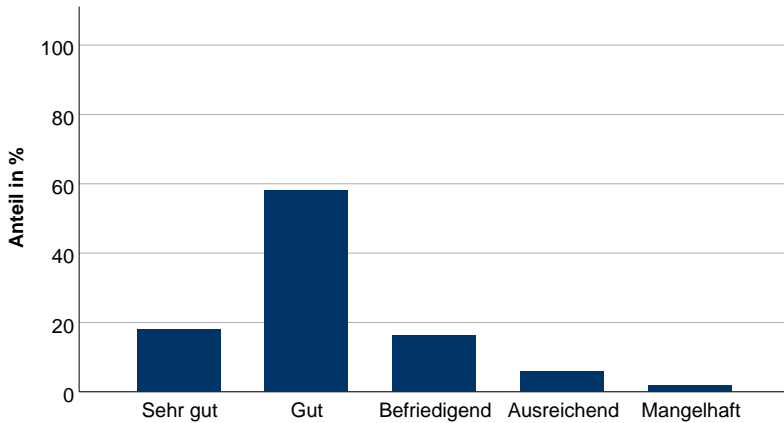


Abbildung 12.17.: Verteilung der für das Lernprogramm inkl. Simulationen vergebenen Schulnoten (Post, alle SuS).

(n.note) Wie Abb. 12.17 zeigt, wurde die Lernumgebung inklusive der Simulationen in Schulnoten als überwiegend *gut* beurteilt. Etwa jeweils 17% wählten die Bewertung *Sehr gut* bzw. *Befriedigend*, während ausreichende bzw. mangelhafte Noten von 7,7% der Schülerinnen und Schülern vergeben wurden. Letzte treten auch nur in zwei Lerngruppen gemeinsam auf, in vier weiteren vereinzelt. Signifikante Unterschiede in den Untergruppen Geschlecht, Zugang, Feedback und Jahrgangsstufe lassen sich nicht nachweisen. Zu beachten ist hier allerdings, dass der Anteil fehlender Werte 9,8% beträgt und nur N=222 gültige Werte vorliegen.

Zusammenfassung & Diskussion

In Kapitel 12.2.1 wurden die Ergebnisse zu nicht-kognitiven Konstrukten wie Interesse, Selbstkonzept und Flow wie auch zur Akzeptanz und Umgang mit der Lernumgebung sowie der Mediennutzung dargelegt. Im Folgenden werden einzelne Aspekte nochmals kompakt dargestellt und in einen Zusammenhang gestellt.

Computer oder Tablets werden im Physikunterricht sehr selten eingesetzt, eher häufiger noch im Mathematikunterricht. Speziell Simulationen werden im Physikunterricht sogar fast nie eingesetzt (bei 10% der SuS immerhin selten). Die unterrichtliche Erfahrung im entsprechenden Umgang mit solchen Medien ist also als gering einzustufen⁶. Die Nutzung

⁶Zu beachten ist hier natürlich, dass die Häufigkeit des Medieneinsatzes an sich kein Qualitätsmerkmal ist, auch der Einsatz der Medien *an sich* ist kein Qualitätsmerkmal

von Computern zur Informationsbeschaffung oder zum Erledigen von Hausaufgaben erfolgt eher im privaten Rahmen, wenn auch mit mäßiger aber regelmäßiger Ausprägung (die Nutzung von digitalen Medien zu Vergnügungszwecken dominiert hier wenig überraschend). Bei Mädchen und Jungen existiert hinsichtlich der privaten Nutzung ein großer Unterschied in der Häufigkeit von Spielen am Computer, was scheinbar auch in der seitens der Jungen größeren, selbsteingeschätzten Kompetenz bei der Bedienung von interaktiven Elementen resultiert. Auch im Verlauf der Intervention – bezogen auf die Bedienungen der Simulationen – bleibt dieser Unterschied bestehen. Generell sind bei der Bedienung von Simulation und Lernumgebungen keine grundlegenden Komplikationen seitens der Schüler bemerkt worden. Diese Datenlage lässt sich durch den Autor zwar weitestgehend bestätigen, jedoch kam es informellen Beobachtungen nach gerade zu Beginn der Unterrichtseinheit (siehe Kapitel 9) mitunter zu Problemen. Ein Art Eingewöhnungsphase sollte – gerade bei nicht allzu intuitiven Bedienschritten – also dringend einkalkuliert werden. Eine inhaltliche Unterforderung durch die Lernumgebung attestieren mit signifikantem Unterschied zu Mädchen vor allem Jungen (vgl. auch Einschätzung zum Selbstkonzept). Letztere wünschen sich auch häufiger als Mädchen einen häufigeren Einsatz von vergleichbaren Simulationen (vgl. Nutzung von Spielen im privaten Bereich).

Das Interesse an Physik(unterricht) der Schüler im Vorfeld zeigte sich zwischen Mädchen und Jungen unterschiedlich, was die Literatur zumindest in Bezug auf das Fachinteresse auch nahelegen (siehe Kapitel 8). Hinsichtlich des Interesses bezogen auf die Intervention zeigt sich, dass dort keine signifikanten Unterschiede zwischen den Geschlechtern mehr nachweisbar sind – die Mädchen also größere Zuwächse zeigen. Wenn auch in deutlich kleinerem Maße, aber dennoch zeigen auch die 1D-Schüler/innen größere Zuwächse als 2D-Schüler/innen. Mädchen in 1D- und 2D-Zugängen sind also grundsätzlich Profiteure. Über die Gründe kann hier nur spekuliert werden, die konsequente Kontextualisierung kann aber durchaus als Faktor in Betracht gezogen werden.

Wie schon im Falle des Interesses bestehen im Vorfeld der Studie prinzipiell Unterschiede im Selbstkonzept zwischen Mädchen und Jungen, wobei die Jungen sich diesbezüglich als selbstsicherer einschätzen. Bezogen auf die Intervention konnte ein Geschlechtereffekt unabhängig vom Unterrichtskonzept abgemildert, wenn auch nicht beseitigt, werden. Nichtsdestotrotz stellte sich absolut bei (2D-)Schülerinnen ein erheblicher Zuwachs des Selbstkonzepts ein. Betrachtet man nämlich die Zugänge, so zeigt sich für 2D-Schüler ein größerer Zuwachs. Über die Gründe des

für guten Physikunterricht. Es geht hier lediglich um Schaffung einer Datenbasis zur Einschätzung der Kompetenz der Schüler.

Rückgangs des Geschlechtereffekts kann auch hier keine sichere Aussage getroffen werden. Ein denkbarer Faktor könnte hier das autonome Arbeiten, weitestgehend ohne Einflüsse anderer Mitschüler/Lehrkräfte, und somit womöglich ein gewisser Grad an Selbstwirksamkeit und Bestätigung bzw. neutraler Rückmeldung durch das Feedback sein. Für das Flow-Konstrukt zeigen sich ebenfalls der bekannte Geschlechtereffekt wie auch ein – minimaler – größerer Zuwachs für 2D-unterrichtete Schüler. Das interventionsbezogene Flow-Erleben zeigt trotz unterschiedlicher Ausgangswerte keine bedeutsamen Unterschiede zwischen den Geschlechtern, Zugängen oder Feedbackvarianten, was unterschiedlich große Zuwächse bedingt.

Erfreulich ist die ausgeprägte, moderne *Fehlerkultur* in den untersuchten Gruppen. Es zeigte sich fast durchweg ein Bild von Klassen, in denen die Physiklehrkraft Fehler akzeptiert⁷ und nicht dramatisiert. Gleichzeitig werden sie als geduldig und individuell hilfsbereit charakterisiert. Das automatische Feedback der Lernumgebung wurde von über der Hälfte der Schüler als hilfreich empfunden. Ebenfalls die Hälfte aller Schüler hätte sich an Stelle des automatischen Feedbacks *kein* ersatzweises Feedback seitens der Lehrkraft gewünscht. Das stützt die These, dass automatisches Feedback von Schülern gerne angenommen wird, da es als neutral und „unbestechlich“ gilt (vgl. dazu auch Äußerung aus Interview „Der Computer weiss ja was richtig ist“). Fast dreiviertel der Schüler zeigen sich von falsch beantworteten MC-Fragen dazu nicht enttäuscht. Dies ist im Sinne formativen Feedbacks und spiegelt auch die beschriebene Feedbackkultur des Unterrichts wider. Hinsichtlich der zweiten Forschungsfrage zeigen sich bei allen Feedback-Items keine signifikanten Unterschiede zwischen den beiden Feedbackvarianten. Dies schließt auch ein, dass speziell von 2D-Schülern Animationsfeedback nicht als hilfreicher, weil entlastender, empfunden wird. An dieser Stelle muss kritisch angemerkt werden, dass nicht untersucht wurde, inwiefern das Feedback, das hier die Arbeit mit (bzw. Beobachtungen an) den Simulationen unterstützt, von den Schülern als *Selbstkontrolle der fachlichen Inhalte* interpretiert wurde. Aufgrund eines solchen Missverständnisses könnten die Schüler fälschlicherweise davon ausgehen, die fachlichen Inhalte verstanden zu haben – dabei betreffen die MC-Fragen und das Feedback hauptsächlich Beobachtungen in den Simulationen.

⁷Lern- vs. Leistungssituation beachten!

12.2.2. Kognitive Fragebögen

In diesem Kapitel werden die Ergebnisse der Vor- und Nachtests dargelegt. Für eine erste Übersicht zeigt Abbildung 12.18 die relativen Lösungshäufigkeiten der Items aus Pre- und Posttest.

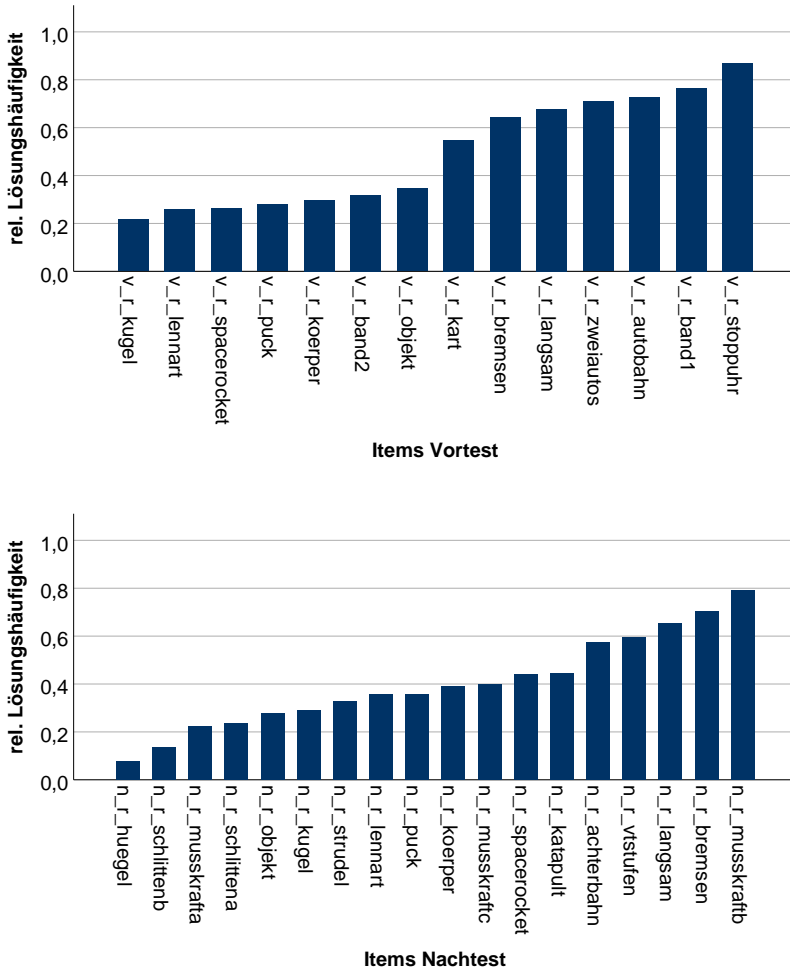


Abbildung 12.18.: Überblick über die relativen Lösungshäufigkeiten im Vor- und Nachtest

Wie schon in der Vorstudie zeigen sich unterschiedliche Lösungshäufigkeiten, so dass ein differenzierender Charakter angenommen werden

kann. Jedoch kann anhand der Grafik ein Fehlen von mittelschweren Items vermutet werden. Weiterhin zeigen sich dabei keine zu schwierigen Items und nur eines etwas zu leicht (*v_stoppuhr*). Im Nachtest fallen zwei Items als zu schwierig auf (*n_huegel* und *n_schlittenb*), ansonsten zeigt sich aber ein ausgewogenes Bild.

Zur Raschanalyse Im Bezug auf die geforderte Eindimensionalität erfüllt der **Pretest** die Bedingungen, da die weiteren ermittelten Komponenten einen zu kleinen Einfluss haben. Auch die Infit-Werte der Items liegen zwischen 0.93 und 1.12 und sind somit für die Auswertung des Fragebogens gut. Die Itemreliabilität beträgt 0.98 und ist somit ebenfalls sehr gut.

Die Personen, die in der ersten Analyse durch zu hohe Outfit- bzw. Outfit-ZSTD-Werte auffielen, wurden zunächst aus einer erneuten Analyse ausgeschlossen. Nach Vergleich der alten mit den neuen Werten mittels Cross-Plot (vgl. Linacre und Wright 2010, ; auch Kapitel 10.3) konnte ein entscheidender Einfluss ausgeschlossen werden, daher wurden die Personen letztlich in der Analyse belassen. Einzig die Reliabilität des Tests ist mit $\alpha_{Cronbach} = .52$ nicht zufriedenstellend. Die spezifische Objektivität ist im Pretest prinzipiell gegeben, da die Rangfolge der Items hinsichtlich ihrer Schwierigkeit weitestgehend beibehalten bleibt.

Abgesehen von zwei gemeinsamen Items, deren Displacement zu groß schien (*v_objekt*, *v_spacerocket*), konnten die verbleibenden sechs Items für die Analyse geankert werden. Im entsprechend geankerten **Posttest** zeigen die Items mit einer Ausnahme sehr gute Infit-Werte von 0.91 bis 1.17, wobei das Item *n_musskrafta* den Wert von 1.2 um 0.04 Punkte überschreitet. Angesichts der doch recht wohlwollenden „Grenzen“ (vgl. Boone, Staver und Yale 2014) verbleibt das Item mit einer derart geringen Abweichung in der Analyse. Die Itemreliabilität beträgt 0.98 und ist daher sehr gut. Die Reliabilität des Tests ist mit $\alpha_{Cronbach} = .64$ gegenüber dem Pretest verbessert. Die spezifische Objektivität ist im Posttest insofern prinzipiell gegeben, als dass die Rangfolge der Items hinsichtlich ihrer Schwierigkeit exakt beibehalten bleibt.

Die Personen mit auffälligen Outfit-Werten wurden auch im geankerten Posttest mit oben beschriebenem Verfahren analysiert. Die betreffenden neun Versuchspersonen konnten letztlich aufgrund zu geringen Einflusses auf die Messung in der Analyse verbleiben.

Ergebnisse In Abbildung 12.19 sind die Personenfähigkeiten aller Schülerinnen und Schüler im Pre- und Posttest dargestellt. Es zeigt sich dabei generell eine Erhöhung der Personenfähigkeit von 0,00⁸ auf 0,30. Dies

⁸gerundet auf zwei Nachkommastellen

V. Ergebnisse

entspricht einer signifikanten Änderung ($t(259) = -4,048; p < .001$) und einem Effekt von $d = .23$ und ist somit klein.

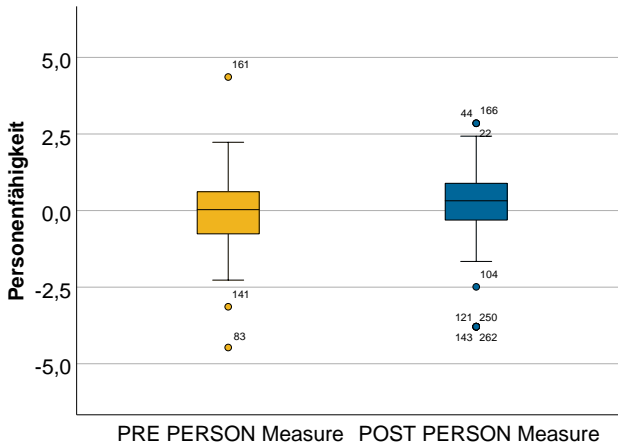


Abbildung 12.19.: Überblick der Personenfähigkeiten aller Schülerinnen und Schüler.

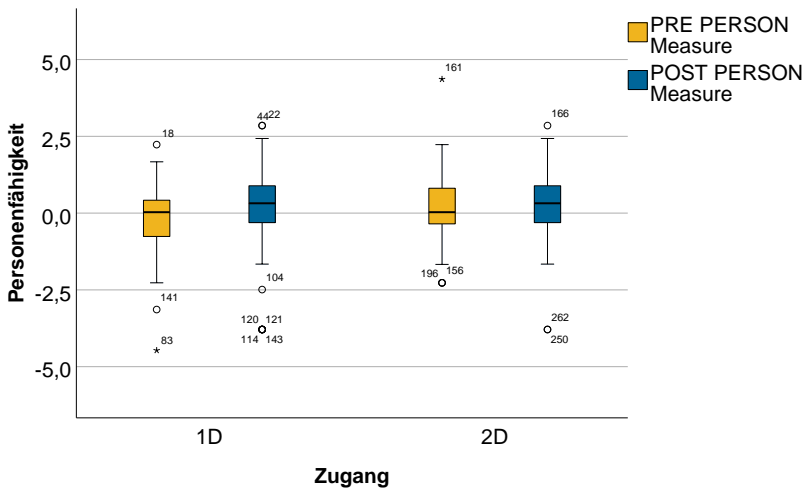


Abbildung 12.20.: Personenfähigkeiten der SuS, aufgeschlüsselt nach Zugängen

In Abbildung 12.20 sind die Personenfähigkeiten nach dem unterrichteten Zugang gesplittet. Es zeigen sich jeweils signifikante Zuwächse (1D: ($t(134) = -2,659; p = .009$); 2D: ($t(124) = -3,077; p = .003$)). Die Effekte sind dabei wiederum im kleinen Bereich angesiedelt (1D: $d = .23$; 2D: $d = .27$). Dabei liegt der Effekt der 2D-Gruppe höher, jedoch wohl nur minimal bedeutsam.

Im Vortest unterscheiden sich die Personenfähigkeiten der beiden Versuchsgruppen nicht signifikant. Wie schon die Abbildung nahelegt, lassen sich auch im Nachtest keine signifikanten Unterschiede zwischen 1D- und 2D-Gruppe feststellen.

In Abbildung 12.21 sind die Personenfähigkeiten hinsichtlich des genutzten Feedbacks dargestellt. Auch hier bestehen in beiden Untergruppen signifikante Unterschiede (Text: ($t(139) = -2,272; p = .025$); Animation: ($t(118) = -3,625; p < .001$)). Dabei sind wiederum jeweils kleine Effekte zu verzeichnen, wobei Schüler mit Animationsfeedback mit $d = .33$ einen größeren Effekt zeigen, als solche mit Textfeedback ($d = .22$).

Im Vortest gibt es keine signifikanten Unterschiede zwischen den Personenfähigkeiten von Schülern mit Text- und Animationsfeedback. Auch im Nachtest zeigt keine der beiden Gruppen gegenüber einer anderen einen signifikanten Zuwachs.

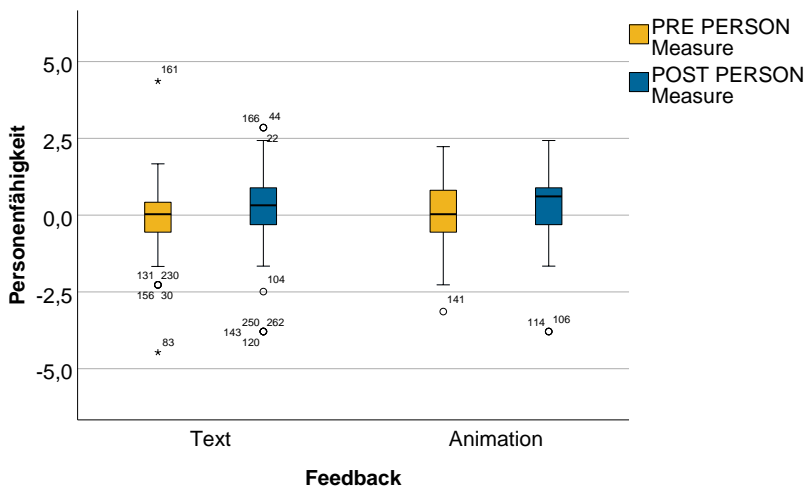


Abbildung 12.21.: Personenfähigkeiten, nach Feedback aufgesplittet.

Unterschiede hinsichtlich des Geschlechts und der Personenfähigkeit sind in Abb. 12.22 dargestellt. Dort sind nur bei den Jungen signifikante

Zuwächse in der Personenfähigkeit zu finden (w: ($t(125) = -1,866; p = .064$); m: ($t(127) = -3,834; p < .001$), Effekt hier $d = .34$). Allerdings zeigt hier ein nicht-parametrischer Wilcoxon-Test auch bei den Mädchen einen signifikanten Unterschied ($z(252) = -3,150; p = .002, r = .19$). Im Vortest zeigen sich signifikante Unterschiede in den Personenfähigkeiten von Mädchen und Jungen ($t(264) = -3,179; p = .002$), insgesamt mit einem kleinen Effekt ($d = .39$). Im Nachtest zeigen sich jedoch deutlich größere Unterschiede zwischen Mädchen und Jungen. Diese sind signifikant mit ($t(255) = 3,766; p < .001$), wobei der Effekt mit $d = .50$ nur knapp als mittelgroß und nicht mehr als klein zu werten ist – aber deutlich größer ausfällt als noch im Vortest.

