

Strukturierung von Lehr-Lern-Sequenzen



DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES DER NATURWISSENSCHAFTEN
(DR. RER. NAT.) DER FAKULTÄT FÜR PHYSIK
DER UNIVERSITÄT REGENSBURG

vorgelegt von

Christian Maurer

aus Bogen

im Jahr 2016

Promotionsgesuch eingereicht am: 18.01.2016

Die Arbeit wurde angeleitet von: Prof. Dr. Karsten Rincke

Prüfungsausschuss:

Prof. Dr. Richter (Vorsitzender)

Prof. Dr. Karsten Rincke (1. Gutachter)

Prof. Dr. Oliver Tepner (2. Gutachter)

Prof. Dr. Huber (weiterer Prüfer)

Termin Promotionskolloquium: 17.03.2016

Inhaltsverzeichnis

1. Einleitung	5
2. Strukturierung	9
2.1. Begriffliche Klärung	9
2.2. Strukturierung als Qualitätsmerkmal von Unterricht	20
2.3. Arten von Strukturvorgaben	28
2.3.1. Strukturvorgaben zum Inquiry-Learning	29
2.3.2. Direct Instruction	40
2.3.3. Forschend-entwickelndes Unterrichtsverfahren	43
2.3.4. Basismodelltheorie	50
2.3.5. Beziehungen zwischen den Strukturvorgaben	61
2.4. Zusammenfassung des Forschungsstandes und Formulierung der Forschungsfragen	69
3. Konzeption und Durchführung der Studie	75
3.1. Studiendesign	75
3.2. Auswahl der Strukturvorgaben	77
3.3. Konzeption der Unterrichtseinheiten	81
3.4. Analyse der Unterrichtseinheiten	89
3.5. Methoden der Datenerhebung	93
3.5.1. Variablen der Untersuchung	94
3.5.2. Ablauf der Untersuchung	95
3.5.3. Wissenstest Impuls (WTI)	96
3.5.4. Test zur Kompetenz zur Erkenntnisgewinnung (WTE)	99
3.5.5. Kognitiver Fähigkeitstest (KFT)	100
3.5.6. Schülerfragebogen	101
3.6. Stichprobe	104
3.7. Methoden der Auswertung	105
4. Ergebnisse der Studie	111
4.1. Ergebnisse zu den Kontrollvariablen	111
4.2. Ergebnisse zu den abhängigen Variablen	114
4.3. Zusammenfassende Ergebnisse	117
4.3.1. Analyse der Gesamtstudie	117
4.3.2. Analyse von Teilgruppen der Gesamtstudie	120
4.3.3. Zusammenfassung der Ergebnisse der Kovarianzanalysen	127

4.3.4.	Effekte der Intervention unter Kontrolle der Kompetenz der Erkenntnisgewinnung (WTE)	128
4.3.5.	Effekte der Intervention unter Kontrolle der kognitiven Grundfähigkeiten (KFTk)	129
4.3.6.	Einfluss des Ergebnisses im Vortest des WTI	131
4.4.	Zusammenhang der Merkmale des Schülerfragebogens mit dem WTI	134
5.	Interpretation und Einordnung der Ergebnisse	139
5.1.	Wirkungen der Interventionen hinsichtlich der Forschungsfragen	139
5.1.1.	Umsetzbarkeit und Lernwirksamkeit des FeU	140
5.1.2.	Lernwirksamkeit der Interventionsstunden	141
5.1.3.	Lernwirksamkeit der Intervention bei Teilgruppen	145
5.1.4.	Schülerwahrnehmung der Interventionsstunden	149
5.1.5.	Abschließender Vergleich der Strukturvorgaben	152
5.2.	Interventionsgruppenunabhängige Studienergebnisse	154
5.2.1.	Geschlechterspezifische Unterschiedlichkeit des generellen Lernzuwachses	154
5.2.2.	Einfluss verschiedener Kontrollvariablen auf das Nachtestergebnis im WTI	155
5.2.3.	Ausprägung der allgemeinen Qualitätsmerkmale und deren Zusammenhang mit dem Nachtestergebnis im WTI	160
6.	Die Studie in der kritischen Rückschau	165
7.	Zusammenfassung und Ausblick	169
A.	Unterrichtsmaterialien	173
B.	Instrumente der Datenerhebung	181
C.	Ergebnisse der Kovarianzanalysen zum Moderatoreneffekt	193
	Danksagung	195
	Literaturverzeichnis	197

1. Einleitung

Wie gestalte ich Unterricht lernwirksam? - dazu gehört auch immer die Überlegung wie die Vermittlung von Informationen erfolgen soll. Diese Informationen müssen dabei „korrekt sein und so *klar und verständlich* präsentiert und *strukturiert* werden, dass sie auf Schülerseite Lernprozesse anregen“ (Helmke, 2009, S. 191; Hervorhebungen im Original). In diesem Zitat zeigen sich grundsätzlich drei in der Realität oft eng zusammenhängende aber unterschiedliche Aspekte, welche für das Lernen im Unterricht von großer Bedeutung sind: Klarheit, Verständlichkeit und Strukturierung. Die Klarheit weist inhaltliche und sprachliche Aspekte auf und ist dabei im Gegensatz zur rein von sprachlichen Merkmalen geprägten Verständlichkeit eher sender- denn empfängerbezogen. Die Strukturierung beschreibt die Planung und Sequenzierung von Unterricht gerade in der Art, dass beim Lernenden eine gut organisierte Wissensbasis aufgebaut wird (Helmke, 2009, S. 191f). Diese Planung und Sequenzierung beinhaltet Überlegungen zu eingesetzten Methoden, aber auch zur Entscheidung, welche Phasen im Unterricht vorkommen sollen, wie diese ausgestaltet und verbunden werden können. Den Rahmenkontext dieser Arbeit bildet die Strukturierung des Unterrichts, wobei der Schwerpunkt des Interesses auf den unterrichtlichen Auswirkungen verschiedener Phasierungen oder Sequenzierungen von Unterricht liegt. Für diese existieren verschiedene Vorschläge.

Einer der vielleicht am weitesten verbreiteten und zugleich missverstandenen Vorschläge ist die »Direct Instruction«. Missverstanden ist diese dahingehend, dass man es fälschlicherweise mit dem gleichsetzt, was man im Deutschen unter Frontalunterricht oder im Englischen schlicht unter »teacher led talking from the front« versteht. »Direct Instruction« hingegen, ist ein sehr lernwirksamer Vorschlag die Phasierung und Sequenzierung von Unterricht zu gestalten (Hattie, 2009, S. 204ff). »Direct Instruction« ist dabei sogar so erfolgreich, dass Adams und Engelmann (1996, S. 98) als Ergebnis ihrer Metastudie zur Untersuchung der Wirksamkeit eben dieser Art der Sequenzierung von Unterricht als Fazit formulieren: „Direct Instruction has no peer in this enterprise“. Dabei ist dieser Vorschlag nicht der einzige, dem eine optimale Gestaltung des Ablaufs von Unterricht zugeschrieben wird. So sprechen Schmidkunz und Lindemann (1992, S. 51) von der „Idealstruktur des Forschend-entwickelnden Unterrichtsverfahrens“, welches Vorgaben über die Gestaltung naturwissenschaftlichen Unterrichts liefert, und Oser und Baeriswyl (2001, S. 1043) davon, dass in ihrer fachunspezifischen Theorie der Basismodelle eben diese durch nichts anderes ersetzt werden könnten. Jede dieser drei exemplarisch ausgewählten Vorgaben

1. Einleitung

erlaubt eine Planung von Physikunterricht und hegt den Anspruch diesen optimal zu gestalten. Eine Entscheidung für oder gegen eine der Vorgaben kann auf dieser Basis nicht getroffen werden. Eine geeignete Basis könnte aber der jeweilige Einfluss auf den Lernerfolg sein.

Ein bedeutsamer Einfluss auf den Lernerfolg scheint generell durch Elemente der Strukturierung gegeben zu sein (z.B. Brophy, 2000; Meyer, 2004; Helmke, 2009). Rakoczy et al. (2007, S. 116) konnten diesen Zusammenhang mit den Ergebnissen einer deutsch-schweizerischen Studie an 37 Schulen auch eindeutig belegen. „With these results, the present study confirms one of the central findings from research on instructional quality: that clear systematic structuring is one of the essential components of good instruction“. Dabei wird der Begriff der Strukturierung aber nicht einheitlich verwendet (Lipowsky, 2009, S. 83). Lipowsky (2009, S. 83) kategorisiert diesen Begriff für eine systematische Betrachtung in drei Bedeutungsfacetten (didaktische Strukturiertheit, Strukturiertheit auf der Verhaltensebene und kognitionspsychologisch verstandene Strukturiertheit). Die Phasierung und Sequenzierung von Unterricht zählt dabei für Lipowsky zur sogenannten didaktischen Strukturierung. Ein Mindestmaß dieser Facette der Strukturierung stellt für Lipowsky dabei eine notwendige Voraussetzung dafür dar, dass Elemente der anderen beiden Bedeutungsfacetten überhaupt ihre Wirkung entfalten können (Lipowsky, 2009, S. 86). Der didaktischen Strukturierung, damit auch der Sequenzierung von Unterricht und der Frage nach der geeignetsten Vorgabe hierzu, wird folglich eine bedeutende Rolle hinsichtlich ihres Einflusses auf die Qualität von Unterricht zuteil. Der Einfluss verschiedener Strukturvorgaben wie der »Direct Instruction«, dem »problem-based learning«, oder dem »inquiry-based teaching« auf den Lernerfolg ist dabei unterschiedlich und abhängig vom jeweiligen Kontext, Fach und Verwendungszweck (Hattie, 2009, S. 201ff). Der Einfluss anderer Vorgaben wie dem »Forschend-entwickelndem Unterrichtsverfahren« oder der »Basismodelltheorie« ist nicht oder noch unzureichend empirisch erforscht, oft fehlen dabei auch faire Vergleiche mit echten Kontrollgruppen, denen ebenfalls eine theoretische Fundierung für die Unterrichtsgestaltung zugrunde liegt.

Unklar ist allerdings, welche dieser Vorgaben zur Strukturierung von Unterricht nun für das Fach Physik oder auch nur für spezielle Kontexte zur lernwirksameren Gestaltung führt. Die Basis für eine Entscheidung bei der Auswahl der im spezifischen Fall geeignetsten Vorgabe ist damit nicht gegeben. Dabei ist die Frage, ob eine optimale Strukturierung überhaupt existiert ebenso offen wie die Frage danach, ob das reine Ausmaß an Strukturierung oder die spezielle Form der Strukturierung für die Qualität des Unterrichts prägend ist. Zur näheren Beleuchtung dieser Fragen will diese Arbeit einen Beitrag leisten. Dieser Beitrag soll speziell Aussagen über unterschiedliche Auswirkungen unterschiedlicher Arten der Sequenzierung ermöglichen, kann dabei aber nicht so grundsätzliche Fragen wie die nach der Existenz einer optimalen Strukturierung beantworten und soll auf der Basis eines Vergleichs zweier didaktisch gut begründeter

Vorgaben für eine unterrichtliche Strukturierung ermittelt werden.

Grundlegend für diesen Vergleich sind zum einen das Forschend-entwickelnde Unterrichtsverfahren nach Schmidkunz und Lindemann (1992), zum anderen die Theorie der Basismodelle nach Oser und Baeriswyl (2001). Diese beiden Vorgaben wurden für die vorliegende Arbeit gewählt, da diese den Unterricht vergleichbar stark, aber abhängig von der Art des angestrebten Lernvorgangs strukturieren und beide ausführlich dokumentiert sind. Gerade die gute Dokumentation der Theorie soll es ermöglichen beide Strukturvorgaben möglichst exakt für die Studie umsetzen zu können. Für die Untersuchung wurden auf Grundlage je einer der beiden ausgewählten Vorgaben zwei 90-minütige Unterrichtseinheiten zum Thema Impuls konzipiert, anschließend Schulklassen in das Lernforschungslabor der Universität Regensburg geladen und unterrichtet. Begleitend wurden die notwendigen Daten erhoben, um Aussagen bezüglich des Lernzuwachses, Aspekten der Motivation und kognitiven Aktivierung und der Strukturierung des Unterrichts ermöglichen zu können.

1. Einleitung

2. Strukturierung

Die Verwendung des Begriffs der Strukturierung erfolgt in der Literatur selten einheitlich (Lipowsky, 2009, S. 83), zudem werden weitere Begriffe in sehr engem Zusammenhang zur Strukturierung thematisiert und oft mit dieser in einem Konstrukt zusammengefasst. Stellvertretend dafür stehen die begrifflichen Festlegungen der *klaren Strukturierung* bei Meyer (2004) und der *strukturellen Klarheit* bei Drollingen-Vetter (2011). Klarheit, Strukturierung und beim Begriff der strukturellen Klarheit zusätzlich die Kohärenz des Inhalts bilden gemeinsam ein Konstrukt. Diese und weitere begriffliche Festlegungen im Kontext des Begriffs der Strukturierung werden im nachfolgendem Abschnitt näher spezifiziert. Aussagen über Effekte und Einflüsse der Strukturierung auf unterrichtliche Aspekte sind aufgrund der Variationen der Bedeutung dieses Begriffs auch hinsichtlich der jeweils gewählten Definition der Strukturierung verschiedener Autoren zu interpretieren. Deshalb soll im Folgenden die begriffliche Variation der Strukturierung an ausgewählten Beispielen dargestellt werden. Auf deren Basis wird anschließend die Arbeitsdefinition der Strukturierung im Rahmen dieser Arbeit angegeben und in nachfolgenden Abschnitten Aussagen zu Wirkungen der Strukturierung bzw. derer Teilkategorien thematisiert.

2.1. Begriffliche Klärung

Bei der Darstellung erfolgreicher Unterrichtsmerkmale aus Sicht der pädagogischen Psychologie liefert Lipowsky (2009) eine grundlegende Systematisierung der begrifflichen Variationen zur Strukturierung anhand von drei Bedeutungsfacetten. Die Facette der *didaktischen Strukturierung* umfasst die inhaltliche Aufbereitung in Teilschritte und eine für die Schüler wahrnehmbare Gliederung und Phasierung des Unterrichts. Die *didaktische Strukturierung* „setzt einen sorgfältig geplanten Unterricht voraus und kann somit als wichtige Voraussetzung für angemessene Anforderungen an die Lernenden begriffen werden“ (Lipowsky, 2009, S. 83). Diese sorgsam vorbereitete Gestaltung des Unterrichts kann dabei beispielsweise nach den Vorgaben der »Direct Instruction« erfolgen. Die Facette der *Strukturierung auf der Verhaltensebene* beschreibt hingegen Aspekte zur Steuerung des Verhaltens und der Disziplin der Lernenden im Klassenzimmer. Ein in diesem Sinne gut strukturierter Unterricht zeichne sich nach Lipowsky durch eine Konsistenz von Regeln, Erwartungen und Grenzen aus und soll eine störungsfreie Lernumgebung sowie ein hohes Maß an echter Unterrichtszeit gewährleisten. Methoden zur Verbindung von Elementen des Vorwissens mit

2. Strukturierung

neuem Wissen, zur Fokussierung der Aufmerksamkeit der Lernenden auf „relevante Aspekte des Unterrichtsgegenstandes“ (Lipowsky, 2009, S. 83) oder zur Verankerung des Wissens (z.B. Strukturierungshilfen wie »Advance Organizer«) bilden die dritte Facette der *kognitionspsychologisch verstandenen Strukturiertheit* (Lipowsky, 2009, S. 83; Lipowsky verwendet die Begriffe Strukturiertheit und Strukturierung dabei synonym). Diese Kategorisierung der gesamten begrifflichen Bandbreite der Strukturierung ermöglicht eine geordnete Betrachtung einzelner Aspekte oder Auffassungen der Bedeutung der Strukturierung von Unterricht. Diese Betrachtung verbleibt dabei aber von der Interpretation und damit vom Autor abhängig, auch da die einzelnen Bedeutungsfacetten in der Darstellung Lipowskys nicht im Detail ausdifferenziert werden. Vielmehr findet in diesem Überblicksbeitrag über Merkmale und Merkmalskonfigurationen erfolgreichen Unterrichts eine heuristische Systematisierung verschiedener Auffassungen der Bedeutung des Begriffs der Strukturierung statt. Insgesamt ermöglicht diese heuristische Systematisierung durch Lipowsky aber einen eben systematischen Blick auf Teilaspekte oder andere Auffassungen des Begriffs der Strukturierung von Unterricht und wird deshalb in den nachfolgenden Abschnitten als Betrachtungsbasis für die begrifflichen Variationen der Strukturierung in der Fachliteratur verwendet.

Auch aus der Sicht des Schulpädagogen Wiater (2012, S. 81) zergliedert sich die Bedeutung des Unterrichtsprinzips der Strukturierung in drei Teilaspekte. Die Bedeutungsunterteilungen von Wiater (2012) und Lipowsky (2009) stimmen allerdings nicht überein. Als erste Komponente der Strukturierung führt Wiater (2012, S. 81) die Durchgliederung komplexer und komplizierter Lerninhalte auf. Zum zweiten beschreibe die Strukturierung Unterrichtsmethoden zum systematischen Erarbeiten von sogenannten Gegenstandsstrukturen und zum dritten müsse bei der Auswahl der Inhalte und Methoden eine Entsprechung der „Denk-, Gefühls-, Könnens- und Wollensstruktur“ (Wiater, 2012, S. 81) der Schüler erreicht werden. Eine Passung der Struktur der Lerninhalte, der Struktur der Methode und der Struktur der Lernenden scheint für Wiater (2012) den Kern einer gelungenen Strukturierung als wichtiges Unterrichtsprinzip darzustellen (Wiater, 2012, S. 81). Die Struktur des Lernenden ist dabei unterteilt in Denk-, Gefühls-, Könnens- und Wollensstruktur. In seinem Buch zum Basiswissen der Schulpädagogik zu Unterrichtsprinzipien differenziert Wiater (2012) dabei nicht weiter aus, wie genau sich diese Teilstrukturen des Lernenden gestalten oder wie die geforderte Passung dieser zur Struktur von Inhalten und Methoden erreicht werden kann. Unklar verbleibt auch, welche Bedeutung Wiater (2012) genau mit der Struktur von Inhalten oder Methoden verbindet. Diese Ungenauigkeiten erschweren eine exakte Einordnung in die Systematisierung nach Lipowsky (2009). Grundsätzlich scheinen aber die ersten beiden Teilaspekte nach Wiater (2012) noch erkennbare Ähnlichkeiten zur didaktischen bzw. kognitionspsychologisch verstandenen Strukturierungen bei Lipowsky aufzuweisen. Aspekte zur Disziplinierung der Lernenden hingegen fehlen bei

der begrifflichen Festlegung der Strukturierung nach Wiater (2012).

Die Strukturierung ist auch wesentliches Element einer Zusammenfassung und Kategorisierung der Ergebnisse empirischer Unterrichtsforschung zu Qualitätsmerkmalen von Unterricht durch den Schulpädagogen Hilbert Meyer (2004). Dieser führt als eines von zehn Merkmalen guten Unterrichts die klare Strukturierung auf, wobei Unterricht gerade dann klar strukturiert sei, „wenn das Unterrichtsmanagement funktioniert und wenn sich ein für Lehrer und Schüler gleichermaßen gut erkennbarer ‚roter Faden‘ durch die Stunde zieht“ (Meyer, 2004, S.26). Meyer (2004, S.26) bezieht diesen roten Faden auf die „didaktisch-methodische Linienführung des Unterrichts“. Der rote Faden ist dabei geprägt durch die Stimmigkeit von Zielen, Inhalten und Methoden, welche daran zu erkennen sei, dass die Unterrichtsstunde „rund“ wirke. Zudem zählt Meyer zur didaktisch-methodischen Linienführung eine Orientierung am methodischen Grundrhythmus (Einstieg - Erarbeitung - Ergebnissicherung) und die methodische Folgerichtigkeit, mit der einzelne Unterrichtsschritte aufeinander aufbauen. Für diese Folgerichtigkeit des methodischen Gangs der gesamten Unterrichtsstunde existierten nach Meyer „keine gesetzmäßigen Vorschriften“ (Meyer, 2004, S. 27), sondern diese gestalte sich abhängig von der „Ziel- und Inhaltsstruktur sowie der methodischen Phantasie des Lehrers und der Schüler“ (Meyer, 2004, S. 27). Damit verbleibt die Festlegung dieses Begriffs bei Meyer etwas unscharf. Etwas schärfer wird dieser Begriff durch eine Abgrenzung zum Begriff der Plausibilität des thematischen Gangs. Dieser beschreibt die logisch gestaltete Abfolge der zu lernenden Sachinhalte und trägt zum Merkmal der inhaltlichen Klarheit bei, welches nach Meyer neben der klaren Strukturierung ein weiteres der zehn wichtigsten Qualitätsmerkmale für guten Unterricht darstellt. Der thematische Gang bezieht sich ebenso wie die methodische Folgerichtigkeit auf die gesamte Unterrichtsstunde und beschreibt bei Meyer unterschiedliche Arten der inhaltlichen Aufarbeitung, welche wohl als verschiedene Wege zum inhaltlichen Kern der Stunde interpretiert werden können. Die methodische Folgerichtigkeit hingegen scheint sich eher auf die methodische Umsetzung (z.B. Wahl der Methoden, Sozialformen und des Niveaus der Lehreraktivität) und der Gliederung des Ablaufs (z.B. nach dem Modell der Direkten Instruktion oder dem Dreischritt Einstieg-Erarbeitung-Ergebnissicherung) der Stunde zu beziehen. Zum Unterrichtsmanagement als zweitem Kernbereich der klaren Strukturierung liefern Regel-, Rollen- und Aufgabenklarheit, fest eingeführte Rituale zum Unterrichtsablauf sowie die Konsequenz der Lehrkraft wesentliche Beiträge (Meyer, 2004, S. 26ff). Deutlich wird, dass in der Definition der klaren Strukturierung nach Meyer die drei Bedeutungsfacetten des Begriffs der Strukturierung nach Lipowsky enthalten sind. Überlegungen zum Ablauf der Unterrichtsstunde, zum Umgang mit den Lernenden und zur Wahl einzelner Methoden sind angegeben, die schrittweise Aufbereitung des Lerninhalts in Teilschritten (Teil der didaktischen Strukturierung bei Lipowsky) ordnet Meyer allerdings dem Merkmal der inhaltlichen Klarheit zu. Unklar bleibt Meyer in

2. Strukturierung

der Angabe, wann genau Einzelaspekte der Strukturierung plausibel, logisch, stimmig oder folgerichtig sind bzw. wie man dies erreichen kann, oft fehlt eine solche Angabe auch. Eine Festlegung, welche Elemente der klaren Strukturierung oder der inhaltlichen Klarheit nun der Strukturierung oder der Klarheit zuzuordnen sind, wird ebenfalls nicht vorgenommen.