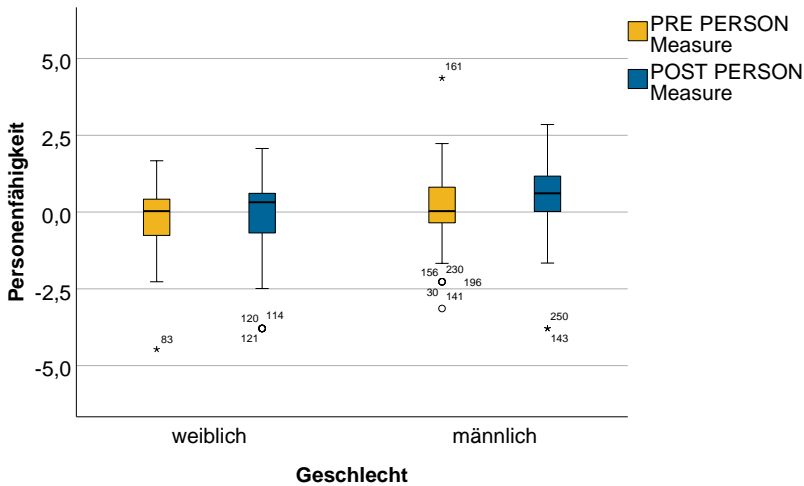


Abbildung 12.22.: Personenfähigkeiten aller SuS, aufgesplittet nach Geschlecht

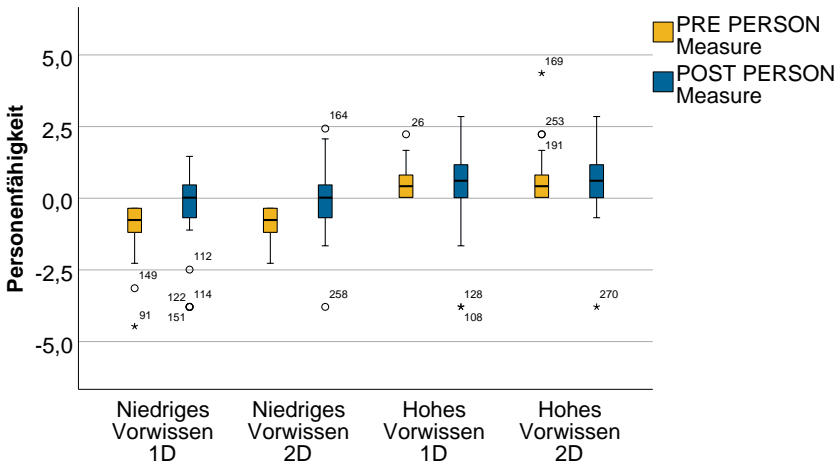


Abbildung 12.23.: Personenfähigkeiten aller SuS, aufgesplittet nach Vorwissen

Zur Unterteilung der Gruppe in Schüler hohen und niedrigen Vorwissens wurde wie schon in der Vorstudie ein Mediansplit der Personenfähigkeit des Vortests durchgeführt. Da eine erhebliche Anzahl von Personen dem Medianwert zugeordnet ist, ergeben sich unterschiedlich große Gruppen.

In der Gruppe niedrigen Vorwissens (Abbildung 12.23) zeigt sich über *alle* Schüler eine signifikante Änderung der Personenfähigkeit ($(t(106) = -5,959; p < .001)$) vom Pre- zum Posttest mit einem mittleren Effekt ($d = .58$). Im Gegensatz dazu zeigt sich solch eine signifikante Änderung bei Schülern hohen Vorwissens nicht. Zwischen den Gruppen bestehen, wie schon die Grafik nahelegt, sowohl im Vortest als auch im Nachtest signifikante Unterschiede mit erheblichem Effekt (Vortest: $(t(270) = -19,320; p < .001)$; Nachtest: $(t(258) = -6,249; p < .001)$). Diese Unterschiede bleiben prinzipiell auch bei veränderter Zuordnung der erwähnten Mediargruppe bestehen.

Schüler geringen Vorwissens zeigen zwischen den Zugängen keine signifikanten Unterschiede im Zuwachs der Personenfähigkeit. Dasselbe gilt für Schüler hohen Vorwissens.

Da die Studie im Doppeljahrgang 7/8 durchgeführt wurde, scheint auch eine Betrachtung der einzelnen Jahrgangsstufen sinnvoll (siehe Abbildung 12.24). Neben dem signifikanten Zuwachs in beiden Jahrgängen (7: $(t(182) = -1,697; p = .093)$; $z(178) = -2,934; p = .003$; 8: $(t(170) =$

$-3,747; p < .001$) wurde hier auch jeweils ein kleiner Effekt erzielt ($r = .29$ bzw. $r = .19$ in Klasse 7).

Im Vortest unterscheiden sich die Personenfähigkeiten von Schülern in 7. und 8. Jahrgangsstufe wenig überraschend signifikant ($(t(270) = -3,603; p < .001)$), der Effekt ist dabei noch klein ($d = .46$). Wie schon im Vortest unterscheiden sich die Personenfähigkeiten der Schüler auch im Nachtest signifikant voneinander ($(t(261) = -4,051; p < .001)$), wobei sich der Effekt leicht vergrößert hat ($d = .55$).

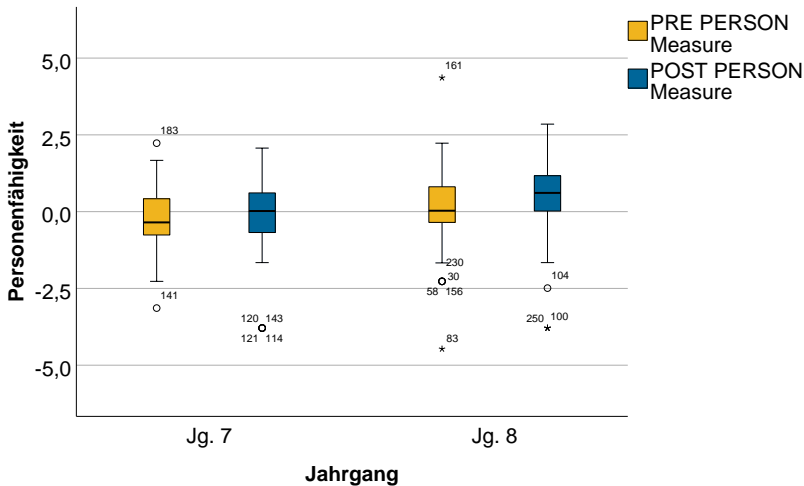


Abbildung 12.24.: Überblick der Personenfähigkeiten der SuS nach Jahrgangsstufe

Nach Versuchsbedingung aufgesplittet zeigen sich in beiden Zugängen mit Animationsfeedback signifikante Unterschiede zwischen Pre- und Posttest (1D+Animation: $(t(106) = -2,173; p = .034)$, $d = .21$; 2D+Animation: $(t(55) = -3,036; p = .004)$, $d = .41$). Die Effektgrößen sind dabei (sehr) klein, aber im 2D-Zugang deutlich größer. Siehe dazu auch Abbildung 12.25. Mittels t-Test lassen sich für die Versuchsbedingungen mit Text-Feedback keine signifikanten Zuwächse nachweisen. Mittels nicht-parametrischem Wilcoxon-Test zeigen sich auch im Textfeedback in beiden Zugängen signifikante Unterschiede (1D: $z(144) = -2,688; p = .007$, $r = .16$; 2D: $z(136) = -2,444; p = .015$, $r = .15$)

Hinsichtlich der Personenfähigkeit im Nachtest unterscheiden sich Schüler beider Feedbackarten weder innerhalb des 1D-Zugangs noch innerhalb des 2D-Zugangs. Innerhalb der Feedbackarten zeichnet sich ein ähnliches Bild. Schüler mit Animationsfeedback zeigen im Nachtest

zwischen 1D- und 2D-Zugang keine signifikanten Unterschiede, ebenso wie Schüler mit Textfeedback. Über alle Versuchsgruppen zeigt jedoch eine ANOVA mit den Faktoren Zugang und Feedback, dass sich die Zuwächse in den jeweiligen Untergruppen nicht unterscheiden.

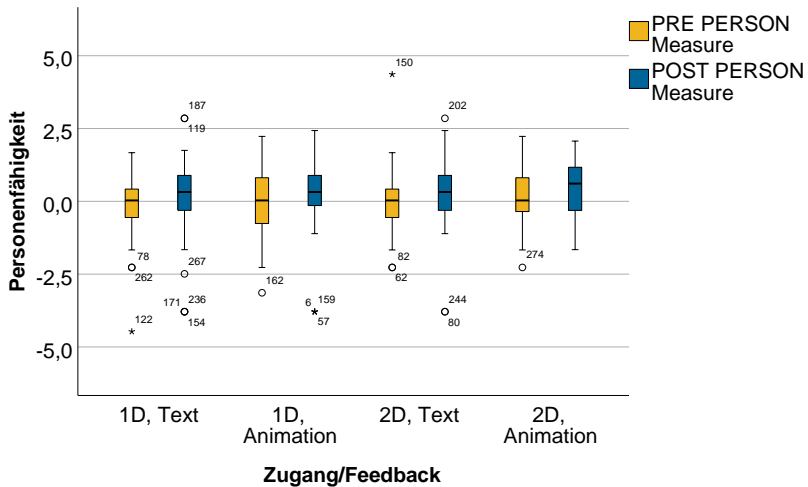


Abbildung 12.25.: Überblick der Personenfähigkeiten nach Versuchsbedingung aufgeteilt

Nach Tobias (2010) waren die traditionell (hier: 1D) unterrichteten Mädchen abschließend hinsichtlich ihrer Leistung nicht mit den Jungen auf einem Niveau. Ein ähnliches Bild scheint sich auch hier abzuzeichnen. Zunächst ist festzustellen, dass in der Gruppe der 1D-unterrichteten Mädchen kein signifikanter Unterschied der Personenfähigkeit zwischen Pre- und Posttest festzustellen ist. Bei den 1D-unterrichteten Jungen hingegen zeigt sich ein signifikanter Unterschied ($(t(72) = -3,752; p < .001)$) und ein Effekt von ($d = .44$), der größte der hier zunächst betrachteten Untergruppen. Siehe hierzu auch Abbildung 12.26.

Im Vortest unterscheiden sich Mädchen und Jungen, die nach dem 1D-Ansatz unterrichtet wurden, signifikant ($(t(135) = -2,417; p = .017)$). Auch im Nachtest bestehen erhebliche Unterschiede in der Personenfähigkeit zwischen Mädchen und Jungen. Die Werte unterscheiden sich signifikant ($(t(130) = -3,806; p < .001)$) und beruhen auf einem kleinen Effekt ($d = .33$).

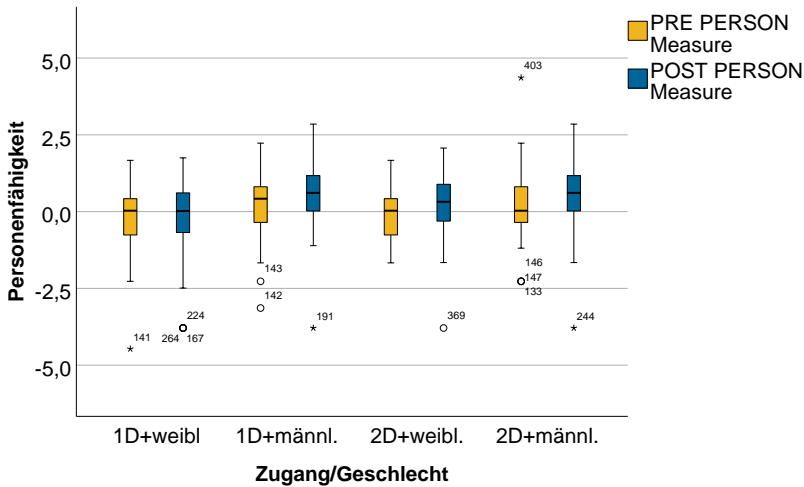


Abbildung 12.26.: Überblick der Personenfähigkeiten, unterteilt nach Geschlecht und Zugang.

Zunächst lässt sich feststellen, dass bei den Mädchen im 2D-Zugang zwischen Pre- und Posttest signifikante Unterschiede festzustellen sind ($t(67) = -2,417; p < .018$). Der Effekt für die Mädchen beträgt $d = .30$. Bei den Jungen hingegen zeigt sich mittels t-Test knapp kein Unterschied ($p = .088$; mittels Wilcoxon-Test: $z(110) = -2,271; p = .023, r = .22$). Siehe dazu auch Abbildung 12.26. Über alle Versuchsgruppen zeigt jedoch eine ANOVA mit den Faktoren Zugang und Geschlecht, dass sich die Zuwächse in den jeweiligen Untergruppen nicht unterscheiden.

Nach Tobias (2010) war eine Stärke des dort getesteten 2D-Ansatzes, dass Leistungsunterschiede zwischen Jungen und Mädchen ausgeglichen werden konnten. In der vorliegenden Studie scheint sich dieser Zusammenhang augenscheinlich zu bestätigen. Im Vortest unterschieden sich die Personenfähigkeiten bei beiden Geschlechtern signifikant ($t(127) = -2,154; p = .033$), wobei dem ein Effekt von $d = .16$ zugrunde liegt und dieser somit sehr klein ist. Im Nachtest hingegen unterscheiden sich Mädchen und Jungen hinsichtlich ihrer Personenfähigkeit nicht mehr signifikant ($t(123) = -1,547; p = .125$.)

Bearbeitung der Tempo-Aufgabe

Die Tempo-Aufgabe (siehe S. 274) wurde außerhalb des raschskalierten Tests ausgewertet, da die Aufgabe für die Teilnehmer je nach Zugang unterschiedlich zu bearbeiten war bzw. unterschiedlich als korrekt gewertet

wurde. Antwortoption *c* bezog sich dabei auf den korrekten Geschwindigkeitsbegriff in einer Dimension, die Optionen *a* und *d* auf die Interpretation nach 2D („(c) Tempo und Geschwindigkeit sind das Gleiche.“ bzw. „(a) Zwei Körper können sich mit gleichem Tempo, aber unterschiedlicher Geschwindigkeit bewegen.“ und „(d) Bewegen sich zwei Körper mit gleicher Geschwindigkeit, so bewegen sie sich automatisch auch in die gleiche Richtung.“). In Tabelle 12.5 sind die relativen Häufigkeiten der Nennung der Optionen dargestellt.

Tabelle 12.5.: Nennung von relevanten Antwortoptionen bei Tempo-Aufgabe (Anteil gültiger Nennungen in %)

	1D		2D	
	v_tempo	n_tempo	v_tempo	n_tempo
<i>c</i> genannt	78	64,3	72,2	34,7
<i>a</i> oder <i>d</i> genannt	5,5	10,2	7,2	25,8

In beiden Versuchsgruppen sind dabei Veränderungen feststellbar, die deutlichste in der 2D-Gruppe. Die Nennungen mindestens einer der korrekten Antwortoptionen haben sich mit 25,8% mehr als verdreifacht (Die einzig komplett korrekte Kombination von *a* und *d* wurde von 12,9% genannt). Jedoch wählten 34,7% der Befragten als Option weiterhin *c* aus. Die 1D-Schüler sollten hier theoretisch keine Änderungen erfahren, aber auch hier sinkt der Anteil der hier eigentlich korrekten Antwort *c*, während der nach 2D-Konzept korrekten Optionen steigt.

Zusammenfassung & Diskussion

In diesem Abschnitt sollen die zentralen Ergebnisse der fachlichen Tests kompakt zusammengefasst und gegebenenfalls bewertet werden.

Zunächst ist zu konstatieren, dass über alle SuS ein grundsätzlicher Lernerfolg (hier: Steigerung der Personenfähigkeit) mit kleinem Effekt messbar ist. In beiden Gruppen, 1D- und 2D-Zugang, ergaben sich signifikante Veränderungen der Personenfähigkeit vom Pre- zum Posttest. Die 2D-Gruppe hat dabei einen leicht größeren Effekt zu verzeichnen. Da im Nachtest aber keine signifikanten Unterschiede festzustellen sind, lässt sich hier über die Gründe spekulieren. Ein Faktor ist hier sicherlich die Kürze der Intervention mit etwa 2 Doppelstunden Nettolernzeit, was keine großen Effekte zu Tage fördern wird (was aber auch absehbar und kalkuliert war). Ein weiterer Faktor könnte die angestrebte Fairness zwischen den Zugängen sein, die sich in den Lernumgebungen widerspiegelt. Letzteres würde jedoch bedeuten, dass die Unterschiede zwischen den Zugängen verschwinden, sobald auf Geschwindigkeitspfeile, Zusatze-

schwindigkeit und 2D-Bewegungen verzichtet wird. Es wäre fatal, daraus voreilig Schlüsse zu ziehen.

Über alle Schüler haben solche mit Animationsfeedback größere Effekte zwischen Pre- und Posttest zu verzeichnen als solche mit Textfeedback, dennoch lassen sich im Posttest keine Unterschiede in der Personenfähigkeit identifizieren. Diese größeren Effekte einer Feedbackvariante lassen sich zunächst schwierig einordnen, weil es im Rahmen der nicht-kognitiven Fragebögen und auch Interviews keine Unterschiede in der Akzeptanz oder empfundenen Nützlichkeit des Feedbacks messbar waren und so keine Hinweise geben könnten.

Betrachtet man die einzelnen Untergruppen des Versuchsdesigns so zeigten sich vor allem in Gruppen des Animationsfeedbacks signifikante Unterschiede zwischen Pre- und Posttest, wobei besonders im 2D-Zugang ein kleiner Effekt zu verzeichnen war. In Kombinationen mit den Zugängen zeigt sich also durchaus ein Unterschied zwischen den Feedbackvarianten. In den nicht-kognitiven Fragebögen konnte ein Unterschied dabei nicht herausgearbeitet oder schlicht nicht abgebildet werden. In den Gruppen mit Textfeedback ließ sich ein Unterschied zwischen Pre- und Posttest in beiden Zugängen nur mittels nichtparametrischer Verfahren nachweisen, wobei sich die Effekte je nach Zugang nicht unterschieden. Die Gruppe mit dem größten Zuwachs stellt mithin die 2D-Gruppe mit Animationsfeedback dar.

Über alle Schüler zeigen Mädchen und Jungen signifikante Unterschiede der Personenfähigkeit mit jeweils kleinen Effekten. Da aber auch schon im Vortest Unterschiede bestehen, können die Mädchen ihre Personenfähigkeit denen der Jungen nicht angleichen (hier liegt sogar ein mittlerer Effekt vor). Bezieht man das Unterrichtskonzept mit ein, lassen sich die Ergebnisse von Tobias (2010) bestätigen. Zuvor bestehende Unterschiede zwischen Mädchen und Jungen im 2D-Konzept ließen sich nach der Intervention nicht mehr als signifikant nachweisen. Im Rahmen des 1D-Lehrgangs ließ sich diese Lücke nicht schließen.

Erhebliche Unterschiede zeigen sich zwischen Schülern unterschiedlichen Vorwissens (hier Abschneiden im Pretest anhand Mediansplit⁹). Zu vermuten wäre – aufgrund der im Theorieteil diskutierten größeren Einfachheit –, dass sich der 1D-Ansatz für Schüler niedrigeren Vorwissen hinsichtlich des Zuwachses der Personenfähigkeit günstiger darstellt. Diese mögliche Vermutung lässt sich anhand der Datenlage nicht bestätigen.

⁹Der bei der Einteilung vorgenommene Mediansplit kann hierbei aber grundsätzlich kritisiert werden. Bei Medianvergleichen werden Schüler ähnlicher Fähigkeit (im mittleren Bereich) evtl. in zwei Gruppen getrennt und Schüler unterschiedlicher Fähigkeiten (mittel und sehr schwach) in eine Gruppe eingeteilt – auch wenn eine gruppeninterne Tendenz der Personenfähigkeit natürlich abzusehen ist. Als Alternative böte sich künftig etwa ein Extremgruppenvergleich an.

Es sind zwischen den Zugängen keine signifikanten Unterschiede festzustellen (ebenso für Schüler hohen Vorwissens). Besondere Auffälligkeit der Daten ist jedoch, dass Schüler geringen Vorwissens im Nachtest kaum die Personenfähigkeit der Schüler hohen Vorwissens im Vortest erreichen. Dies legt die Vermutung des Vorwissens als *erheblichen* Faktor für den Wissenszuwachs innerhalb dieses Lernsettings nahe!

12.2.3. Vergleich zwischen Haupt- und Vergleichsstudie

Selbstverständlich sollen die Ergebnisse der Hauptstudiengruppe mit denen der Vergleichsgruppe verglichen werden. Zur Erinnerung: Die Vergleichsgruppe füllte lediglich die Fachtests aus ohne einer Lernumgebung oder einem bestimmten Lernkonzept gefolgt zu sein.

Zu beachten ist, dass die Vergleichsgruppe einen umfangreicheren Fachtest bearbeitete als die spätere Hauptgruppe (siehe Anhang A.2.1). Dies sollte bei der späteren Interpretation berücksichtigt werden. Um jedoch von vornherein eine gewisse Fairness zu gewährleisten, wird die Vergleichsgruppe im Folgenden auf die Items reduziert, die sie mit der Hauptgruppe gemein hat. Eine Aufstellung der relativen Lösungshäufigkeiten *aller* bearbeiteten Items befindet sich im Anhang auf S. 304. Die dortigen Ergebnisse wurden auch für die Auswahl von Items der Vor- und Nachtests der Hauptstudie genutzt.

Die Daten der insgesamt $N=375$ Versuchspersonen¹⁰ werden nun in einer Raschanalyse in einem *gemeinsamen* Analyseschritt ausgewertet. Dieses bietet sich gegenüber Anchoring an, da alle Items identisch sind. Es handelt sich also um eine Betrachtung von Untergruppen (Haupt- und Vergleichsstudie) innerhalb der selben Analyse.

Qualitätssicherung Da es sich um eine neue Analyse handelt, müssen Aspekte zur Qualitätssicherung erneut dargelegt werden.

Das Konstrukt zeigt sich in der Winsteps PCA für eine Analyse als ausreichend eindimensional. Bei der Analyse der Outfit-MNSQ -Werte der **Personen** ergaben sich lediglich für 5 Personen nicht modellkonforme Werte, sodass diese aus einer ersten Analyse entfernt wurden. Nach Gegenüberstellung mit der bereinigten Analyse zeigten sich keine nennenswerten Unterschiede, sodass letztlich alle Personen verbleiben konnten. Alle **Items** zeigen (sehr) gute Infit-Werte zwischen 0.88 und 1.26. Die Reliabilität des Tests beträgt $\alpha_{Cronbach}=.62$, was nicht gut, aber für dieses Projekt im akzeptablen Rahmen ist.

Ergebnisse Abbildung 12.27 zeigt den Vergleich zwischen den Personenfähigkeiten der Vergleichs- und Hauptgruppe des Nachtests. Schon eine Inaugenscheinnahme des Diagramms zeigt keine deutlichen Unterschiede, weder in Verteilung, Lage oder Spannbreiten der Quartile. Sowohl ein t-also auch ein Mann-Whitney-U-Test ($z(380) = -1,517; p = .605; (t(378) = 1,497; p = .135)$) bestätigen diese Annahme.

¹⁰Für die genaue Beschreibung der jeweiligen Stichproben, siehe entsprechende Kapitel. Man beachte jedoch zunächst, dass der Nachtest nicht von allen Schülern bearbeitet wurde, so dass die Anzahl der VPN in diesem Abschnitt nicht der Summe aller Versuchspersonen in beiden Erhebungen entspricht.

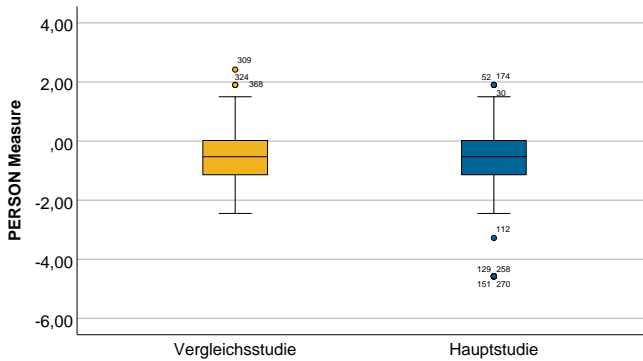
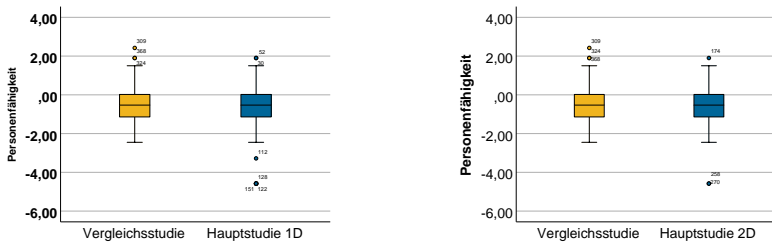


Abbildung 12.27.: Personenfähigkeiten der SuS in Vergleichs- und Hauptstudie.

Da eine Unter- oder Überlegenheit eines Unterrichtsansatzes durchaus im Rahmen des Möglichen zu sein scheint, wird die Vergleichsgruppe im Folgenden mit den 1D- und 2D-Schülern der Hauptstudie verglichen (siehe Abbildung 12.28). Völlig unabhängig vom Zugang zeigt sich hier aber ein ähnliches Bild wie beim Vergleich von allen Schülern ungeachtet des Zugangs. Es scheinen keine Unterschiede in den Personenfähigkeiten zu bestehen. Auch hier zeigt sich der t-Test jeweils eindeutig. Der zugehörige verteilungsfreie Test liefert das gleiche Ergebnis.



(a) Personenfähigkeiten der SuS in Vergleichs- und Hauptstudie, Teilpopulation 1D. (b) Personenfähigkeiten der SuS in Vergleichs- und Hauptstudie, Teilpopulation 2D.

Abbildung 12.28.: Gegenüberstellung Vergleichsgruppe und Hauptstudie mit jeweiligem Zugang

Ein Vergleich mit den Feedbackvarianten zeigte ebenfalls keine Unterschiede und wird hier nicht mehr gesondert aufgeführt.

Diskussion Die Ergebnisse des Vergleichs überraschen zunächst. Es scheint, dass mit mehr zeitlichem Aufwand im Unterricht (Versuchsgruppe) im Test ähnliche Ergebnisse erzielt werden, wie in der kurzen Intervention dieser Arbeit. Wie lässt sich dies deuten?

Aus kritischer Perspektive kann das Testinstrument für den Vergleichszweck angezweifelt werden. Im KC-konformen Unterricht können auch andere thematische Schwerpunkte als die der Lernumgebung (z. B. auch Trägheit, Beschleunigung), die der Test nicht erhebt, gesetzt worden sein. In ähnlicher Weise könnte argumentiert werden, dass sich der Test eng an diesen Inhalten der Lernumgebung orientiert und die Schüler der Intervention daher im Vorteil gegenüber der Vergleichsgruppe waren. Weiterhin könnte ein Testlängeneffekt aufgetreten sein, da der Test der Vergleichsgruppe etwa doppelt so lang war wie die hier verglichenen Items. Darüber hinaus lag der Unterricht bei zwei Klassen *etwas* längere Zeit zurück¹¹.

Aus mutiger Perspektive könnte evtl. (auch im Bezug auf Tobias (2010)) konstatiert werden, dass mehr Unterrichtszeit wie in der Vergleichsgruppe nicht zwingend effektiver sein muss.

Eine Ursachenzuschreibung ist mit den vorliegenden Mitteln nicht zu treffen, womit der Leser oder die Leserin dieser Arbeit die eigenen Gedanken anhand der obigen Interpretationen weiterführen kann.

¹¹Leider ohne dies rückwirkend genau beziffern zu können.

12.2.4. Cognitive Load

In Kapitel 4.4 wurde dargelegt, weshalb die Betrachtung der kognitiven Belastung¹² beim Bearbeiten der Simulationen in beiden Lernumgebungen sinnvoll sein könnte. Im Rahmen der Hauptstudie wurde in vier Klassen begleitend mit entsprechendem Erhebungsinstrument die empfundene geistige Anstrengung abgefragt. Die Verteilung der Versuchspersonen und Zugänge ist dabei leider unvorteilhaft ausgefallen. Von insgesamt 98 Teilnehmern entfallen nur 28 – eine Klasse – auf den 2D-Ansatz. Die Ergebnisse sind also auch hier als Hinweise zu verstehen, die evtl. weiterer Forschung bedürfen. Die Ergebnisse der separaten Studie werden im Anschluss schlaglichtartig dargestellt.

Die Tabellen 12.6 und 12.7 zeigen für jede Simulation den Mittelwert der Schülerangaben von 1 – 9. Es zeigt sich, dass alle Simulationen der 2D-Lernumgebung (MW: 3,36) als geistig herausfordernder bewertet werden, als alle Simulationen der 1D-Lernumgebung (MW: 2,66). Schließt man die gemeinsamen Simulationen (Freifallturm, Achterbahnkontrolle, Zielfoto) aus der Analyse aus, bleibt der große Unterschied bestehen (3,42 vs. 2,56). Zieht man die Spannweite der Skala in Betracht, zeigt sich aber dass die empfundene Belastung, der mental Effort, meist im unteren Bereich der Skala angesiedelt ist. Selbst die am belastendsten Simulationen erreichen im Maximum die Mitte der Skala (Rafting: 4.92; Zielfoto: 3.00). Für eventuelle weitere Bearbeitungen der Simulation nach den Multiple-Choice-Fragen wurde der Mental Effort durch die Schüler wiederholt eingeschätzt. Für die 1D-Simulationen traf dies auf 31 – 41 Schüler zu. Der gesamte Mittelwert bei wiederholten Bearbeitungen ist mit 2,16 über alle Simulationen etwas niedriger als bei der Erstbearbeitung. Die 2D-Simulationen wurden von 17 – 24 Schülern wiederholt bearbeitet. Auch hier zeigt sich über alle Simulationen ein leicht gesenkter Mittelwert von 3,00.

Für weitere statistische Berechnungen wird im Folgenden die durchschnittliche kognitive Belastung pro Versuchsperson bestimmt (Kriterium: Mindestens für 7 Simulationen Angaben gemacht). Für 1D-Schüler beträgt diese im Mittel 2,64; für 2D-Schüler 3,34. Ein Man-Whitney-U-Test zeigt hier einen signifikanten Unterschied mit einem gerade mittleren Effekt ($z(89) = -2,306$; $p = .021$; $r = .24$). Inwiefern zeigen sich Besonderheiten innerhalb des Zugangs hinsichtlich Geschlecht oder Vorwissen? Weder im 1D-Zugang noch im 2D-Zugang (hier geringe Fallzahl beachten!) lassen sich mittels nicht-parametrischer Tests Unterschiede in den genannten Untergruppen feststellen.

¹²Auch: Cognitive Load. Hier im speziellen der Mental Effort.

Abschließend lässt sich konstatieren, dass sich Hinweise auf deutlich unterschiedliche Mental Loads beim Bearbeiten von 1D- und 2D-Simulationen ergeben. Die Absolutwerte der angegebenen Belastung sind jedoch für beide Zugänge als gering einzustufen. Bei wiederholter Bearbeitung der Simulationen, z. B. nach Feedback, sinkt die eingeschätzte Belastung pro Simulation leicht. Dies deutet zumindest nicht darauf hin, dass der kognitive Load – z. B. bei der Suche nach eigenen Fehlern in Simulationen – beim Bearbeiten steigt.

Tabelle 12.6.: Mittelwerte der empfundenen kognitiven Belastung im 2D-Ansatz

Simulation	N	MW	SD
CCL.Zielfoto	28	3,14	1,40
CCL.Achterbahnkontrolle	28	3,75	1,75
CCL.Freifallturm	26	2,65	1,54
CCL.Achterbahnx	27	1,92	1,38
CCL.Scooter_A	28	2,82	1,72
CCL.Rafting	28	4,92	2,24
CCL.Bremscrash	25	3,56	1,68
CCL.Scooter1	25	2,40	1,29
CCL.Scooter2	25	3,84	1,99
CCL.Airhockey	22	4,36	2,08
CCL.Ventilator	22	3,63	1,96
CCL.Neuepucks	21	3,38	1,68
MW alle Sim		3,36	
MW ohne gemeinsame		3,42	

Tabelle 12.7.: Mittelwerte der empfundenen kognitiven Belastung im 1D-Ansatz

Simulation	N	MW	SD
CCL_Rollercoaster	68	2,42	1,37
CCL_Scooter	64	2,56	1,51
CCL_Zielfoto	67	3,00	1,71
CCL_Achterbahnkontrolle	65	2,96	1,61
CCL_Diagrammscooter	63	2,60	1,55
CCL_Freifallturm	63	2,71	1,47
CCL_Mia	63	2,25	1,48
CCL_Achterbahn	64	2,70	1,46
CCL_Testfahrt	62	2,56	1,27
CCL_Wildwasser	62	2,82	1,50
MW alle Sim		2,66	
MW ohne gemeinsame		2,56	

Schlaglicht: Vorstudie zum Cognitive Load

Die Gestaltung der Studie ist näher in Kapitel 9.1.3 beschrieben. Die weiteren Ausführungen dazu beziehen sich auf Mitschker (2017). Nur weitere Quellen oder eigene Schätzungen werden im Folgenden kenntlich gemacht.

Im Rahmen der Vorstudie wurden mehrere Hypothesen untersucht, von denen drei hier vorgestellt werden. Die erste Hypothese, „Die zweidimensionalen Simulationen zum Thema Kinematik verursachen einen höheren Cognitive Load als die eindimensionalen Simulationen, da sie eine größere Informationsmenge bereit stellen, die im Arbeitsgedächtnis verarbeitet werden muss“, wurde anhand des auch in dieser Arbeit verwendeten Items untersucht. Es zeigte sich, dass die geistige Anstrengung beim Bearbeiten der 1D-Simulationen im Mittel einen Skalenteil geringer ausfiel als bei den 2D-Simulationen (2,57 vs. 3,53). Dabei wurde keine 1D-Simulation als anstrengender empfunden als eine 2D-Simulation. Auch im Rahmen dieser Arbeit werden die Absolutwerte der Anstrengung nicht als kritisch bewertet. Als ausgleichendes Moment stellt Mitschker bei den 2D-Simulationen nach Selbstauskünften höhere Aufmerksamkeit und Freude bei 2D-Simulationen fest.

In einer zweiten Hypothese wurde überprüft, ob die geistige Anstrengung positiv mit dem Wissenszuwachs korreliert (jeweils Selbstauskünfte). Entsprechende Untersuchungen zeigten jedoch nur bei 2 Simulationen (Freifallturm, Ventilator) signifikante, positive Korrelationen (je $r = .19$).

Mitschker reflektiert im Zuge der Diskussion die Erhebung des Wissenszuwachses, da die Einschätzung dessen sich nicht nur auf physikalische Inhalte sondern auch auf die Bedienung der Simulationen beziehen haben könnte. Hinsichtlich dieser Hypothese ergeben sich also nach Einschätzung des Autors dieser Arbeit durchaus Anknüpfungspunkte für weitere Forschung.

Die Untersuchung der dritten Hypothese – dass bei erneuter Bearbeitung die geistige Anstrengung niedriger sei – zeigte uneinheitliche Ergebnisse. Die Abweichung von der Einschätzung der ersten Bearbeitung ist sehr gering, im Mittel 0,28 Skaleneinheiten. Nach Mitschker ist die Streuung um den Ursprungswert aber relativ groß, weshalb zu dieser Hypothese „keine pauschale Aussage möglich zu sein“ scheint.

Teil VI.

Zusammenfassung

13. Diskussion

13.1. Forschungsfragen

Im folgenden Abschnitt werden die Forschungsfragen diskutiert (siehe Kapitel 6).

1. Inwiefern hat der Einsatz unterschiedlicher Unterrichtsansätze (1D bzw. 2D) Einfluss auf den Lernerfolg?

Mangels einheitlicher Definition für den Lernerfolg in der Mechanik, wird der Begriff in dieser Arbeit wie folgt verstanden: Für das latente Konstrukt „Mechanikverständnis“ wird auf Grundlage der Pre-Post-Fachtests mittels Raschanalyse für jeden Schüler eine Personenfähigkeit bestimmt (siehe dazu Kapitel 10.3). Ein Zuwachs dieser Personenfähigkeit zwischen Pre- und Posttest wird im Sinne dieser Arbeit als Lernzuwachs bezeichnet. Lässt man die Zugänge außer Acht zeigt sich über alle Schüler im Mittel ein signifikanter Zuwachs der Personenfähigkeit, der mit einem kleinen Effekt einhergeht (Kapitel 12.2.2). Gleichermäßen lassen sich signifikante Zuwächse innerhalb der Zugänge 1D und 2D identifizieren. Auch diese unterliegen jeweils einem kleinen Effekt, wobei der der 2D-Gruppe minimal größer bestimmt wurde als derjenige der 1D-Gruppe. Während im 2D-Ansatz im Vortest noch signifikante Unterschiede zwischen Mädchen und Jungen bestehen, lassen sich diese im Nachtest nicht mehr nachweisen. Im 1D-Ansatz fand diese Angleichung nicht statt.

2. Inwiefern hat Textfeedback oder Animationsfeedback Einfluss auf den Lernerfolg?

Unterteilt man die Population der Hauptstudie nicht nach unterschiedlichen Zugängen sondern nach Feedbackarten Text- oder Animationsfeedback, so zeigen sich auch in diesen beiden Untergruppen signifikante Zuwächse. Dabei zeigt die Gruppe, die das Animationsfeedback erhalten hat, leicht größere Effekte. Die jeweiligen Personenfähigkeiten zeigen im

Vergleich zwischen den Varianten sowohl im Vor- als auch im Nachtest keine signifikanten Unterschiede.

3. Inwiefern interagiert das Unterrichtskonzept mit der Feedbackvariante hinsichtlich des Lernerfolgs?

Für diese Analyse wurden zunächst die so erhaltenen vier Untergruppen intern verglichen. Mittels parametrischer bzw. nicht-parametrischer Verfahren lassen sich in allen Untergruppen signifikante Zuwächse messen. Die beiden Textfeedback-Gruppen (1D und 2D) zeigen dabei kleine Effekte, die ähnlich groß sind. Auch in den beiden Animationsfeedback-Gruppen sind die Effekte dabei klein. Auffällig ist jedoch, dass der Effekt in der 2D-Gruppe doppelt so groß ausfällt wie in der 1D-Gruppe ($d = .21$ vs. $d = .41$).

4. Wie ist die Akzeptanz von Lernumgebung und Simulationen für Schülerinnen und Schüler im Anfangsunterricht der Mechanik?

Die Akzeptanz der Lernumgebung wurde explorativ mithilfe von Fragebögen (alle SuS) und schlaglichtartig durch Interviews erhoben. Die überwiegende Tendenz dieser Erhebungen zeigt, dass die Lernumgebung von den Schülern eher gut angenommen wurde. *Überwiegend* bedeutet aber, dass es auch kritische Stimmen gab. Bemängelt wurde mitunter beispielsweise eine sich mit der Zeit einstellende Langeweile, bisweilen umständliche Bedienung (Starten der Simulationen) aber auch Unverständlichkeiten in der Instruktion und den Texten. Letztere lassen sich aber durch die quantitative Befragung als nicht wesentlich interpretieren, da die Texte und Instruktionen überwiegend als verständlich bezeichnet wurden. Auch in den Interviews zeigte sich überwiegend eine Offenheit gegenüber der praktizierten Lernform, wenn auch zum Teil mit Einschränkungen. Zu beachten ist aber auch, dass ständig 1 – 2 Hilfskräfte zur Unterstützung der Schüler bei technischen Problemen (vor allem zu Beginn) zugegen waren, was eine Situation darstellt, die nicht die Unterrichtsrealität darstellt. Die grundsätzliche Akzeptanz spiegelt sich auch im Wunsch nach häufigerem Einsatz von Simulationen und Lernumgebungen wider (von jeweils Zweidritteln der Befragten (eher) Zustimmung). Hier sind jedoch auch wiederum Gendereffekte zu beobachten, wobei Jungen sich häufiger vergleichbare Simulationen im Unterricht wünschen als

Mädchen. Dennoch sei davor gewarnt, die Ergebnisse mit Überschwang zu betrachten! Wie schon in der Theorie dargelegt könnten auch die im Rahmen dieser Studie Neuigkeitseffekte (siehe Kapitel 4.2.1) – mit unbekannter Ausprägung – aufgetreten sein. Unterstützt wird diese Annahme mitunter durch Kommentare aus freien Schülerantworten (z. B. „Ich fand es war eine gute Abwechslung zum normalen Unterricht.“). Daher sollten keine Schlüsse gezogen werden, die diese oder ähnliche Lernumgebungen grundsätzlich und dauerhaft für motiviertes Mitarbeiten für tauglich erklären, gerade auch vor dem Hintergrund der vereinzelt Erwähnung von aufkommender Langeweile!

Zur Akzeptanz der Lernumgebung seitens der beobachtenden Lehrkräfte können anhand der Daten keine Aussagen getroffen werden, auch wenn vereinzelt positive Aspekte („Begeisterung“) wie auch negative Aspekte („untrennbare Verknüpfung von Lernen mit Spaß“, „unterschiedliche Qualität der Simulationen“) genannt werden.

5. Inwiefern beeinflussen sich affektive Merkmale (Interesse, Selbstkonzept, Flow) die Unterrichtsansätze und Feedback?

Zunächst muss vor der Beantwortung dieser Forschungsfrage betont werden, dass sich die Konstrukte (Interesse etc.) jeweils auf den Zeitraum vor der Intervention mit den Konstrukten auf den Interventionszeitraum beziehen und verglichen werden. Dabei kann davon ausgegangen werden, dass die Konzepte des Vorunterrichts relativ stabil angesehen werden können. Es wäre also unseriös zu schlussfolgern, dass durch die Intervention das Interesse an Physik dieser oder jenen Gruppe langfristig gesteigert worden wäre.

Zunächst ist festzustellen, dass sich die betrachteten Konstrukte vor der Intervention zwischen den Zugängen unterscheiden. Diese Unterschiede ließen sich im Vorfeld mitunter auf Klasseneffekte zurückführen, die oft nach Ausschluss einer auffälligen Klasse nicht mehr festzustellen waren. Unterschiede sind hingegen konsequent zwischen Jungen und Mädchen zu messen. Sowohl im Interesse als auch im Selbstkonzept zeigen sich signifikante Unterschiede mit mittleren Effekten, im Flow noch mit kleinem Effekt. Bezogen auf die Intervention zeigen sich für den Flow in keiner Untergruppe mehr signifikante Unterschiede. Im Bereich des Selbstkonzepts konnte der Effekt deutlich abgemildert werden, wenn auch noch Unterschiede bestehen. Im Interesse sind erfreulicherweise keine Unterschiede mehr festzustellen. Damit verknüpft ist der Zuwachs des interventionsbezogenen Interesses bei den Mädchen deutlich größer

als bei den Jungen. Auch für den interventionsbezogenen Flow lassen sich im Nachfragebogen keine Unterschiede mehr feststellen, mithin ist ein größerer Zuwachs der Mädchen zu verzeichnen.

Hinsichtlich der Zugänge lassen sich in den jeweiligen Nachfragebögen in keinem der Konstrukte signifikante Unterschiede feststellen. In Bezug auf die erfahrenen Zuwächse fallen die Effekte zwischen Pre- und Postfragebogen für den 2D-Zugang jedoch durchweg größer aus.

Grundsätzlich stellte sich also heraus, dass im besonderen Maße Mädchen, besonders im 2D-Zugang, hinsichtlich der affektiven Komponenten profitieren. Eine Frage, die mithilfe dieser Studie nicht beantwortet werden kann, ist die nach den Ursachen. Grundsätzlich stellt sich nämlich die Frage: Sind die Ursachen für den Profit der Mädchen in der Lernumgebung/den Simulationen/der Anlage der Studie zu finden oder eher im Wegfall von Faktoren, die nur im üblichen Unterricht auftreten?

6. Inwiefern beeinflusst die Arbeit mit den Simulationen *unterschiedlicher* Mechanik-Zugänge die kognitive Belastung der Schüler?

Die Untersuchung zur kognitiven Belastung beim Bearbeiten der Simulationen wurde in Rahmen dieser Studie begleitend in 4 Klassen – mit leicht ungleicher Aufteilung auf die Mechanik-Zugänge – durchgeführt. Die erhaltenen Ergebnisse stellen deshalb und wegen möglicher Kritik an der Erhebung (siehe Abschnitt 13.2) im besten Falle *Hinweise* auf bestimmte Sachverhalte dar.

Der Cognitive Load wurde auf Basis des Mental Efforts erhoben, für den wiederum die empfundene geistige Anstrengung ein möglicher Indikator ist. Es zeigte sich, dass deutliche Unterschiede in der empfundenen Anstrengung sowohl auf Ebene der Simulationen als auch zwischen Versuchspersonen vorherrschen. Die Simulationen der 1D-Lernumgebung wurden auf einer neunstufigen Skala etwa 0,7 Skaleneinheiten weniger anstrengend geratet als die der 2D-Lernumgebung (ohne gemeinsame Simulationen 0,8 Skaleneinheiten). Der Unterschied in der Belastung auf Personenebene ist signifikant und zeigt einen mittleren Effekt. Bei der wiederholten Durchführung der Simulationen zeigen sich insgesamt leicht niedrigere Angaben der Schüler. Unterschiede konnten in diesem spezifischen Setting mit diesem Mittel gezeigt werden – jedoch sind die Werte der Skala in solch einem Bereich angesiedelt, sodass von einer Überlastung im Allgemeinen nicht die Rede sein kann. Diese niedrigen Werte können mitunter darauf zurückzuführen sein, dass die Lernumgebung und damit auch die Simulationen in relativ kleine thematische

Abschnitte unterteilt sind. Entsprechend gibt es in diesen Simulationen weniger Freiheitsgrade als in Simulationen, die umfangreichere Interaktionen und Eingriffsmöglichkeiten in einer Simulation ermöglichen (siehe auch Abschnitt 13.3).

13.1.1. Einordnung der Studie

In diesem Kapitel soll die vorliegende Untersuchung in einen Zusammenhang hinsichtlich der Erkenntnisse und Methodik zu bisherigen Studien eingeordnet werden. In besonderem Maße soll dabei neben anderen die hier bereits vielzitierte Studie von Tobias (2010) berücksichtigt werden, da auch in jenem Vorhaben der Anfangsunterricht zur Mechanik mit zwei Unterrichtsansätzen verglichen wurde.