Helmke (2009, S. 191) räumt diesbezüglich ein, dass diese Begriffe in der Realität sehr oft eng zusammenhängen, fasst beide auch in einem Konstrukt zusammen, fordert aber explizit ein begriffliches Auseinanderhalten von Klarheit und Strukturiertheit. Die Begriffe der Strukturiertheit und der Strukturierung werden auch bei Helmke nicht differenziert und scheinen erneut synonym verwendet zu werden. Den Begriff der Klarheit gliedert Helmke (2009, S. 191) in seiner Übersicht von Bereichen, Merkmalen und Prinzipien zur Unterrichtsqualität grundsätzlich in vier Komponenten auf. Die Ausgestaltung dieser vier Komponenten präzisiert Helmke (2007, S. 45) in einem vorhergehenden Zeitungsartikel. Insgesamt lassen sich so die vier von Helmke differenzierten Komponenten der Klarheit wie folgt darstellen:

- Aspekte der Verstehbarkeit (akustische Komponente; Merkmale der Sprache und des Sprechens wie Artikulation, Stimmmodulation oder Lautstärke)
- Aspekte der Prägnanz (sprachliche Komponente; z.B. Vermeidung von Vagheitsausdrücken, inkorrektur Grammatik oder störenden Sprechverzögerungen und Füllwörtern)
- Aspekte der Kohärenz (inhaltliche Komponente; inhaltlicher Zusammenhang eines Textes oder einer Rede)
- Aspekte der Korrektheit (fachliche Komponente; Vermeidung fachlicher Fehler z.B. in einem Lehrervortrag)

Der Strukturierung weist Helmke (2009, S. 192) zwei Bedeutungen zu: „Aus *gedächtnispsychologischer* Sicht umfasst sie alle Merkmale des Informationsangebots, die darauf abzielen, den Aufbau einer gut organisierten Wissensstruktur zu unterstützen. Aus *didaktischer* Perspektive heißt Strukturierung, dass der Unterricht so geplant und sequenziert ist, dass er dieses Ziel erreicht“. Für die Sequenzierung könne man sich entweder an der Struktur des Lerninhaltes oder in erster Linie am Lerner orientieren. Dann träten deren Vorkenntnisse, Fähigkeiten und Interessen in den Vordergrund der Planungen. Vergleicht man die Begriffe der Klarheit oder der Strukturierung bei Helmke mit Lipowskys Kategorisierung der Strukturierung, zeigt sich eine deutliche Abgrenzung der Klarheit von Lipowskys Facetten der Strukturierung. Betrachtet man die Strukturierung nach Helmke, so kennzeichnet diese die Ausgestaltung der Planung und die Sequenzierung der Unterrichtsphasen, worin sich eine deutliche Passung zur Facette der didaktischen Strukturierung offenbart, konkrete Maßnahmen und Handlungen zum Aufbau einer geeigneten Wissensstruktur (Facette der ko-

gnitionspsychologisch verstandenen Strukturiertheit) sind darin eher implizit enthalten. Die Auffassung des Begriffs der Strukturierung bei Helmke (2009) bzw. Lipowsky (2009) scheint sich im Wesentlichen aber nur um den Aspekt der Strukturierung der Verhaltensebene zu unterscheiden, welcher bei Helmke nicht explizit zur Strukturierung gezählt wird. Zudem verweist Helmke (2009, S. 197) darauf, dass der Begriff der Strukturierung unterschiedlich konzeptualisiert wird. So herrsche in der Schulpädagogik ein recht breites Verständnis des Strukturierungsbegriffes (vgl. Meyer, 2004 oder Wiater, 2012) vor, wohingegen in der pädagogisch-psychologisch orientierten Unterrichtsforschung (vgl. Lipowsky, 2009) und der kognitionspsychologisch orientierten Lehr-Lern-Forschung ein etwas enger gefasstes Konzept mit mehreren Teilkategorien dominiere.

Rakoczy et al. (2007, S. 101) weisen aber darauf hin, dass auch in den Studien zur Lehr-Lern-Forschung nicht immer ausreichend zwischen Teilkategorien der Strukturierung unterschieden wird. Diese Teilkategorien könnten auf unterrichtliche Merkmale wie die Motivation oder den Lernzuwachs unterschiedliche Auswirkungen aufweisen. Diese zu erfassen ist Ziel einer empirischen Studie von Rakoczy et al. (2007) im Rahmen der Fachdidaktik Mathematik. Sie differenzierten hierzu hinsichtlich des Begriffs der Strukturierung Aspekte der *strukturierten Organisation der Lernumgebung* von Aspekten der *strukturierten Präsentation der Lerninhalte*. Zur *strukturierten Präsentation der Lerninhalte* tragen nach Rakoczy allgemein eine klar organisierte Präsentation der Sachinhalte und eine systematische didaktische Gestaltung bei (Rakoczy et al., 2007, S. 105). Der Begriff der systematischen didaktischen Gestaltung wird dabei nicht explizit näher präzisiert, sondern findet sich implizit in den Bewertungsrichtlinien von Rakoczy für die Qualität des Teilkonstrukts der *strukturierten Präsentation der Lerninhalte* wieder. Dieses Teilkonstrukt der Strukturierung bewerten Rakoczy et al. (2007, S. 108) nach dem Ausmaß, mit welchem die Präsentation der Unterrichtsstunde (a) Schritt für Schritt folgerichtig strukturiert wurde, (b) als Ganzes nachvollziehbar war, (c) in präzisen Ausdrücken formuliert wurde, (d) Wichtiges von Unwichtigem getrennt wurde und (e) die theoretischen Inhalte nachvollziehbar dokumentiert wurden. Wie ein großes Ausmaß in den einzelnen Kategorien zu erreichen ist, ist dabei nicht näher angegeben. Die Einzelaspekte der *strukturierten Präsentation der Lerninhalte* zeigen bei Rakoczy eine begriffliche Übereinstimmung mit dem Konstrukt der Klarheit bei Helmke. In (a), (b), (c), (e) und mit kleinen Einschränkungen auch (d) lassen sich Elemente der Klarheit wiederfinden. Teilweise zeigen sich auch Ähnlichkeiten zu den Facetten nach Lipowsky. (a) scheint sich nach meiner Interpretation der didaktischen und (d) der kognitionspsychologisch verstandenen Strukturierung zuordnen zu lassen. Rakoczy, Klieme, Lipowsky und Drollinger-Vetter (2010, S. 232) hingegen sortieren in einem späteren Artikel die Teilkategorie der *strukturierten Präsentation der Lerninhalte* (dort auch inhaltliche Strukturierung genannt) komplett in die Facette der kognitionspsychologisch verstandenen Strukturierung ein. In diesem Folgeartikel sollen die 2007 „vorgestellten Ana-

2. Strukturierung

lysen fortgesetzt und vertieft werden“ (Rakoczy et al., 2010, S. 231). Im Zuge dessen findet auch eine zusätzliche Bezugnahme auf andere Kategorisierungen des Begriffs der Strukturierung statt. Eine dieser Kategorisierungen ist dabei die Unterscheidung der drei Bedeutungsfacetten der Strukturierung nach Lipowsky. Die Bedeutungsfacetten der didaktischen Strukturierung und der Strukturierung auf der Verhaltensebene werden in der zweiten Studie von Rakoczy et al. (2010, S. 232) „unter Unterrichtsquantität und organisatorischer Strukturierung gefasst“ und entsprechen damit dem Konstrukt der *Organisation der Lernumgebung* ihres Artikels von 2007. Letztgenanntes Konstrukt umfasst die Dimensionen »Disziplinprobleme« und »classroom management«, weist große Ähnlichkeiten zu Meyers Unterrichtsmanagement auf und befasst sich demnach mit Störungen des Unterrichtsverlaufs und auch Präventivmaßnahmen um eben diese zu verhindern. Insgesamt ordnen damit Rakoczy et al. (2010) in ihrem nachfolgenden Artikel die ursprünglich *strukturierte Präsentation der Lerninhalte* genannte Teilkomponente der Strukturierung (2010: *inhaltliche Strukturierung*) der Facette der kognitionspsychologisch verstandenen Strukturierung zu, wohingegen sie ihr Konstrukt der *Organisation der Lernumgebung* mit den beiden verbleibenden Bedeutungsfacetten der Strukturierung nach Lipowsky gleichsetzt. Damit weicht Rakoczy (2010) in ihrer Bezugnahme auf die Systematisierung nach Lipowsky von meiner Einordnung ihrer Konstrukte vor allem durch ihre Zuordnung der didaktischen Strukturierung zur *Organisation der Lernumgebung* ab. Eine eindeutige Unterscheidung der Begriffe der Klarheit und der Strukturierung findet in den beiden angeführten Artikeln Rakoczys dabei nicht statt. Die grundsätzliche Zweiteilung des Begriffs der Strukturierung entsprechend ihrem Artikel aus dem Jahr 2007 wird durch die Untersuchungsergebnisse aus der ersten Studie von 2007 gestützt, da Rakoczy et al. (2007, S. 115f) unterschiedliche Effekte beider Teilbereiche unter anderem auf die Motivation nachweisen konnten.

In einer weiteren Studie der Fachdidaktik der Mathematik analysiert Drollinger-Vetter (2011) die fachdidaktische Qualität der Anleitung von mathematischen Verstehensprozessen im Unterricht und definiert hierzu das Konstrukt der *strukturellen Klarheit*. Die „Strukturelle Klarheit bezieht sich auf die Qualität des Strukturaufbaus im zeitlichen Verlauf des Unterrichts“ (Drollinger-Vetter, 2011, S. 213) und umfasst folgende Teilaspekte:

1. Vorkommen und Deutlichkeit der Verstehenselemente und Repräsentationen im Verlauf
2. Verständliche und sachlogische Reihenfolge und Widerspruchsfreiheit der vorkommenden konzeptspezifischen Verknüpfungen (Beziehungen zu weiteren mathematischen Konzepten)
3. Vorkommen eines inhaltlichen roten Fadens
4. Kohärenz der Verstehenselemente über längere Zeiträume hinweg, besonders bei Repräsentationswechsel

5. Vernetzung der Verstehenselemente

6. Fachliche Korrektheit

Diese Teilaspekte enthalten dabei Begrifflichkeiten, welche nachfolgend näher ausgeführt werden. Verstehenselemente kennzeichnet Drollinger-Vetter (2011, S. 183) dabei als „diejenigen Teilkonzepte eines zu verstehenden Konzepts, die man verstanden haben muss, um [...] das Konzept als Ganzes verstehen zu können“, wobei sie mit Konzepten in ihrer Studie wiederum klar eingegrenzte mathematische Sachverhalte wie den Satz des Pythagoras oder den Cosinussatz bezeichnet. Etwas unklar bleibt, wie eine verständliche Reihenfolge der konzeptspezifischen Verknüpfungen geartet ist. Der inhaltliche rote Faden hingegen wird durch Drollinger-Vetter näher spezifiziert und betrifft die Verstehenselemente und die Repräsentationen des jeweiligen Konzepts und soll dann gewährleistet sein, wenn in der Unterrichtsstunde das Konzept als solches früh erkennbar wird und die einzelnen Verstehenselemente nicht zeitlich zu kurz im Unterricht thematisiert werden oder zu dicht aufeinander folgen. Der rote Faden bei Meyer (2004) bezieht sich hingegen auf Aspekte der *methodischen* Gestaltung der Stunde im Gegensatz zum *inhaltlichen* roten Faden bei Drollinger-Vetter, dieser basiert vielmehr auf Meyers Begriff der Plausibilität des thematischen Gangs, wird aber durch obige Beschreibung begrifflich ausgeschärft. Der Begriff der „Kohärenz hat [...] mit dem Vorkommen der Verstehenselemente und mit Verknüpfungen zwischen den Repräsentationen zu tun, welche via die zentralen Verstehenselemente verlaufen. [...] Je nachdem, welche Verstehenselemente und Repräsentationen in der ersten Phase vorgekommen sind, braucht es zusätzliche Verstehenselemente und Verknüpfungen, um die zweite Phase zu verstehen“ (Drollinger-Vetter, 2011, S. 208).

Insgesamt weist die von Drollinger-Vetter definierte *strukturelle Klarheit* weitgehende Überschneidungen zum Merkmal der *Klarheit und Strukturiertheit* bei Helmke auf, verzichtet jedoch auf akustische und sprachliche Aspekte der Klarheit und weist nicht nur bei der Phasierung des Unterrichtsverlaufs einen starken inhaltlichen Fokus auf. Als Bedeutungsfacetten der Strukturierung im Sinne Lipowskys lassen sich Elemente der didaktischen Strukturierung (Teilaspekte 2, 3) und der kognitionspsychologisch verstandenen Strukturierung (Teilaspekt 5) in Ansätzen wiederfinden, nicht jedoch Elemente der Struktur der Verhaltensebene.

Auch im Bereich der Didaktik der Naturwissenschaften findet sich eine Konzeptualisierung mit einer ebenfalls stark inhaltlich geprägten begrifflichen Auffassung der Strukturierung. In einer Studie zur „Entwicklung eines standardisierten, reliablen und validen Instruments zur Erfassung der professionellen Wahrnehmung der inhaltlichen Strukturierung im naturwissenschaftlichen Grundschulunterricht“ (Meschede, 2014, S. v) weist das Konstrukt der *inhaltlichen Strukturierung* einen starken inhaltlichen Fokus und eine Abgrenzung zur Facette der Strukturierung auf der Verhaltensebene auf. Die *inhaltliche*

2. Strukturierung

Strukturierung umfasst dabei die Strukturierung des Lerngegenstandes (Elementarisierung des Sachinhaltes oder Gliederung des Gesamtthemas) und die Strukturierung des Interaktionsgeschehens (z.B. Maßnahmen zum Betonen wichtiger Ergebnisse). Meschede (2014, S. 68) gibt fünf Indikatoren für ihre Konzeptualisierung der Strukturierung an, welche dabei nicht als trennscharfe Teilkategorien verstanden werden sollen:

1. Eine Zielklarheit schaffen
2. Das Gespräch durch Maßnahmen des Zusammenfassens strukturieren
3. Das Gespräch durch Maßnahmen des Hervorhebens strukturieren
4. Eine inhaltliche Klarheit der Lehrer- und Schüleräußerungen sicherstellen
5. Mündliche Gesprächsbeiträge durch geeignete Veranschaulichungen unterstützen.

Der Indikator der Zielklarheit ist dann deutlich ausgeprägt, wenn ein gemeinsames Ziel expliziert wird, welches als Orientierungsrahmen während des Unterrichts fungiert und im Verlauf transparent gehalten wird. „Im naturwissenschaftlichen (Grundschul-) Unterricht werden Ziele häufig durch (zu bearbeitende) Fragestellungen („Forscherfragen“) formuliert, wie z.B. „Wie kommt es, dass ein Schiff schwimmt?““ (Meschede, 2014, S. 70). Die Indikatoren zwei und drei zeigen sich durch das Vorkommen entsprechender Maßnahmen in der beobachteten Unterrichtssequenz. Die inhaltliche Klarheit der Lehrer oder Schüleräußerungen „zeichnet sich allgemein durch eine sprachlich verständliche, prägnante und fachlich korrekte Darstellung und Entwicklung des Unterrichtsgegenstandes aus“ (Meschede, 2014, S. 75) und wird anhand von Einzelbeispielen aus Unterrichtsbeobachtungen präzisiert. So bezieht sich die inhaltliche Klarheit auf Aspekte zur Einführung der Fachsprache, einer korrekten Verwendung der Fachbegriffe, einer Einforderung einer präzisen und korrekten Ausdrucksweise der Lernenden und einer unmissverständlichen Aufgabenstellung. Die Unterstützung durch Veranschaulichungen erfolgt dabei im naturwissenschaftlichen Grundschulunterricht meist durch enaktive (Erwerb von Wissen im konkret-handelnden Umgang) oder ikonische Repräsentationen. Diese sollen insbesondere das inhaltliche Nachvollziehen erleichtern und am Vorwissen der Lernenden anknüpfen. In dieser Konzeptualisierung der inhaltlichen Strukturierung findet damit weder die Sequenzierung des Lerngegenstandes im Sinne einer didaktischen Strukturierung noch eine Strukturierung auf der Verhaltensebene Berücksichtigung, Elemente der kognitionspsychologisch verstandenen Strukturierung hingegen prägen dieses Konstrukt. Die Indikatoren eins, zwei, drei und fünf können als Maßnahmen zum Aufbau einer komplexen und geordneten Wissensstruktur unter letztgenannter Facette der Strukturierung subsumiert werden. Legt man die begriffliche Differenzierung von Strukturierung und Klarheit wie beispielsweise bei Helmke (vgl. S. 12) zugrunde, ergänzt der Indikator der inhaltlichen Klarheit der Lehrer- und Schü-

leräußerungen die inhaltliche Strukturierung noch um einen Teilaspekt der Klarheit.

Im Rahmen einer durch das IPN (Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik) in Kiel geleiteten Videostudie zu Lehr-Lern-Prozessen im Physikunterricht der Sekundarstufe in Deutschland war ein Zielbereich die Erfassung des Merkmals der *Zielorientierung* und dessen unterrichtliche Auswirkungen. Trepke, Seidel und Dalehefte (2003, S. 201ff) erfassten die *Zielorientierung* in einer ersten Projektphase über die Teilkategorien der Orientierung am Ziel, der Klärung der Anforderungen, der Organisation des Lernstoffs und der Schlüssigkeit im Unterrichtsverlauf. In einer zweiten Projektphase ist dieses Merkmal „um weitere aus theoretischer Sicht relevante Aspekte zielorientierten Unterrichts erweitert worden. Neben der Klarheit und Schlüssigkeit im Unterricht sind hier nun instruktionale Aspekte wie strukturierende Äußerungen im Unterrichtsverlauf und die Einbettung der Inhalte und der Unterrichtsstunde in längerfristige Ziele hinzugekommen“ (Herweg, 2008, S. 59f). Hieraus wird auch deutlich, dass das Merkmal der *Zielorientierung* wesentliche Überschneidungen zu den Konstrukten der Klarheit und der Strukturierung aufweist, teilweise werden diese Begriffe fast synonym zur Zielorientierung verwendet. Nachfolgend werden Teilkomponenten der *Zielorientierung* ausführlicher dargestellt und anschließend den Facetten der Strukturierung sowie Aspekten der Klarheit zugeordnet.

Herweg (2008, S. 17) differenziert insgesamt vier Teilkomponenten eines zielorientierten Unterrichts. Das Unterrichtsziel (1) soll zu Beginn des Unterrichts explizit durch die Lehrkraft thematisiert werden. Dies kann durch eine bloße Nennung des Themas, einer Frage- oder Problemstellung oder einer ausführlichen Beschreibung des Erkenntnisprozesses erfolgen. Um während dieses Prozesses das Unterrichtsziel präsent zu halten, soll durch Maßnahmen der expliziten Zielorientierung (2) wie Zielerückgriffen (regelmäßiger Rückbezug auf ein vorab benanntes Unterrichtsziel), regelmäßigen und abschließenden Zusammenfassungen sowie klaren Anforderungen und Zielen bei Experimenten und Aufgaben gewährleistet werden. Die dritte Teilkomponente heißt implizite Kohärenz des Unterrichtsziels (3) und bezeichnet Aspekte zur Sequenzierung der Lerninhalte, insbesondere der fachlich zusammenhängenden Inhaltserarbeitung. Weiterhin tragen eine schlüssige Einbettung von Experimenten (Beobachtungen aus Experimenten können mit Lerninhalten verknüpft werden), eine angemessene Informationsfülle und eine Berücksichtigung des Vorwissens (u.a. passendes Vokabular der Lehrkraft und inhaltliche Orientierung am Vorwissen) zu dieser Teilkomponente bei. Als vierte abgrenzbare Komponente führt Herweg (2008, S. 30) Anker und Integrationshilfen (4) für „einen bedeutungshaltigen und organisierten Aufbau neuer Wissensstrukturen“ auf. Dies kann durch die Integration der neuen Inhalte in bestehende Vorwissensstrukturen, durch die Schaffung von Anwendungsmöglichkeiten, durch Maßnahmen wie »Advance Organizer« zur Transparenz des roten Fadens oder durch die

2. Strukturierung

Einordnung der Inhalte in übergeordnete Kontexte erreicht werden (Herweg, 2008, S. 17-35). Mit Ausnahme der impliziten Kohärenz des Unterrichtsziels (3) lassen sich die Komponenten der Zielorientierung wiederum als Maßnahmen zur Erleichterung des Aufbaus einer geordneten Wissensstruktur interpretieren und so der Facette der kognitionspsychologisch verstandenen Strukturierung nach Lipowsky zuordnen. Bei (3) finden sich vor allem inhaltliche Aspekte der Klarheit (u.a. die kohärente Inhaltserarbeitung), aber auch Elemente der didaktischen Strukturierung (z.B. Sequenzierung der Inhalte). Die Facette der Strukturierung auf der Verhaltensebene ist nicht in dem Merkmal der *Zielorientierung* enthalten. Somit finden sich in der *Zielorientierung* nur zwei der drei Facetten der Kategorisierung nach Lipowsky.