Zunächst muss trotz des Replikationscharakters (siehe Kapitel 3) die Unterschiedlichkeit der Studien herausgestellt werden. Dabei sind vor allem die Interventionsdauer (ca. 3 Doppelstunden vs. komplette Unterrichtseinheit), der Forschungsansatz (Felduntersuchung mit Ausschluss der Lehrervariable und vorgegebenen Material vs. Design-Based-Research-Ansatz mit minimaler Steuerung des Unterrichts) und die Auswertung der Fachtests (Rasch-Analyse vs. klassisch) zu nennen. Vor diesem Hintergrund ist zunächst zu konstatieren, dass in beiden Studien generell ein Lernzuwachs in beiden Gruppen festzustellen war. In den Nachtests der Studie von 2010 unterscheiden sich (hinsichtlich vergleichbarer Aufgaben) die Ergebnisse von Schülern der Kontroll- zu Treatment-Gruppe mit einem mittleren Effekt ($d = .54$). In der vorliegenden Studie zeigte sich kein Unterschied zwischen den Gruppen, der Effekt vom Pre- zum Posttest fiel in der 2D-Gruppe jedoch etwas größer aus als in der 1D-Gruppe ($d = .27$ vs. $d = .23$). Bezüglich Gender-Effekten berichtet Tobias (ebd.), dass Mädchen insofern vom 2D-Zugang mehr profitieren, als dass im Vortest bestehende Unterschiede im Nachtest „verschwinden“, während sie in der Kontrollgruppe anwachsen. Ähnliches lässt sich aus der vorliegenden Studie berichten: Während in der 1D-Gruppe auch im Nachtest signifikante Unterschiede zwischen den Geschlechtern bestehen, so sind diese in der 2D-Gruppe nicht mehr nachzuweisen. Insgesamt bestätigt sich aber auch in dieser Studie die von R. Wodzinski (1996), wonach das 2D-Konzept nicht gravierend schwieriger einzuschätzen ist, als das 1D-Konzept (siehe Abschnitt 3.6.1).

Zu einzelnen fachlichen Aspekten wie dem Begriff der Geschwindigkeit bzw. Zusatzgeschwindigkeit berichtet Tobias (2010), dass zumindest der erste Begriff bei den (2D)-Schülern kaum auf Verständnisschwierigkeiten stößt und auch die Zusatzgeschwindigkeit in Argumentationen mitunter auch von selbst genutzt wird. Im Geschwindigkeits-Item, das beide

Studien gemein haben, nannten 2010 68,2% aller 2D-Schüler min. eine der korrekten Antwortoptionen, in der vorliegenden Untersuchung nur knapp 26% – während in Interviews der Vektorcharakter der Geschwindigkeit durch die Schüler erwähnt wurde. Einzig die Zusatzgeschwindigkeit wird in den Interviews der vorliegenden Studie von den Schülern zur Argumentation bei Bewegungsveränderungen *nicht* herangezogen.

Zeitverzögerte Nachtests (Follow-Up Tests) wie bei Tobias (2010) wurden hier nicht durchgeführt. Dies ist vor allem mit der Kürze der Intervention und dem darauf folgenden Unterricht begründbar. Letzterer, der inhaltlich darüber hinaus auch noch unbekanntes Konzepte folgte (und so nicht sichergestellt ist, ob 1D, 2D oder sonstige Konzepte verfolgt wurden), dürfte die Dauer der Intervention um ein Vielfaches überdauern und somit die Aussagekraft verwässern.

Ein Einfluss der Gruppenzugehörigkeit auf affektive Konstrukte wie Interesse oder Selbstkonzept konnte bei Tobias (ebd.) nicht aufgezeigt werden. Auch in dieser Studie zeigen sich im Nachtest bei diesen Variablen keine Unterschiede zwischen den 1D- und 2D-Versuchsgruppen, im Bereich des *Zuwachses* des Selbstkonzepts zeigen 2D-Schüler einen größeren Effekt als die 1D-Schüler. Generell zeigten sich in dieser Studie Zuwächse im Interesse und Selbstkonzept (mehr dazu siehe oben, Kapitel 13.1).

13.2. Kritik

Ein universelles Messinstrument für den Lernerfolg im Anfangsunterricht der Mechanik existiert nicht (wobei für den Autor auch fraglich ist, ob dies überhaupt möglich sei). Somit muss man sich damit abfinden, dass *der* Lernerfolg nun nicht absolut und objektiv bestimmbar ist, sondern lediglich das ist, was der jeweilige Test misst. Diese Problematik ist nicht neu und soll hier auch nicht dramatisiert sondern nur angemerkt werden. Als schwach ist lediglich die Reliabilität des Fachtests zu bewerten. Hier wäre eine ausgiebigere Pilotierung mit größerem Aufgabenpool zur Auswahl von Aufgaben wünschenswert gewesen. Das Dilemma der Testfairness wurde in dieser Arbeit an unterschiedlicher Stelle dargelegt, und soll daher hier nur der Vollständigkeit halber erwähnt werden.

Die Messinstrumente zu nicht-kognitiven Aspekten stellten sich für den Zweck entsprechend als brauchbar dar. Der Nachfragebogen enthält relativ viele selbst erstellte Einzelitems zur Usability – speziell – dieser Lernumgebung. Hier wäre ein Rückgriff auf etablierte Skalen grundsätzlich wünschenswert gewesen, standen aber so passgenau nicht zur Verfügung. Das Erheben des Cognitive Load mit nur einem Item entsprang zwar sinnvollerweise der Überlegung, die Schüler von ihrer Arbeit nicht zu sehr

abzulenken und zu unterbrechen. Jedoch wirft sich so evtl. eine Frage nach der Validität auf.

Die Interviews als kognitive und nicht-kognitive Erhebungsinstrumente zeigten sich ebenfalls zweckentsprechend brauchbar. Zu bemängeln wäre hier lediglich die relativ geringe Fallzahl, die den quantitativen Aspekten der Auswertung nicht zu Gute kam.

Hinsichtlich des Forschungsdesigns ist kritisch anzumerken, dass das Bestreben, in 2 – 3 Doppelstunden Mechanikkonzepte zu vergleichen, die Wirkung unterschiedlicher Feedbackvarianten wie die Akzeptanz des Lernsettings zu evaluieren sowie in Teilen die Kognitive Belastung zu untersuchen ambitioniert ist und prinzipiell auch möglich ist. Jedoch zeigte sich nach Ansicht des Autors durchaus, dass die Fokussierung innerhalb des jeweiligen Aspekts nicht immer die gewünschte Tiefe erreichte, weshalb sich auch für folgende Forschung Anknüpfungspunkte bieten (siehe Kapitel 6). Die Kürze der Intervention im Vergleich zur recht hohen inhaltlichen Dichte wurde im Rahmen dieser Arbeit gut begründet. Eine kleinere, dafür inhaltlich tiefergehende Themenauswahl, könnte jedoch später eine größere Trennschärfe zwischen Personen großer und kleiner Fähigkeit erreichen, da das Erarbeiten neuer Inhalte intensiver erfolgen kann und Unterschiede so sichtbarer werden können.

Die Schülerinnen und Schüler fertigen während der Bearbeitung der Lernumgebung Notizen zu – ihrer Auffassung nach – wichtigen Inhalten an, um eine gewisse Nachhaltigkeit zu gewährleisten. Diese Aufzeichnungen wurden in dieser Arbeit nicht ausgewertet oder begutachtet. Informelle Beobachtungen während der Studiendurchführung oder auch Bemerkungen der Lehrkräfte zeigen jedoch eine deutlich unterschiedliche Qualität dieser Mitschriften. Dies kann ein Hinweis darauf sein, dass einige Schüler mit dem selbstständigen Auswählen von relevanten Informationen mit anschließender Verschriftlichung oder Skizze deutlich überfordert waren (siehe dazu auch Forschungsperspektive).

Die Bedienung der Tablets durch die Schüler funktioniert zwar, war aber gerade zu Beginn der Einheit in einigen Klassen von einer nötigen Eingewöhnung geprägt (speziell das Öffnen/Schließen der Simulationen, der Lernumgebung). Hier böten sich Änderungen zweierlei Art an. Zum einen müssten die Simulationen langfristig in die Lernumgebung implementiert werden. Zum anderen könnte die Einführung auf andere Weise erfolgen, z. B. in Form eines kurzen Tutorial-Videos, das durch die Schüler bei Notwendigkeit individuell betrachtet werden und auf das die Lehrkraft zur eigenen Entlastung verweisen kann. Generell zeigen aber die Ergebnisse der Fragebögen in eine Richtung, sodass davon ausgegangen werden kann, dass es keine strukturellen Probleme bei der Bedienung gab. Gerade gehäuft auftretende Probleme zu Beginn einer Einheit, die

im Zweifel von einer Lehrkraft im Alleingang gemanagt werden müssen, wirken sich wohl kaum akzeptanzsteigernd aus – ohne diese Behauptung allerdings belegen zu können.

13.3. Ausblick

13.3.1. Forschung

In diesem Abschnitt wird ein Ausblick auf unterschiedliche Forschungsperspektiven gegeben, die sich aus dieser Arbeit ergeben.

In dieser Arbeit war das Feedback als Absicherung für die Bearbeitung der Simulationen gedacht. Einzelne Feedbackitems nach den Simulationen bezogen sich nämlich in der vorliegenden Lernumgebung auch auf *Beobachtungen*, ohne explizit fachlich zu sein (fik. Beispiel: „Wenn der Wasserstrahl eingeschaltet ist, bewegt sich das Boot“). Seitens der Schüler könnte ein solches Item mit dem gewünschten Lernzuwachs verwechselt werden, da Schüler informellen Beobachtungen nach vereinzelt zu Aussagen wie „Ich habe gelernt, dass das Boot vom Wasserstrahl angetrieben wird“ [passendes fik. Beispiel] bewegt wurden. Daher wäre es denkbar ein rein fachliches Feedback am Ende eines thematischen Abschnitts zu etablieren und „Beobachtungs-Items“ zu reduzieren. Dieser Punkt benötigt weitere Absicherung!

Mit anderen Forschungsdesigns sollte – und das ausdrücklich nicht nur in der Folgeforschung zu diesem Projekt – die fachliche Implementation von Tablets erprobt werden, wenn sie in den Unterricht herein gegeben und von den Lehrkräften in deren Unterricht nicht in einer Lernumgebung sondern *je nach Bedarf* eingebunden werden. Dies ist vor allem hinsichtlich eines Unterrichtseinsatz noch realistischer und vor allem auch flexibler. Dies wäre aber ein anders gelagerter Forschungsschwerpunkt, sodass von paralleler Erforschung von Unterrichtszugängen seitens des Autors dieser Arbeit eher abgeraten wird!

Durch Hinweise eines Schülers im Fragebogen könnte eine Peerberatung an bestimmten Stellen einer Lernumgebung, z. B. im Anschluss an Simulationen, erprobt werden. Dies hätte nach Ansicht des Autors auch den Vorteil, dass die monotone Arbeitsform der Einzelarbeit aufgebrochen würde und die Schüler die Möglichkeit hätten sich auszutauschen. Hier bietet sich womöglich eine Forschungsperspektive.

Wie schon im Abschnitt 13.2 erwähnt, fiel den Schülern das Filtern von bedeutenden Inhalten schwer. Dass Schüler in der siebten und achten Klasse das Trennen von physikalischen und nicht-physikalischen Inhalten erlernen müssen ist aus praktischer Sicht relativ unbestritten und auch im KC als zu erlernende Kompetenz aufgeführt (Niedersächsischen

Kultusministerium 2015). Doch lassen sich dabei für erfolgreiches „Notizenmachen“ Prädiktoren, Muster oder Zusammenhänge entdecken, die zur Verbesserung des Lernerfolgs genutzt werden können? Auch hier bieten sich Anknüpfungspunkte für weitere Forschung.

Der Cognitive Load beim Arbeiten mit den unterschiedlichen Mechanikkonzepten wurde nur randständig erforscht, um erste Eindrücke dazu zu gewinnen. Auch gerade im Zuge von aktuellen Inklusionsdebatten, scheint es sinnvoll die geistige Anstrengungen beim Bearbeiten der Konzepte näher zu beleuchten, und so auch verstehen zu können, welche Elemente des Lernmaterials oder Konzepts tatsächlich zu einer hohen kognitiven Belastung beitragen. Dies kann sich an die Diskussion anknüpfen, ob es im einzelnen Fall günstiger ist, Lernschwierigkeiten im späteren Verlauf zu vermeiden oder eine große Einfachheit zu Beginn eines Themas zu gewährleisten. Weiterhin bietet es sich an, in der Erhebung des kognitiven Loads auf leicht komplexere Simulationen zurückzugreifen, um das Auftreten von Bodeneffekten zu verhindern bzw. um die Trennschärfe zu verbessern.

Ein erheblicher Teil fachdidaktischer Forschung zur Mechanik wurde bislang am Gymnasium betrieben. Bei der zunehmenden Verbreitung von Gemeinschaftsschulen scheint daher weitere Forschung in anderen Schulzweigen wie eben Gesamtschulen, Oberschulen, Stadtteilschulen oder den verbliebenen Haupt- und Realschulen – den Möglichkeiten bzw. Curricula entsprechend – sinnvoll.

13.3.2. Praktische Implikationen

In diesem Abschnitt werden praktisch orientierte Aspekte für Unterricht und Entwicklung mittel- bis langfristiger Perspektive, die sich aus dieser Arbeit ergeben, dargestellt. Der Anspruch ist es hier daher nicht, sofort umsetzbare Anregungen für die „morgige Unterrichtsstunde“ zu bieten.

Wie schon erwähnt wäre es mehr als fahrlässig, aus dem Vergleich zwischen den Schülern der Hauptstudie und der Vergleichstudie den Schluss zu ziehen, dass mit einem zeitlichen Aufwand von 2 – 3 Doppelstunden der gleiche Lerneffekt zu erzielen sei, wie im Rahmen einer ganzen Unterrichtseinheit (vgl. 12.2.3). Da aber auch schon bei Tobias (2010) von ähnlichen Lernzeiteffekten berichtet wird, scheint es zumindest lohnenswert, die optimale Ausnutzung der Unterrichtszeit zu hinterfragen. Denkbar wäre beispielsweise, die Kernkonzepte eines Themengebietes überblicksartig – wie hier in der Lernumgebung – darzustellen und im Anschluss die weitere Unterrichtszeit in differenzierenden Unterrichtset-

tings zu vertiefen¹. Solche Innovationen benötigen selbstverständlich eine wie auch immer geartete Form der Evaluation oder Beforschung.

Die Arbeit am Tablet bereitete den Schülern überwiegend Freude, wobei sie auch den Lernerfolg hoch einschätzen². Die in Kapitel 4.2.1 beschriebenen Neuigkeitseffekte und Verstehensillusionen sollten in der Praxis der Lehrkraft bewusst sein. So ist nicht zu erwarten, dass die *hohe* Motivation bei der Arbeit mit Tablets von außergewöhnlich langer Dauer sein wird (s. auch u.). Verstehensillusionen können gegebenenfalls auch direkt im Unterricht thematisiert werden. Was kann man an den immer populäreren Lernvideos mit welchen Strategien lernen? Was kann man anhand dieser Simulation, die zunächst bunt und einsichtig erscheint tatsächlich lernen?

Dass eine passgenaue methodische Abstimmung hinsichtlich Lerngruppe und Lernzielen erfolgen muss, ist unstrittig. In dieser Arbeit wurde dies ergänzend im Rahmen von einzelnen Schüleräußerungen in Fragebögen oder Interviews deutlich, dass die lange Tätigkeit am Bildschirm für die Augen unangenehm sein kann. Damit zusammenhängend ist es nicht mehr als Spekulation, dass Schülern die Arbeit am *eigenen* Tablet dennoch Freude bereitet, weil sie so immer eigenes Material zur Verfügung haben und nichts mit anderen Schülern teilen müssen. Die eigentliche selbstverständliche Monokultur betrifft auch das Verwenden von Simulationen anstelle von Realexperimenten, deren Fehlen in den geführten Interviews durchaus für die Dauer der Intervention aus Schülersicht nicht gefehlt haben, langfristig aber Unterrichtsbestandteil sein sollten. Beim Einsatz von Simulationen sollten aus der Teiluntersuchung zum Cognitive Load nicht der Schluss gezogen werden, dass Arbeiten mit Simulationen in der Mechanik *grundsätzlich* nicht stark belastend sei (Abschnitt 13.1).

¹Wobei natürlich auch Anleihen eines klassischen *Wochenplans* nicht von der Hand zu weisen sind.

²Wobei diese beiden Faktoren nach Ansicht des Autors im Physikunterricht auf Dauer bevorzugt gemeinsam auftreten sollten!

Literatur

- Amenda, Thomas (2017). *Bedeutung fachlicher Elementarisierungen für das Verständnis der Kinematik*. Bd. 230. Studien zum Physik- und Chemielernen. Berlin: Logos Verlag.
- Amenda, Thomas und Schecker, Horst (2014). „Moment mal ... (7): Ort, Ortsverschiebung, Weg – wofür steht eigentlich das s ?“ In: *Praxis der Naturwissenschaften im Unterricht Physik* 63.2, S. 40–42.
- Aufenanger, Stefan (2017). „Zum Stand der Forschung zum Tableteinsatz in Schule und Unterricht aus nationaler und internationaler Sicht“. In: *Tablets in Schule und Unterricht*. Hrsg. von Jasmin Bastian und Stefan Aufenanger. Wiesbaden: Springer VS, S. 119–138.
- Aufenanger, Stefan und Bastian, Jasmin (2017). „Einführung: Tableteinsatz in Schule und Unterricht – wo stehen wir?“ In: *Tablets in Schule und Unterricht*. Hrsg. von Jasmin Bastian und Stefan Aufenanger. Wiesbaden: Springer VS, S. 1–10.
- Aufschnaiter, Claudia von (2014). „Laborstudien zur Untersuchung von Lernprozessen“. In: *Methoden in der naturwissenschaftsdidaktischen Forschung*. Hrsg. von Dirk Krüger, Ilka Parchmann und Horst Schecker. Springer Spektrum, S. 81–94.
- Autorengruppe Paducation (2014). *Evaluation eines Modellversuchs mit Tablets am Hamburger Kurt-Köbler-Gymnasium*. URL: https://www.ifib.de/publikationsdateien/paducation_bericht.pdf.
- Bader, Franz und Dorn, Friedrich (2008). *Physik 7/8 Gymnasien Niedersachsen*. Braunschweig: Schroedel.
- Barth, Johannes Maximilian (2014). *Experimentieren im Physikunterricht der gymnasialen Oberstufe*. Bd. 164. Studien zum Physik- und Chemielernen. Berlin: Logos Verlag.
- Baumert, Jürgen u. a. (1997). *Bildungsverläufe und psychosoziale Entwicklung im Jugendalter: Dokumentation - Band 1. Skalen Längsschnitt I, Wellen 1-4: Unveröffentlichtes Paper*. Berlin. URL: https://pure.mpg.de/pubman/faces/ViewItemFullPage.jsp?itemId=item_2293704_2&view=EXPORT.
- Beichner, Robert J. (1994). „Testing student interpretation of kinematics graphs“. In: *American Journal of Physics* 62.8, S. 750. DOI: 10.1119/1.17449.

- Berger, Veit (2006a). „Im Physikunterricht experimentieren“. In: *Physik-Didaktik*. Hrsg. von Helmut F. Mikelskis. Berlin: Cornelsen Scriptor, S. 149–182.
- (2006b). „Mit dem Computer im Unterricht modellieren“. In: *Physik-Didaktik*. Hrsg. von Helmut F. Mikelskis. Berlin: Cornelsen Scriptor, S. 139–148.
- Berghaus, Margot, (Hrsg.) urlfrom (1999). *Interaktive Medien — interdisziplinär vernetzt*. Wiesbaden: VS Verlag für Sozialwissenschaften. DOI: 10.1007/978-3-322-85123-9.
- Bergius, Rudolf (2018). *Feldstudie, Feldforschung*. In M. A. Wirtz (Hrsg.), *Dorsch - Lexikon der Psychologie*. URL: <https://portal.hogrefe.com/dorsch/feldstudie-feldforschung/>.
- Betrancourt, Mireille (2005). „The Animation and Interactivity Principles in Multimedia Learning“. In: *The Cambridge handbook of multimedia learning*. Hrsg. von Richard E. Mayer. Cambridge Univ. Press, S. 287–296. DOI: 10.1017/CBO9780511816819.019.
- Betrancourt, Mireille und Tversky, Barbara (2000). „Effect of computer animation on users’ performance: A review“. In: *Le travail humain* 63.4, S. 311–329.
- Bienefeld, Marc (2017). „Replikationsstudien in der Erziehungswissenschaft“. In: *Forschungspraxen der Bildungsforschung. Zugänge und Methoden von Wissenschaftlerinnen und Wissenschaftlern*. Hrsg. von Heinrich u. a. Münster: Waxmann.
- Black, P. (2003). „The Nature and Value of Formative Assessment for Learning“. In: *Improving Schools* 6.3, S. 7–22.
- Blackmore, Karen, Wilkie, O. und Manders, D. (2015). „To What Extent Can Tablet Computers (iPads) be Used to Support the Acquisition of Investigative Skills in Primary Science?“ In: *INTED2015 Proceedings*, S. 2791–2794.
- Blake, C. und Scanlon, E. (2007). „Reconsidering simulations in science education at a distance: Features of effective use“. In: *Journal of Computer Assisted Learning* 23.6, S. 491–502. DOI: 10.1111/j.1365-2729.2007.00239.x. (Zuletzt geprüft am 10. 10. 2016).
- Bleichroth, Wolfgang (1991). *Fachdidaktik Physik*. Köln: Aulis Verlag.
- Boczianowski, Franz Karl Joachim (2011). *Pfeile als mentales Werkzeug*. Berlin. DOI: 10.18452/16273. (Zuletzt geprüft am 26. 10. 2016).
- Boone, William J., Staver, John R. und Yale, Melissa S. (2014). *Rasch analysis in the human sciences*. Dordrecht: Springer.
- Bortz, Jürgen und Döring, Nicola (2006). *Forschungsmethoden und Evaluation*. 4., überarb. Aufl., [Nachdr.] Heidelberg: Springer-Medizin-Verlag.

- Bos, Wilfried (2009). *KESS 7: Skalenhandbuch zur Dokumentation der Erhebungsinstrumente*. Bd. 4. HANSE - Hamburger Schriften zur Qualität im Bildungswesen. Münster: Waxmann.
- Bredthauer, Wilhelm u. a. (2010). *Impulse Physik*. 1. Aufl., [Dr.] 1. Klett.
- Brenneke, Rudolf und Schuster, Günter, Hrsg. (1965). *Physik für Gymnasien*. Braunschweig: Vieweg.
- Bresges, Andre u. a. (2013). *Das Reichshofer Experimentierdesign“ zur Entwicklung und Überprüfung des Einsatzes von Tablet-PC im Physikunterricht*. *Phydid B*. URL: <http://www.phydid.de/index.php/phydid-b/article/view/467/607> (zuletzt geprüft am 16. 10. 2020).
- Bühner, Markus und Ziegler, Matthias (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München u. a.: Pearson - Higher Education.
- Clausen, M. (2004). *Online Rater-Inventar zur Unterrichtsqualität: 3. Revision*.
- Claussen, Ute (1989). „Die Schnittstelle zwischen Simulation und Animation: ein Diskussionsbeitrag“. In: *GI - 19. Jahrestagung I*. Hrsg. von M. Paul. Informatik-Fachberichte. Berlin und Heidelberg: Springer, S. 474–485.
- Deimann, Markus (2002). „Motivationale Bedingungen beim Lernen mit Neuen Medien“. In: *Medienunterstütztes Lernen – Beiträge von der WissPro-Wintertagung 2002*. Hrsg. von WissPro, S. 61–70.
- Demtröder, Wolfgang (2018). *Experimentalphysik 1*. Berlin, Heidelberg: Springer Berlin Heidelberg. DOI: 10.1007/978-3-662-54847-9.
- Dresing, Thorsten und Pehl, Thorsten (2010). „Transkription“. In: *Handbuch qualitative Forschung in der Psychologie*. Hrsg. von Günter Mey und Katja Mruck. Wiesbaden: VS, Verl. für Sozialwiss, S. 723–733.
- Duit, Reinders (2004a). „Alltagsvorstellungen berücksichtigen!“ In: *Schüler-Vorstellungen in der Physik*. Hrsg. von Rainer Müller. Aulis-Verlag, S. 3–7.
- (2004b). „Schülervorstellungen - von Lerndefiziten zu neuen Unterrichtsansätzen“. In: *Schüler-Vorstellungen in der Physik*. Hrsg. von Rainer Müller. Aulis-Verlag, S. 8–14.
- Duit, Reinders und Wodzinski, Christoph (2010). *Merkmale guten Physikunterrichts*. URL: <https://www.ipn.uni-kiel.de/de/das-ipn/abteilungen/didaktik-der-physik/piko/pikobriefe032010.pdf>.
- Eckhardt, Marc u. a. (2013). „How effective is instructional support for learning with computer simulations?“ In: *Instructional Science* 41.1, S. 105–124. DOI: 10.1007/s11251-012-9220-y.

- Erdfelder, Edgar und Ulrich, Rolf (2018). „Zur Methodologie von Replikationsstudien“. In: *Psychologische Rundschau* 69.1, S. 3–21. DOI: 10.1026/0033-3042/a000387.
- Finkenberg, Frank (2018). „Flipped Classroom im Physikunterricht“. Dissertation. Würzburg: Universität Würzburg. URL: <http://nbn-resolving.de/urn:nbn:de:bvb:20-opus-164146>.
- Franzen, Axel (2014). „Antwortskalen in standardisierten Befragungen“. In: *Handbuch Methoden der empirischen Sozialforschung*. Hrsg. von Nina Baur und Jörg Blasius. Wiesbaden: Springer VS, S. 701–711. DOI: 10.1007/978-3-531-18939-0{\textunderscore}51.
- Fridberg, Marie, Thulin, Susanne und Redfors, Andreas (2018). „Preschool children’s Collaborative Science Learning Scaffolded by Tablets“. In: *Research in Science Education* 48, S. 1007–1026. DOI: 10.1007/s11165-016-9596-9.
- Fritz, Catherine O., Morris, Peter E. und Richler, Jennifer J. (2012). „Effect size estimates: current use, calculations, and interpretation“. In: *Journal of experimental psychology. General* 141.1, S. 2–18. DOI: 10.1037/a0024338.
- Genz, Florian und Bresges, Andre (2017). „Projektbeispiele für Design-Based-Research im naturwissenschaftlichen Unterricht: Weiterentwicklung des Reichshofer Experimentierdesigns mit Tablets an Schulen“. In: *Tablets in Schule und Unterricht*. Hrsg. von Jasmin Bastian und Stefan Aufenanger. Wiesbaden: Springer VS, S. 63–86.
- Girwidz, Raimund (2010). „Medien im Physikunterricht“. In: *Physikdidaktik*. Hrsg. von Ernst Kircher, Raimund Girwidz und Peter Häußler. Springer-Lehrbuch. Berlin: Springer.
- (2015). „Multimedia und lerntheoretischen Aspekten“. In: *Physikdidaktik*. Hrsg. von Ernst Kircher, Raimund Girwidz und Peter Häußler. Springer Lehrbuch. Berlin und Heidelberg: Springer Spektrum, S. 843–872.
- Gläser, Jochen und Laudel, Grit (2010). *Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen*. 4. Auflage. Lehrbuch. Wiesbaden: VS Verlag. URL: <http://d-nb.info/1002141753/04>.
- Gross, Dietmar u. a. (2016). *Technische Mechanik 1: Statik*. 13., aktualisierte Auflage. Berlin und Heidelberg: Springer Vieweg. DOI: 10.1007/978-3-662-49472-1. URL: <http://dx.doi.org/10.1007/978-3-662-49472-1>.
- Hagmann, Stefan (2011). *Cognitive Load bei Entscheidungsprozessen in Qualitätszirkeln. Eine empirische Studie*. Bd. 7. München und Mering: Rainer Hampp Verlag.

- Hahn, Karl (1955). *Methodik des physikalischen: Unterrichts*. [2. Aufl.] Heidelberg: Quelle & Meyer.
- Hammer, Hildegard und Hammer, Karl (1995). *Grundkurs der Physik 1: Mechanik - Wärmelehre*. 7., bearb. Aufl. Berlin/Boston: De Gruyter und De Gruyter Oldenbourg.
- Hattie, John und Timperley, Helen (2007). „The Power of Feedback“. In: *Review of Educational Research* 77.1, S. 81–112. DOI: 10.3102/003465430298487.
- Häußler, Peter und Hoffmann, Lore (1995). „Physikunterricht - an den Interessen von Mädchen und Jungen orientiert“. In: *Unterrichtswissenschaft* 23.2, S. 107–126.
- Helmke, Andreas (2017). *Unterrichtsqualität und Lehrprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts : Franz Emanuel Weinert gewidmet*. 7. Auflage. Schule weiterentwickeln, Unterricht verbessern Orientierungsband. Seelze-Velber: Klett/Kallmeyer.
- Hestenes, David und Wells, Malcolm (1992). „A mechanics baseline test“. In: 30.3, S. 159. DOI: 10.1119/1.2343498.
- Hoffmann, Lore, Häußler, Peter und Lehrke, Manfred (1998). *Die IPN-Interessenstudie Physik*. Bd. 158. IPN. Kiel: IPN.
- Höttecke, Dietmar (2010). „Forschend-entdeckender Physikunterricht“. In: *Naturwissenschaften im Unterricht Physik* 21.119, S. 4–12.
- Ivanjek, Lana (2016). *Verständnis der Mechanik messen: (unveröffentlichter Vortrag)*. Frankfurt.
- Janczyk, Markus und Pfister, Roland (2013). *Inferenzstatistik verstehen*. Berlin, Heidelberg: Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-34825-9.
- Janvier, Claude (1981). „Use of situations in mathematics education“. In: *Educational studies in mathematics* 12.1, S. 113–122.
- Jong, Ton de und van Joolingen, Wouter R. (1998). „Scientific Discovery Learning with Computer Simulations of Conceptual Domains“. In: *Review of Educational Research* 68.2, S. 179–201.
- Jung, Walter (1980). *Mechanik für die Sekundarstufe I*. Bd. 2. Beiträge zur Methodik und Didaktik der Physik Phänomene und Begriffe. Frankfurt: Diesterweg.
- Jung, Walter, Reul, Horst und Schwedes, Hannelore (1977). *Untersuchungen zur Einführung in die Mechanik in den Klassen 3-6*. 1. Aufl. Beiträge zur Methodik und Didaktik der Physik. Frankfurt: Diesterweg.
- Karsenti, Thierry und Fievez, Aurelien (2013). *The iPad education: uses, benefits, and challenges: A survey of 6,057 students and 302 teachers in Quebec, Canada: Preliminary Report of Key Findings*. Montreal.

- Kircher, Ernst, Girwidz, Raimund und Häußler, Peter, Hrsg. (2010). *Physikdidaktik: Theorie und Praxis*. 2. Aufl. Springer-Lehrbuch. Berlin: Springer. DOI: 10.1007/978-3-642-01602-8.
- Hrsg. (2015). *Physikdidaktik: Theorie und Praxis*. 3. Auflage. Springer Lehrbuch. Berlin und Heidelberg: Springer Spektrum.
- Kramer, Ulrich und Neculau, Mihaela (2000). „Modellbildung und Simulation“. In: *Taschenbuch der Informatik*. Hrsg. von Uwe Schneider, Dieter Werner und Joachim Ebert. München: Fachbuchverl. Leipzig im Carl-Hanser-Verl., S. 445–466.
- Kreiten, Marga (2010). „Möglichkeiten von interaktiven 3d-Simulationen zur Unterstützung von Versuchen im physikalischen Praktikum“. In: *Phydid B*.
- Krell, Moritz (2017). „Evaluating an instrument to measure mental load and mental effort considering different sources of validity evidence“. In: *cogend education* 4, S. 1–10. DOI: 10.1080/2331186X.2017.1280256.
- Krüger, Dirk, Parchmann, Ilka und Schecker, Horst, Hrsg. (2014). *Methoden in der naturwissenschaftsdidaktischen Forschung*. Springer Spektrum.
- Krummeck, Vanessa (2008). *Multimediale, multicodele, multimodale und interaktive Komponenten in mathematischen Lernumgebungen: Dissertation*. München.
- Kultusministerkonferenz (2004). *Bildungsstandards im Fach Physik Bildungsstandards im Fach Physik für den Mittleren Schulabschluss*.
- Leisen, Josef (2005). „Kinematik ohne Dynamik?“ In: *Praxis der Naturwissenschaften im Unterricht Physik* 54.5, S. 2–3.
- Linacre, John M. (2004). „Equating/Linking with Anchors“. In: *Rasch Measurement Transactions* 18.3, S. 993. URL: <https://www.rasch.org/rmt/rmt183.pdf>.
- (2011). *Practical Rasch Measurement: Tutorial 3*. URL: <https://www.winsteps.com/a/winsteps-tutorial-further-3.pdf>.
- (2019). *A user's guide to Winsteps*.
- Linacre, John M. und Wright, D., Hrsg. (1994a). *Rasch Measurement Transactions*. Bd. 8.3.
- (1994b). „Reasonable mean-square fit values.“ In: *Rasch Measurement Transactions*. Hrsg. von John M. Linacre und D. Wright. Bd. 8.3, S. 370.
- Hrsg. (2010). *Rasch Measurement Transactions*. Bd. 23.4.
- Mathan, S. A. und Koedinger, Kenneth R. (2002). „An empirical assessment of comprehension fostering features in an intelligent tutoring system.“ In: *Intelligent Tutoring Systems ITS 2363*, S. 330–343.

- Mayring, Philipp (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. 12., überarb. Aufl. Beltz Pädagogik. Weinheim: Beltz.
- Meschede, Dieter (2015). *Gerthsen Physik*. 25. Aufl. Springer-Lehrbuch. Berlin: Springer Spektrum. DOI: 10.1007/978-3-662-45977-5.
- Messmer, Roland (2015). „Stimulated Recall als fokussierter Zugang zu Handlungs- und Denkprozessen von Lehrpersonen“. In: *Forum Qualitative Sozialforschung* 16, S. 1–20.
- Mézes, C., Erb, R. und Schröter, E. (2012). „Der Einfluss von Videoexperimentieranleitungen auf die Motivation von Schülerinnen und Schülern“. In: *Phydid B*, S. 17–27.
- Mie, Klaus (2002). „Multiple-Choice-Aufgaben im Unterricht“. In: *Unterricht Physik* 13.67, S. 8–11.
- Ministerium für Kultus, Jugend und Sport Baden Württemberg (2016). *Bildungsplan des Gymnasiums Physik*. URL: <http://www.bildungsplaene-bw.de/Lde/LS/BP2016BW/ALLG/GYM/PH> (zuletzt geprüft am 15.04.2018).
- Ministerium für Schule und Weiterbildung NRW (2008). *Kernlehrplan Physik*.
- Mitschker, Lea (2017). „Kognitive Belastung beim Lernen mit Simulationen“. Masterarbeit LUH (unveröffentlicht).
- Moosbrugger, Helfried und Kelava, Augustin (2012). *Testtheorie und Fragebogenkonstruktion*. 2., aktualisierte und überarbeitete Auflage. Springer-Lehrbuch. Springer-Verlag Berlin Heidelberg.
- Müller, Andreas und Ditton, Hartmut (2014). „Feedback: Begriff, Formen und Funktionen“. In: *Feedback und Rückmeldungen*. Hrsg. von Hartmut Ditton und Andreas Müller. Münster und New York: Waxmann, S. 11–28.
- Müller, Rainer, (Hrsg.) urlfrom (2004). *Schüler-Vorstellungen in der Physik*. Aulis-Verlag.
- (Hrsg.) urlfrom (2017). *Dorn.Bader - Einführungsphase Physik Sek II*. Druck A1. Braunschweig: Schroedel westermann.
- Narciss, Susanne (2006). „Informatives tutorielles Feedback“. Zugl.: Dresden, Techn. Univ., Habil.-Schr., 2005. Münster u. a.: Waxmann.
- Neumann, Knut (2014). „Rasch-Analyse naturwissenschaftsbezogener 28 Leistungstests“. In: *Methoden in der naturwissenschaftsdidaktischen Forschung*. Hrsg. von Dirk Krüger, Ilka Parchmann und Horst Schecker. Springer Spektrum, S. 355–369.
- Niebert, Kai und Gropengießer, Harald (2014). „Leitfadengestützte Interviews“. In: *Methoden in der naturwissenschaftsdidaktischen Forschung*. Hrsg. von Dirk Krüger, Ilka Parchmann und Horst Schecker. Springer Spektrum, S. 121–132.

- Niedersächsischen Kultusministerium (2015). *Kerncurriculum Kerncurriculum für das Gymnasium Schuljahrgänge 5-10 Naturwissenschaften*. URL: <https://cuvo.nibis.de/cuvo.php?p=download&upload=18> (zuletzt geprüft am 17. 10. 2020).
- Niegemann, Helmut M. (2008). *Kompendium multimediales Lernen*. X media press. Springer.
- Niegemann, Helmut M. u. a. (2008). *Kompendium multimediales Lernen*. Xmediapress. Berlin und Heidelberg: Springer.
- Nistor, Nicolae (2013). „Etablierte Lernmanagementsysteme an der Hochschule: Welche Motivation ist dabei wünschenswert?“ In: *E-Learning zwischen Vision und Alltag*. Hrsg. von Claudia Bremer und Detlef Krömker. Münster: Waxmann, S. 181–191.
- Nitsch, Renate (2015). *Diagnose von Lernschwierigkeiten im Bereich funktionaler Zusammenhänge*. Wiesbaden: Springer Fachmedien Wiesbaden. DOI: 10.1007/978-3-658-10157-2.
- NLQ (2015). *Mobiles Lernen mit Tablet-Computern an niedersächsischen Schulen – Projektabschlussbericht*. URL: <https://wordpress.nibis.de/mobileslernen/files/Abschlussbericht-Tablet-Projekt-NLQ.pdf> (zuletzt geprüft am 18. 10. 2020).
- Nolting, Wolfgang (2013). *Grundkurs Theoretische Physik 1*. Berlin, Heidelberg: Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-29937-7.
- Obst, David (2013). *Interaktive Tafeln im Physikunterricht*. Bd. 160. Studien zum Physik- und Chemielernen. Berlin.
- O'Neill, T. u. a. (2013). „How much item drift is too much?“ In: *Rasch Measurement Transactions* 27.3, S. 1423–1424.
- Paas, Fred G. W. C. und van Merriënboer, Jeroen J. G. (1994). „Instructional control of cognitive load in the training of complex cognitive tasks“. In: *Educational Psychology Review* 6.4, S. 351–371. DOI: 10.1007/BF02213420.
- Pavla, Hanzalova und Stepan, Hubalovsky (2016). „Tablet as a Tool for Creating Pupils ePortfolios in Science Subjects“. In: *International Journal of education and technologies* 10, S. 193–198.
- Rabe, T., Meinhardt, C. und Krey, O. (2012). „Entwicklung eines Instruments zur Erhebung von Selbstwirksamkeitserwartungen in physikdidaktischen Handlungsfeldern“. In: *Zeitschrift für Didaktik der Naturwissenschaften* 18, S. 293–315.
- Reinhold, Peter (2006). „Elementarisierung didaktische Rekonstruktion“. In: *Physik-Didaktik*. Hrsg. von Helmut F. Mikelskis. Berlin: Cornelsen Scriptor, S. 86–101.

- Reinmann, Gabi und Mandl, Heinz (2006). „Unterrichten und Lernumgebungen gestalten“. In: *Pädagogische Psychologie*. Hrsg. von Andreas Krapp und Bernd Weidenmann. Weinheim: Beltz PVU, S. 614–658.
- Renkl, Alexander u. a. (2003). „Cognitive load beim Lernen aus Lösungs-Beispielen“. In: *Zeitschrift für pädagogische Psychologie* 17, S. 93–101.
- Rey, Günter Daniel (2009). *E-Learning: Theorien, Gestaltungsempfehlungen und Forschung*. 1. Aufl. Psychologie Lehrbuch. Bern: Huber.
- Rheinberg, Falko (2004). *Motivationsdiagnostik*. Bd. 5. Kompendien psychologische Diagnostik. Göttingen: Hogrefe.
- Rheinberg, Falko, Vollmeyer, Regina und Engeser, Stefan. (2003). „Die Erfassung des Flow-Erlebens“. In: *Diagnostik von Motivation und Selbstkonzept*. Hrsg. von J. Stiensmeier-Pelster und Falko Rheinberg. Göttingen: Hogrefe, S. 261–279.
- Richtberg, Stefan (2018). *Elektronenbahnen in Feldern. Konzeption und Evaluation einer webbasierten Lernumgebung*. Bd. 258. Studien zum Physik- und Chemielernen. Logos Verlag.
- Rieber, Lloyd (2005). „Multimedia Learning in Games, Simulations, and Microworlds“. In: *The Cambridge handbook of multimedia learning*. Hrsg. von Richard E. Mayer. Cambridge Univ. Press, S. 549–568.
- Riese, Josef und Reinhold, Peter (2014). „Entwicklung eines Leistungstests für fachdidaktisches Wissen“. In: *Methoden in der naturwissenschafts-didaktischen Forschung*. Hrsg. von Dirk Krüger, Ilka Parchmann und Horst Schecker. Springer Spektrum, S. 257–268.
- Ristic, Nikola (2006). „Neue Medien verändern die Schule.“ In: *Forum E* 58.2, S. 9–12.
- Rode, Henning (2016). „Prototypen evidenzbasierten Physikunterrichts“. Dissertation. Berlin: Logos Verlag.
- Rodriguez, Michael C. (2005). „Three Options Are Optimal for Multiple-Choice Items: A Meta-Analysis of 80 Years of Research“. In: (*Keine Angabe*) 24.2, S. 3–13. DOI: 10.1111/j.1745-3992.2005.00006.x.
- Rost, Jürgen (2004). *Lehrbuch Testtheorie - Testkonstruktion*. 2., vollst. überarb. und erw. Aufl. Aus dem Programm Huber: Psychologie-Lehrbuch. Bern u. a.: Huber.
- Rutten, Nico, van Joolingen, Wouter R. und van der Veen, Jan T. (2012). „The learning effects of computer simulations in science education“. In: *Computers and Education* 58.1, S. 136–153. DOI: 10.1016/j.compedu.2011.07.017. (Zuletzt geprüft am 18.05.2016).
- Scheck, Florian (1990). *Mechanik: Von den Newtonschen Gesetzen zum deterministischen Chaos*. Zweite, erweiterte Auflage. Springer-Lehrbuch.