Es existiert aber auch eine grundsätzlich andere Auffassung des Begriffs der Strukturierung, so kann dieser auch in *Sicht-* und *Tiefenstruktur* zweigeteilt werden. Diese Zweiteilung basiert auf der fachunspezifischen Lehr-Lern-Theorie der Basismodelle nach Oser und Baeriswyl (2001), welche die Strukturierung von Unterricht thematisiert. Für die Fachdidaktik Physik wurde diese Theorie und damit die zweiteilige Kategorisierung des Begriffs der Strukturierung u.a. durch Reyer (2004), Trendel (2007) und Wackermann (2008) adaptiert. Nach Wackermann geht es in der Tiefenstruktur von Unterricht „um das Denken und Lernen von Schülern und Lehrern und um ihre Intentionen und sozialen Beziehungen“ (Wackermann, 2008, S. 8). Elemente der Tiefenstruktur wie die Denk- und Lernschritte der Schüler sind dabei schwer direkt zu beobachten und müssen meist aus beobachtbaren Aspekten durch Interpretation konstruiert werden, bilden aber die Grundlage der unterrichtlichen Gestaltung der Lehrkräfte. Alle direkt beobachtbaren Elemente dieses Unterrichtsgeschehens bilden die Sichtstruktur. Diese umfasst beispielsweise die tatsächlich beobachteten Sozialformen, die Redeanteile oder das Lehrer- und Schülerverhalten (Reyer, 2004; Wackermann, 2008). Eine Trennung in Sicht- und Tiefenstrukturen findet sich auch im Modell der Unterrichtsqualität der COACTIV-Studie wieder. Die COACTIV-Studie verwendet als Basis Untersuchungen unterschiedlicher Aspekte der Berufstätigkeit von Lehrkräften. Diese Untersuchungen erfolgten dabei aus verschiedenen Blickwinkeln der empirischen pädagogischen Forschung. „Ziel von COACTIV war es, diese unterschiedlichen Ansätze zu ordnen, in einem übergreifenden Modell, das Erkenntnisse aus verschiedenen Forschungsrichtungen zusammenführt, zu integrieren und empirisch zu prüfen“ (Baumert & Kunter, 2011, S. 29). Ebenso stellt „eine Analyse der Unterrichtstätigkeit der Lehrkräfte und eine Beschreibung entsprechender Qualitätskriterien [...] auch einen Kernbereich von COACTIV dar“ (Kunter & Voss, 2011, S. 85). Im Zuge der Analyse der Qualität des beobachteten Mathematikunterrichts erfolgte eine Zuordnung von Unterrichtsmerkmalen zur *Sicht-* oder zur *Tiefenstruktur*. Grundsätzlich basiert auch bei Kunter und Voss (2011) diese Zweiteilung der Strukturierung auf den begrifflichen Festlegungen von Oser und Baeriswyl (2001). Kunter und Voss bzw. Reyer differieren aber in

der Kategorisierung von Elementen der Sicht- oder Tiefenstruktur. Zählt Reyer (2004, S. 126) die Klassenorganisation und Disziplin zur Sichtstruktur, ordnen Kunter und Voss (2011, S. 98) die Klassenführung der Tiefenstruktur zu. Eine Eindeutigkeit in der Auffassung und Interpretation dieser Zweiteilung der Strukturierung kann also nicht festgestellt werden.

Generell lässt sich keine eindeutige Konzeptualisierung des Begriffs der Strukturierung durch die hier aufgeführten Autoren erkennen. In der Schulpädagogik (vgl. Wiater, Meyer oder Helmke) zeigt sich eher eine allgemeine Definition des Begriffs, welche nur teilweise Aspekte der Strukturierung auf der Verhaltensebene enthält, teilweise noch versucht die Struktur des Lernenden zu berücksichtigen, ohne diese dabei zu konkretisieren. Konkretere Angaben finden sich in den Arbeiten der Fachdidaktiken der Mathematik und auch der Naturwissenschaften. Mit den dort definierten Konstrukten (inhaltliche Strukturierung, strukturelle Klarheit oder Zielorientierung) legen beide Fachbereiche den Fokus auf inhaltliche Komponenten der Strukturierung und Maßnahmen zum Aufbau einer komplexen und geordneten Wissensstruktur und damit hauptsächlich auf Elemente der Sichtstruktur. Beide verzichten auf Elemente der Strukturierung auf der Verhaltensebene, unterscheiden sich aber auf der Ebene der Messung der Ausprägung der jeweils definierten Konstrukte und weisen dabei Elemente der Begriffe der Strukturierung und der Klarheit auf. Eine differenzierte Betrachtung dieser Begrifflichkeiten ist aus Sicht der fachdidaktischen Lehr-Lern-Forschung notwendig, um letztlich auf Basis empirischer Studien präzise Aussagen über die tatsächlichen Stellschrauben wirksamen Unterrichts zu ermöglichen. In einem ersten Schritt konnten Rakoczy et al. (2007, S. 115) diese Notwendigkeit der Differenzierung auch empirisch nachweisen, indem sie für die von ihnen unterschiedenen Teilkonstrukte der Strukturierung echt unterschiedliche unterrichtliche Auswirkungen nachweisen konnten. Untersucht man demnach die Bedeutung der Strukturierung auf unterrichtliche Aspekte wie den Lernerfolg, treten Zusammenhänge besonders dann klar hervor, wenn nur einzelne Komponenten der Strukturierung untersucht werden und klar angegeben ist, welche Komponenten davon im Kontext der jeweiligen Untersuchung von zentraler Bedeutung sind.

Der im Kontext dieser Arbeit geplante Vergleich verschiedener theoretischer Vorgaben zur Strukturierung der Lehr-Lern-Prozesse legt die Theorie der Basismodelle nach Oser und Baeriswyl (2001), bzw. deren Adaption für den Physikunterricht durch Reyer, Trendel und Wackermann, sowie das Forschend-entwickelnde Unterrichtsverfahren nach Schmidkunz und Lindemann (1992) zu Grunde. Beide Vorgaben zur Sequenzierung von Lehr-Lern-Prozessen formulieren diese Prozesse auf Grundlage der jeweils angenommenen Lernwege der Lernenden, sind inhaltlich universell anwendbar und methodisch flexibel gestaltbar. Dies erlaubt eine zu den theoretischen Vorgaben konforme Konzeption von Unterrichtsstunden, welche sich bezüglich Inhalt und Methoden nicht wesentlich unterscheidet. Teilaspekte der Strukturierung können so kontrolliert

2. Strukturierung

oder konstant gehalten werden. Dies erlaubt eine weitergehende Spezifizierung des Strukturbegriffs im Rahmen dieser fachdidaktischen Studie und ermöglicht letztlich einen tiefer gehenden Blick auf die Wirkungen der Vorgaben zur Strukturierung der verwendeten Modelle. Der Fokus dieser Arbeit liegt damit nicht auf einzelnen Maßnahmen, Handlungen oder Inhalten (Ebene der Sichtstruktur). Der Fokus liegt auf der Unterschiedlichkeit der Vorgaben beider Modelle zur Verknüpfung einzelner Unterrichtsabschnitte auf Basis der Lernprozesse, welche die Lernenden durchlaufen müssen (Ebene der Tiefenstruktur). Die Grundlage für das Verständnis der Strukturierung einer Unterrichtsstunde in dieser Arbeit ist damit Teil der Facette der didaktischen Strukturierung bei Lipowsky und entspricht weitestgehend Helmkes Beschreibung der Strukturierung aus Sicht der Lehr-Lern-Forschung (Helmke, 2009, S. 198), welche deshalb als Arbeitsdefinition ausgewählt wird.

Festlegung: *Die **Strukturierung** bezieht sich auf die Schlüssigkeit, mit der im Unterricht funktional unterschiedliche Phasen aufeinander folgen und miteinander verknüpft sind.*

2.2. Strukturierung als Qualitätsmerkmal von Unterricht

Dieser Abschnitt soll die Bedeutung der Strukturierung insbesondere auf den Lernerfolg beleuchten. Hierzu werden zunächst die Aussagen einzelner Arbeiten über Strukturierung als allgemeines Qualitätsmerkmal dargestellt und im Anschluss weitere Aussagen zu Wirkungen von Teilaspekten der Strukturierung entsprechend den Bedeutungsfacetten nach Lipowsky sowie der Zweiteilung in Sicht- und Tiefenstruktur kategorisiert. Abschließend wird der Einfluss der verschiedenen Bedeutungsfacetten der Strukturierung auf den Lernerfolg und affektiv-motivationale Merkmale zusammengefasst.

Die begriffliche Festlegung der Strukturierung im Rahmen dieser Arbeit umfasst dabei nur Teilaspekte der Strukturierung, wenn diese als allgemeines Qualitätsmerkmal von Unterricht betrachtet wird. Aussagen in der Literatur zu Wirkungen der (allgemeinen) Strukturierung können demzufolge nicht direkt auf Wirkungen der für diese Arbeit relevanten Teilaspekte übertragen werden, dienen hier aber als Orientierungsbasis für die nachfolgende spezifischere Aufarbeitung der Wirkungen von Teilaspekten der Strukturierung.

Strukturierung als allgemeines Qualitätsmerkmal

Die Strukturierung als allgemeines Qualitätsmerkmal ist Teil der Prinzipien effektiven Unterrichtens, welche Brophy (2000) in einem Booklet als Synthese aus der Analyse von Forschungsergebnissen zum Unterrichten formuliert. Brophy (2000, S. 6) führt dabei vorrangig den Lernzuwachs oder Leistungsstand, aber auch das Erreichen eines guten Klassenklimas und positiver Einstellungen der Schüler gegenüber der Schule, den Lehrern und den Klassenkameraden als maß-

gebend für die Qualität von Unterricht auf. Auf Basis dieses Verständnisses von Unterrichtsqualität formuliert Brophy zwölf Prinzipien effektiven Unterrichts. „Much of the research support for these principles comes from studies of relationships between classroom processes (measured through observation systems) and student outcomes (most notably, gains in standardized achievement tests). However, some principles are rooted in the logic of instructional design (e.g. the need for alignment among a curriculum’s goals, content, instructional methods and assessment measures)” (Brophy, 2000, S. 6). Zwei der von Brophy so herausgearbeiteten Prinzipien stehen in Verbindung zur Strukturierung. Mit *establishing learning orientations* beschreibt er Maßnahmen zur Vorstrukturierung der Lernprozesse im Unterricht, z.B einer vorausgreifenden Zusammenfassung. Diese Maßnahmen sollen ein zielorientiertes Vorgehen und eine Verknüpfung zum Vorwissen der Schüler/innen ermöglichen und auf diese Weise das nachfolgende Lernen vorbereiten, aber auch die Motivation der Lernenden steigern (Brophy, 2000, S. 15f). Das Prinzip *establishing learning orientations* kann dabei der Bedeutungsfacette der kognitionspsychologisch verstandenen Strukturierung nach Lipowsky (vgl. S. 9) zugeordnet werden. Elemente der Facette der didaktischen Strukturierung finden sich in einem weiteren von Brophy genanntem Prinzip effektiven Unterrichts. „To facilitate meaningful learning and retention, content is explained clearly and developed with emphasis on its structure and connections” (Brophy, 2000, S. 17). Dieses Prinzip nennt Brophy *coherent content*. Es umfasst allerdings neben Elementen der didaktischen Strukturierung (Sequenzierung von Unterricht; inhaltliche Aufbereitung in Teilschritte) auch Aspekte (Gestik, Kommunikationsfähigkeit der Lehrkraft) welche der Klarheit oder der Verständlichkeit (Definition nach Helmke vgl. S. 12) zuzuordnen sind (Brophy, 2000, S. 17f). Verschiedene Komponenten der (allgemeinen) Strukturierung sind damit zwei Prinzipien effektiven Unterrichts zuzuordnen. Beiden Prinzipien attestiert Brophy (2000, S. 15ff) eine positive Wirkung vorrangig auf den Lernprozess.

Die von Brophy formulierten zwölf Kriterien dienen auch Meyer (2004) als Orientierung für dessen *Kriterienmix*. Damit bezeichnet Meyer seine Zusammenstellung von zehn Merkmalen guten Unterrichts. Grundlage für deren Ausarbeitung stellt die Auswertung einzelner Studien und Metastudien, sowie „eine eigene normative Orientierung“ (Meyer, 2004, S. 17) aus der Sicht eines Didaktikers dar. Das im Kriterienmix enthaltene Merkmal der *klaren Strukturierung* (Definition siehe S. 11) soll einen hohen Anteil an echter Lernzeit, eine hohe Schüleraufmerksamkeit, eine Reduzierung von Störungen, die Transparenz der Leistungserwartungen und damit einen erhöhten Lernerfolg ermöglichen. Auf den Lernerfolg hat nach Meyer aus Sicht der Empiriker kein anderes Merkmal einen stärkeren Einfluss (Meyer, 2004, S. 35). Als Grundlage für diese Einschätzung dienen Meyer die Ergebnisse verschiedener Metaanalysen z.B. derer von Scheerens (1992) oder Wang, Haertel und Walberg (1993).

Helmke (2009) thematisiert in seinem Buch *Unterrichtsqualität - erfassen, bewerten,*

2. Strukturierung

verbessern insbesondere fächerübergreifende Merkmale der Unterrichtsqualität. Dabei hebt er insgesamt zehn Qualitätsmerkmale von Unterricht als besonders relevant hervor (Helmke, 2009, S. 169). Mehrere Einzelstudien (Wellenreuther, 2005; Schnotz, 2006), mehrere Metaanalysen und auch andere zusammenfassende Arbeiten im deutschen oder angloamerikanischen Sprachraum, darunter auch die von Brophy (2000) und Meyer (2004) bilden dabei die Grundlage seiner Einschätzung zur Wirksamkeit einzelner Unterrichtsmerkmale. Zwar unterscheidet Helmke (2009, S.191f) begrifflich scharf zwischen Klarheit und Strukturiertheit, fasst diese allerdings als eines der zehn Qualitätsmerkmale von Unterricht zusammen. Als Teil des Merkmals der *Klarheit und Strukturiertheit* (Definition siehe S. 12) führt damit auch Helmke (2009, S. 197f) die Strukturierung als allgemeines Qualitätsmerkmal von Unterricht auf, welches aber speziell auf die Förderung der Informationsverarbeitung und damit den Lernerfolg abziele. „Als lern- und leistungsförderlich haben sich eine gute Wissensstrukturierung und zusätzliche Lernhilfen herausgestellt“ (Helmke, 2009, S. 197). Damit bleibt Helmke in den Aussagen zur Wirkung eher vage, bescheinigt insgesamt eine lernförderliche Wirkung, präzisiert diese nicht näher hinsichtlich verschiedener kognitiver oder affektiver Komponenten wie der Motivation, aber hinsichtlich einer Fach- und Personenspezifität. So profitieren Schüler mit geringeren Vorkenntnissen in besonderem Maße von klar strukturiertem Unterricht (Helmke, 2009, S. 200). Dieser „ist vor allem in Fächer mit hierarchisch strukturierter Wissensbasis (prototypisch: Mathematik) ein Schlüsselmerkmal für erfolgreichen Unterricht“ (Helmke, 2009, S. 200).

Dass eine gelungene Strukturierung den Unterricht positiv beeinflusst, scheint unstrittig. Besondere Einigkeit besteht bei der Wirkung auf den Lernerfolg der Schüler. Wie aber eine gute Strukturierung von Unterricht gelingen kann, ist bezüglich der Abfolge und der Verbindungen einzelner unterrichtlicher Phasen, sowie einer geeigneten Zerlegung der Lerninhalte in den fachunspezifischen Beiträgen zur Strukturierung als Qualitätsmerkmal kaum thematisiert. Für Meyer (2004, S. 27) gibt es dabei auch keine eindeutigen Regeln für eine erfolgreiche Sequenzierung, diese hänge zu stark von der jeweiligen unterrichtlichen Ausgangssituation ab. In den meisten Fällen könne aber mit dem Dreischritt (Einstieg, Erarbeitung, Ergebnissicherung) als Grundrhythmus des Unterrichts eine grundlegende Orientierung geschaffen werden. Mehr als der Status einer Orientierung kann diesem Dreischritt dabei nicht zugeschrieben werden. Weder sind detaillierte Vorgaben über die Ausgestaltung der angegebenen Phasen oder deren Verbindung untereinander angegeben, noch liegt ein Anspruch auf deren Vollständigkeit vor. Weitere unterrichtliche Phasen könnten notwendig sein, um eine transparente Sequenzierung des Unterrichts und einen schrittweisen Aufbau der Unterrichtsinhalte zu erreichen. Ungenau bleiben die hier angeführten Zusammenfassungen wichtiger Qualitätskriterien von Unterricht auch in den Auswirkungen auf verschiedene Teilbereiche des Unterrichts. Führt Brophy beispielsweise einen Zusammenhang der Strukturierung mit der Motivation

der Lernenden an, findet sich bei Helmke oder Meyer diesbezüglich keine vergleichbare Aussage wieder. Dies kann auch in der jeweils unterschiedlichen Zusammensetzung der betreffenden Qualitätsmerkmale aus verschiedenen Teilaspekten der Strukturierung begründet sein. Rakoczy et al. (2007, S. 115f) konnten nachweisen, dass verschiedene Teilaspekte der Strukturierung auch unterschiedliche Wirkungen auf verschiedene kognitive und affektive Merkmale der Lernenden aufweisen, weshalb diese Wirkungen im Folgenden für die einzelnen Bedeutungsfacetten der Strukturierung jeweils getrennt angegeben werden.

Wirkung von Elementen der didaktischen Strukturierung

Lipowsky (2009, S. 83) führt »Direct Instruction« und das »mastery learning« als Elemente der didaktischen Strukturierung an. Die Lernwirksamkeit dieser beiden Elemente kennzeichnet Hattie (2009, S. 297) jeweils mit einer Effektstärke von circa $d = 0.6$ (Cohens d). Mit Hilfe dieses Maßes für die Stärke des Effekts erfasst und ordnet Hattie in seiner auf über 800 einzelnen Metastudien basierenden fachunspezifischen (Meta-)Metastudie die Lernwirksamkeit einer Vielzahl unterrichtlicher Bedingungsfaktoren. Der Einfluss dieser Bedingungsfaktoren kann dabei grundsätzlich als bedeutsam angesehen werden, wenn das Effektstärkemaß d einen Wert von $d = 0.2$ überschreitet. Zusätzlich kann zwischen einem kleinen ($d > 0.2$), mittleren ($d > 0.5$) oder großem Effekt ($d > 0.8$) differenziert werden (Rasch, Friese, Hofmann & Naumann, 2006a, S. 68). Hattie (2009) gibt allerdings einen kritischen Wert von $d = 0.4$ an, ab welchem im Unterricht wesentliche Veränderungen eintreten. Die Effektstärken für die »Direct Instruction« und das »mastery learning« liegen über dem von Hattie angegebenen kritischen Wert von $d = 0.4$, und sind somit auch unterrichtlich bedeutsam. Allerdings finden sich unter den von Hattie untersuchten Merkmalen auch weitere Elemente der didaktischen Strukturierung, welche diesen kritischen Wert für deren Wirksamkeit unterschreiten (z.B. »problem-based learning« mit $d = 0.15$ oder »inquiry-based teaching« mit $d = 0.31$) (Hattie, 2009, S. 201). Damit wird der didaktischen Strukturierung ein allgemein positiver, in seiner Ausprägung dabei deutlich unterschiedlicher und damit nach Hattie nicht immer bedeutsamer Effekt auf den Lernerfolg zugeschrieben. Aussagen zu Auswirkungen auch auf affektive Merkmale der Lernenden finden sich bei Hattie nicht, werden aber u.a. in Metastudien zum »inquiry-based teaching« von Bybee et al. (2006) oder zur »Direct Instruction« von Adams und Engelmann (1996) angeführt. Adams und Engelmann (1996, S. 97) schreiben beispielsweise der »Direct Instruction« einen positiven Einfluss auf das Selbstvertrauen der Lernenden zu. Dies bestätigt Carnine (2000, S. 7) in einer Art Plädoyer für die Verwendung der »Direct Instruction« und deutet auch eine positive Auswirkung auf weitere affektive Merkmale an, präzisiert diese Aussage aber nicht weiter. Bybee et al. (2006, S. 26) heben zudem hervor, dass eine spezielle Form des »inquiry-based teachings« das Interesse am und die allgemeinen Einstellungen zum Fach verbessert. Positive Auswirkungen des »problem-based learnings« erwartet Hmelo-Silver (2004)

2. Strukturierung

in einem Überblicksartikel zur grundlegenden Natur und empirisch bestätigten Wirkungen des Problem-basierten Lernens auch auf die Motivation. Die von ihr ausgewerteten Studien liefern für den vermuteten Wirkzusammenhang zur Motivation allerdings nur wenig belastbare Informationen. Nur als Ergebnis einer Studie konnte ein Zusammenhang des Problem-basierten Lernens zur Motivation festgestellt werden, wobei sich dieser Zusammenhang wiederum von anderen Variablen abhängig zeigt und damit eben wenig belastbar erscheint. Allgemein stellt Hmelo-Silver (2004, S. 259f) fest, dass zahlreiche Studien zu finden sind, welche die Lernwirksamkeit im Fokus haben, allerdings wenige zur Motivation. Eine Wirkung auf affektive Merkmale der Lernenden kann damit nur für einzelne Elemente der didaktischen Strukturierung konstatiert werden, wohingegen die Wirkung dieser Elemente auf die Lernwirksamkeit als einheitlich positiv beschrieben wird. Allerdings erweist sich diese generell positive Wirkung zwischen einzelnen Elementen der didaktischen Strukturierung als deutlich unterschiedlich ausgeprägt.

Wirkung von Elementen der Strukturierung auf der Verhaltensebene

„Eine Strukturierung auf der Verhaltensebene begünstigt eine störungsfreie Lernumgebung, fördert die Aufmerksamkeit der Lernenden und sorgt dafür, dass mehr Unterrichtszeit für die Auseinandersetzung mit den Unterrichtsthemen zur Verfügung steht“ (Lipowsky, 2009, S. 83). Die Variable »time for learning« in der Metastudie von Seidel und Shavelson (2007) kann dieser Facette zugeordnet werden und zeigt kleine positive Effekte auf den Lernerfolg und motivational-affektive Merkmale („e.g., development of stable interests, motivational orientations, attitudes or belief systems“; Seidel & Shavelson, 2007, S. 470f). Auch in der Metastudie von Hattie (2009) können dort untersuchte Merkmale der Strukturierung der Verhaltensebene zugeordnet werden. Dies ist zum einen das Merkmal »decreasing disruptive behavior«, welches eine allgemeine Reduzierung der Anzahl und Dauer der unterrichtlichen Störungen durch die Lernenden umfasst und einen Einfluss von $d = 0.34$ auf den Lernerfolg aufweist. Zum anderen ist dies das »classroom management« welches in Hatties Studie eher im Sinne eines Sammelmerkmals vieler Elemente der Strukturierung auf der Verhaltensebene kategorisiert wird, damit eine große Ähnlichkeit zur hier betrachteten Facette der Strukturierung aufweist und welcher ein mittlerer Effekt ($d = 0.52$) auf den Lernzuwachs zugeschrieben wird (Hattie, 2009, S. 102ff). Eine positive Korrelation mit dem Lernzuwachs können Rakoczy et al. (2007) in ihrer Studie zum Einfluss der Strukturierung als Qualitätsmerkmal im Mathematikunterricht auch für das Konstrukt der strukturierten Organisation der Lernumgebung nachweisen. Dieses Konstrukt umfasst dabei die Dimension der »disciplinary problems« und damit den Umgang mit allen den Ablauf des Unterrichts störenden Vorkommnissen sowie die Dimension des »classroom managements«, welche in dieser Studie im wesentlichen Maßnahmen zur Prävention von Störungen beinhaltet. Wenn Lehrkräfte das Ausmaß an Störungen gering halten konnten, beobachteten Rakoczy et al. (2007)

zudem, dass die Lernenden über ein hohes Maß an kognitiver Aktivität und positiven emotionalen Erfahrungen (im Sinne der grundlegenden Bedürfnisse der Lernenden nach der Selbstbestimmungstheorie von Deci & Ryan, 1993) berichteten. „These two variables describe the quality of the individual use of the learning opportunities provided, which mediates the effect of a structured learning environment on current motivation“ (Rakoczy et al., 2007, S. 116). Mediatorvariablen erklären dabei grundsätzlich warum und wie eine unabhängige Variable (hier: Strukturierung der Lernumgebung) auf eine abhängige Variable (hier: Aktuelle Motivation) Einfluss nimmt (Frazier, Barron & Tix, 2004, S. 116). Deshalb vermuten Rakoczy et al. (2007, S. 116), dass Maßnahmen zur Strukturierung der Lernumgebung nur dann eine Wirkung auf die aktuelle Motivation der Lernenden entfalten, wenn diese Maßnahmen die individuellen kognitiven und emotionalen Prozesse verstärken. Lipowsky (2009, S. 84) berichtet zusammenfassend zudem von positiven Effekten von Elementen der Strukturierung auf der Verhaltensebene auf die Interessensentwicklung, das Engagement und das Selbstkonzept. Ähnlich wie bei der didaktischen Strukturierung zeigt sich hier insgesamt ein einheitlich positiver Effekt auf den Lernzuwachs. Eine positive Wirkung auch auf affektiv-motivationale Aspekte des Lernens wird zudem für eine größere Anzahl (als bei der Bedeutungsfacette der didaktischen Strukturierung) einzelner Merkmale der Lernenden bescheinigt.