- Berlin und Heidelberg: Springer. DOI: 10.1007/978-3-662-08595-0. URL: <http://dx.doi.org/10.1007/978-3-662-08595-0>.
- Schecker, Horst (1985). *Das Schülervorverständnis zur Mechanik: Dissertation*. Universität Bremen.
- Schecker, Horst und Gerdes, Jörn (1999). „Messung von Konzeptualisierungsfähigkeit in der Mechanik: Zur Aussagekraft des Force Concept Inventory“. In: 5.1, S. 75–89. (Zuletzt geprüft am 28.09.2016).
- Schecker, Horst, Wilhelm, Thomas u. a., Hrsg. (2018). *Schülervorstellungen und Physikunterricht: Ein Lehrbuch für Studium, Referendariat und Unterrichtspraxis*. [1. Auflage]. Lehrbuch. Berlin: Springer Spektrum.
- Schmit, Stefan (2014). *Schulbücher als Lehr- und Lernmaterialien: Das Thema 'Bewegungsbeschreibung' in Physikschulbüchern der Sekundarstufe I*. Logos Berlin.
- Schnotz, Wolfgang und Lowe, Richard (1998). „A Unified View of Learning from Animated and Static Graphics“. In: *Kommt noch* 6.
- Shavelson, Richard J., Hubner, Judith J. und Stanton, George C. (1976). „Self-Concept: Validation of Construct Interpretations“. In: *Review of Educational Research* 46.3, S. 407–441. DOI: 10.2307/1170010.
- Shute, Valerie J. (2008). „Focus on Formative Feedback“. In: *Review of Educational Research* 78.1, S. 153–189.
- Staatsinstitut für Schulqualität und Bildungsforschung (2008). *Lehrplan Natur und Technik Jgst. 7*.
- Starauschek, Erich (2006). „Im Physikunterricht kommunizieren“. In: *Physik-Didaktik*. Hrsg. von Helmut F. Mikelskis. Berlin: Cornelsen Scriptor, S. 183–196.
- Strobl, Carolin (2015). *Das Rasch-Modell: Eine verständliche Einführung für Studium und Praxis*. Sozialwissenschaftliche Forschungsmethoden. Mering: Rainer Hampp Verlag.
- Sweller, John (1988). „Cognitive Load During Problem Solving: Effects on Learning“. In: *Cognitive Science* 12.2, S. 257–285. DOI: 10.1207/s15516709cog1202{\textunderscore}4.
- Sweller, John, van Merriënboer, Jeroen J. G. und Paas, Fred G. W. C. (1998). „Cognitive architecture and instructional design.“ In: *Educational Psychology Review* 10.3, S. 251–296.
- Tesch, Maike (2005). „Das Experiment im Physikkunterricht“. Kiel, Univ., Diss., 2005.
- Theyßen, Heike (2014). „Methodik von Vergleichsstudien zur Wirkung von Unterrichtsmedien“. In: *Methoden in der naturwissenschaftsdidaktischen Forschung*. Hrsg. von Dirk Krüger, Ilka Parchmann und Horst Schecker. Springer Spektrum.
- Thornton, Ronald K. (1998). „Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation

- of Active Learning Laboratory and Lecture Curricula“. In: (*Keine Angabe*) 66.4, S. 338. DOI: 10.1119/1.18863. (Zuletzt geprüft am 30.09.2016).
- Tinnefeld, Thomas (2013). *Dimensionen der Prüfungsdidaktik: Analysen und Reflexionen zur Leistungsbewertung in den modernen Fremdsprachen*. Bd. 1. Saarbrücker Schriften zu Linguistik und Fremdsprachendidaktik A, Monographien. Saarbrücken: htw saar.
- Tipler, Paul Allen, Mosca, Gene und Wagner, Jenny (2015). *Physik: Für Wissenschaftler und Ingenieure*. 7. Aufl. 2015. Berlin und Heidelberg: Springer Berlin Heidelberg.
- Tobias, Verena (2010). *Newton'sche Mechanik im Anfangsunterricht: Die Wirksamkeit einer Einführung über die zweidimensionale Dynamik auf das Lehren und Lernen*. Bd. Bd. 105. Studien zum Physik- und Chemielernen. Logos.
- Urhahne, Detlef u. a. (2000). „Computereinsatz im naturwissenschaftlichen Unterricht: Ein Überblick über die pädagogisch-psychologischen Grundlagen und ihre Anwendung“. In: *Zeitschrift für Didaktik der Naturwissenschaften* 6, S. 157–186.
- Varnai, Agnes Szabone (2006). *Unterstützung des Problemlösens in Physik durch den Einsatz von Simulationen und die Vorgabe eines strukturierten Kooperationsformats: Zugl.: Paderborn, Univ., Diss. 2006*. Bd. Bd. 57. Studien zum Physik- und Chemielernen. Berlin: Logos-Verl. URL: http://deposit.d-nb.de/cgi-bin/dokserv?id=2888420&prov=M&dok_var=1&dok_ext=htm.
- Viennot, Laurence (1979). „Spontaneous Reasoning in Elementary Dynamics“. In: *European Journal of Science Education* 1.2, S. 205–221. DOI: 10.1080/0140528790010209.
- Wackermann, Rainer (2008). *Überprüfung der Wirksamkeit eines Basismodell Trainings für Physiklehrer: Zugl.: Duisburg, Essen, Univ., Diss., 2007*. Bd. 75. Studien zum Physik- und Chemielernen. Berlin: Logos-Verl.
- Weidenmann, Bernd (2009). „Multimedia, Multicodierung und Multimodalität beim Online-Lernen“. In: *Online-Lernen: Handbuch für Wissenschaft und Praxis*. Hrsg. von Ludwig Issing. München: Oldenbourg, S. 73–86.
- Welling, Stefan (2017). „Methods matter: Methodisch-methodologische Perspektiven für die Forschung zum Lernen und Lehren mit Tablets“. In: *Tablets in Schule und Unterricht*. Hrsg. von Jasmin Bastian und Stefan Aufenanger. Wiesbaden: Springer VS.
- Wenzel, Michael (2018). „Computereinsatz in Schule und Schülerlabor“. Dissertation. Logos Verlag Berlin GmbH.

- Wiesner, Hartmut (2004). „Verbesserung des Lernerfolgs in der Unterricht über Mechanik“. In: *Schüler-Vorstellungen in der Physik*. Hrsg. von Rainer Müller. Aulis-Verlag, S. 114–119.
- (2011). *Kraft und Geschwindigkeitsänderung*. Bd. 5. Unterricht Sek. I Physik.
- Wiesner, Hartmut, Schecker, Horst und Hopf, Martin, Hrsg. (2011). *Physikdidaktik kompakt*. Hallbergmoos: Aulis Verlag.
- Wilhelm, Thomas (2005). *Konzeption und Evaluation eines Kinematik Dynamik Lehrgangs zur Veränderung von Schülervorstellungen mit Hilfe dynamisch ikonischer Repräsentationen und graphischer Modellbildung: Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades der Bayerischen Julius-Maximilians-Universität Würzburg*. Bd. Bd. 46. Studien zum Physik- und Chemielernen. Logos Verlag.
- Winkelmann, Jan (2015). *Auswirkungen auf den Fachwissenszuwachs und auf affektive Schülermerkmale durch Schüler- und Demonstrationsexperimente im Physikunterricht: Zugl.: Frankfurt (Main), Univ., Diss., 2014*. Bd. 179. Studien zum Physik- und Chemielernen. Berlin: Logos-Verl.
- Wissenschaftsrat (2014). *Bedeutung und Weiterentwicklung von Simulation in der Wissenschaft*. Hrsg. von Wissenschaftsrat. Dresden.
- Wodzinski, Rita (1996). *Untersuchungen von Lernprozessen beim Lernen Newtonscher Dynamik im Anfangsunterricht*. Bd. 25. Naturwissenschaften und Technik - Didaktik im Gespräch. LIT.
- (2004). „Lernschwierigkeiten in der Mechanik“. In: *Schüler-Vorstellungen in der Physik*. Hrsg. von Rainer Müller. Aulis-Verlag, S. 107–113.
- (2015). „Mädchen im Physikunterricht“. In: *Physikdidaktik*. Hrsg. von Ernst Kircher, Raimund Girwitz und Peter Häußler. Springer Lehrbuch. Berlin und Heidelberg: Springer Spektrum, S. 583–604.
- Wright, D. und Douglas, D. A. (1976). „Rasch item analysis by hand“. Research Memorandum No. 21. University of Chicago.
- Zang, Markus (2012). „Der Simulationsbaukasten Algodoo: Beschreibung und Beispiele“. Schriftliche Hausarbeit für die erste Staatsprüfung. Würzburg: Julius-Maximilians-Universität Würzburg. URL: http://www.thomas-wilhelm.net/arbeiten/Zula_Algodoo.pdf.
- Zwiorek, Sigrid (2006). „Mädchen und Jungen im Physikunterricht“. In: *Physik-Didaktik*. Hrsg. von Helmut F. Mikelskis. Berlin: Cornelsen Scriptor, S. 73–85.

Teil VII.

Anhang

A. Erhebungsinstrumente

A.1. Fragebögen und Tests

A.1.1. Fragebogen Lehrkräfte

- Jahre Berufserfahrung als Lehrkraft (inkl. Vorbereitungsdienst):
- Mein weiteres Unterrichtsfach:
- Beschreiben Sie Auffälligkeiten im Verhalten der Lerngruppe während der Arbeit mit dem Lernprogramm (bitte OHNE Nennung von Namen)
- Beschreiben Sie besondere Schwierigkeiten, die sich den Schülerinnen und Schülern der Lerngruppe während der Arbeit mit dem Lernprogramm/den Simulationen ergaben (falls Sie entsprechende Beobachtungen gemacht haben).
- (nur 2D) Geben Sie an, inwiefern Sie die frühe Einführung einer zweidimensionalen Dynamik in Jahrgang 7/8 eher als Chance oder als Schwierigkeit für die Lernenden sehen (gerne an konkreten Beispielen).
- (nur 1D) Beschreiben Sie, inwiefern Sie die traditionelle Einführung des Geschwindigkeitsbegriffs als reine Betragsgröße als langfristig lernhinderlich oder lernförderlich einschätzen (gerne an konkreten Beispielen).

- Geben Sie an, welche Kriterien den erfolgreichen Einsatz von Simulationen im Physikunterricht Ihrer Erfahrung nach ausmachen (z. B. Eigenschaften der Simulation, der Lerngruppe. . .).
- Was ich noch sagen möchte

A.1.2. Fragebogen Schüler

Tabelle A.1.: Übersicht der Fragen zum Interesse in Pre- und Postfragebogen der Hauptstudie

Pretest	
Abkürzung ¹	Fragentext
[v.freudephysik]	Auf die Beschäftigung mit Physik würde ich nur ungerne verzichten, einfach weil sie mir Freude macht.
[v.effekte]	Mich interessieren physikalische Effekte/Phänomene.
[v.wissenschaftler]	Ich mag in Physik die Rolle des Wissenschaftlers, der Zusammenhänge entdeckt.
Posttest	
[n.freude]	Während der Arbeit mit dem Lernprogramm hatte ich Freude daran, mich mit den Inhalten zu beschäftigen.
[n.effekteint]	Mich haben die physikalischen Effekte/Phänomene aus dem Lernprogramm/den Simulationen interessiert. Der Physikunterricht hat mir während der Arbeit mit dem Lernprogramm Freude bereitet. ²
[n.entdecken]	Ich mochte es, in den Simulationen/dem Lernprogramm physikalische Zusammenhänge zu entdecken. Es hat mir Freude bereitet, in den Simulationen physikalische Zusammenhänge zu entdecken. ²

¹ Den Abkürzungen ist zur besseren Unterscheidung grundsätzlich ein *v.* für Vortest oder ein *n.* für Nachtest vorangestellt.

² Version Vorstudie

Tabelle A.2.: Übersicht der Fragen zum Selbstkonzept in Pre- und Postfragebogen der Hauptstudie

Pretest	
Abkürzung ¹	Fragentext
[v_liegt]	Physik liegt mir nicht besonders.
[v_mühe]	Auch wenn ich mir Mühe gebe fällt mir Physik schwerer als vielen anderen.
[v_begabung]	Kein Mensch kann alles. Für Physik habe ich einfach keine Begabung!
Posttest	
[n_lagen]	Die Inhalte der letzten Stunden lagen mir nicht besonders.
[n_mühe]	Auch wenn ich mir Mühe gegeben habe, fielen mir die Inhalte der letzten Stunden schwerer als anderen.
[n_begabung]	Kein Mensch kann alles. Für die Inhalte der letzten Stunden hatte ich einfach keine Begabung!
[n_groberklären] ²	Ich bin zuversichtlich, dass ich das Gelernte jemandem grob erklären könnte.
[n_geschlagen] ²	Bei der Arbeit mit dem Lernprogramm habe ich mich gut geschlagen.

¹ Den Abkürzungen ist zur besseren Unterscheidung grundsätzlich ein *v.* für Vortest oder ein *n.* für Nachtest vorangestellt.

² Nur im Nachfragebogen der Vorstudie enthalten! Dort einzige Fragen zum Selbstkonzept!

Tabelle A.3.: Übersicht der Fragen zum Flow in Pre- und Postfragebogen der Hauptstudie

Pretest	
Abkürzung ¹	Fragentext
[v_gedanken]	Im Physikunterricht bin ich mit meinen Gedanken häufig woanders
[v_zeitvergeht]	Im Physikunterricht merke ich oft gar nicht wie die Zeit vergeht.
[v_vertieft]	Im Physikunterricht bin ich ganz in das Experimentieren vertieft.
Posttest	
[n_gedanken]	Im Physikunterricht mit den Tablets war ich mit meinen Gedanken häufig woanders. ²
[n_zeit]	Im Physikunterricht mit den Tablets habe ich oft gar nicht bemerkt, wie die Zeit vergeht. ²
[n_vertieft]	Im Physikunterricht mit den Tablets war ich ganz in das Lernprogramm vertieft. ²

¹ Den Abkürzungen ist zur besseren Unterscheidung grundsätzlich ein *v.* für Vortest oder ein *n.* für Nachtest vorangestellt.

² In Vorstudie statt „Im Physikunterricht mit...“ „Während der Arbeit am Tablet...“

Tabelle A.4.: Übersicht der Fragen zur Fehlerkultur in Pre- und Posttest

Pretest	
Abkürzung ¹	Fragentext
[v_geduld]	Unser Lehrer/in ist geduldig, wenn einer von uns im Physikunterricht einen Fehler macht.
[v_schlimmes]	Bei unserem Lehrer/in ist Fehlermachen nichts Schlimmes.
[v_hilfteinzeln]	Unserer Lehrer hilft uns auch einzeln, wenn wir bei einer Aufgabe nicht weiter wissen.
[v_kümmert]	Unserer Lehrer kümmert sich darum, wenn einer von uns Schülern Probleme beim Lösen von Aufgaben hat.
Posttest	
[n_anklickfalsch]	Ich fand es nicht schlimm, wenn ich eine „Anklickfrage“ im Lernprogramm falsch beantwortet habe. ²
[n_unterbrochen]	Die Hinweise nach den „Anklickfragen“ haben meine Gedanken nur unterbrochen und gestört.
[n_empfunden]	Wenn ich Probleme beim Lösen der Ankreuzfragen hatte, habe ich die Hinweise oder Animationen als Hilfe empfunden.

¹ Den Abkürzungen ist zur besseren Unterscheidung grundsätzlich ein *v.* für Vortest oder ein *n.* für Nachtest vorangestellt.

² In Vorstudie statt „Ich fand es nicht schlimm...“ „Ich war sehr enttäuscht...“

Tabelle A.5.: Übersicht der Fragen zur sonstigen Aspekten im Prefragebogen

Abkürzung ¹	Fragentext
	Wie oft benutzt du außerhalb der Schule Computer/Tablet/Smartphone für folgende Aktivitäten?
[v_spielen]	Spielen
[v_vergnügen]	Zum Vergnügen im Internet surfen (Youtube, Facebook etc.)
[v_info]	Suche nach praktischen Informationen, Nachrichten
[v_hausauf]	Hausaufgaben erledigen, für die Schule recherchieren
	Wie oft benutzt du außerhalb der Schule Computer/Tablet/Smartphone für folgende Aktivitäten?
[v_matheunterricht]	Im Mathematikunterricht (z. B. zum Graphen zeichnen, Tabellen anlegen, Taschenrechner)
[v_physikunterricht]	Im Physikunterricht (z. B. zum Graphen zeichnen, Tabellen anlegen, Taschenrechner)
[v_simulationen]	Im Physikunterricht für Simulationsprogramme
[v_a]	Anderes, nämlich:
[v_touchscreen]	Die Bedienung von Geräten mit Touchscreen (z. B. Smartphones) bereitet mir keine Probleme.
[v_spielebedien]	Ich finde bei Spielen oder Simulationen schnell heraus, welche Dinge bedienbar sind und welche nicht.

¹ Den Abkürzungen ist zur besseren Unterscheidung grundsätzlich ein *v.* für Vortest vorangestellt.

Tabelle A.6.: Übersicht der Fragen zur sonstigen Aspekten im Posttest

Abkürzung ¹	Fragentext
[nlernhäufiger]	Ich wünsche mir, dass ähnliche Lernprogramme im Physikunterricht häufiger als bisher eingesetzt werden.
[n.verständlich]	Die Texte waren verständlich.
[n.simhäufiger]	Ich wünsche mir, dass Simulationen (ohne ganzes Lernprogramm) im Physikunterricht häufiger als bisher eingesetzt werden.
[n.bewusst]	Mir ist bewusst, was ich mit Lernprogramm und Simulationen gelernt habe.
[n.simprob]	Die Bedienung der Simulationen mithilfe des Touchscreens bereitete mir keine Probleme.
[n.lernprob]	Die Bedienung des Lernprogramms mithilfe des Touchscreens bereitete mir keine Probleme.
[n.autom]	Ich hätte lieber Tipps/Hinweise von meinem Lehrer/in bekommen statt automatisch vom Lernprogramm.
[n.bedienbar]	Ich fand bei den Simulationen schnell heraus, welche Dinge bedienbar sind und welche nicht.
[n.steuern]	Das Steuern des Lernprogramms (vor, zurück, speichern. . .) ging einfach.
[n.ohnesim]	Ich hätte genauso gut ohne Simulationen und nur mit den Texten gelernt.
[n.zueinfach]	Lernprogramm/Simulationen waren viel zu einfach!
[n.note]	(Antwortmöglichkeiten von 1 – 6) Das Lernprogramm mit Simulationen erhält von mir die Schulnote:
[n.kommentar]	(Offene Frage) Was ich noch zu Simulationen/Lernprogramm/der Studie sagen möchte:

¹ Den Abkürzungen ist zur besseren Unterscheidung grundsätzlich ein *n.* für Nachtest vorangestellt.

A.2. Tests

A.2.1. Aufgaben

Zur besseren Übersicht sind die Tests (Vor- Nachtests in Haupt und Vorstudie und Vergleichsstudie) nicht einzeln aufgeführt, sondern es ist der komplette Aufgabenpool dargestellt. Die Zugehörigkeit der Aufgaben zu den jeweiligen Tests ist gekennzeichnet.

Legende

- ◀ Hauptstudie, Vortest
- ▶ Hauptstudie, Nachtest
- Vorstudie
- ◊ Vergleichsstudie

stoppuhr Name der Aufgabe

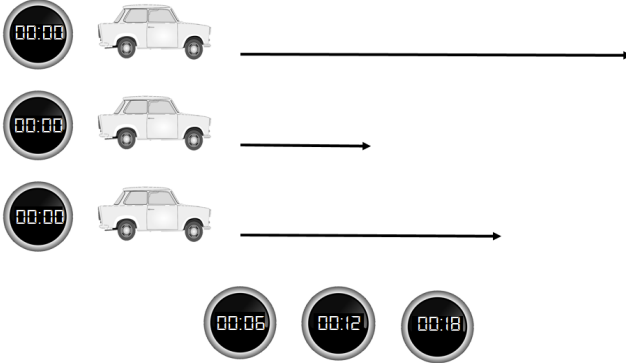
autobahn ◀ ■ ◊

Für eine 70km lange Strecke auf der Autobahn benötigt ein Wagen eine halbe Stunde. Wie groß ist das Tempo des Autos?

- 70km/h
- 35km/h
- 140km/h
- 70,5km/h

stoppuhr ◀ ◆ nach Jung, Reul und Schwedes (1977)

Die Autos sind gleich schnell und fahren die Streckenlänge, die der Pfeil anzeigt. Die Zeit, die jedes Auto braucht, hat ein Beifahrer mit der Stoppuhr gemessen. Welche Stoppuhr gehört zu welchem Auto



Die linke Stoppuhr gehört zum O oberen O mittleren O unteren Auto.

Die mittlere Stoppuhr gehört zum O oberen O mittleren O unteren Auto.

Die rechte Stoppuhr gehört zum O oberen O mittleren O unteren Auto.

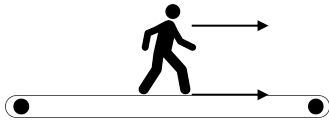
kart ◀ ■ ◆

Ein leichtes und ein schweres Kart erreichen eine Geschwindigkeit von 40km/h innerhalb von 5 Sekunden. Auf welches wirkt dabei die größere Kraft?

- Auf das leichte Kart
- Auf das schwere Kart
- Auf beide wirkt die gleiche Kraft, weil beide in 5 Sekunden die gleiche Geschwindigkeit erreichen.

band1+band2 ◀ ◆ nach Jung, Reul und Schwedes (1977)

Auf Flughäfen gibt es oft "Förderbänder", auf denen man entlang gehen kann (wie das Männchen auf dem Bild.) Das Band bewegt sich mit einem Tempo von 2m/s nach rechts.



a) Wie schnell ist das Männchen insgesamt, wenn es auf dem Förderband selbst mit einem Tempo von 3m/s nachts rechts läuft?

- 2m/s
- 5m/s
- 6m/s
- 1m/s

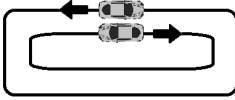


b) Wie schnell muss das Männchen auf dem Band laufen, damit es von außen gesehen scheinbar still an einer Stelle steht?

- 2m/s
- -3m/s
- -2m/s
- 3m/s

zweiautos ◀ ◆ (ebd.)

Die beiden Autos fahren in entgegengesetzter Richtung auf ihrer Bahn. Sie treffen sich immer genau oben in der Mitte.

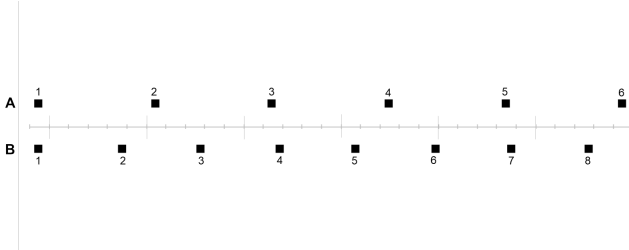


Was kannst du über die Geschwindigkeit der Autos sagen?

- Beide Autos sind gleich schnell.
- Das obere Auto ist schneller.
- Das untere Auto ist schneller.
- Da die Autos in unterschiedliche Richtungen fahren kann man sie nicht vergleichen.

koerper ◀ ■ ▶ ◆ (Hestenes und Wells 1992)

Die Abbildung zeigt die Position zweier Körper im zeitlichen Abstand von jeweils 1 Sekunde. Die Körper bewegen sich nach rechts.



Wie ändert sich die Geschwindigkeit der Körper A und B?

- Die Geschwindigkeiten ändern sich nicht.
- Die Geschwindigkeit von Körper B ändert sich stärker als die von A.
- Die Geschwindigkeit von Körper A ändert sich stärker als die von B.
- Die Geschwindigkeiten von Körper A und B ändern sich.
- Nicht genügend Informationen zum Antworten.

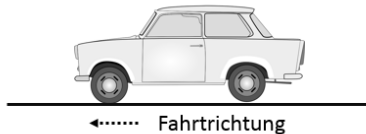
tempo ◀ ■ ▶ ◆ (Tobias 2010)

Kreuze alle Aussagen an, die dir sinnvoll erscheinen.

- Zwei Körper können sich mit gleichem Tempo, aber unterschiedlicher Geschwindigkeit bewegen.
- Zwei Körper können sich mit gleicher Geschwindigkeit, aber unterschiedlichem Tempo bewegen.
- Tempo und Geschwindigkeit sind das Gleiche.
- Bewegen sich zwei Körper mit gleicher Geschwindigkeit, so bewegen sie sich automatisch auch in die gleiche Richtung.

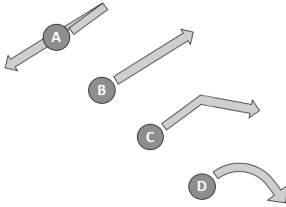
bremsen ◀ ■ ▶ ◆

Ein Auto fährt und muss abbremsen. Zeichne die Richtung ein, in die eine Kraft dafür wirken muss (du sollst direkt ins Bild zeichnen).



kugel ◀ ■ ▶ ◆

Eine Kugel rollt über einen glatten Tisch. Der Pfeil zeigt dir die Richtung, in die die Kugel sich bewegt.

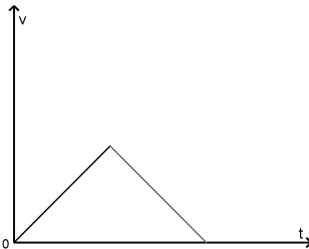


Gib an, in welchen der Situationen jeweils eine Kraft auf die Kugel wirken muss (mehrere Antworten)!

- A
- B
- C
- D

langsam ◀ ■ ▶ ◆

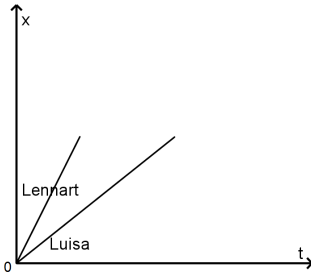
Du siehst das t-v-Diagramm einer Autofahrt. Welche Aussage kannst du mithilfe des Diagramms sicher treffen?



- Das Auto fährt bis zu einer bestimmten Stelle, dann kehrt es um.
- Das Auto fährt einen steilen Berg hoch und anschließend genauso steil wieder herunter.
- Zuerst wird das Auto immer schneller, dann wird es wieder langsamer.
- Im ersten Teil fährt das Auto schnell, im zweiten Teil langsam.

lennart ◀ ■ ▶ ◆

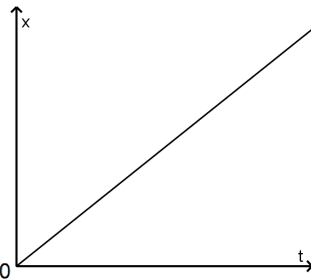
Du siehst das t-x-Diagramm von zwei Fußgängern. Kreuze die richtigen Aussagen an, es können mehrere richtig sein!



- Luisa ist schneller als Lennart.
- Lennart ist schneller als Luisa
- Lennart ist eine weitere Strecke gelaufen als Luisa.
- Luisa ist eine weitere Strecke gelaufen als Lennart.
- Beide sind genau gleich weit gelaufen.

objekt ◀ ■ ▶ ◆ (Beichner 1994)

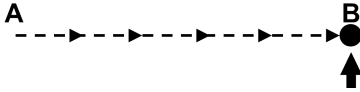
Du siehst das t-x-Diagramm (Zeit-Ort-Diagramm) einer Bewegung eines Objektes. Welche Aussage ist die beste Interpretation?



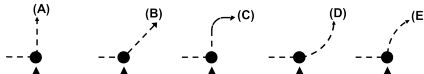
- Das Objekt wird konstant immer schneller.
- Das Objekt bewegt sich nicht.
- Die Schnelligkeit des Objekts erhöht sich gleichmäßig.
- Das Objekt bewegt sich mit konstanter Geschwindigkeit.

puck ◀ ■ ▶ ◊ (Hestenes und Wells 1992; Tobias 2010)

Der schwarze Puck gleitet von links über einen glatten Tisch in das Bild hinein.



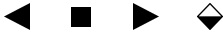
Wenn er den Punkt B erreicht, bekommt er einen kurzen "Kick" senkrecht nach oben (in Richtung des dicken Pfeils).



Welche der gestrichelten Linien beschreibt die Bewegung des Pucks nach dem "Kick"?

- A
- B
- C
- D
- E

spacerocket



Im Fahrgeschäft "Space Rocket" werden die Fahrgäste in einer kleinen Gondel mit einer Kraft abgeschossen, die bei jedem Abschuss exakt gleich ist - egal wie voll die Gondel besetzt ist. Welche der Aussagen über die Geschwindigkeit stimmt?

- Die Gondel wird schneller, wenn Sie leerer ist.
- Die Gondel wird immer gleich schnell, da die Kraft immer gleich ist.
- Die Gondel wird schneller, wenn sie voll besetzt ist, da sie dann mehr Schwung hat.
- Nichts von allem.

achterbahn

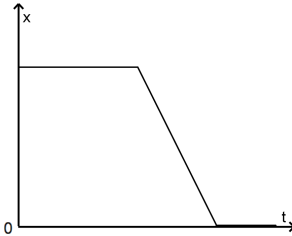


Der 30 Meter lange Zug der Achterbahn benötigt 2 Sekunden um an Max vorbei zu rasen. Wie groß ist das Tempo des Zuges?

- 30km/h
- 60m/s
- 28km/h
- 15m/s

huegel ► ■ ◆ (Beichner 1994)

Du siehst das t-x-Diagramm einer Bewegung eines Objektes. Welche Aussage ist die richtige Interpretation?

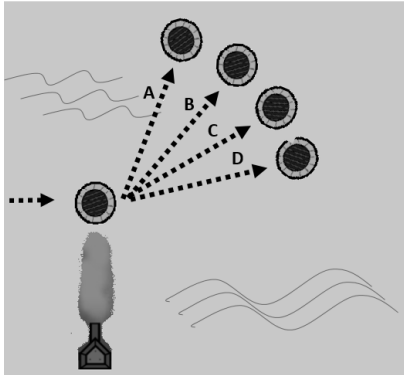


- Das Objekt rollt über eine gerade Oberfläche. Dann rollt es einen Hügel herunter und hält an.
- Das Objekt bewegt sich zunächst nicht. Dann rollt es einen Hügel herunter und hält an.
- Das Objekt bewegt sich mit konstanter Geschwindigkeit, dann wird es langsamer und hält an.
- Das Objekt bewegt sich zunächst nicht, dann bewegt es sich zurück und hält wieder an.
- Das Objekt rollte über eine glatte Fläche, dann bergab, danach rollt es unten weiter.

strudel ► ■ ◆

Zu sehen ist eine Wildwasserbahn, in der Boote von links nach rechts an einem Wasserstrahl vorbeifahren. Durch den Wasserstrahl wirkt für einen Moment eine Kraft senkrecht auf das jeweilige Boot.

Das Bild zeigt mehrere Boote, die an dem Wasserstrahl vorbei gefahren sind.



Welche der folgenden Aussagen können zutreffen? Es könnten mehrere sein!

- Boot A hat eine größere Masse als Boot C

oder

- Boot C hat eine größere Masse als Boot A

oder

- Boot B ist schneller am Wasserstrahl vorbeigefahren als Boot D

oder

- Der Wasserstrahl hat bei Boot B "stärker" auf das Boot eingewirkt als bei Boot D.

schlittena+schlittenb ► ■ ◆ (Thornton 1998)

Ein Schlitten bewegt sich über eine Eisfläche (es ist so glatt, dass er vom Eis nicht gebremst wird). Entscheide dich für die Kraft, die den Schlitten wie in den Aufgaben beschrieben bewegt!



a) Wird eine Kraft benötigt um den Schlitten nach rechts bewegen, so dass er dabei gleichmäßig immer schneller wird? Wenn ja, welche?

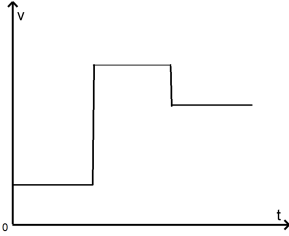
- Die Kraft zeigt nach rechts (wie im Bild) und wird immer stärker.
- Die Kraft zeigt nach rechts (wie im Bild) und bleibt immer gleich stark.
- Die Kraft zeigt nach rechts (wie im Bild) und wird immer schwächer.
- Es wird keine Kraft benötigt (wie im unteren Bild).

b) Wird eine Kraft benötigt um den Schlitten mit gleichbleibender Geschwindigkeit nach rechts bewegen? Wenn ja, welche?

- Die Kraft zeigt nach rechts (wie im Bild) und wird immer stärker.
- Die Kraft zeigt nach rechts (wie im Bild) und bleibt immer gleich stark.
- Die Kraft zeigt nach rechts (wie im Bild) und wird immer schwächer.
- Es wird keine Kraft benötigt (wie im unteren Bild).

vtstufen ► ■ ◆ (Bader und Dorn 2008)

Du siehst das t-v-Diagramm einer Autofahrt. Welche Aussage kannst du mithilfe des Diagramms sicher treffen?



- Das Auto entfernt sich, kommt aber wieder ein Stück zurück.
- Das Auto fährt erst sehr langsam, dann schnell, und am Ende etwas langsamer.
- Das Auto parkt - jeweils zu unterschiedlichen Zeitpunkten.
- Das Auto fährt über drei unterschiedliche Straßen.

katapult ► ■ ◆

Ein Achterbahnwagen wird auf ebener Strecke für 2 Sekunden mit einer bestimmten Kraft von einem Stahlseil gezogen und so auf seine Höchstgeschwindigkeit gebracht. Welche Möglichkeit gibt es, den Wagen noch schneller zu bekommen? (Mehrere Antworten können richtig sein)

Man könnte...

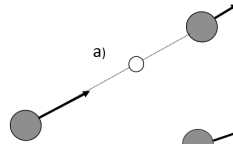
- ...einen leichteren Wagen benutzen.
- ...kürzer als 2 Sekunden ziehen (Bahn kann sich früher frei bewegen).
- ...mit gleicher Kraft länger als 2 Sekunden ziehen.
- ...einen schwereren Wagen benutzen (hat mehr Schwung)

musskrafta+musskraftb+musskraftc ► ■ ◆ (Tobias 2010)

Unten siehst du jeweils die Geschwindigkeiten eines Balles zu zwei Zeitpunkten. Gib jeweils an, ob in der Zwischenzeit eine Kraft auf den Ball ausgeübt werden musste. Wenn du glaubst es muss eine Kraft wirken, so zeichne ihre Richtung an den markierten Stellen (Punkt zwischen den Fußbällen) ein.

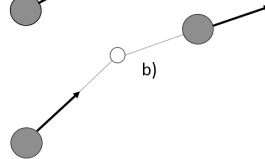
a) Muss hier eine Kraft wirken?

- Ja.
- Nein.



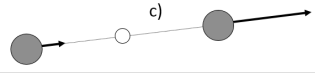
b) Muss hier eine Kraft wirken?

- Ja.
- Nein.



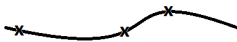
c) Muss hier eine Kraft wirken?

- Ja.
- Nein.



vektor ■ ◆ (Wiesner 2011)

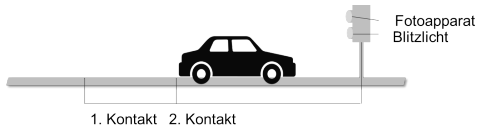
Hier siehst du eine Eisenbahnstrecke von oben. Ein Zug fährt gleich schnell von links nach rechts.



Zeichne an den drei markierten Stellen je einen passenden Geschwindigkeitspfeil ein!

blitzer ■ (Bredthauer u. a. 2010)

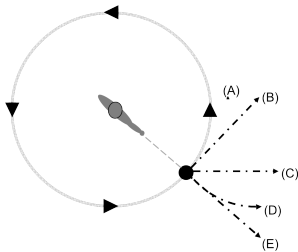
Wie erkennt der Blitzer, ob man zu schnell gefahren ist? Erkläre die Funktionsweise!



Meine Antwort:

fcikreis ■ ◆ (Hestenes und Wells 1992; Tobias 2010)

Ein schwerer Ball ist am Faden befestigt und wird, wie in der Abbildung gezeigt, im Kreis horizontal herum geschwungen. An dem gekennzeichneten Punkt reißt plötzlich der Faden. Der Vorgang wird von oben betrachtet.

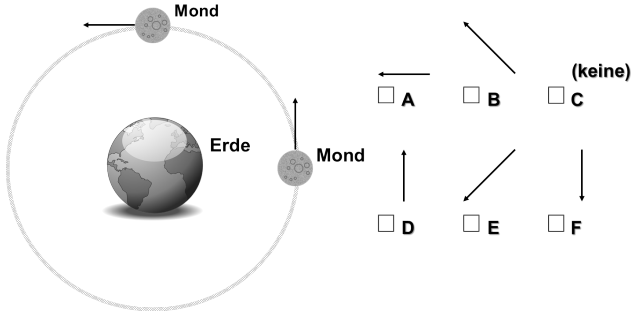


Welchen Weg nimmt der Ball, nachdem der Faden gerissen ist?

- A
- B
- C
- D
- E

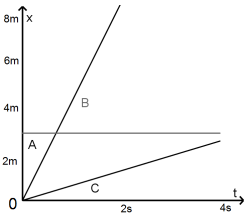
mond ■ ◆ (Tobias 2010)

Die Zeichnung links zeigt den Ort des Mondes zu zwei Zeitpunkten, die ungefähr sieben Tage auseinander liegen. Welche Zusatzgeschwindigkeit erhält der Mond in diesen sieben Tagen?



polizist ■ ◆ (Bader und Dorn 2008)

Du siehst das t-x-Diagramm der Bewegung von drei Objekten.



Welcher Graph gehört zum welchem Objekt?

Zum **Fußgänger** gehört:

- A
- B
- C

Zum **Auto** gehört:

- A
- B
- C

Zum **Polizisten, der am Straßenrand steht**, gehört:

- A
- B
- C

kreisbew ■

Du weisst: Bewegt sich ein Körper auf einer Kreisbahn und benötigt für jede Umdrehung die gleiche Zeit, dann nennt man dies eine gleichförmige Kreisbewegung. Kreuze alle zutreffenden Aussagen an!

- Die Geschwindigkeit während einer gleichförmigen Kreisbewegung ist konstant
- Geschwindigkeit bedeutet physikalisch dasselbe wie Tempo oder Schnelligkeit
- Die Geschwindigkeit ändert sich während einer gleichförmigen Kreisbewegung dauernd
- Während einer gleichförmigen Kreisbewegung wird der Körper beschleunigt.

zentri ■

“Im Kettenkarussell spüre ich, wie mich eine Kraft nach außen zieht! Ist doch klar, dass bei einer Kreisbewegung also eine Kraft nach außen wirkt!“ - Erkläre, weshalb diese Aussage falsch ist!

Meine Antwort:

beschleunigung ■ ◆ (Hestenes und Wells 1992; Tobias 2010)

Die Abbildung zeigt die Position zweier Körper im zeitlichen Abstand von jeweils 1 Sekunde. Die Körper bewegen sich nach rechts.



Haben die beiden Körper irgendwann einmal die gleiche Geschwindigkeit?

- Nein
- Ja, zum Zeitpunkt 2
- Ja, zum Zeitpunkt 5
- Ja, zum Zeitpunkt 2 und 5
- Ja, irgendwo im Bereich zwischen 3 und 4

mond2 ■

Der Mond benötigt für ein Umkreisen der Erde immer 28 Tage.

a) Welche der eingezeichneten Kräfte wirken auf den Mond?

- A
- B
- C
- D

b) Die auf den Mond wirkende Gesamtkraft ist

- 0
- nicht 0, wirkt in Richtung der Bewegung
- nicht 0, wirkt nicht in Richtung der Bewegung

A.3. Interviewleitfaden

- ... Frage/Impuls
- ▶ ... Folgefrage/-impuls, Vertiefung
- ➡ ... mögl. Ergänzung / Anregung [Anm.: Nur wenn es stockt, bzw. extrem knappen Antworten.]

Begrüßung

Danke, dass du bei dem Interview mitmachen möchtest!. Da ich ja gar nicht so schnell mitschreiben könnte, habe ich ein **Aufnahmegerät** dabei, um unser Gespräch **aufzuzeichnen**. Das können wir dann in Ruhe abschreiben! Natürlich verrätst du uns auf der Aufnahme **nicht wie du heißt und auf welche Schule du gehst**. Alles ist **anonym** und die Tondatei ist **nur für uns Forscher** zugänglich. ***Bist du einverstanden? Ja/Nein! Nenne uns deinen Code!***

Einstieg

- Nenne mir die „**Top 3**“ **der Begriffe**, über die du in den letzten Stunden etwas gelernt hast. Auf Platz 1 landet . . . , auf Platz 2 landet . . . und auf Platz 3 . . . !
 - ➡ Es ist überhaupt nicht schlimm, wenn dir nur ein oder zwei einfallen!

Lernumgebung

- Beschreibe, wie du die **Arbeit am Tablet** – in Einzelarbeit – mit Lernprogramm und Simulationen **empfunden** hast!
 - ➡ Konzentration schwierig? // Ungestörtheit angenehm? // Austausch fehlend?
 - ▶ **Erkläre** uns, ob du dir vorstellen kannst, auch **in Zukunft genauso mit Simulationen und Lernprogramm etwas über Physik zu lernen!**
 - ➡ Würdest du nur mit Simulationen lernen wollen, ohne das Lernprogramm? Oder nur ohne die Ankreuz-Fragen? Oder . . . ?
 - ▶ Bei den Simulationen gab es kein Versuchsmaterial! Erkläre uns inwiefern du dir in den letzten Stunden **zusätzlich echte Experimente gewünscht** hättest!
 - ➡ Hättest du mit “echten“ Versuchen oder Simulationen hier besser gelernt?
 - ➡ Bereiten dir die Simulationen oder echte Experimente mehr Freude?
 - ▶ (Als Text vorlegen, siehe S. 292) Begründe, **welcher Aussage du dich eher zuordnen** würdest.
 - ➡ Natürlich soll Lernen Freude machen – aber war es manchmal vielleicht zu verlockend *nur* zu spielen?

Umgang mit Simulationen

○ Unser Testkandidat wollte herausfinden, **welche Eigenschaften die Höchstgeschwindigkeit des Bootes verändern**. Betrachte das Video und gib uns **deine Meinung zu seinem Vorgehen** (aufgezeichnetes Video zeigen)!

➡ Welche Eigenschaften können verändert werden?

▮ Erkläre, welche Fehler gemacht wurden!

➡ Er möchte zunächst z. B. den Einfluss der Masse / des "Gewichts" untersuchen. Gib an, ob seine Einstellungen dafür geeignet sind!

▮ Beschreibe eine korrekte Vorgehensweise!

Inhalt

○ Erkläre, was ein Physiker (oder eine Physikerin) meint, wenn er (oder sie) von **Geschwindigkeit** redet!

▮ (2D) Stelle die beiden Scooter so ein, dass sie (1) die gleiche Geschwindigkeit, (2) das gleiche Tempo aber unterschiedliche Geschwindigkeit haben (Screencapture)

▮ Bearbeite folgende Aufgabe.

○ 1D ("Scooter"-Simulation): **Erkläre den Begriff Kraft mit Hilfe dieser Simulation!**

➡ Überlege, woran man erkennen/sehen kann, dass eine Kraft wirkt! (Ändert sich z. B. die Geschwindigkeit/die Richtung?)

○ 2D ("Mias Kart"-Simulation): **Erkläre den Begriff Kraft mit Hilfe dieser Simulation!**

➡ Überlege, woran man erkennen/sehen kann, dass eine Kraft wirkt / es eine Einwirkung gibt!

Feedback

○ *Simulations-Feedback*: Nach den Simulationen hast du jeweils zwei Ankreuz-Aufgaben bearbeitet, um deine Beobachtungen zu überprüfen. Nach **jeder Antwort hast du einen Hinweis bekommen. Beschreibe** wie du diese **Hinweise genutzt** hast!

▮ Gib an, ob du die Hinweise – speziell auch nach richtigen Antworten – genutzt hast! Waren sie hilfreich?

○ *Animations-Feedback*: Nach den Simulationen hast du jeweils zwei Ankreuz-Aufgaben bearbeitet, um deine Beobachtungen zu überprüfen. Je nach Antwort hast du einen Hinweis bekommen oder konntest dir als Hilfe eine Animation (kurzes Video) anschauen. **Beschreibe** wie du diese **Hinweise oder Animationen genutzt** hast!

▮ Gib an, ob du die Hinweise – speziell auch nach richtigen Antworten – genutzt hast! Waren sie hilfreich?

▮ Erkläre, ob du manches lieber selbst hättest ausprobieren wollen statt nur zuzuschauen!

▮ Gib an, ob es schwer für dich zu erkennen war, was genau dir in der Animation gezeigt wird und wie dir das weiterhelfen sollte!

- Überlege dir, ob du es hilfreicher gefunden hättest, die Rückmeldung z. B. von deinem Lehrer, deinen Mitschülern – oder garnicht – zu erhalten!
 - Ein Lehrer könnte dir ja auch noch auf weitere Fragen antworten – er kann aber z. B. nicht immer bei dir sein.

Sonstiges

- Die Unterrichtseinheit war recht kurz. Ist es vorgekommen, dass dir **inhaltlich etwas unklar** geblieben ist?
 - Gib an, was dir beim Verstehen geholfen hätte. Mehr Zeit? // Hilfe vom Lehrer? // Mehr/andere Texte? Bilder? // Experimente?
- Beschreibe, was wir **verändern** sollten, damit **die Simulationen/das Lernprogramm** – deiner Meinung nach – **besser** werden!
- Jetzt hast du die Chance uns **alles zu sagen, was du sonst noch zu den letzten Stunden sagen möchtest!**

A.4. Ratingkategorien offene Frage

A: Positiver Gesamteindruck In dieser Kategorie werden Bemerkungen zusammengefasst, die keiner speziellen anderen Kategorie wie „Software“ zuzuordnen sind und sich somit eher unspezifisch äußern. Das Positive markierende Wendungen können sein: *Spaß, gut, Freude, toll, wiederkommen nochmal, gefallen*, auch Abschwächungen wie *ganz gut*.

B: Negativer Gesamteindruck In dieser Kategorie werden Bemerkungen zusammengefasst, die keiner speziellen anderen Kategorie wie „Software“ zuzuordnen sind und sich somit eher unspezifisch äußern. Das Negative markierende Wendungen können sein: *keinen Spaß, langweilig, nicht so gut, nichts gelernt, cool*, auch Abschwächungen wie *hätte besser sein können*.

C: Positiver Eindruck der Helfer In dieser Kategorie werden Äußerungen zu den betreuenden Personen zusammengefasst. Positive Wendungen werden exemplarisch etwa durch *nett, hilfreich* oder *freundlich* angezeigt.

D: Negativer Eindruck der Helfer In dieser Kategorie werden Äußerungen zu den betreuenden Personen zusammengefasst. Negative Wendungen werden exemplarisch etwa durch *unfreundlich* oder *nicht weitergeholfen* angezeigt.

E: Pos. Anmerkungen zu allg. (Tablet)arbeit inkl. technischer Aspekte

In dieser Kategorie werden Äußerungen zum Tablet an sich zusammengefasst. Diese beziehen sich auf technische Aspekte, wie etwa eine angenehme Bedienung oder das allgemeiner Setting (z. B. Einzelarbeit). Davon ausgeschlossen sind Äußerungen zum *Gehalt* des Lernprogramms, wie etwa dem empfunden Lernzuwachs oder der Schwierigkeit. Typische Indikatoren für diese Kategorie sind Wendungen wie *Arbeit mit Tablets war gut*.

F: Neg. Anmerkungen zu allg. (Tablet)arbeit inkl. technischer Aspekte

In dieser Kategorie werden Äußerungen zum Tablet an sich zusammengefasst. Diese beziehen sich auf technische Aspekte, wie etwa eine angenehme Bedienung oder das allgemeiner Setting (z. B. Einzelarbeit). Davon ausgeschlossen sind Äußerungen zum *Gehalt* des Lernprogramms, wie etwa dem empfunden Lernzuwachs oder der Schwierigkeit. Typische Indikatoren für diese Kategorie sind Wendungen wie *Speicher hat nicht funktioniert, Einzelarbeit war doof*.