Wirkung von Elementen der kognitionspsychologisch verstandenen Strukturierung

Zur Facette der kognitionspsychologisch verstandenen Strukturierung zählen verschiedene Maßnahmen und Handlungen, welche neu erlernte Wissensselemente mit dem Vorwissen der Lernenden verbinden sollen oder allgemein zum Aufbau einer geordneten Wissensstruktur beitragen. Lipowsky (2009, S. 83) führt als Beispiele von Maßnahmen dieser Facette der Strukturierung u.a. Maßnahmen zur Herstellung von Zusammenhängen, zur Schaffung von Übersichten (z.B. »advance Organizer«), oder auch Lehrerfragen auf. In einer Metaanalyse listen Marzano, Gaddy und Dean (2000) in einer Auswahl von neun besonders lernwirksamen Kategorien drei Kategorien auf, welche Maßnahmen umfassen, die sich dieser dritten Facette der Strukturierung zuordnen lassen. »Identifying similarities and differences« ($d = 1.61$), »summarizing and note taking« ($d = 1.00$) sowie »activating prior knowledge« ($d = 0.59$) weisen dabei einen mittleren bis sehr großen Einfluss auf den Lernerfolg auf (Marzano et al., 2000, S. 4). Auch Hattie (2009) ordnet in seiner (Meta-)Metastudie verschiedenen Merkmalen einen Einfluss auf den Lernzuwachs zu. Für Merkmale, welche der kognitionspsychologisch verstandenen Strukturierung zuzuordnen sind, gibt Hattie (2009, S. 297ff) eine im Vergleich zu Marzano et al. (2000) geringere, aber erneut einheitlich positive Effektstärke an. Bis auf »learning hierarchies« ($d = 0.19$) weisen diese Merkmale eine Effektstärke von $d > 0.4$ auf und werden von Hattie damit als bedeutsam eingestuft. So kennzeichnet Hattie (2009, S. 297ff) die Merkmale »advance organizer« mit der Effektstärke $d = 0.41$, »questioning« mit $d = 0.46$, »goals« mit $d = 0.56$ und »concept mapping« mit $d = 0.58$. Auch Seidel

2. Strukturierung

et al. (2006, S. 811) berichten als Ergebnis der IPN-Videostudie zur Analyse von Physikunterricht positive Auswirkungen auf den Lernerfolg durch Maßnahmen zur klaren Darstellung und Transparenz von Zielen. Diese Maßnahmen können dabei der Facette der kognitionspsychologisch verstandenen Strukturierung zugewiesen werden. Seidel et al. (2006) führen weiterhin an, dass ein Unterricht mit hoher Zielklarheit und Transparenz Vorteile bei der Motivation und der von den Schülerinnen und Schülern wahrgenommenen Unterstützung aufweist. Elemente der kognitionspsychologisch verstandenen Strukturierung können neben positiven Auswirkungen auf den Lernzuwachs also auch Einflüsse auf affektiv-motivationale Aspekte aufweisen. Zu diesem Ergebnis gelangt auch Meschede (2014) in einer Literaturanalyse zur inhaltlichen Strukturierung. Das Konstrukt der inhaltlichen Strukturierung weist in ihrer Dissertation dabei große Übereinstimmungen zur Facette der kognitionspsychologisch verstandenen Strukturierung auf, weshalb eine Übertragung der Aussagen von Meschede (2014) auch auf die Wirkung dieser Facette der Strukturierung gerechtfertigt erscheint. „Zusammenfassend bestätigen die empirischen Befunde, dass strukturierende Maßnahmen sich [...] vor allem auf leistungsbezogene Zielkriterien [...], aber ebenso auch auf affektiv-motivationale Aspekte positiv auswirken“ (Meschede, 2014, S. 90).

Insgesamt kann damit ein bedeutsamer Einfluss von Elementen aller Bedeutungsfacetten der Strukturierung auf den Lernerfolg festgestellt werden. Unter anderem Bybee et al. (2006); Rakoczy et al. (2007); Lipowsky (2009); Seidel et al. (2006) und auch Meschede (2014) berichten zudem von positiven Auswirkungen von Elementen verschiedener Bedeutungsfacetten der Strukturierung auch auf affektiv-motivationale Merkmale. Diese Autoren listen Zusammenhänge von Elementen der Strukturierung auf eine Vielzahl verschiedener nicht-kognitiver Merkmale auf, dabei werden Zusammenhänge zu spezifischen Merkmalen aber selten von mehreren dieser Autoren aufgeführt. Insgesamt erscheint deshalb die Befundlage für Zusammenhänge zu nicht-kognitiven Aspekten weniger einheitlich und in der Gesamtheit auch weniger breit (Hmelo-Silver, 2004, S. 259f.) zu sein. Für die Beziehungen der einzelnen Bedeutungsfacetten der Strukturierung untereinander nimmt Lipowsky (2009, S. 86) eine hierarchische Ordnung an. „Es kann angenommen werden, dass ein Mindestmaß an didaktischer Strukturierung eine notwendige Voraussetzung für eine wirksame Klassenführung darstellt, die wiederum als wichtige Voraussetzung dafür angesehen werden kann, dass inhaltsbezogene Strukturierungen und Hinweise Wirkungen entfalten können“ (Lipowsky, 2009, S. 86). Diesbezüglich betrachten Rakoczy et al. (2010, S. 230) es als erwiesen, dass die strukturierte Organisation der Lernumgebung (Facette der Strukturierung auf der Verhaltensebene) als Vorbedingung für positive Wirkungen weiterer Unterrichtsqualitätsmerkmale (Elemente der kognitionspsychologisch verstandenen Strukturierung) fungiert. Die Rolle der didaktischen Strukturierung als Voraussetzung positiver Auswirkungen von Elementen der beiden weiteren Facetten erscheint allerdings weniger klar zu

sein, zumindest wird für diese Annahme bei Lipowsky (2009) keine dies bestätigende empirische Studie angegeben. Ebenso unklar hinsichtlich der didaktischen Strukturierung ist die Frage, worin sich die deutlich unterschiedliche Lernwirksamkeit verschiedener Vorgaben zur Gestaltung der unterrichtlichen Abläufe begründet. Offen bleibt auch der Grad des Einflusses von strukturierenden Maßnahmen aller Facetten auf affektiv-motivationale Merkmale. Finden sich zahlreiche Angaben zu Effektstärken einzelner Elemente auf die Lernwirksamkeit, gelingt dies nicht für affektiv-motivationale Merkmale.

Wirkung von Elementen der Tiefen- oder Sichtstruktur

Kunter und Voss (2011) betrachten eine nach Elementen der Sicht- und Tiefenstruktur (vgl. S. 18) getrennte Untersuchung der Qualität von Unterricht grundsätzlich als sinnvoll, da „die Forschung zur Effektivität von Unterricht zeigt, dass das Vorliegen bestimmter Sichtstrukturen und die Qualität der Tiefenstrukturen weitgehend unabhängig voneinander variieren, dass also innerhalb der gleichen Sichtstruktur Aufgabenstellungen oder die Interaktion zwischen Lehrenden und Lernenden völlig unterschiedlich gestaltet sein können“ (Kunter & Voss, 2011, S. 87). Kunter und Voss (2011, S. 107) geben zudem an, dass bei ihrer Studie alle drei untersuchten Dimensionen der Tiefenstruktur einen systematischen und positiven Einfluss auf die Lernzuwächse aufwiesen. Die drei untersuchten Dimensionen waren dabei *Effizienz der Klassenführung* (entspricht weitestgehend der Facette der Strukturierung der Verhaltensebene), *Potenzial zur kognitiven Aktivierung* (Facette der kognitionspsychologisch verstandenen Strukturierung) und die *Konstruktive Unterstützung* (enthält mit einem schrittweisen Aufbau des Sachinhalts der Stunde Teilaspekte der Facette der didaktischen Strukturierung, aber auch Aspekte des Lehrerverhaltens und der Beziehung zw. Lehrer und Schülern). Auswirkungen von Elementen der Sichtstruktur wurden in dieser Studie nicht explizit mit der Lernleistung in Verbindung gesetzt (Kunter & Voss, 2011, S. 86ff.), aber auf die Ergebnisse anderer Studien (u.a. Hattie, 2009) verwiesen. Die Ergebnisse dieser Studien weisen Elementen der Sichtstruktur generell eine geringere Erklärungsmacht hinsichtlich der Lernzuwächse bei Schülern zu (Kunter & Voss, 2011, S. 87). Reyer (2004) hingegen konnte in seiner Studie zu Oberflächenmerkmalen und Tiefenstrukturen im Unterricht die leistungsrelevanten Unterrichtsmerkmale auf der Oberfläche des Unterrichts (objektiv beobachtbare Aspekte des Unterrichtsgeschehens Reyer, 2004, S. 59; Sichtstruktur) identifizieren. Empirisch gefundene Leistungsunterschiede zwischen Schülern oder Klassen lassen sich nicht eindeutig auf Elemente der Tiefenstruktur des beobachteten Physikunterrichts zurückführen. „Unter Bezug auf die geringe kodierte Tiefenstruktur-Bandbreite muss dies aber nicht bedeuten, dass die Tiefenstruktur ungeeignet wäre, den Unterrichtserfolg zu erhöhen“ (Reyer, 2004, S. 279). Defizite in der Unterrichtsqualität lassen sich als Defizite in der Tiefenstruktur identifizieren, dies deutet zumindest ein Potential zur Verbesserung der Unterrichtsqualität außerhalb der Sichtstruktur an (Reyer, 2004, S. 279). Die Ergebnisse zur unterrichtlichen Wirksamkeit von

2. Strukturierung

Elementen der Sicht- oder Tiefenstruktur differieren damit stark und scheinen primär von der Wahl der Festlegung einzelner Elemente der beiden Teilkategorien in den jeweils betrachteten Studien abzuhängen, weshalb diese Zweiteilung des Begriffs der Strukturierung nicht als systematische Betrachtungsbasis im Rahmen dieser Arbeit verwendet wird.

Die vorliegende Studie strebt einen Beitrag zur Klärung der Frage an, welchen Einfluss verschiedene Arten der Sequenzierung von Lehr-Lern-Prozessen primär auf Lernerfolg aufweisen. Zudem werden mögliche Effekte bestimmter Sequenzierungsarten auf ausgewählte affektiv-motivationale Merkmale exploriert. Der Fokus dieser Untersuchung liegt damit auf der Wirkung verschiedener Gestaltungen der Sequenzierung von Unterricht auf Basis theoretischer Vorgaben. Deshalb beleuchtet der nachfolgende Abschnitt verschiedene solcher Strukturvorgaben hinsichtlich ihrer Konzeption und ihren unterrichtlichen Auswirkungen insbesondere auf den Lernerfolg.

2.3. Arten von Strukturvorgaben

Festlegung: *Der Begriff der **Strukturvorgabe** beschreibt konzeptionelle Modelle zur Strukturierung von Unterricht, welche Art, Abfolge und Verbindung funktional unterschiedlicher Phasen vorgeben.*

Die Existenz von theoretisch begründeten Vorgaben zur Organisation von Lehr- und Lernprozessen stellt dabei keine hinreichende Bedingung für deren Verwendung dar. Als Ergebnis der IPN-Videostudie zur Analyse von Physikunterricht konnten Seidel et al. (2006, S. 805) feststellen, dass der im deutschen Physikunterricht dominierende Demonstrationsunterricht sich durch eine Erarbeitung der Inhalte im Klassengespräch, einer Veranschaulichung physikalischer Phänomene durch Demonstrationsexperimente, durch meist induktives Vorgehen, aber nicht durch eine theoretische Fundierung auszeichnet. Es fehlt eine Systematik in der Unterrichtsgestaltung. Auch „im Bereich des Experimentierens ließ sich zeigen, dass Lernprozesse in Experimentalphasen oft nur unzureichend und wenig systematisch unterstützt wurden“ (Seidel et al., 2006, S. 812), obwohl Vorgaben hierzu in mehreren, weit verbreiteten Strukturvorgaben sowohl im deutsch-, als auch im angloamerikanischen Sprachraum existieren. Der nachfolgende Abschnitt der Arbeit beleuchtet zunächst einzelne, ausgewählte Strukturvorgaben aus dem angloamerikanischen Sprachraum. Im Zuge dessen werden diese Vorgaben dargestellt und hinsichtlich zweier grundsätzlich verschiedener theoretischer Ausrichtungen kategorisiert. Ausgewählt wurden drei Vorgaben zum *Inquiry-Learning* und damit zur Kategorie der *Erarbeitung* der Inhalte sowie die *Direct Instruction*, welche primär Abläufe zur *Darstellung* oder *Vermittlung* der Inhalte angibt. Anschließend werden zwei Strukturvorgaben aus dem deutschsprachigen Raum bezüglich Konzeption und ihre Effekten auf den Unterricht beschrieben und mit den Strukturvorgaben aus dem angloamerikani-

schen Raum in Beziehung gesetzt. Abschließend erfolgt eine Zusammenfassung des Forschungsstandes zu den Arten der Strukturvorgaben und die Angabe der Forschungsfragen dieser Arbeit.

2.3.1. Strukturvorgaben zum Inquiry-Learning

Im Rahmen einer geplanten Reform des naturwissenschaftlichen Unterrichts Ende der neunziger Jahre formulierte die von der Regierung der Vereinigten Staaten von Amerika beauftragte Organisation des National Research Councils Vorgaben und Empfehlungen für Unterricht, welche in den National Science Education Standards festgehalten sind. Diese Standards sollen zur Entwicklung einer fundierten naturwissenschaftliche Grundbildung (*scientific literacy*) beitragen. „Scientific literacy is the knowledge and understanding of scientific concepts and processes required for personal decision making, participation in civic and cultural affairs, and economic productivity“ (Council, 1996, S. 22). Von zentraler Bedeutung für das Lernen in den Naturwissenschaften ist zudem der Prozess der *Inquiry* (Council, 1996, S. 2). Dem Begriff des *Inquiry* können verschiedene Bedeutungen zugeschrieben werden, diese variieren entsprechend der vorliegenden Kontexte. Sieht man *Inquiry* als eine Vorgabe des Lehrwegs so weist dieses fünf grundlegende Merkmale auf (Council, 2000, S. 24ff.):

1. Learners are engaged by scientifically oriented questions.
2. Learners give priority to evidence, which allows them to develop and evaluate explanations that address scientifically oriented questions.
3. Learners formulate explanations from evidence to address scientifically oriented questions.
4. Learners evaluate their explanations in light of alternative explanations, particularly those reflecting scientific understanding.
5. Learners communicate and justify their proposed explanations.

Für eine gelungene Integration dieser Merkmale im Unterricht stehen dabei verschiedene Strukturvorgaben zur Verfügung. Als grundlegend für diese Vorgaben führen die National Science Education Standards Experimentier- oder andere Erfahrungsprozesse durch die Lernenden auf, welche auf realen Phänomenen oder Problemen basieren (Council, 1996, S. 31).

Discovery-Learning

Einen Vorschlag zur Integration von *Inquiry* im Unterricht formuliert Bruner bereits 1961. In seinem Text „The Act of Discovery“, setzt sich Bruner (1961) auf einer theoretischen Ebene mit der Bedeutung des Entdeckungsprozesses bei den Lernenden auseinander. Die Lernenden sollen dabei nicht Aspekte entdecken, welche tatsächlich eine neue wissenschaftliche Erkenntnis darstellen, sondern vielmehr ihr eigenes Wissen durch eigene Gedankengänge aufbauen. Von die-

2. Strukturierung

sem Entdeckungsprozess erwartet Bruner (1961) und nach dessen Aussage auch zahlreiche Mathematiker, Physiker oder auch Historiker dieser Zeit deutliche und positive Effekte auf die Lernenden. Bruner (1961) vermutet eine Verbesserung der Lernfähigkeiten der Schüler insbesondere auch außerhalb schulischer Kontexte, sowie eine Verschiebung von mehrheitlich extrinsisch zu intrinsisch motivierten Schülern und damit eine stärkere Verinnerlichung der Lernmotivation. Durch den Fokus auf den *Inquiry-Prozess* soll zudem die Fähigkeit ausgebildet und trainiert werden, sinnvolle und gut geplante Untersuchungen durchführen zu können. Die aus den durchgeführten Untersuchungen gewonnenen Informationen schätzt Bruner (1961) weiterhin als schneller aus dem Gedächtnis abrufbar ein. Diese Gedanken bilden die Grundlage des *Discovery Learnings*, welches ausgehend von der Basis Bruners beständig weiterentwickelt wurde. Ein Ergebnis dieser Weiterentwicklung stellt das sogenannte *Scientific Discovery Learning* dar. Hierin werden wesentliche Aspekte des wissenschaftlichen Erkenntnisprozesses mit aufgenommen und durch einen dreistufigen zyklischen Ablauf grob die Struktur des Unterrichts vorgegeben. Dieser Zyklus beinhaltet das Planen und Entwerfen von Experimenten, deren Durchführung und Evaluation. Eine weitere, feingliedrigere Unterteilung ist möglich, die Grundstruktur bleibt dabei aber erhalten (Jong & Joolingen, 1998, S. 180f.).

Ein Überblick zur Literatur zum (*Scientific-*) *Discovery Learning* im Rahmen zweier Metastudien zum Discovery Learning durch Alfieri, Brooks, Aldrich und Tenenbaum (2011, S. 1) zeigt, dass der Begriff nicht scharf und ebenso wenig einheitlich definiert wird. Auf Basis einer Literaturübersicht formulieren Alfieri et al. (2011, S. 2) grundsätzliche Merkmale des Discovery Learnings. Sie ordnen dem Discovery Learning Lernvorgänge zu, bei welchen die zu lernenden Inhalte oder Konzepte nicht direkt präsentiert oder zugänglich gemacht werden, sondern von den Lernenden nur mit Hilfe des zur Verfügung gestellten Materials selbst erarbeitet werden. Über das Ausmaß der Lenkung und Leitung durch die Lehrkraft können Alfieri et al. (2011, S. 2) keine Einigkeit bei in der zugrunde gelegten Literatur erkennen. Sie berichten über ein Spektrum von minimaler bis hin zu sehr intensiver Betreuung bei unterschiedlichen Studien zum Discovery Learning (Alfieri et al., 2011, S. 2). Andere Autoren ergänzen obigen Grundkonsens zum Discovery Learning noch um die Forderung, dass die Lernenden ihre eigenen Experimente planen und entwickeln müssen (van Joolingen, 1999; Jong & Joolingen, 1998). Insgesamt zeigt sich eine sehr starke Betonung eines erfahrungsbasierten Lernens rund um die durch die Lehrkraft vorgegebenen Voraussetzungen oder Probleme und damit eine Berücksichtigung der wesentlichen Vorgaben der National Science Education Standards für naturwissenschaftlichen Unterricht.

Das 5E-Modell

Diese grundlegenden Voraussetzungen erfüllen neben dem Discovery Learning noch weitere Strukturvorgaben. Eine dieser Vorgaben ist der von Atkin und Karplus (1962) vorgeschlagene dreistufige *Learning Cycle*. Der Zyklus startet mit

einer relativ ungerichteten Explorationsphase, welche dazu dient, die Ausgangssituation näher zu beleuchten und erste Informationen zu sammeln. Ausgehend davon werden in der zweiten Phase formale Aspekte, oft Definitionen eines neuen Konzepts thematisiert und im abschließenden Schritt wird das neu Erlernte in anderen Kontexten angewendet. Die Ausgestaltung des *Learning Cycles* unterliegt ebenso wie die des *Discovery Learnings* einer Weiterentwicklung im Sinne einer Anpassung an verschiedene unterrichtliche Situationen sowie der Integration hinzugewonnener Forschungsergebnisse oder politischer bzw. gesellschaftlicher Forderungen (zitiert nach Bybee, 1997, S. 173f.). Für das in den National Science Education Standards als Ziel vorgegebene Erreichen einer naturwissenschaftlichen Grundbildung (*Scientific Literacy*) schlägt Bybee (1997) in einem Buchbeitrag unter dem Kapitel »Improving Instruction« eine Weiterentwicklung des *Learning Cycles* als unterrichtliche Strukturvorgabe vor. Deren Kern bildet der ursprünglich von Atkin und Karplus (1962) formulierte dreistufige Zyklus. Bybee setzt diesem eine Aktivierungs- bzw. Motivationsphase voran und endet zusätzlich mit der ausführlichen Evaluation der gewonnenen Erkenntnisse. Insgesamt gliedert Bybee den Ablauf seiner Vorgabe in fünf Phasen: *engagement*, *exploration*, *explanation*, *elaboration* und *evaluation* und formuliert dadurch den *5E-Learning Cycle*. Mit dessen Hilfe kann das Programm eines ganzen Schuljahres, eine zusammenhängende Sequenz oder auch einzelne Stunden in sich immer wiederholende Sequenzen von fünf aufeinander aufbauenden Phasen strukturiert werden (Bybee, 1997, S. 176).