G: positiver Eindruck des Lernprogramms

Hier werden nun die Anmerkungen zum Programm an sich und/oder der Auseinandersetzung mit Inhalten beschrieben. Exemplarische, positive Indikatoren sind hier etwa *viel mit Programm gelernt, verständliche Texte, Simulationen haben Spaß gemacht* oder *die Videos haben geholfen*.

H: negativer Eindruck des Lernprogramms

Hierunter fallen Bemerkungen zum Lernprogramm, in denen es um die Auseinandersetzung mit Inhalten und/oder dem Programm an sich geht. Exemplarische, negative Kommentare wären z. B. *habe mit Programm nichts gebracht, Die Simulationen waren verwirrend* oder *Das Programm war viel zu lang!*.

I: Konstruktive Verbesserungsvorschläge

In dieser Kategorie werden Vorschläge seitens der Schülerinnen und Schüler aufgenommen. Fiktive Beispiele wären etwa *Bitte wichtige Stellen markieren, eine Zusammenfassung alle 3 Seiten wäre cool*. Dazu gehören aber auch eher implizite Wünsche wie *Mit Maus wäre es einfacher*, oder *Gruppenarbeit hat gefehlt*.

Sonstige Regel: Einordnung in 2 Kategorien nur, wenn strikte Trennung tatsächlich möglich ist. Bei I hieße das, dass zusätzlich nur negativ kodiert wird, wenn die SuS auch klare Kritik formulieren (*Die Texte waren viel zu kompliziert. Bitte verständlichere Texte!*)



Abbildung A.1.: Fiktive Schüleräußerungen zur Vorlage im Interview, zwei Versionen (Bilderquelle: Pixabay, gemeinfrei)

B. Simulationen

ACHTERBAHN

1D: 2D: ✓

Inhalt:

Bewegungsbeschreibung
und Einwirkungen



Beschreibung: Seitliche Ansicht einer Achterbahn mit “Drop” und stationärer Bremse. Höhen- bzw. Intensitätsverstellbar

“Forschungsauftrag”: *Finde mit der Simulation etwas über die Bewegung der Wagen heraus!*

Didaktischer Kommentar: Sensibilisierung für Aspekte der Bewegungsbeschreibung; aber auch sich durch niedrige Hemmschwelle mit Simulationen als solchen vertraut machen.

SCOOTER

1D: 2D: ✓

Inhalt:

Bewegungsbeschreibung



Beschreibung: Draufsicht auf einen Autoscooter, wobei nur ein Fahrzeug steuerbar ist, aber mit anderen interagieren kann.

“Forschungsauftrag”: *Finde mit der nächsten Simulation etwas über die Bewegung der Wagen heraus!*

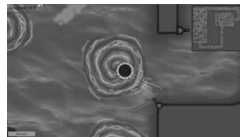
Didaktischer Kommentar: Sensibilisierung für Aspekte der Bewegungsbeschreibung; aber auch sich durch niedrige Hemmschwelle mit Simulationen als solchen vertraut machen.

RAFTING (MIT PFEIL)

1D: 2D: ✓

Inhalt:

Bewegung und Geschwindigkeitsbegriff



Beschreibung: Raftingbahn in Draufsicht – Steuerung des kleinen, runden Bootes mithilfe diverser Wasserstrudel und -düsen.

“Forschungsauftrag”: *In der nächsten Simulation kannst du die Bewegung des Bootes der Wildwasserbahn beeinflussen! Finde heraus, mit welchen Begriffen Max die Bewegung des Bootes durch die Bahn so genau beschreiben kann, dass sie sich sein kranker Freund Tobi genau vorstellen kann!*

Didaktischer Kommentar: Durch die Gegebenheiten und den Forschungsauftrag wird die Aufmerksamkeit auf die Faktoren *Richtung* (nach oben, nach unten, um den Stein herum...) und *Tempo* (schnell(er), langsam(er) werdend...) gelenkt. In einer weiteren Variante der Simulation wird sogleich der Geschwindigkeitspfeil mit angezeigt.

ZIELFOTO

1D: ✓ 2D: ✓

Inhalt:

Geschwindigkeits- bzw. Tempobegriff



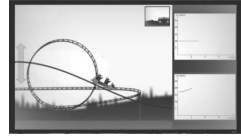
Beschreibung: Achterbahn in Seitenansicht. Einfahrender Zug kann durch stationäre Bremse in unterschiedlicher Intensität verlangsamt werden. Jede Sekunde wird dabei im Stil einer Stroboskop-Kamera eine Aufnahme gemacht.

“Forschungsauftrag”: *In der Simulation ist die Bahn und die Kamera dargestellt. Finde heraus, wie **und warum** du anhand der Fotoserien – auch ohne “wehende Haare” – erkennen kannst, ob die Bahn langsam oder schnell gefahren ist.*

Didaktischer Kommentar: Je nach Lernumgebung wird zwischen Geschwindigkeit und Tempo unterschieden. Zentral ist jeweils das Verhältnis von zurückgelegter Strecke zu einem festen Zeitintervall (hier durch die Fotos im Sekundentakt).

ACHTERBAHN- 1D: ✓ 2D: ✓ KONTROLLE

Inhalt:
t-v-Diagramme



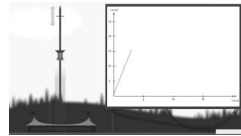
Beschreibung: Achterbahn in zwei unterschiedlichen Seitenansichten. Zug kann durch diverse Parameter unterschiedlich schnell fahren gelassen werden. Das jeweilige Tempo bzw. Geschwindigkeit wird dabei direkt im jeweiligen Diagramm auf der rechten Seite laufend aufgezeichnet und dargestellt.

“Forschungsauftrag”: *In der Simulation auf der nächsten Seite schlüpfst du in die Rolle eines Mitarbeiters, der die Bahn bedient. Finde heraus, was er aus dem Diagramm über die Bewegung der Bahn erfahren kann!*

Didaktischer Kommentar: Je nach Lernumgebung wird zwischen Geschwindigkeit und Tempo unterschieden. Durch die Variation des Geschwindigkeitsbetrages wird es möglich die Analogie zwischen der grafischen Darstellung im Diagramm und der “Realität” zu erkennen.

FREIFALLTURM 1D: ✓ 2D: ✓

Inhalt:
t-x-Diagramme



Beschreibung: Freifallturm in Seitenansicht. Die Ausklinkhöhe sowie die Länge der Bremsstrecke kann variiert werden.

“Forschungsauftrag”: *In dieser Simulation versetzt du dich in die Rolle des Mitarbeiters, der den Turm bedient! Finde heraus, welche Informationen über die Bewegung er aus dem Diagramm bekommen kann!*

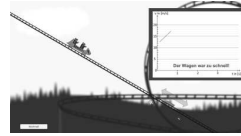
Didaktischer Kommentar: Anbahnung von t-x-Diagrammen. Durch Variation unterschiedlicher Parameter und Beobachtung der aktuellen Höhe kann die Analogie zwischen der grafischen Darstellung und der “Realität” erkannt werden.

BREMS ODER CRASH

1D: 2D: ✓

Inhalt:

t-v-Diagramm ablesen



Beschreibung: Seitenansicht einer Achterbahn. Die Intensität der Bremse ist variierbar.

“Forschungsauftrag”: Hier bist du nun selbst für die Sicherheit der Bahn verantwortlich! Aber natürlich auch für den Spaß der Fahrgäste! Der Wagen darf nicht schneller werden als 18m/s ! Versuche die Bremse so einzustellen und am Diagramm die Geschwindigkeit zu überprüfen.

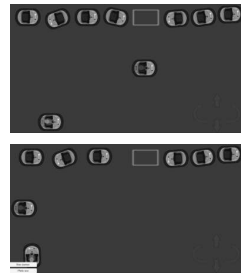
Didaktischer Kommentar: Das Ablesen eines t-v-Diagramm soll hier noch einmal geübt werden – mit der Motivation, dass der Zug ein bestimmtes Tempo nicht überschreiten darf.

SCOOTER 1 & SCOOTER 2

1D: 2D: ✓

Inhalt:

Einwirkung & Zusatzgeschwindigkeit



Beschreibung: Draufsicht eines Autoscooters. Ein Scooter (“Mia”) ist dabei steuerbar und kann mit einem anderen – zunächst ruhenden – Scooter (“Max”) durch Stöße interagieren.

“Forschungsauftrag”: Für Scooter 1: Finde heraus, durch welche Einwirkung von Mia ihr Freund Max in die Lücke geschubst werden kann. Spielregel: Mia darf Max nur einmal schubsen/stoßen! Nicht schieben! Für Scooter 2: Versetze dich wieder in die Rolle von Mia und versuche Max mit der selben Einwirkung wie vorher (oder wie im Bild) senkrecht in die „Parklücke“ zu stoßen! Spielregel: Mia darf Max nur einmal schubsen/stoßen! Nicht schieben! Für Scooter 2 (Teil 2): Na los! Finde heraus, durch welche Einwirkung du Max in die Lücke stoßen kannst! Spielregel: Du darfst nur einmal schubsen/stoßen!

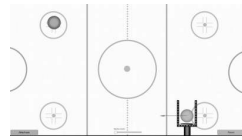
Didaktischer Kommentar: Zu beachten ist hier die Vernachlässigung diverser Einflussfaktoren. (hier näher erläutern). Grundsätzlich ist im Autoscooter kein Rutschen in dem Maße wie in der Simulation möglich. Geschwindigkeitspfeile sind zuschaltbar.

AIRHOCKEY

1D: 2D: ✓

Inhalt:

Einwirkungsstärke & Zusatzgeschwindigkeit



Beschreibung: Draufsicht auf einen Airhockeytisch¹. Ein Puck kann dabei aus einer sich bewegenden Vorrichtung senkrecht zur Bewegungsrichtung mittels einer Feder unterschiedlicher "Stärke" abgeschossen werden.

"Forschungsauftrag": *Finde etwas über die Wirkung der unterschiedlichen Abschussmöglichkeiten beim Airhockey heraus!*

Didaktischer Kommentar: Reibungsfreiheit, Geschwindigkeitspfeile zuschaltbar, bedingte Gamification (Treffer des grünen Pucks).

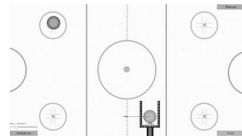
¹ Zum Spiel Airhockey: "Gespielt wird auf einem Spezialtisch in Billardtischgröße. Eine ebene, glatte Platte dient als Spielfeld. Bei manchen Tischen wird durch viele kleine Löcher Luft geblasen, wodurch unter dem Spielpuck ein Luftkissen entsteht, auf dem der Puck praktisch ohne Reibungsverluste sehr schnell gleitet." Aus: Wikipedia-Artikel *Air-Hockey*, abgerufen am 19.4.2018

VENTILATOR

1D: 2D: ✓

Inhalt:

Einwirkungsdauer & Zusatzgeschwindigkeit



Beschreibung: Basis wie *Airhockey*. Abschuss hier nicht über Feder sondern über Ventilator/Gebläse mit variabler Wirkungsdauer. Angelehnt an Simulationssoftware „Kraftstoß“. www.thomas-wilhelm.net/simu_stoss.zip Zuletzt geprüft am 30.11.2020

"Forschungsauftrag": *Finde etwas über die neue Abschussmöglichkeit und deren Einfluss auf die Zusatzgeschwindigkeit heraus!*

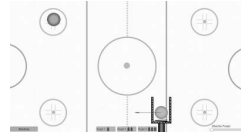
Didaktischer Kommentar: Reibungsfreiheit, Geschwindigkeitspfeile zuschaltbar, bedingte Gamification (Treffer des grünen Pucks).

NEUE PUCKS

1D: 2D: ✓

Inhalt:

Masse des Körpers
& Zusatzgeschwindigkeit



Beschreibung: Prinzipiell wie *Airhockey*, hier aber Variation der Massen der Pucks möglich.

“Forschungsauftrag”: *Untersuche den Einfluss der unterschiedlichen Pucks und der Federn auf die Zusatzgeschwindigkeit!*

Didaktischer Kommentar: Reibungsfreiheit, Geschwindigkeitspfeile zuschaltbar, bedingte Gamification (Treffen des grünen Pucks).

ROLLER-COASTER

1D: ✓ 2D:

Inhalt:

Bewegungen beschreiben



Beschreibung: Seitenansicht einer horizontalen Achterbahnstrecke mit Zug, der per Katapultstart gestartet und per stationärer Bremse verlangsamt werden kann (jeweils mit Intensitätsparametern).

“Forschungsauftrag”: In diesen Simulationen kannst du Attraktionen im Freizeitpark selbst ausprobieren. Finde mit den Simulationen etwas über die Bewegungen der Wagen und Boote heraus!

Didaktischer Kommentar:

SCOOTER (1D- UMGEBUNG)

1D: ✓ 2D:

Inhalt:

Bewegungen beschreiben



Beschreibung: Seitenansicht eines Autoscooters. Nach links und rechts in unterschiedlicher Geschwindigkeit steuerbar

“Forschungsauftrag”: In diesen Simulationen kannst du Attraktionen im Freizeitpark selbst ausprobieren. Finde mit den Simulationen etwas über die Bewegungen der Wagen und Boote heraus!

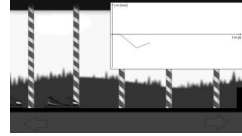
Didaktischer Kommentar:

DIAGRAMM UND SCOOTER

1D: ✓ 2D: e

Inhalt:

Bewegungen beschreiben



Beschreibung: Seitenansicht eines Autoscooters. Nach links und rechts in unterschiedlicher Geschwindigkeit steuerbar

“Forschungsauftrag”: In der nächsten Simulation siehst du das t-v-Diagramm des Scooters. Finde heraus, welche Informationen man diesem t-v-Diagramm entnehmen kann!

Didaktischer Kommentar: Je nach Bewegungsrichtung werden im t-v-Diagramm negative bzw. positive Werte für die Geschwindigkeit angezeigt.

MIAS KART

1D: ✓ 2D:

Inhalt:

Kraftwirkungen



Beschreibung: Seitenansicht eine Karts, das aufgrund eines „Motorschadens“ zunächst liegenbleibt. Mit div. Einflüssen kann die Geschwindigkeit verändert werden.

“Forschungsauftrag”: In der nächsten Simulation kannst du Mias Fahrt nachstellen. Finde heraus, wie sich Mias Bewegung beeinflussen lässt! Öffne dazu die Simulation

Didaktischer Kommentar: Geschwindigkeitsanzeige zuschaltbar. Nach gewisser Zeit stellt sich minimaler Reibungseffekt ein.

WILD- WASSERBAHN

1D: ✓ 2D:

Inhalt:

2. Newtonsches Axiom



Beschreibung: Seitenansicht einer Wildwasserbahn. Masse (“Beladung”)

“Forschungsauftrag”: In der nächsten Simulation kannst du die Wildwasserbahn bedienen. Untersuche die Auswirkung der Beladung sowie des Wasserstrahls auf die Geschwindigkeit!

Didaktischer Kommentar: Der Abstand zwischen Wasserkanone und Boot verändert sich, weshalb die wirkende Kraft nicht konstant sein kann, sofern sich die Intensität des Wasserstrahls nicht ändert. Dieser Faktor wird vernachlässigt.

C. Weiteres Material



Ein Tag im Freizeitpark – Bewegungen entdecken

„Spickzettel“

Teil 1: _____ (Die Überschrift für den Teil bestimmst du!)

Das habe ich im letzten Teil des Lernprogramms gelernt / Das habe ich mir gemerkt (du darfst schreiben, malen, zeichnen.....):

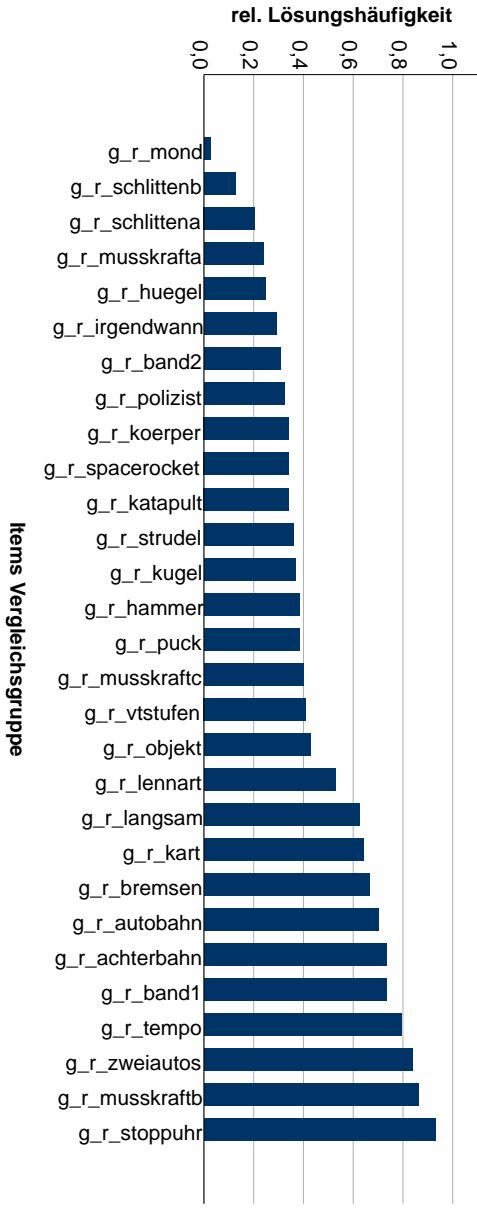
Hinweis: Versuche bitte deine eigenen Worte und Zeichnungen zu verwenden!

© LUH, IDMP, AG Physikdidaktik

Abbildung C.1.: Exemplarischer Notizzettel, der von Schülerinnen und Schülern während der Lerneinheit an unterschiedlichen Stellen ausgefüllt werden sollte.

D. Ergänzungen zu Ergebnissen

Abbildung D.1.: Relative Lösungshäufigkeit aller Items der Vergleichsgruppe



D.1. Codierung der fachlichen Interviewinhalte

Code	Bezeichnung	Fiktives Beispiel
Geschwindigkeit 1D-SuS		
G1A	Geschwindigkeit als Schnellsein	"Gibt an wie schnell jemand ist"
G1B	nur Nennung notwendiger Begriffe	"Hat was mit Strecke und Zeit zu tun"
G1C	(nahezu) korrekte oder prinzipielle Beispielerklärung	"Also wenn ich 4km in einer Stunde laufe, dann bin ich 4km/h schnell"
G1D	Geschwindigkeit als zurückgelegte Strecke pro Zeit	"Die Geschwindigkeit gibt an wie viel zurückgelegter Strecke in welcher Zeit an"
G1E	sonstige unspezifische Anmerkungen (evtl. mit Fachberiffen)	"Dann also ist Geschwindigkeit ziemlich wichtig"
G1F	(teilweise) falsch Erklärung	"Wenn eine Kraft wirkt hat man eine Geschwindigkeit!"
Woran erkennt man Kraftwirkung?		(1D erklärt 2D)
K1A	Bewegung	"Der Autoscooter bewegt sich"
K1B	Bewegungsänderung (umgangssprachlich)	"Der Scooter steht erst auf der Stelle, dann rutscht er weg"
K1C	Richtung verändert sich / Geschwindigkeit verändert sich (evtl. mit Beispiel)	"Der Scooter wird schneller / wird abgelenkt"
K1D	Bewegungszustand ändert sich	"Der Scooter verändert seine Bewegung/Bewegungszustand"
K1E	sonstige unspezifische Anmerkung (evtl. mit Fachbegriffen)	"Geschwindigkeit und Kraft sind ja wichtige Dinge"

K1F	(falsche) Erklärung	Es muss immer eine Kraft wirken, damit es sich bewegt
Variablenkontrolle 1D-SuS		
V1A	Fehler wird nicht erkannt	”Würde ich auch so machen”
V1B	im Beispiel nicht alle Kombinationen ausprobiert	”Er hat ja garnicht alles getestet”
V1C	Mangelnde Variablenkontrolle am Beispiel erläutert	”Da gleicht sich wieder was aus, weil er größere Masse und größere Zeit
V1D	Mangelnde Variablenkontrolle	”Er verändert mehrere Einstellungen gleichzeitig”
V1E	Sinnvoller Verbesserungsvorschlag	„Er sollte nur eine Sache verändern!“
V1F	falsche oder unspezifische Verbesserungsvorschläge	

Code	Bezeichnung	Fiktives Beispiel
Geschwindigkeit 2D-SuS		
G2A	Geschwindigkeit als Schnellsein	”Gibt an wie schnell jemand ist”
G2B	Korrekte Beispielrechnung Tempo	”Also wenn ich 4km in einer Stunde laufe, dann bin ich 4km/h schnell”
G2C	nur Nennung notwendiger Begriffe Tempo und/oder Richtung	”Hat was mit Schnelligkeit und Richtung zu tun”
G2D	gerichtete Größe mit Betrag als Tempo, Tempo als Strecke pro Zeit	”Geschwindigkeit gibt an, wie schnell sich etwas in welche Richtung bewegt”

G1E	sonstige unspezifische Anmerkungen (evtl. mit Fachberiffen)	”Dann also ist Geschwindigkeit ziemlich wichtig”
G1F	(teilweise) falsch Erklärung	”Wenn eine Kraft wirkt hat man eine Geschwindigkeit!”
Woran erkennt man Kraftwirkung?		
K2A	Bewegung	”Das Kart bewegt sich!”
K2B	Bewegungsänderung (umgangssprachlich)	”Das Kart verändert seine Bewegung”
K2C	Richtung verändert sich / Geschwindigkeit verändert sich (evtl. mit Beispiel)	”Das Kart verändert seine Geschwindigkeit / Tempo”
K2D	„Entstehung“ einer Zusatzgeschwindigkeit	”Das Kart erhält eine Zusatzgeschwindigkeit”
K2E	sonstige unspezifische Anmerkung (evtl. mit Fachbegriffen)	”Geschwindigkeit und Kraft sind ja wichtige Dinge”
K2F	(falsche) Erklärung	”Es muss immer eine Kraft wirken, damit es sich bewegt”
Variablenkontrolle 2D-SuS		
V2A	Fehler nicht erkannt	„Würde ich auch so machen“
V2B	im Beispiel nicht alle Kombinationen ausprobiert	”Er hat ja garnicht alles getestet”
V2C	Mangelnde Variablenkontrolle am Beispiel erläutert	”Da gleicht sich wieder was aus, weil er größere Masse und größere Zeit
V2D	Mangelnde Variablenkontrolle	”Er verändert mehrere Einstellungen gleichzeitig”
V1E	Sinnvoller Verbesserungsvorschlag	”Er sollte nur eine Sache verändern!”

V2F falsche oder unspezifische Verbesserungsvorschläge evtl trennen