Bybee (1997, S. 177ff.) beschreibt dabei insbesondere die Ausgestaltung dieser fünf Phasen noch etwas präziser: Die erste Phase (*engagement*) dient der Präsentation des Kontexts des Lerninhalts. Dies kann mit Hilfe von Fragen, einer Problemformulierung oder einem überraschendem Phänomen bewerkstelligt werden. Ziel dieses eher kurzen und einfach zu haltenden Abschnitts ist es, die Lernenden in einen Zustand zu versetzen, in welchem sie verblüfft durch und motiviert für die Lerngelegenheit sind. Ist dieser Zustand erreicht, soll in der zweiten Phase (*exploration*) Gelegenheit und Raum für die Lernenden geschaffen werden, um eigenständig ihre eigenen Vorstellungen erkunden zu können. Dabei können Zusammenhänge aufgedeckt, Muster entdeckt, Variablen identifiziert oder Ereignisse in Frage gestellt werden. Diese Exploration ermöglicht allen Lernenden gemeinsame und konkrete Erfahrungen, welche die Basis für eine später erfolgende, formale Behandlung der Lerninhalte bilden. In der folgenden Phase (*explanation*) werden Erklärungen zur gewählten Ausgangssituation von den Lernenden auf Grundlage der gemeinsamen Erfahrungen aus der Exploration eingefordert, im Anschluss daran wissenschaftliche Erklärungen durch die Lehrkraft vorgestellt. In dieser Phase sollen die allgemeinen Konzepte und Theorien aus den Beobachtungen hervorgehoben und nachvollziehbar gemacht werden. Verfügen die Lernenden über eine Erklärung (auf Basis der Konzepte und Theorien) der Lernaufgabe, werden ihnen in der vierten Phase (*elaboration*) weitergehende Erfahrungen ermöglicht, im

2. Strukturierung

Rahmen welcher sie die Konzepte oder Theorien anwenden, erweitern oder ausschärfen können. Die Anwendung des Erlernten auf neue Kontexte oder Probleme und der Austausch zwischen den Lernenden über deren Verständnis der Inhalte stellen wesentliche Aspekte dieser Phase dar. Abschließend erhalten die Lernenden auch eine Rückmeldung (*evaluation*) über die Angemessenheit der erworbenen Erklärungen von Seiten der Lehrkraft. Dies kann durchaus prozessbegleitend und eher informell gehandhabt werden oder schriftlich in Form eines Wissenstests erfolgen. Wesentlich ist, dass die Erklärungen oder Problemlösungen der Lernenden den Blicken der Peers oder der Lehrkraft standhalten müssen. Damit ist der fünfstufige Lernzyklus abgeschlossen, welcher auch für nachfolgende oft in Schrittzahl oder Gewichtung leicht veränderter Lernzyklen die Grundlage darstellt. Eisenkraft (2003) schlägt beispielsweise vor, Bybees Zyklus um zwei Stufen zu erweitern. Forschungsergebnisse der Kognitionspsychologie hätten deutlich gemacht, dass das Hervorheben und Thematisieren der Schülervorstellungen, sowie das Üben des Transfers des Erlernten, notwendige Komponenten im Lernprozess darstellten. Beide Aspekte sieht Eisenkraft (2003, S. 57) in Bybees *5E-Learning cycle* nicht ausreichend eingegliedert, weshalb Eisenkraft diesem eine Phase (*elicit*) voranstellt, um die Schülervorstellungen und das Vorwissen zu aktivieren und eine Phase (*extend*) anfügt, um den Transfer in neuen Kontexten anzuwenden. Durch das Anfügen dieser beiden Phasen entsteht so aus dem 5E ein 7E-Modell.

Durch dieses bewusste Strukturieren und Sequenzieren einzelner Abläufe soll bereits im 5E-Modell eine Brücke von den aktuellen Schüler(wohl)vorstellungen zu den wissenschaftlichen Konzepten geschlagen werden (Bybee, 1997, S. 184). Für die Ausgestaltung des *5E-Learning Cycles* betrachtet Bybee (1997, S. 185f.) die Unterrichtsforschung zum Konzeptwechsel, die Übereinstimmung der Strukturvorgabe mit den Prozessen der wissenschaftlichen Forschung (*scientific inquiry*), die Nützlichkeit der Vorgabe zur Planung und Entwicklung von Unterrichtsmaterialien und den praktischen Nutzen für Lehrkräfte der Naturwissenschaften als grundlegend. Zusätzlich dazu versucht das 7E-Modell neuere bzw. weitere Forschungsergebnisse zur Kognition zu berücksichtigen. Die von Eisenkraft (2003) vorgeschlagene Erweiterung um zwei Phasen kann diesbezüglich als eine stärkere Verankerung des Brückenschlages von den Schülervorstellungen zu den wissenschaftlichen Konzepten angesehen werden.

Problem-Based-Learning

Ein weiterer, alternativer Vorschlag zur Strukturierung von Lehr-Lern-Sequenzen ist das *Problem Based Learning*. „Problem Based Learning represents a major development and change in educational practice that continues to have a large impact across subjects and disciplines around the world“ (Gijbels, Dochy, den Bossche & Segers, 2005, S. 27). Dabei erkennen Gijbels et al. (2005, S. 29) in ihrer Metaanalyse zum Problem Based Learning, dass dieses in verschiedenen Domänen und im Hinblick auf oft sehr unterschiedliche Zielsetzungen kaum einheitlich umgesetzt wird. Als ein Ergebnis ihrer Literaturrecherche

für eine Metastudie zur Wirksamkeit des Problem Based Learnings erkennen Gijbels et al. (2005, S. 29), dass zahlreiche durchaus differierende Strukturvorgaben existieren, welche sich aber alle auf das Lehren und Lernen in konkreten Problemsituationen beziehen. Viele dieser Vorgaben zur Strukturierung von Unterricht würden dabei in der Fachliteratur als Problem Based Learning-Ansätze beschrieben. Von zentraler Bedeutung für diese Ansätze sei dabei der Fokus auf das Ausgangsproblem, welches den Lernprozess initiieren soll. Eine einheitliche Definition oder Realisierung von Problem Based Learning werde auch durch die große Bandbreite an Auffassungen des Problem Based Learnings erschwert. Nach Borrowes (1996) lassen sich aber sechs Kerncharakteristiken des Problem Based Learnings formulieren (zitiert nach Gijbels et al., 2005, S. 29f.):

1. Lernen ist schülerzentriert.
2. Lernen vollzieht sich in Kleingruppen von Schülern.
3. Ein anwesender Tutor fungiert als Moderator oder Leiter.
4. Authentische Probleme werden zu Beginn einer Lernsequenz präsentiert, noch bevor eine Vorbereitung oder Untersuchung begonnen hat.
5. Die aufgetretenen Probleme sind Hilfsmittel zum Aufbau des benötigten Wissens und zum Aufbau der zum Lösen des Problems notwendigen Fähigkeiten.
6. Neue Informationen werden durch selbstgesteuertes Lernen erworben.

Um einen Überblick zu schaffen, wie Unterricht mit Hilfe der Vorgabe des Problem Based Learnings strukturiert werden kann, wird an dieser Stelle eine gut dokumentierte Interpretation des Problem Based Learnings exemplarisch dargestellt. Als Beispiel wurde der *Problem-based Learning Cycle* nach Hmelo-Silver (2004) ausgewählt. Auch im Überblicksartikel zum Problem Based Learning von Hmelo-Silver (2004) zeichnet sich diese Strukturvorgabe durch Kleingruppenarbeit aus, in welcher die Lernenden diejenigen Inhalte erwerben sollen, welche sie befähigen ein eingangs formuliertes Problem zu lösen. Das gewählte Problemszenario soll dabei authentisch und variabel bzgl. der Lösung und des Lösungsweges sein (Hmelo-Silver, 2004, S. 236f.). Hmelo-Silver (2004, S. 235 und S. 239) sieht die Aufgabe des Lehrenden primär in der Begleitung der Lernenden auf deren Lernwegen entlang des *Problem-based Learning Cycles* (Siehe Abb. 2.1).

Dieser Lernweg wird auch als der *Problem Based Learning Tutorial Process* beschrieben und beginnt mit der Präsentation einer komplexen Problemsituation. Ausgehend von einer minimalen Menge an Informationen zu diesem Problem müssen sich die Lernenden aktiv weitergehende Inhalte zur Lösung der Situation beschaffen. Diese können durch Fragestellungen an die betreuende Lehrkraft, durch das Ausführen von Experimenten oder einer Literaturrecherche erhalten werden. Während dieses Prozesses werden die gesammelten Informationen

die Lernenden dazu befähigt, metakognitive Strategien aus dem *Problem-based Learning Cycle* auf Lernsituationen aus allen Lebensbereichen übertragen zu können. Zusätzlich sei das Problem Based Learning konzipiert um soziales, gemeinschaftliches Lernen und Zusammenarbeiten zu schulen, sowie um die intrinsische Motivation der Lernenden zu stärken.

Sowohl das Problem Based Learning, das Discovery Learning als auch das 5E-Modell sind Strukturvorgaben, welche ein erfahrungsbasiertes Lernen anhand einer wohl gewählten Ausgangssituation ermöglichen sollen und Elemente des Inquiry Ansatzes integrieren. Das Zusammenspiel zwischen den Untersuchungen (*inquiry*) der Lernenden und dem Aufbau der Lerninhalte unterscheidet sich aber dabei. Beim Discovery Learning stellt sich nicht die Frage wie die Unterrichtszeit zwischen primär an Untersuchungen orientieren Phasen und eher traditionell Inhalte-vermittelnden Phasen gestaltet wird. Im Fokus steht vielmehr, wie schnell oder wie stark die Lernenden durch Begleitmaterialien oder die Lehrkraft auf ihrem Entdeckungsprozess angeleitet werden, bzw. welcher Zeitraum für die Entdeckungen zur Verfügung gestellt wird (Hammer, 1997, S. 489). Im Gegensatz dazu können sowohl im 5E-Modell als auch beim Problem Based Learning Informationen den Lernenden direkt zugänglich gemacht werden. Während beim Problem Based Learning diese Inhalte idealerweise in Form einer selbst gesteuerten Lernaufgabe zur Verfügung gestellt werden, sind beim 5E Modell Phasen direkter Instruktion vorgesehen (Hmelo-Silver, 2004, S. 239). Ausgeschlossen ist deren Existenz aber auch in einer Strukturierung nach dem Problem Based Learning-Muster nicht, wie Hmelo-Silver, Duncan und Chinn (2007, S. 100) in einer späteren Veröffentlichung explizit angeben. Als Anpassung des Problem Based Learnings an den jeweiligen Entwicklungsstand der Lernenden ist direkte Instruktion auf einer *just-in-time* Basis möglich (Hmelo-Silver et al., 2007, S. 100), aber eben nicht üblich (Hmelo-Silver, 2004, S. 260).

Obwohl die Verwendung von Strukturvorgaben zum *Inquiry Learning* in den National Science Education Standards empfohlen wird und der Einsatz des *5E-Learning Cycles*, des *Problem Based Learnings* oder des *Discovery Learnings* sich auch didaktisch begründen lassen, obliegt die faktische Umsetzung im Unterricht weiterhin der volitionalen Entscheidung der Lehrenden für oder gegen die Verwendung der Vorgabe des *Inquiry Learnings* sowie der einzelnen zugehörigen Strukturvorgaben. Zudem muss berücksichtigt werden, dass auch der Einsatz didaktisch gut begründbarer Strukturvorgaben nicht zwangsläufig zu einer tatsächlichen, also auch empirisch beobachtbaren Verbesserung des Lernens führt. Die empirische Überprüfung der Wirkung von Unterricht nach den Vorgaben einzelner dieser Gestaltungsvorschläge ist notwendig. Zahlreiche Untersuchungen wurden diesbezüglich gerade im angloamerikanischen Sprachraum auch durchgeführt. Der nachfolgenden Abschnitt stellt ausgewählte Ergebnisse der Untersuchungen zur Wirksamkeit der Strukturvorgaben zum Inquiry Learning dar.

2. Strukturierung

Aussagen zur Wirksamkeit der Strukturvorgaben zum Inquiry Learning

Aus Sicht von Kirschner, Sweller und Clark (2006, S. 2) wäre eine differenzierte Betrachtung dieser drei Vorgaben für Unterricht (Discovery Learning, 5E-Modell, Problem Based Learning) wohl nicht notwendig gewesen, betrachten sie diese doch als grundsätzlich pädagogisch gleichgeartet. So ordnen Kirschner et al. (2006, S. 1) diese und noch weitere Vorgaben in ihrem Artikel zur Analyse des Misserfolgs von „Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching“ dem *Minimally Guided Approach* zu. Minimally Guided bedeutet dabei, dass den Lernenden die Kerninhalte nicht präsentiert werden, sondern diese selbständig entdeckt oder konstruiert werden müssen. Als wesentliche Grundaussage der Autoren kann bereits der Titel des zu diesem Thema verfassten Artikels angesehen werden: „Why minimally guidance during instruction does not work“. Kirschner et al. (2006, S. 3) begründen die Formulierung und die Aussage dieses Titels auf Basis empirischer Untersuchungen der letzten 50 Jahre, welche überwältigenden und unmissverständlichen Beweis hierfür erbracht hätten. Die Ursache für die mangelnde Lerneffizienz schreiben die Autoren dabei der inquiry-orientierten Grundausrichtung zu. Diese beanspruche das Arbeitsgedächtnis der Lernenden durch die Suche nach Problemlösungen in einem so starken Ausmaß, dass das Arbeitsgedächtnis für weitere Prozesse nicht mehr zur Verfügung stehe und so nicht zum eigentlichen Lernvorgang benutzt werden könne (Kirschner et al., 2006, S. 5). Die im Artikel formulierte Kritik geht dabei noch über die fehlende Lernwirksamkeit hinaus: „Not only is unguided instruction normally less effective, there is evidence that it may have negative results when students acquire misconceptions or incomplete and/or disorganized knowledge“ (Kirschner et al., 2006, S. 15).

Vor allem die von durch Kirschner et al. (2006) sehr eng gefasste Kategorisierung verschiedener und selten einheitlich interpretierter Instruktionsvorgaben als ein zusammengehörender Ansatz (*Minimally Guided Approach*) ruft Widerspruch hervor. Hmelo-Silver et al. (2007) kritisieren in ihrer Antwort auf oben dargestellten Artikel nicht die generelle Aussage des Artikels, wohl aber die Zuordnung des Problem Based Learnings und der Ansätze des Inquiry-Learnings zum *Minimally Guided Approach*. „The teacher plays a key role in facilitating the learning process and may provide content knowledge“ (Hmelo-Silver et al., 2007, S. 100), zudem strukturiere die Lehrkraft schwierige Aufgaben und Prozesse und versuche die kognitive Belastung zu reduzieren. Wird die Aufgabe der Lehrkraft in Lernumgebungen des Problem Based Learnings bzw. verschiedener Ansätze des Inquiry-Learnings so interpretiert, erscheint eine Einordnung eben dieser Ansätze zum *Minimally Guided Approach* auf Basis der Definition dieses Begriffes bei Kirschner et al. (2006) unzulässig zu sein. Für beliebige Lernumgebungen zum Problem Based Learning und dem 5E-Modell kann demnach die Grundaussage der geringen Effizienz nicht einfach pauschal übernommen werden, weshalb nachfolgend eine differenziertere Betrachtung der Wirkung der verschiedenen Strukturvorgaben zum Inquiry-Learning erfolgen soll.

Aussagen zur Wirksamkeit des Discovery Learnings

Mit „discovery, like surprise, favors the well prepared mind“ formuliert Bruner (1961, S. 1) bereits eine Vorbedingung für effektives Lernen nach den Vorgaben des Discovery Learning. Berücksichtigt man diesen Hinweis bei der Umsetzung, verspricht sich Bruner (1961) jedoch deutlich positive Auswirkungen durch die Verwendung dieser Strukturvorgabe. Im darauf folgenden Zeitraum von über 50 Jahren entwickelten sich immer wieder verschiedene Ausprägungen des Discovery Learnings. Mayer (2004) kategorisiert diese Vielzahl an verschiedenen Ausprägungen in einem Review-Artikel zum Discovery Learning in zwei Kategorien ein. Die erste Kategorie bildet die sogenannte *Pure Discovery*, bei welchem die Lernenden mit einem Problem konfrontiert werden und dieses ohne oder mit wenig Hilfe der Lehrkraft lösen sollen. Dem gegenüber steht die Kategorie der *Guided Discovery*, die sich dadurch auszeichnet, dass die Lehrkraft die Lernenden aktiv beim Problemlöseprozess unterstützt, indem sie Hinweise, Rückmeldungen oder sogar die korrekte Antwort beisteuert. Mit Hilfe dieser Einteilung verschiedener Ausprägungen der Strukturvorgabe des Discovery Learnings führt Mayer (2004) eine historische Aufarbeitung der Literatur zum Discovery Learning durch. Strukturvorgaben entsprechend des *Pure Discovery* zeigen sich dabei in allen untersuchten Zeiträumen und Teilaspekten den Ansätzen gemäß *Guided Discovery* unterlegen (Mayer, 2004). Vergleicht man diese Aussage mit denjenigen von Kirschner et al. (2006), besteht Einigkeit in der Unwirksamkeit von generell unangeleitetem Unterricht (*minimally guided Instruction*), Uneinigkeit wiederum in der pauschalen Zuordnung des Discovery Learnings zu eben dieser Kategorie durch Kirschner et al. (2006). Alfieri et al. (2011, S. 7) konnten in Übereinstimmung zu den Aussagen von Mayer (2004) in einem Vergleich von 108 Studien zeigen, dass Discovery Learning ohne aktive Steuerung (*hier: unassisted Discovery*) durch die Lehrkraft einem Unterricht mit deutlicher Lehrersteuerung hinsichtlich des Lernzuwachses unterlegen ist. Dabei zeigt sich eine mittlere negative Effektstärke von $d = -0.38$, wobei einzelne Studien stark von diesem Wert abweichen. Insgesamt ergibt sich ein inhomogenes Bild, welches Alfieri et al. (2011) auf die Unterschiedlichkeit in den Untersuchungen den verwendeten Tests und den Lernaufgaben zurückführen. Deutlich wirkungsvoller ist Discovery Learning, wenn die Lehrkraft den Unterricht bewusst lenkt (*Enhanced Discovery*). Im Vergleich zu anderen in den Studien vorkommenden Instruktionsformen weisen die Ansätze zur *Enhanced Discovery* im Mittel eine größere Lernwirksamkeit ($d = 0.3$) auf (Alfieri et al., 2011, S. 7). Auch dies ist allerdings kein generell gültiger Effekt. Bei einem Teil der Studien zur *Enhanced Discovery* mussten die Lerner z.B. Regeln oder Strategien entwickeln, oder Antworten auf experimentelle Fragen finden. Diese Studien bilden bei (Alfieri et al., 2011, S. 5) die Unterkategorie der *Generation*. Ansätze zum Discovery Learning aus dieser Kategorie erwiesen sich im Vergleich zu den in den jeweils verwendeten Kontrollgruppen eingesetzten Unterrichtsformen als weniger lernwirksam ($d = -0.15$). Alfieri et al. (2011,

2. Strukturierung

S. 5) legen in ihrer Arbeit noch zwei weitere Kategorien für *Enhanced Discovery* fest. Die Unterkategorie der *elicited explanation* kennzeichnet Discovery Learning Ansätze, bei welchen Erklärungen durch die Lernenden eingefordert werden, die Unterkategorie der *guided discovery* hingegen subsumiert Ansätze, bei denen die Lehrkraft mittels Instruktion oder Feedback den Entdeckungsvorgang begleitet. Diesen beiden Unterkategorien ordnen Alfieri et al. (2011, S. 11) wiederum positive Effektstärken von $d = 0.36$ für *elicited explanation* und $d = 0.50$ für *guided discovery* zu. Im Hinblick auf die in den jeweils untersuchten Studien zugrunde liegenden Fachbereiche geben Alfieri et al. (2011, S. 9) die Effektivität des lehrergesteuerten Discovery Learnings zwar als einheitlich positiv an, in der Ausprägung allerdings schwankten die Effektstärken zwischen $d = 0.11$ für die Naturwissenschaften und $d = 1.05$ für das Erlernen motorischer Fähigkeiten. „Overall, results seem to favor enhanced-discovery methods over other forms of instruction“ (Alfieri et al., 2011, S. 11). Dabei muss aber beachtet werden, dass auch *Enhanced Discovery* sich bei Alfieri et al. (2011) nicht bei allen berücksichtigten Studien als die effizienteste Strukturvorgabe erwies.

Aussagen zur Wirksamkeit des 5E-Modells

Den Kern des 5E-Modells bildet der dreistufige *Learning Cycle* (vgl. S. 30) nach Atkin und Karplus (1962). Eine Aussage über dessen Wirksamkeit ist damit auch ein Hinweis auf die Wirksamkeit des 5E-Modells, zu welchem die Literaturbasis im Vergleich zum dreistufigen Zyklus noch nicht so breit untersucht ist (Bybee et al., 2006, S. 29). In einem zusammenfassenden Bericht über die Ursprünge und die Auswirkungen des 5E-Modells zeichnen Bybee et al. (2006) mit Hilfe der von ihnen verwendeten empirischen Studien zunächst ein sehr eindeutiges Bild von den Effekten des *Learning Cycles* und damit des Kerns des 5E-Modells auf das Lernen. Einen insgesamt positiven Einfluss bescheinigen Bybee et al. (2006, S. 29f.) dem Learning Cycle bezüglich des Lernzuwachses (*mastery of science matter*), der Fähigkeit zum wissenschaftlichen Argumentieren (*scientific reasoning*), sowie der Entwicklung von Interesse an und positiven Einstellungen zu den Naturwissenschaften (*interest and attitudes toward science*). Keine eindeutigen Effekte konnten Bybee et al. (2006) auf die Kategorien *Understanding of the Nature of Science*, *Understanding of the Complexity and Ambiguity of Empirical work*, sowie *Development of Practical Skills* oder *Development of Teamwork Skills* erkennen. Bybee et al. (2006, S. 30) geben für die Effekte des 5E-Modells in ihrem Bericht zu den Effekten des *Learning Cycles* gleichgeartete Effekte und Ausrichtungen an. Einzelne dieser Effekte konnten für das 5E-Modell jedoch nicht als statistisch signifikant identifiziert werden. Vor diesem Kontext betonen Bybee et al. (2006, S. 30) die weniger breite Dokumentationsbasis für Effekte des 5E-Modells, wodurch eine entsprechende Identifikation signifikanter Effekte erschwert sein könnte. Bybee et al. (2006, S. 35) geben zusätzlich an, dass sich die Wirksamkeit des 5E-Modells in Einzelstudien dabei als nicht immer einheitlich positiv erwiesen hat. In der BSCS (Biological Sciences Curriculum Study) »Science for Life and Living« wurde die Wirksamkeit des 5E-Modells bezüglich

des Lerninhalts und des Wissens über naturwissenschaftliche Untersuchungen mit der Wirksamkeit des Unterrichts einer Kontrollgruppe gegenüber gestellt. Die Art des Unterrichts in dieser Kontrollgruppe wurde dabei nicht konstant gehalten. Der Vergleich beider Gruppen deutet eine größere Lernwirksamkeit des 5E-Modells an. Allerdings zeigte sich diese größere Lernwirksamkeit nicht in jeder der untersuchten inhaltlichen Teilkategorien. Bei drei von insgesamt elf untersuchten Teilkategorien konnte kein signifikanter Unterschied im Lernzuwachs zwischen den Gruppen gefunden werden. Bei einer Teilkategorie ergab sich sogar eine Überlegenheit der Kontrollgruppe. Weitere Untersuchungen über die Wirksamkeit des 5E-Modells, insbesondere über die Gelingensbedingungen sind notwendig. Insgesamt erkennen Bybee et al. (2006, S. 41) allerdings eine wachsende Literaturbasis, welche die Wirksamkeit des Modells nahe legt. Auch in der (Meta-)Metaanalyse von Hattie (2009) finden sich indirekt Aussagen zur Wirksamkeit des 5E-Modells wieder. Das 5E-Modell kann ebenso wie der dreistufige Zyklus nach Atkin und Karplus (1962) der Kategorie des *Inquiry-based Teachings* in Hatties Studie über die Lernwirksamkeit unterrichtlicher Bedingungsfaktoren zugeordnet werden. „Overall, inquiry-based instruction was shown to produce transferable critical thinking skills as well as significant domain benefits, improved achievement, and improved attitude towards the subject“ (Hattie, 2009, S. 211). Die gesamte Effektstärke auf den Lernzuwachs des *Inquiry-based Teachings* gibt Hattie (2009, S. 208f.) mit $d = 0.31$ an. Dabei zeigen sich auch bei Hattie (2009, S. 208f.) deutliche Unterschiede bezüglich der Wirksamkeit von Strukturvorgaben des *Inquiry-based Teachings* zwischen den einzelnen ausgewerteten Metastudien.

Aussagen zur Wirksamkeit des Problem Based Learnings

Hmelo-Silver et al. (2007, S. 102f.) geben an, dass Untersuchungen zur Wirksamkeit des Problem Based Learnings meist auf Bereichen der medizinischen Ausbildung basieren. Dies erschwert eine Übertragung der Aussagen auch auf andere Disziplinen. Zudem zeigten Studien und Metastudien ein uneinheitliches Bild über die Lernwirksamkeit des Problem Based Learnings, oft existiere kein Unterschied zu den Vergleichsgruppen in den Studien, nur gelegentlich zeigten die Problem Based Learning-Gruppen einen größeren Lernzuwachs. Als wirkungsvoller habe sich das Problem Based Learning hinsichtlich der Entwicklung von Fähigkeiten zur Argumentation, zum selbst gesteuerten Lernen oder zur Vorbereitung auf zukünftige Lernaufgaben erwiesen. Auch in der Metastudie von Gijbels et al. (2005) weist nur eine von insgesamt 40 Studien keinen medizinischen Kontext auf. Gijbels et al. (2005) verwenden dabei zur Auswahl der Studien die sechs Kerncharakteristiken (Siehe S. 33) des Problem Based Learnings nach Barrows (1996) und untersuchen die Wirkung des Problem Based Learnings auf drei Teilbereiche der Wissensstruktur. Diese unterteilen Gijbels et al. (2005, S. 34f.) dazu in das Verstehen von Wissenselementen (*declarative knowledge*, 1.Stufe), das Verbinden der Wissens Elemente oder Größen, z.B. mit Regeln, oder Zusammenhängen (*procedural knowledge*, 2.Stufe) und der

2. Strukturierung

Organisation der Wissensstruktur (*conditional knowledge*, 3.Stufe). Zeigt sich bezüglich der ersten Stufe nur ein sehr schwach positiver und zudem nicht signifikanter Vorteil zu Gunsten des Problem Based Learnings, so weisen Gijbels et al. (2005, S. 45f.) für die zweite Stufe eine große und signifikante sowie für die dritte Stufe eine kleine bis mittlere, jedoch wiederum nicht signifikante positive Effektstärke aus. In seiner (Meta-)Metastudie legt Hattie (2009) ebenfalls obige sechs Kerncharakteristiken für das Problem Based Learning zur Auswahl und Kategorisierung der von ihm zusammengefassten Arbeiten zugrunde. In Übereinstimmung zu Gijbels et al. (2005) schätzt Hattie (2009, S. 210f.) das Problem Based Learning zum Aufbau von Wissenselementen (*surface knowledge*) nur als eingeschränkt wirksam ein und weist darauf hin, dass selbst negative Effekte in Studien berichtet werden. Für tiefer gehende Lernprozesse, bei welchen der Aufbau von Wissenselementen bereits abgeschlossen ist, konstatiert Hattie (2009, S. 210f.) allerdings deutlich positive Effekte des Problem Based Learnings. Den insgesamt positiven aber kleinen Gesamteffekt auf die Lernwirksamkeit kennzeichnet Hattie (2009, S. 210f.) mit einer Effektstärke von $d=0.15$. Dieser Gesamteffekt des Problem Based Learnings, liegt für Hattie unterhalb einer für die Praxis wünschenswerten Grenze von $d=0.4$. Liegt die unterrichtliche Zielsetzung des Problem Based Learnings allerdings primär auf der Verbindung oder Anwendung von bereits erlernten Wissenselementen liegt die Wirksamkeit dieser Strukturvorgabe über der Grenze von $d=0.4$ (Hattie, 2009, S. 210f.). „The application of knowledge, not development of knowledge, is the heart of the success of problem-based-learning“ (Hattie, 2009, S. 211).

Insgesamt zeigt sich in der Literatur ein sehr differenziertes Bild zu den einzelnen Strukturvorgaben zum Inquiry-Learning. Eine allgemein lernförderliche Tendenz wird konstatiert, jedoch nur, wenn der Unterricht eine aktive Steuerung durch die Lehrkraft erfährt und oft auch nur für spezifische Zwecke. Beispielsweise scheint das Problem Based Learning für die Verbindung von Wissenselementen, nicht aber für deren Aufbau geeignet zu sein. Für diesen Zweck eignet sich bei erkennbarem Bedarf die Verwendung bestimmter Phasen Lehrer-gesteuerter Instruktion (Hmelo-Silver et al., 2007, vgl. S. 100).

2.3.2. Direct Instruction

Grundkonzeption

Eine allgemein stark lehrer-gesteuerte Instruktion wird im Englischen oft auch mit dem Begriff der *teacher-directed instruction* oder dem der *direct instruction* beschrieben. Trotz sehr ähnlichem oder sogar identischen Wortlauts unterscheidet sich die Strukturvorgabe der *Direct Instruction* deutlich von einer schlicht stark lehrer-gesteuerten Instruktion. Die Direct Instruction gibt sehr klare Gestaltungsvorgaben zum Unterricht, nicht in jeder Phase steht dabei aber die Lehrperson im Vordergrund (Hattie, 2009, S. 204f.). In einem Überblicksartikel

zur Direct Instruction beschreiben Magliaro, Lockee und Burton (2005) die von Bereiter und Engelmann (1966) formulierte Basisform der Direct Instruction als einen dreistufigen, systematischen Ablauf des Lehrprozesses, welcher von einer durchgehenden Einschätzung und Evaluation des Lernfortschritts geprägt ist. Zunächst erfolge dabei eine Vorstellung und Einführung des neuen Themengebiets, daran schließe die Präsentation der wesentlichen Inhalte an, welche nachfolgend in Übungsphasen angewendet würden. Die Anwendung des Gelernten erfolge dabei schrittweise vom angeleiteten zum unabhängigen Üben und werde von einer unmittelbaren Rückmeldung über den Lernfortschritt der Lernenden begleitet. Magliaro et al. (2005, S. 44) halten zusätzlich fest, dass der Begriff der Direct Instruction zur Beschreibung einer gewissen Bandbreite an Instruktionsmodellen verwendet werde, welche sich durch folgende Kernkomponenten auszeichneten:

1. Der Lerninhalt wird in bewusst angeordneten, kleinen Schritten unterrichtet.
2. Ziele über die Leistungsmessung oder zu erwartende Fähigkeiten müssen klar und transparent gehalten werden.
3. Die Lernenden erhalten Gelegenheiten, das neu Gelernte mit ihrem Vorwissen zu verbinden.
4. Die Lernenden erhalten schrittweise Gelegenheit zur Anwendung.
5. Die Lernenden erhalten zusätzliche Anwendungsmöglichkeiten, welche eine zunehmend eigenverantwortliche und unabhängige Übung anregen.
6. Rückmeldung erfolgt schrittweise, begleitend zur Anwendung.

Engelmann (1980, S. 4) schreibt der Direct Instruction größere Anforderungen an die Lehrkräfte als anderen Ansätzen zu, da die Vorgabe der Direct Instruction alle kontrollierbaren, leistungsrelevanten Teilaspekte des Unterrichts auch wirklich kontrollieren will. Für die Lehrkräfte bedeute dies, eine nahezu choreografiert erscheinende, fest vorgeschriebene Sequenz durchzuführen. Orientiert an der ursprünglich dreistufigen Basisform, differenziert Engelmann (1980) sein Modell der Direct Instruction zu einer fünfstufigen Sequenz weiter aus.

Die Struktur der Direct Instruction

Als erster Schritt erfolgt der Einstieg in ein neues Konzept auf der Basis zuvor errungener Fähigkeiten oder Wissensselementen. Danach folgt die Hauptpräsentation der Unterrichtsstunde in Form einer kleinschrittigen, systematisch wohl durchdachten Erklärung oder Demonstration, entwickelt um nur *eine* Interpretation des Konzepts klar hervortreten zu lassen. Die Präsentation des Konzepts muss dabei mit geeigneten Beispielen und Gegenbeispielen untermauert werden. Anschließend erhalten die Lernenden Gelegenheit verbal auf Fragen oder Aufgaben einzugehen, um aufzuzeigen wie weit fortgeschritten ihr Lernweg ist. Daran zeigt sich auch, ob die Lernenden bereits befähigt sind das

2. Strukturierung

Konzept auf weitere Beispiele zu übertragen. Im vierten Teilschritt erfolgt eine Rückmeldung an die Lernenden auf deren verbale Beiträge. Fehlerhafte Beiträge werden korrigiert, sind inhaltliche Lücken zu erkennen, folgt eine Wiederholung der nicht berücksichtigten Aspekte als Vorbereitung für die nachfolgende Phase der unabhängigen Einübung. Diese Phase verläuft weitestgehend von den Lernenden selbst gesteuert, bei Bedarf kann die Lehrkraft aber regulierend eingreifen (zitiert nach Magliaro et al., 2005, S. 46f.). Im Detail wurde dieser Ablauf an spezifische Ansprüche verschiedener Fachbereiche (*Good and Grows Strategies for Effective Teaching model* oder *Hunters Design of Effective Lessons model*), der Unterschiedlichkeit einzelner Lernenden oder dem Komplexitätsgrade verschiedener Kontexte (beides: *Rosenshines Explicit Teaching model*) angepasst. Die Umsetzung der Kernkomponenten und die Orientierung an der Basisform stellen dabei für Magliaro et al. (2005, S. 43ff.) unabhängig von der zeitlichen Anpassung ein gemeinsames Merkmal der verschiedenen Ansätze zur Direct Instruction dar.

Aussagen zur Wirksamkeit der Direct Instruction

Adams und Engelmann (1996) verwenden in ihrer Metastudie nur Einzelstudien zur Lernwirksamkeit von Strukturvorgaben, welche für eine der mindestens zwei untersuchten Gruppen, den von Engelmann (1980) dargestellten Ansatz der Direct Instruction als theoretische Grundlage aufweisen. Diese Bedingung erfüllen insgesamt 173 Studien. Für 109 dieser Studien listen Adams und Engelmann (1996, S. 42) eine signifikante Überlegenheit der Strukturvorgabe der Direct Instruction im Vergleich zur jeweiligen Kontrollgruppe auf. Nur in zwei Fällen zeigt sich ein signifikant geringerer Lernzuwachs. Alle Studien zusammengefasst, ergibt sich im Vergleich eine sehr große Effektstärke zu Gunsten der Direct Instruction (Adams & Engelmann, 1996, S. 43). Die Vergleichsgruppe setzt sich dabei aus allen in den Studien zur Planung des Unterrichts in den Kontrollgruppen verwendeten Strukturvorgaben zusammen und repräsentiert eine große Bandbreite verschiedener Unterrichtsabläufe. Adams und Engelmann (1996, S. 43f.) stellen weiterhin fest, dass der als Ergebnis der Auswertung aller Studien festgestellte große positive Effekt der Direct Instruction, weitestgehend unabhängig davon ist, ob an Sonderschulen oder an Schulen des regulären Bildungssystems unterrichtet wurde. Abhängig zeigt sich der Effekt in ihrer Metastudie allerdings vom Alter der Lernenden und des Unterrichtsfaches. So profitieren jüngere Schüler (Kindergarten) stark, aber weniger als ältere (Gesamtschulen, weiterführende Schulen und Erwachsenenbildung) und für naturwissenschaftlichen Unterricht zeigt sich Direct Instruction effizienter als für das Erlernen von Sprachen. Adams und Engelmann (1996, S. 43f.) weisen für die Wirkung der Direct Instruction bei allen untersuchten Teilaspekten positive Effekte nach. Für einzelne Teilgebiete finden Adams und Engelmann (1996) sogar besonders positive Effekte der Direct Instruction. „It produces superior results for basic skills and for higher-order cognitive skills in reading and in math“ (Adams & Engelmann, 1996, S. 97). Zudem geben Adams und Engel-

mann (1996, S. 97) an, dass Direct Instruction einen deutlich positiven Einfluss auf das Selbstvertrauen der Lernenden zeigt. Auf Basis dieser Aussagen, kommen Adams und Engelmann (1996, S. 98) bezüglich der Wirksamkeit des Direct Instruction als Strukturvorgabe zu folgendem Fazit: „Direct Instruction has no peer in this enterprise“. Auch in Hatties (Meta-)Metastudie (2009) erzielt keine Strukturvorgabe für Unterricht eine höhere mittlere Lerneffektivität. Direct Instruction erreicht bei Hattie (2009, S. 204f.) eine Effektstärke von $d = 0.59$ und liegt damit in einem Bereich, von welchem sich Hattie im Unterricht merklige Auswirkungen erwartet. Die Datengrundlage für seine Aussagen stellen Studien dar, denen eine etwas breitere Definition der Direct Instruction als bei Adams und Engelmann (1996) zu Grunde liegt, sich aber dennoch klar von *direct instruction* im Sinne eines nicht näher spezifizierten, lehrerzentrierten Unterrichtsgesprächs unterscheidet. Auch Hattie (2009, S. 204ff.) ermittelt für die Wirkung der Direct Instruction bei allen in den untersuchten Metastudien aufgeführten Teilgruppen eine deutlich positive Effektstärke bezüglich des Lernerfolges. Insbesondere eine öffentlich oft vermutete einseitige Eignung der Direct Instruction nur für jüngere oder weniger leistungsfähigere Schüler kann Hattie (2009, S. 204ff.) auf Basis seiner Datenanalyse widerlegen. Hattie (2009, S. 204ff.) bezeichnet die Verwendung von Direct Instruction als eines der erfolgreichsten Kriterien für Lernerfolg. Vorbehalte gegen Direct Instruction seien dabei meist nur begrifflichen Missverständnissen geschuldet (Hattie, 2009, S. 205ff.). Ein ebensolches würde die Gleichsetzung von Direct Instruction mit Frontalunterricht darstellen. Wird Frontalunterricht oft nur im Sinne eines eher unreflektierten Frage-Antwort-Spiels durchgeführt (Duit, Hepp & Rincke, 2013, S. 5), gibt Direct Instruction sehr genaue Vorgaben über Ablauf, Aufbau und auch Ausgestaltung einer Unterrichtsstunde und will alle Stellschrauben des Lernerfolges gleichzeitig kontrollieren (Engelmann, 1980, S. 4).

2.3.3. Forschend-entwickelndes Unterrichtsverfahren

Obwohl Seidel et al. (2006) im Physikunterricht in Deutschland kaum eine theoretische Fundierung beobachten können, existieren auch im deutschsprachigen Raum entsprechende Vorgaben zur Strukturierung des Unterrichts. Eine dieser Vorgaben ist das *Forschend-entwickelnde Unterrichtsverfahren* (FeU) welches ursprünglich von Fries und Rosenberger (1967) für den mathematischen und naturwissenschaftlichen Unterricht formuliert und von Schmidkunz und Lindemann (1992) spezifisch für die naturwissenschaftlichen Fächer weiterentwickelt wurde.

Grundkonzeption

Grundlegend für den Phasen- oder Stufenverlauf des FeU nach Fries und Rosenberger (1967, S.14ff) ist der von Fries und Rosenberger angenommene, bei der Aneignung rationaler Erkenntnisinhalte durchschrittene, formale Ablauf

2. Strukturierung

der Denkprozesse des Schülers. Diese Denkprozesse geben die Strukturierung der Unterrichtsstunde fest vor, die Entscheidung über didaktisch-methodische Einzelmaßnahmen bleiben der Lehrkraft überlassen. Die Gesamtstruktur ist in drei Arbeitsschritte unterteilt. Der erste Abschnitt A dient der Neuerarbeitung von Erkenntnissen. In den Abschnitten B und C werden die so neu gewonnenen Erkenntnisse gefestigt. Im Fokus von Abschnitt B liegen den Schüler aktivierende Übungen „im unmittelbaren Zusammenhang mit dem vorausgegangenen Arbeitsabschnitt [...] im Sinne einer Mechanisierung und Automatisierung“ (Fries & Rosenberger, 1967, S. 39). Erst anschließend soll sich der Lernende in Arbeitsabschnitt C mit den neu erworbenen und geübten Erkenntnissen ihm zugängliche reale Anwendungssituationen erschließen. Den Ablauf der Arbeitsabschnitte A und C gibt die grundlegende Gliederung der Denkprozesse vor. Diese Denkprozesse beginnen in den Arbeitsabschnitten A und C mit der Stufe der Problemgewinnung (I), welche in die Phasen des Problemgrundes (Ia), der Problemfindung (Ib) und der Problemerkkenntnis (Ic) unterteilt sind. Danach folgt die Stufe der Problemlösung (II) mit Überlegungen zur Problemlösung (IIa), deren Planung (IIb), Durchführung (IIc) und Diskussion (IID). Einzelne Phasen können entfallen, die Reihenfolge der Denkstufen und -phasen soll aber nicht verändert werden. Für den Arbeitsabschnitt B sehen Fries und Rosenberger (1967, S.38f) keine entsprechende Untergliederung in einzelne Denkphasen vor. Insgesamt geben Fries und Rosenberger (1967, S. 16-39) damit einen stark phasierten Verlauf für ihre Vorgabe zum forschend-entwickelnden Unterricht vor, welcher in seiner Grundform auch in das FeU nach Schmidkunz und Lindemann (1992) integriert wird.

Lindemann (1989) sieht die Legitimation für das FeU und dessen Weiterentwicklung auf zwei Ebenen begründet. Auf der Ebene der psychologischen und pädagogischen Erkenntnisse zum Lernprozess versucht das FeU mehreren didaktischen Lernprinzipien gerecht zu werden. Zum einen ist dies das Prinzip der Motivation mit dem Ziel eine Lernbereitschaft bei den Lernenden zu erzeugen, zum anderen soll sich die Arbeit der Lernenden durch ein hohes Maß an Eigenaktivität und Möglichkeiten für Erfolgserlebnisse auszeichnen. Das Lernen soll grundlegend aus Problemsituationen heraus erfolgen und gleichzeitig „die kognitiven, affektiven und psychomotorischen Fähigkeitsbereiche des Menschen“ (Lindemann, 1989, S. 21) ansprechen. Dies deckt sich in weiten Teilen mit den didaktischen Prinzipien, welche Fries und Rosenberger (1967, S.14f) auch in ihrem Ansatz des forschend-entwickelnden Unterrichts berücksichtigt glauben. Zusätzlich zu dieser gemeinsamen didaktischen Basis geht Lindemann (1989) davon aus, dass mit Hilfe des FeU auch das Abstraktionsvermögen der Lernenden geschult werden könne und sieht darin eine grundlegende Zielsetzung des FeU. Diese Zielsetzung prägt auch die Umgestaltung des ursprünglichen Stufen- und Phasenmodells in wesentlichen Zügen. Die zweite Begründungsebene bildet der Erkenntnisprozess in der Wissenschaft, welchem Lindemann (1989, S. 20) Gemeinsamkeiten zum Erkenntnisgewinn der Schüler

zuschreibt und der deshalb die Grundlage des Stufenmodells nach Schmidkunz und Lindemann (1992) bildet. Die psychologisch-pädagogische und die fachwissenschaftliche Ebene sollen im Forschend-entwickelnden Unterrichtsverfahren gleichsam vereint sein. *Forschend* bedeutet dabei, dass der Lernende die neuen Erkenntnis auf Basis seines Vorwissens und seiner Fähigkeiten in der Regel selbstständig erwerben soll. Mit *entwickelnd* verbindet Lindemann (1989, S. 22) die Notwendigkeit einer bedarfsabhängigen Lehrersteuerung des Lern- und Forschungsprozesses. Schmidkunz und Lindemann (1992, S. 24ff.) legen dabei die Rolle der Lehrkraft in verschiedenen Denkstufen und Denkphasen des FeU nicht grundsätzlich fest. Die Rolle kann sich sehr unterschiedlich gestalten und zeigt sich in ihrer Ausgestaltung abhängig vom thematisierten Inhalt, der Schülergruppe und auch der Art des Problems. Gerade die Art des Problems ist dabei für den Verlauf des FeU von entscheidender Bedeutung. „Für einen Lernenden entstehen immer dann Problemsituationen, wenn er mit Gegenständen, Situationen, Geschehnissen oder Vorgängen konfrontiert wird, die er mit seinem Vorwissen und seinen bisherigen Erfahrungen nicht erklären bzw. nicht nach bekannten Verhaltensmustern lösen kann“ (Schmidkunz & Lindemann, 1992, S. 16f.). Diese Definition beinhaltet zwei im Unterricht grundsätzlich unterschiedliche Ausgangssituationen. Möglich ist eine Problemsituation, zu der die Lernenden bereits über alle notwendigen Kenntnisse verfügen und damit in der Lage sind das Problem weitestgehend theoretisch zu lösen. Möglich ist aber auch, dass ein Sachverhalt dargestellt, oder ein Phänomen beobachtet wird, zu welchem die wesentlichen inhaltlichen Grundlagen fehlen. Beide Möglichkeiten bedingen bei Schmidkunz und Lindemann (1992, S. 23) jeweils einen unterschiedlichen unterrichtlichen Ablauf. Schmidkunz und Lindemann (1992, S. 23) benennen diese erste Variante eines solchen Ablaufs mit der Bezeichnung *Deduktiver Zweig*. Diesen Ablauf empfehlen Schmidkunz und Lindemann (1992, S. 23) bei einer Problemsituation, zu der die zur Lösung notwendigen fachlichen Kenntnisse bereits erlernt wurden. Mit *Induktiver Zweig* bezeichnen Schmidkunz und Lindemann (1992, S. 23) den zweiten unterrichtlichen Ablauf im FeU. Dieser soll eingesetzt werden, wenn im Unterricht ein Phänomen betrachtet werden soll, zu welchem die relevanten inhaltlichen Kenntnisse bei den Lernenden noch nicht vorhanden sind.

Der *Induktive Zweig* stellt aus Sicht der Autoren den Normalfall des FeU dar (Schmidkunz & Lindemann, 1992, S. 35). Sowohl für den *Deduktiven* als auch den *Induktiven Zweig* geben Schmidkunz und Lindemann (1992) klare Richtlinien zur Gestaltung von aufeinanderfolgenden Lehr- und Lernphasen vor. „Die aufeinanderfolgenden Lernphasen sind dabei keinesfalls als unabänderlicher Algorithmus zu sehen“ (Schmidkunz & Lindemann, 1992, S. 21), Phasen können modifiziert, verkürzt oder weggelassen werden. Der Phasenprozess wird dabei insgesamt in fünf Denkstufen und diese jeweils in drei Denkphasen gegliedert. Die resultierende Stufen- und Phasenstruktur des FeU wird im folgenden Abschnitt näher dargestellt.

2. Strukturierung

Die Struktur des FeU

Als Orientierung für die formulierten Vorgaben fungiert der Prozess der naturwissenschaftlichen Erkenntnisgewinnung. Für die Erkenntnisgewinnung der Schüler rückt damit das Experiment in den Mittelpunkt der systematischen Strukturierung des Unterrichtsverlaufs. Das Forschend-entwickelnde Unterrichtsverfahren (FeU) geht dabei aber über die rein wissenschaftsmethodische Struktur des Experimentierens hinaus, da es Phasen der Wissenssicherung, der Anwendung und der Übung enthält (di Fuccia & Ralle, 2010, S. 297f.). Eine Übersicht zur Struktur des FeU ist in Abb. 2.2 dargestellt.

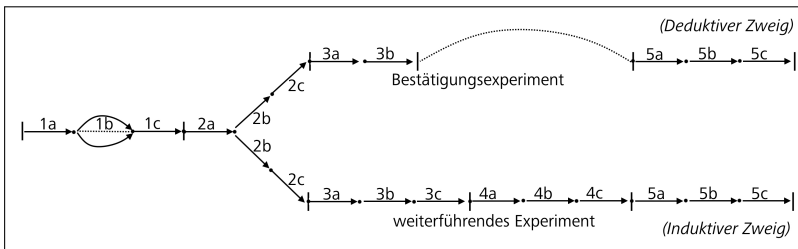


Abbildung 2.2.: Vorgaben des FeU zu den Möglichkeiten der didaktischen Strukturierung des Unterrichts nach unterschiedlichen Vorkenntnissen (Schmidkunz & Lindemann, 1992, vgl. S. 36)

Problemgewinnung (1) als erste Denkstufe des *induktiven Zweigs* beginnt mit der Denkphase des *Problemgrundes* (1a). In dieser Denkphase soll die Basis gelegt werden, um später inhaltlich zentrale Aspekte aus der ausgewählten Problemsituation heraus arbeiten zu können. Die Problemsituation als solche und die zu behandelnden Inhalte sind zuerst nur der Lehrkraft bekannt. In der Phase des *Problemgrundes* soll es der Lehrkraft zunächst gelingen, dass das Problem von den Lernenden aufgenommen wird. In der zweiten Phase der *Problemerkennung* (1b) wird das eigentliche Problem mit Hilfe des Vorwissens der Lernenden ausdifferenziert und Nebenaspekte rücken in den Hintergrund. Gelingt dies den Lernenden selbständig, sprechen Schmidkunz und Lindemann (1992, S. 25) von einer Problemfindung (oberer Pfeil in Abb. 2.2 bei Denkphase 1b). Die Problemfindung betrachten Schmidkunz und Lindemann (1992, S. 23) dabei als Idealfall. Alternativ „muss der Lehrende das Problem durch mehr oder weniger starke eigene Aktivität aufzeigen“ (Schmidkunz & Lindemann, 1992, S. 25), in diesem Fall sprechen Schmidkunz und Lindemann (1992, S. 25) von der Problemstellung (unterer Pfeil in Abb. 2.2 bei Denkphase 1b). Als wichtig betrachten die Autoren es dabei, dass in der anschließenden Denkphase der *Problemerkennung* (1c) das zentrale Problem genau und eindeutig formuliert wird. Diese Phase endet erst, wenn alle Lernenden die Problemsituation genau erfassen und damit als Grundlage für den nachfolgenden aktiven Problemlöse-

prozess zur Verfügung haben. Die zweite Denkstufe umfasst *Überlegungen zur Problemlösung* (2), welche als erste Phase die *Analyse des Problems* (2a) enthält. Darin sollen alle einzelnen, verfügbaren und notwendigen Vorkenntnisse und Vorerfahrungen im Hinblick auf das zu lösende Problem identifiziert und bereit gestellt werden. Erst nach dieser Phase weisen *induktiver* und *deduktiver* Zweig leicht unterschiedliche Phasenverläufe auf (Schmidkunz & Lindemann, 1992, S. 26).

In diesem Absatz erfolgt nun eine Darstellung des Verlaufs nur des *induktiven Zweiges*. Die Ausgestaltung des *deduktiven Zweiges* ist hingegen Inhalt des nachfolgenden Abschnitts. Auf Basis des in (2a) bereit gestellten Wissens werden im *induktiven Zweig* in der Denkphase der *Lösungsvorschläge* (2b) verschiedene, zumindest teilweise begründbare Lösungsansätze formuliert, welche mit Hilfe von Experimenten realisierbar sein sollen. In der anschließenden Phase der *Entscheidung für einen Lösungsvorschlag* (2c) soll sich die Gruppe auf einen der möglichen Lösungsansätze festlegen. Einigt man sich auf einen Vorschlag, welcher nicht zur Lösung des Problems führt, laufen dennoch anschließend alle weiteren Denkprozesse zur Problemlösung ab, die Gruppe muss sich danach aber neu für einen alternativen Lösungsansatz entscheiden und vollzieht einen Kreisprozess innerhalb des FeU. Die *Durchführung des Lösungsvorschlages* (3) erfolgt in der dritten Stufe und beginnt mit der *Planung des Experimentes* (3a). Im Rahmen dieser Planung wird festgelegt wie groß die einzelnen Gruppen sind und ob innerhalb derer arbeitsteilig oder arbeitsgleich gearbeitet werden soll. Zudem ist auch die materielle Ausgestaltung und Vorbereitung des Experiments Teil dieser Planung. Die Ausführung des Experiments wird in der *praktischen Durchführung des Lösungsvorhabens* (3b) vollzogen. In diesem Teilzweig des FeU sprechen Schmidkunz und Lindemann (1992, S. 29) von einem weiterführenden, echten Forschungsexperiment, „weil der Lernprozess an dieser Stelle nicht abgeschlossen wird, sondern durch weitere Denkstufen, insbesondere durch die Stufe der Abstraktion weitergeführt wird“. Die so neu gewonnenen Erkenntnisse werden anschließend in der Phase der *Diskussion der Ergebnisse* (3c) näher betrachtet. Die vorformulierten Lösungsvorschläge werden durch den Experimentierausgang entweder verifiziert oder falsifiziert. Wird ein Lösungsansatz falsifiziert, so durchlaufen die Lernenden die Denkphasen ab der Entscheidung für einen Lösungsansatz erneut, wird einer der vermuteten Ansätze verifiziert, folgt die Stufe der *Abstraktion der gewonnenen Erkenntnisse* (4) zunächst in Form der *ikonischen Abstraktion* (4a). Hier wird in Form von Skizzen oder Experimentieranordnungen versucht, möglichst anschaulich die Ergebnisse der Experimente darzustellen, welche darauf folgend in der Phase der *verbalen Abstraktion* (4b) durch die Lernenden in kurzen, klaren Aussagen formuliert werden. Schmidkunz und Lindemann (1992, S. 30) vermuten, dass das Ausmaß an Abstraktionsvermögen auf Seiten der Lernenden in dieser Denkstufe in zunehmenden Maße eingefordert wird. In der Phase der *symbolischen Abstraktion* (4c) seien die Anforderungen an das Abstraktionsvermögen der

2. Strukturierung

Lernenden maximal. Im Chemieunterricht würde die *symbolische Abstraktion* zum Beispiel einer Angabe von Reaktionsgleichungen entsprechen, möglich ist allerdings auch eine quantitative mathematische Darstellung der Sachverhalte. Bei manchen der thematisierten Inhalte haben einzelne der Abstraktionsphasen keine Bedeutung und können entfallen. Ebenso kann eine Anpassung an spezielle Schulstufen oder Situationen erfolgen. In der fünften Denkstufe der *Wissenssicherung* (5) schließt eine Suche von *Anwendungsbeispielen* (5a) zum Inhalt der Unterrichtsstunde an. „Hier soll nicht eine Anwendung betrieben werden, um zu üben und zu gewissen Fertigkeiten zu gelangen, sondern es soll ein beziehungsreiches Wissen mit vielen Querverbindungen aufgebaut werden“ (Schmidkunz & Lindemann, 1992, S. 31). Der Bezug zu anderen Unterrichtsfächern, oder zu Situationen aus Alltag, Technik oder Umwelt steht im Fokus dieser Denkphase. Ziel ist der fächerübergreifende Transfer des Gelernten. Das Gelernte soll wiederum in der Phase der *Wiederholung* (5b) zusammengefasst und neu aufbereitet werden. Zudem soll wiederholt werden, wie die Lernenden zu den Ergebnissen gelangten. Oft endet an dieser Stelle das Unterrichtsverfahren, kann aber durch eine *Lernzielkontrolle* (5c) ergänzt werden, welche der Überprüfung des aufgebauten Wissens und somit des Unterrichtserfolges dient (Schmidkunz & Lindemann, 1992, S. 31).

Grundsätzlich gilt obiger Stufen- und Phasenverlauf in weiten Teilen auch für den *deduktiven Zweig* des FeU. In diesem Fall verfügen die Lernenden aber bereits über die relevanten Sachinhalte, welche im *induktiven Zweig* erst aufgebaut und abstrahiert werden müssen. Demnach können die Phase der Diskussion der Ergebnisse, sowie die Stufe der Abstraktion entfallen. Die Abstraktion der Inhalte stellt für Schmidkunz und Lindemann (1992, S. 33f.) aber auch ein wesentliches Element des *deduktiven Zweiges* des FeU dar. Bei diesem Zweig des FeU erfolgt eine Abstraktion allerdings schon während der *Analyse des Problems* (2a). Diese Phase wird zudem weiter ausdifferenziert. In (2a₁) erfolgt die Bereitstellung des Vorwissens, wonach in (2a₂) die Entwicklung von Prognosen und in (2a₃) eben eine Abstraktion und Symbolisierung erfolgt. Die Phasen (2b) und (2c) wiederum verlaufen analog zum *induktiven Zweig*. Allerdings verändert sich die Funktion des Experiments. Im *deduktiven Zweig* wird das Experiment zur Bestätigung der theoretischen Überlegungen eingesetzt. Entsprechend bezeichnen Schmidkunz und Lindemann (1992, S. 32ff.) diese Art des Experiments als Bestätigungsexperiment. „Mit diesem Bestätigungsexperiment schließt in der Regel der Lernprozess in diesem Zweig des Verfahrens ab“ (Schmidkunz & Lindemann, 1992, S. 34), die abschließende Wissenssicherung erfolgt in beiden Zweigen analog.

Aussagen zur Wirksamkeit des FeU

Der Strukturvorgabe des FeU attestieren di Fuccia und Ralle (2010, S. 297f.) einen hohen Stellenwert und Beliebtheit im schulischem Alltag. Zudem werde das FeU in nahezu jeder naturwissenschaftsdidaktischen Ausbildung von Lehrkräften sowohl in der ersten als auch in der zweiten Phase der Lehrerbildung

thematisiert. Dennoch existieren zur Untersuchungen der Wirksamkeit des FeU nur sehr wenige empirische Studien (Sommer, 2014). Verschiedene Aussagen zur Wirkung existieren, sind auf Unterrichtsbeobachtungen und -analysen begründet, letztlich aber kaum empirisch überprüft. Lindemann (1989, S. 27) vermutet in einem Artikel anlässlich des 60. Geburtstags von Heinz Schmidknecht eine verringerte Wirksamkeit des FeU immer dann, wenn einzelne Phasen dieses Verfahrens entfallen. Unabhängig davon, ob das Fehlen einzelner Phasen bewusst so geplant oder unbewusst übergangen wurde, soll dadurch die didaktische Wirksamkeit der „Idealstruktur“ des FeU verringert werden (vgl. Abb. 2.3).

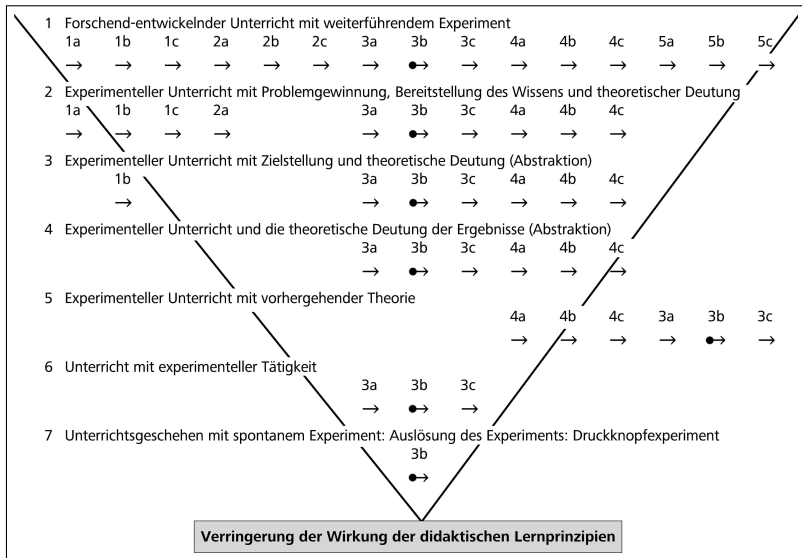


Abbildung 2.3.: Vermutete Verringerung der Wirkung der didaktischen Lernprinzipien des FeU (Lindemann, 1989, vgl. S. 310)

Diese *erwartete* Wirksamkeit begründet Lindemann (1989, S. 27) dabei mit der Berücksichtigung vieler didaktischer Lernprinzipien, der genauen Strukturierung des Unterrichts und der sinnvollen Einbettung des Experiments in den Verlauf der Stunde, welche mit einer Verwendung des FeU zur Gestaltung des Unterrichts einher gehe. „Auf diesen Grundlagen kann zusammenfassend festgestellt werden, und die Praxis bestätigt dies, dass das forschend-entwickelnde Unterrichtsverfahren zum Verständnis und zur Handhabung des experimentellen Chemieunterrichts einen wesentlichen Beitrag leistet“ (Lindemann, 1989, S. 27). Einen empirisch gefundenen Nachweis für diese Überzeugung listet

2. Strukturierung

Lindemann (1989) allerdings nicht auf. Eine empirisch abgesicherte Wirksamkeit des FeU ist auch allgemein kaum dokumentiert. Als eine der wenigen Einzelstudien vergleicht die Dissertation von Rudolf Herbers (1990) anhand einer Unterrichtssequenz zu einem ausgewähltem Thema den kurz- und langfristigen Lernerfolg von FeU gegenüber Frontalunterricht. Letzterer ist dabei nur dahingehend definiert, als dass der Lehrer ohne Diskussion den Unterrichtsverlauf bestimmt (Herbers, 1990, S. 12). Der Lernerfolg wird in dieser Studie in acht Kategorien gemessen. In drei dieser Kategorien zeigt sich die frontal unterrichtete Gruppe beim kurzfristigen Lernerfolg derjenigen deutlich überlegen, welche Unterricht gemäß des FeU erfuhren. In den anderen fünf Kategorien treten diesbezüglich keine großen Unterschiede auf. Für die Bestimmung des langfristigen Lernerfolgs wurde jährlich über einen Zeitraum von sechs Jahren der Kenntnisstand zu allen acht Kategorien erhoben. Die letzte Datenerhebung lässt erkennen, dass der Kenntnisstand der FeU-Gruppe bei sieben der acht Kategorien nun höher als in der Vergleichsgruppe lag, weshalb Herbers (1990) eine langfristig stabilere Lernwirkung des FeU vermutet. Zudem weist Herbers (1990) auf eine positive Wirkung bezüglich des Interesses der Lernenden hin. Eine statistische Überprüfung der Signifikanz oder der Effektstärken der beobachteten Mittelwertsunterschiede der Gruppen erfolgt in keiner der Kategorien.

Insgesamt scheinen die Aussagen zu Wirkungen des FeU mehr von Überzeugungen als von empirischen Nachweisen geprägt zu sein. In der Studie von Herbers (1990) deutet sich zumindest aber an, dass Unterricht mit Hilfe des FeU lernwirksam gestaltet werden kann. Das Ausmaß der Lernwirksamkeit bleibt sowohl in Größe als auch bezüglich Signifikanz und der Eignung für verschiedene Inhalte ungeklärt. Ebenso bleibt offen, ob durch das FeU auch allgemeine Lernziele wie das Ausbilden einer Abstraktionsfähigkeit auf Seiten der Lernenden gezielt gefördert werden. Gerade das Training der Abstraktion repräsentiert aber eine der wesentlichen Grundideen (Lindemann, 1989, S. 20) des FeU und zudem einen der wesentlichen Kritikpunkte an der Unterrichtspraxis im Fach Chemie.

2.3.4. Basismodelltheorie

Grundkonzeption

Auch die Basismodelltheorie (BMT) wird deutlich von der Kritik an der aktuellen unterrichtlichen Praxis der Lehrkräfte beeinflusst. So formulieren Oser und Baeriswyl (2001, S. 1032) in einem Artikel zur Zusammenfassung ihrer theoretischen Schlussfolgerungen aus einer mehrjährigen Studie an schweizerischen Schulen über die beobachteten Lehrkräfte: „They organize what is visible (the sight-structure of learning) and neglect to consider the more important question of what is happening in the mind of the student (the basis-model)“. Mit *Basismo-*

dell bzw. *Tiefenstruktur des Lernens* bezeichnen Oser, Patry, Elsässer, Sarasin und Wagner (1997, S. 1 bzw. 8f.) nicht direkt beobachtbare Vorgänge im Unterricht. Die *Tiefenstruktur* beinhaltet Aspekte wie den tatsächlichen Lernvorgang im Kopf der Lernenden, also deren geistige Aktivitäten. Im Gegensatz dazu beschreibt der Begriff *Sichtstruktur* alle Vorgänge, welche gut beobachtet werden können. Dies umfasst beispielsweise die Wahl der Sozialform, der Medien oder Methoden. Die Fokussierung auf Elemente der Sichtstruktur gerade bei der Gestaltung von didaktischen Modellen führt aus Sicht von Oser et al. (1997, S. 6) dazu, dass „die theoretischen Modelle der Didaktik alles Fuzy-Theorien sind, die keinen handlungsweisenden Charakter haben“. Die Überbetonung des Konzepts der Methodenfreiheit führt zu einer nahezu willkürlichen Gestaltung unterrichtlicher Abläufe und bedingt die Nichtberücksichtigung vorhandener, sehr detaillierter Beschreibungen der Lernabläufe u.a. auf Basis entwicklungs- und gedächtnispsychologischer Erkenntnisse. Diese Erkenntnisse legen aber nahe, dass beispielsweise Begriffe eben nicht in willkürlicher Art und Weise erfolgreich aufgebaut werden können, sondern vielmehr eine gewisse Schrittfolge beachtet werden muss. Ohne einen Weg, der zumindest ein minimales Maß an Struktur vorgibt, ist Lernen chaotisch. Deshalb fordern Oser et al. (1997) eine Umkehr der pädagogischen Sichtweise, keine Fokussierung auf Methoden, Sozialformen oder Inhalte, sondern eine Fokussierung auf die Aktivierung geistiger Aktivitäten, da „Lernen aus *äußeren* Aktivitäten besteht, die *geistige* Aktivitäten in Bewegung setzen“ (Oser et al., 1997, S. 1). Das Problem der Verbindung beider Aktivitätsarten, ist dabei genau das Grundproblem des Brückenschlags zwischen Lehren und Lernen. Durch das Lehren können nur die Bedingungen für das Lernen gelegt werden, die geistigen Aktivitäten der Lernenden sind dabei oft schwer erfassbar, bilden aber dennoch die Grundlage für die Planung der Lehraktivitäten. Diese Aktivitäten unterscheiden sich hinsichtlich ihrer Ziele. Der Aufbau von Wissen, das Lernen aus Erfahrungen oder das Abspeichern von Episoden, zielen dabei jeweils auf eine andere, unterschiedliche Art des Lernens ab und erfordern unterschiedliche Ketten an Handlungen. Diese Handlungsketten geben die Schritte im Lernprozess der Lernenden genau vor und können damit als Lernskripts angesehen werden. Ein Lernskript für eine spezifische Art des Lernens bildet ein Basismodell. Die Struktur eines Basismodells (Lernskript) ist dabei nicht auf den ersten Blick zu erkennen, sondern muss „hypothetisch erschlossen oder vom Lernenden erfragt werden“ (Oser et al., 1997, S. 9), kann dabei nicht einfach frei erfunden werden, sondern entspricht vielmehr einer pädagogisch-psychologischen Tradition. Die Entwicklung einzelner Basismodelle erfolgt bei Oser et al. (1997, S. 11f.) in fünf Schritten oder Stufen. In einer ersten Stufe zur Definition der Basismodelle haben Oser et al. (1997, S. 11) aus wissenschaftlicher und didaktischer Literatur entsprechende Lernskripts „herausdestilliert“. Als zweite Stufe fordern Oser et al. (1997, S. 11) ein, dass es möglich sein muss, auf theoretischer Basis zu erklären, wie Veränderungen im Lernenden durch die Verwendung eines Basismodells hervorgerufen werden. Als dritte Stufe muss wiederum die

2. Strukturierung

Möglichkeit gegeben sein, einen theoretisch vorgegebenen Ablauf auch tatsächlich im Unterricht wiederzufinden und beschreiben zu können. Weiterhin soll die Angabe einer Schrittfolge in beobachtbaren Unterteilungen des Unterrichts erfolgen können (vierte Stufe). Abschließend nennen Oser et al. (1997, S. 11) in einem fünften Schritt die empirische Überprüfung der definierten Schrittfolge als notwendige aber nicht hinreichende Bedingung für die Ausgestaltung der Basismodelle. Die Autoren gehen in ihrer Arbeit von einer begrenzten Zahl an Basismodellen aus, garantieren aber keine Vollständigkeit ihrer Auflistung von Basismodellen, sondern empfehlen eine empirische Überprüfung sowohl hinsichtlich der Beschränkung der Anzahl der Modelle, als auch hinsichtlich einer eventuellen Vollständigkeit des angegebenen Basismodellkatalogs. In diesem Katalog unterscheiden die Autoren insgesamt zwölf voneinander unabhängige, nicht vermischbare Basismodelle (vgl. Tabelle 2.1).

Lernen d. Eigenerfahrung und entdeckendes Lernen	Entwicklung als Erziehungsziel	Problemlösen
Konzeptaufbau	Kontemplatives Lernen	Lernen durch Motilität
Strategielernen	Routinebildung	Entwicklung von Sozialbeziehungen
Entwicklung von Wertesystemen	Lernen mit Hypertext	Verhandeln

Tabelle 2.1.: Die Basismodelle des Lehrens und Lernens

Die verschiedenen Basismodelle repräsentieren verschiedene Zieltypen des Lernens mit einem jeweils besonderen Merkmal. Beispielsweise schreiben Oser et al. (1997, S. 14) dem Basismodell des *Konzeptaufbaus* den Aufbau von vernetztem Wissen als Lernzieltyp und als besonderes Merkmal größere Sach- und Fachzusammenhänge und Analysebildung zu. Diese verschiedenen Elemente und Ziele in den einzelnen Basismodellen führen auch zu einer jeweils unterschiedlichen Ausgestaltung der zugehörigen Lernsequenz. Für das Basismodell des *Konzeptaufbaus* ergibt sich folgender unterrichtlicher Ablauf (Oser et al., 1997, S. 9):

1. Direkte oder indirekte Bewusstmachung der bisher erworbenen Wissensstruktur
2. Vorstellen eines Beispiels mit prototypischen Charakter
3. Präsentation bzw. Repräsentation einer oder mehrerer neuer Elemente, die der alten Wissensstruktur fremd waren, aber beim Prototyp auftauchen
4. Eingliederung der neuen Elemente durch Aktivitäten wie Vergleichen, in Beziehung setzen, Einschließen, Trennen etc.

5. Schaffung einer optimalen Koordination innerhalb der neuen Wissensstruktur; durch Anwendung des neuen Wissens auf ein anderes Gebiet

Die Basismodelle geben den Grundrhythmus des Unterrichts vor, welcher fest und nicht veränderbar ist. Veränderbar bleiben die Elemente der Sichtstruktur, welche so die freie und frei gestaltbare Natur des Lernens widerspiegeln. Im Idealfall beeinflussen die Elemente der Basismodelle die Wahl der Methoden, nicht umgekehrt (Oser et al., 1997, S.20ff.). Die Grundannahme von Oser und Baeriswyl (2001, S. 1043) bei der Formulierung ihrer Theorie zum Lehr-Lern Prozess ist, dass jede Sequenz des schulischen Lernens auf einer Choreographie basiert, welche die Freiheit der Methoden- und Sozialformwahl (Sichtstruktur) mit einer absolut strikten Abfolge an notwendigen inneren Lernschritten (Basismodelle) verbindet. Diese Lernschritte sind dabei so konzipiert, dass man keinen einzelnen Schritt entfallen lassen darf und deren Abfolge bei jedem Basismodell und damit für jeden Lernzieltyp eindeutig festgelegt ist. Allerdings kann ein Lerngegenstand auch mit verschiedenen Zieltypen verknüpft werden, damit stehen verschiedene Lernskripts für ein Thema zur Verfügung. Eine Kombination von Basismodellen halten Oser et al. (1997, S. 23f.) bei solchen Lerngegenständen für möglich. Sinnvoll wird eine Kombination immer dann, wenn sie der Vereinfachung des Lernens dient. So können mehrere Basismodelle hintereinander (additiv) im Unterricht folgen, oder beispielsweise das Basismodell der *Routinebildung* in das Basismodell des *Konzeptaufbaus* eingeschoben werden. Die Begründung für diese Art der Kombination der Basismodelle sehen Oser und Baeriswyl (2001, S. 1049) darin, die kognitive Last bei den abschließenden Anwendungen auf neue Situationen im Basismodell Konzeptaufbau zu minimieren, indem die Anwendung des neuen Konzepts vorher routinisiert wurde. Basismodelle können dabei nur komplett zwischen zwei einzelne Schritte eines anderen Modells eingefügt werden. Dies soll eine Desorientierung des Lernenden beziehungsweise einen Bruch des Lernwegs verhindern. Entscheidend bei der Umsetzung der BMT ist dabei immer der Brückenschlag vom Lehren zum Lernen, welcher stets mit einem Wechsel von Lehrer- und Schülerperspektive verbunden ist. Dadurch wird eine einseitige Sichtweise der Lehrkräfte auf den Lerninhalt und somit eine Hemmung des Lernprozesses insgesamt vermieden (Oser & Baeriswyl, 2001, S. 1061). Die Steuerung des Lernprozesses mit Hilfe der BMT ist dabei nicht spezifisch für ein Unterrichtsfach formuliert, die BMT verfolgt vielmehr den Anspruch universell anwendbar zu sein.

Modifikation für den Physikunterricht

Will man die BMT auf einzelne Unterrichtsfächer übertragen, so sehen Trendel, Wackermann und Fischer (2007, S. 28) damit die Notwendigkeit einer Anpassung an fachliche Strukturen verbunden. Hierfür schätzen Trendel et al. (2007, S. 28) die BMT als „an manchen Stellen wenig spezifisch“ ein und betrachten eine Weiterentwicklung auf theoretischer Ebene als erforderlich. Diese Weiter-

2. Strukturierung

entwicklung erfolgte im Rahmen zweier Dissertationsprojekte. Die Modifikation der BMT für den Physikunterricht wurde in einem ersten Schritt durch Reyer (2004) und darauf aufbauend von Wackermann (2008) geleistet. Diese Modifizierung umfasste sowohl Überlegungen zur Notwendigkeit spezieller Abläufe innerhalb eines Basismodells, als auch zur Auswahl von für die Physik wesentlicher Basismodelle. Reyer (2004) führt dabei eine Diskussion und Modifikation aller Basismodelle vor physikdidaktischem Hintergrund durch, ohne dabei „die Grundidee der Theorie“ (Reyer, 2004, S. 32) zu berühren, verzichtet bei der Übertragung auf die Physik aber auf zwei Basismodelle. Das Modell *Strategielernen* ordnet Reyer (2004, S. 37) als systematisches Strategie-Lernen im Sinne einer „automatisierten Anwendung des Wissens über das eigene Lernen“ dem *Routinelernen* zu. Elemente des Basismodells *Verhandeln* interpretiert Reyer (2004) als innerhalb geeigneter Sichtstrukturen oder anderer Basismodelle (*Entwicklung von Sozialbeziehungen oder Wertebeziehungen*) erlernbar, weshalb das Modell *Verhandeln* in seiner Adaption der BMT für den Physikunterricht entfällt. Zudem erfasst Reyer (2004) die Häufigkeit der Basismodelle im Unterricht. Als zeitlich dominant erweisen sich *Lernen durch Eigenerfahrung* und *Konzeptaufbau*. „Sehr auffällig ist die geringe Relevanz der anspruchsvollen Lernzieltypen ‚Problemlösen‘ und ‚Konzeptwechsel‘“ (Reyer, 2004, S. 278). Trendel et al. (2007) verwenden in ihrer Studie zur Untersuchung von Bedingungen für erfolgreiche Lernprozesse im Physikunterricht nur drei der von Oser vorgeschlagenen Basismodelle. Neben den nach Reyer (2004) häufig vorkommenden Modellen *Lernen durch Eigenerfahrung* und *Konzeptaufbau* setzen Trendel et al. (2007) noch das Basismodell *Problemlösen* ein, obwohl Reyer (2004, vgl. S. 222) das Basismodell *Hypertextlernen* häufiger im Unterricht identifizieren kann. Die normative Festlegung des *Problemlösens* als eines von drei Standardbasismodellen erfolgt, da Trendel et al. (2007, S. 14) die Fähigkeit komplexe Problem zu lösen als „ein wesentliches Ziel naturwissenschaftlichen Kompetenzerwerbs“ betrachten. Trendel et al. (2007, S. 14) führen an, dass diese drei für den Physikunterricht grundlegenden Basismodelle sich in guter Passung zu den fachlichen Strukturen darbieten und einen großen Teil des Unterrichts abdecken. Des weiteren erweise sich eine Differenzierung des Physikunterrichts hinsichtlich dieser drei unterschiedlichen Basismodelle als sinnvoll. „Betrachtet man die notwendigen Handlungsschritte zu jedem Basismodell fällt auf, dass Lernwege und deren Voraussetzungen sich nicht nur in Details, sondern grundsätzlich unterscheiden können“ (Trendel et al., 2007, S. 14). Als Ergebnis theoretischer Vorüberlegungen zu seiner Studie halten Trendel et al. (2007) zudem fest, dass das Basismodell *Problemlösen* nur dann erfolgreich verlaufen kann, wenn das zum Lösen erforderliche Wissen bereits vorhanden ist. Das Ergebnis beziehungsweise das Lösungskriterium ist bei diesem Basismodell vorgegeben, der Weg dorthin allerdings nicht. Anders verhält sich dies beim *Lernen durch Eigenerfahrung*. Hier ist der von den Lernenden beschrittene Weg während ihrer Erfahrungen oft vorgegeben, die Ergebnisse sind aber anfangs unbestimmt. Zudem sind große Vorkenntnisse nicht notwendigerweise erforderlich für einen erfolgreichen,

erfahrungsbasierten Aufbau des Wissens. Als notwendig erachten Trendel et al. (2007, S. 14f.) die Vorkenntnisse der Lernenden allerdings, liegt wie beim *Konzeptaufbau* die Integration des Neuen in ein bestehendes Wissenssystem im Fokus. Auch bei diesem Basismodell ist das Ziel der Aufbau neuen Wissens, nur ist dieses Wissen anders geartet als bei *Lernen durch Eigenerfahrung*. Wird Wissen aus Erfahrungen aufgebaut, so führen diese nicht notwendigerweise zu einem neuen Konzept, „weil Konzepte im Allgemeinen nicht entdeckt, sondern erfunden werden“ (Trendel et al., 2007, S. 15). Der Sinngehalt eines Konzepts kann dabei dahingehend verstanden werden, als dass dieses Begriffe beschreibt, welche ihre Bedeutung erst vor einem fachlichen Hintergrund erhalten. Der Komplexitätsgrad bzw. das Abstraktionsniveau von derartigen Konzepten kann stark variieren. So zählen Wackermann, Hater, Simic, Pieper und Priemer (2012, S. 11) einfachere Fachbegriffe wie *Strecke* ebenso wie komplexere Begriffe wie *Spannung*, *Kraft* oder *Impuls* zu der Klasse der Konzepte. Beim *Konzeptaufbau* kann auch eine Erweiterung oder Anpassung bestehender Vorstellungen stattfinden. Sind diese Präkonzepte aber sehr stabil und schwer durch andere ersetzbar, ist für Wackermann et al. (2012) ein Konzeptwechsel notwendig. Weshalb Wackermann et al. (2012) die Liste der wesentlichen Basismodelle (vgl. Tabelle 2.2) für den Physikunterricht um das Basismodell des *Konzeptwechsels* erweitern, welches auf dem ursprünglichen Modell der *Entwicklung als Erziehungsziel* von Oser und Baeriswyl basiert.

Zudem fügen Wackermann und Priemer (2012, S. 218) beim ersten Handlungskettenschritt des Basismodells *Lernen durch Eigenerfahrung* die Einführung der Kontextes hinzu, da sich dieser Einschub als „empirisch notwendig“ erwiesen habe. In welcher Form diese empirische Notwendigkeit dabei erkannt oder nachgewiesen wurde, wird allerdings nicht weiter ausgeführt. Insgesamt bilden bei Wackermann und Priemer (2012) damit vier verschiedene Basismodelle die Grundlage der Planung für die Thematisierung physikalischer Inhalte oder Zielsetzungen. Die Unterschiedlichkeit der Zielsetzungen findet sich in der Unterschiedlichkeit der Basismodelle wieder. Dies ermöglicht eine Zuordnung der Basismodelle zur jeweiligen unterrichtlichen Zielsetzung. Erfolgt eine Anwendung bzw. Reorganisation von bereits erworbenen Wissens-elementen zur Lösung eines vorliegenden Problems, zu welchem der Lösungsweg dem Lernenden unbekannt ist, geben Wackermann und Priemer (2012) vor, den Unterricht im Sinne des Basismodells Problemlösen zu sequenzieren. Es ist nicht das Ziel dieses Basismodells, Wissen zu generieren. Sollen Wissens-elemente aufgebaut werden, welche von den Lernenden *gefunden* werden können, so soll die Strukturierung der Stunde mit Hilfe des Basismodells *Lernen durch Eigenerfahrung* erfolgen. Geht es um die Vermittlung *erfundener* Elemente dient bei Wackermann und Priemer (2012) der *Konzeptaufbau* als Strukturvorgabe. „Wenn belastbare Alltagsvorstellungen im Widerspruch zu physikalischen Konzepten stehen oder wenn richtige physikalische Vorstellungen, die im Unterricht beigebracht wurden, ergänzt/ersetzt/erweitert werden müssen“ (Wackermann et al., 2012,

2. Strukturierung

	Lernen durch Eigenerfahrung	Konzeptaufbau	Problemlösen	Konzeptwechsel
1.	0. Einführen des Kontexts ——— 1. Inneres Vorstellen und Planen	Bewusstmachen des Vorwissens	Problem-generierung	Vergegenwärtigung bestehender Konzepte
2.	Handeln im Kontext	Durcharbeiten eines prototypischen Musters	Problem-präzisierung	Disäquilibration „Erschüttern“
3.	Erste Ausdifferenzierung, Reflexion	Darstellen der wesentlichen Merkmale und Prinzipien	Lösungs-vorschläge	Aufkommen des Neuen
4.	Generalisierung der Ergebnisse	Aktiver Umgang mit neuem Konzept	Prüfen der Lösungs-vorschläge	Wichtigkeit des Neuen Sezieren des Alten
5.	Übertragung auf größere Zusammenhänge	Vernetzung mit bekanntem Wissen	Vernetzung, Transfer auf andere Problemklassen	Integration von Neu und Alt

Tabelle 2.2.: Der Ablauf der modifizierten Basismodelle für den Physikunterricht (Wackermann & Priemer, 2012)

S. 7), findet das Basismodell des Konzeptwechsels Anwendung.

Wirkungen der BMT

Wackermann (2008) verwendet in seiner Dissertation die Modifikation durch Reyer (2004) und Trendel et al. (2007) als Grundlage für die Schulung an der Studie teilnehmender Lehrkräfte. Im Rahmen dieser Schulung werden die Lehrkräfte im Umgang mit drei Basismodellen (Lernen durch Eigenerfahrung; Problemlösen; Konzeptaufbau) geschult. Das Basismodell Konzeptwechsel ist zu diesem Zeitpunkt noch nicht in die Modifikation der BMT integriert. Ziel der Arbeit von Wackermann (2008) ist die Überprüfung der Wirksamkeit eines Basismodell-Trainings für Physiklehrer, welche mit Hilfe der Fortbildung befähigt werden sollen Basismodelle theoriekonform in Schulklassen der Jahrgangsstufen 8 bis 13 zu unterrichten. Die Auswirkungen dieses Unterrichts wurden in einem Prä- und einem Posttest sowohl in der Interventions- als auch in der Kontrollgruppe hinsichtlich einer Vielzahl unterrichtlicher Merkmale erhoben. Auf Schülerebene konnte Wackermann (2008, S. 77) im Bereich der Unterrichtswahrnehmung mehrere kleine, signifikante Effekte nachweisen. Die

Einschätzung der Lernenden zum Tempo beim Voranschreiten im Unterricht (Pacing), zur Klarheit und Strukturiertheit, zur Verwendung von Strategien zum Verständnisaufbau und zur Fehlerkultur weisen im Vergleich zur Kontrollgruppe wünschenswertere Kennwerte auf. Aspekte der Schüleremotionen wie Interesse oder Motivation zeigen keine signifikanten Effekte. Allerdings konnte Wackermann (2008, S. 77) die Hypothese „Klassen mit besonders basismodellhaftem Unterricht zeigen auch die größten Effekte“ bestätigen. Eine Untergruppe der Fortbildungslehrkräfte (IG Plus) setzte die BMT besonders exakt um und Schüler der IG Plus Lehrkräfte bewerteten den Unterricht auf allen Bereichen der Unterrichtswahrnehmung signifikant besser. Die Stärke der Effekte ist dabei größer als in der Gruppe aller Fortbildungslehrkräfte und liegt im kleinen bis mittleren Bereich. Kleine signifikante Effekte zeigen sich bei Schülern der IG Plus auch hinsichtlich des Fach- und Sachinteresses Physik und der intrinsischen Motivation. Eine möglichst theoriekonforme unterrichtliche Realisation der BMT erweist sich folglich als besonders wirkungsvoll. Wackermann (2008, S. 84) folgert daraus, dass die Studie Hinweise darauf gibt, „dass die Basismodelle des Lehrens und Lernens den Physikunterricht verbessern können“ (Wackermann, 2008, S. 84). Zudem betont Wackermann (2008, S. 84), dass sich die BMT als für Lehrende erlernbar und im Unterricht umsetzbar erwiesen habe.

Zander, Krabbe und Fischer (2013) überarbeiteten das Design der Fortbildung für eine weiterführenden Studie, um mögliche Effekte der BMT auch auf das Fachwissen der Lernenden erfassen zu können. Zentrales Element der Fortbildung blieb weiterhin die Sequenzierung des Unterrichts nach Vorgabe der BMT und damit die Fokussierung auf Lernprozesse der Schülerinnen und Schüler. Lernende aus achten Klassen von 14 verschiedenen Gymnasien in Nordrhein-Westfalen wurden dazu einer Vergleichsgruppe beziehungsweise einer Interventionsgruppe zugeordnet. Zugehörig zur Interventionsgruppe sind alle Schüler, deren Lehrer eine Fortbildung zur Theorie der Basismodelle erhielten und diese im Unterricht umsetzten. Der Unterricht der Schüler in der Vergleichsgruppe wird nicht konkreter angegeben, die Lehrkräfte dieser Schüler erhielten keine Fortbildung und wurden nicht videographiert, weshalb Zander, Krabbe und Fischer (2015) diese Gruppe in einem späteren Artikel auch nicht mehr als Kontrollgruppe kennzeichnen. Inhaltliche Grundlage beider Gruppen bildet die lehrplankonforme Umsetzung des Themas *Mechanik*, zu welchem mit einem Fachwissenstest am Anfang und Ende des Schuljahres der Wissenszuwachs der Lernenden ermittelt wurde. Die Interventionsgruppe zeigt dabei einen deutlich größeren Lernzuwachs als die Vergleichsgruppe. Die zugehörige Effektstärke wird mit $d = 0.32$ (kleiner bis mittlerer Effekt) angegeben. Zander et al. (2013) sprechen trotzdem von einem großen Effekt, da die Interventionsgruppe einen um über fünfzig Prozent größeren Lernzuwachs aufweist. Die untersuchten Gruppen unterscheiden sich zudem in den Vorbedingungen für Lernerfolg. Der Lernerfolg der teilnehmenden Schüler hängt in der Vergleichsgruppe stärker

2. Strukturierung

vom Vorwissen und den kognitiven Fähigkeiten ab, bezüglich des Interesses am Fach Physik kann allerdings kein Unterschied im Vor- und Nachtest festgestellt werden. Insgesamt weist bei Zander et al. (2013) die Vergleichsgruppe einen höheren Kennwert für das Fachinteresse als die Interventionsgruppe auf. Damit kann die Annahme von Wackermann (2008), dass eine förderliche Wirkung von basismodellkonformen Unterricht auf das Interesse existiert, nicht bestätigt werden. Bei Wackermann (2008) erweist sich dieser angenommene Einfluss auch nur in der Teilgruppe IG Plus als statistisch signifikanter kleiner Effekt. Eine vergleichbare Unterteilung in Teilgruppen wird bei Zander et al. (2013) hingegen nicht vorgenommen, wäre aber für eine genaue Untersuchung dieses möglichen Einflusses förderlich gewesen. In einer zusätzlichen Auswertung der Daten des Wissenstests mittels einer latenten Klassenanalyse konnten Zander et al. (2015, S. 391f) zudem nachweisen, dass insbesondere die im Vortest schwachen sowie durchschnittlichen Schülerinnen und Schüler hinsichtlich des Lernerfolgs von der durchgeführten Lehrerfortbildung und dem damit verbundenen basismodellkonformen Unterricht profitieren konnten. In der Gruppe der im Vortest stärkeren Schüler traten keine erkennbaren Unterschiede zwischen den Untersuchungsgruppen auf.

Sowohl die Studie von Wackermann (2008) als auch die von Zander et al. (2013) basieren damit auf der Schulung von Lehrkräften im Rahmen einer Fortbildung. Die Ergebnisse beider Studien lassen sich deshalb nicht ohne weiteres nur auf die Unterschiedlichkeit in der Sequenzierung des Unterrichts nach der BMT zurückführen. Die empirisch gefundenen Effekte können unter anderem auch durch die Fortbildung an sich oder durch das Video-Coaching der Lehrer der Interventionsgruppen verursacht sein. Wackermann, Trendel und Fischer (2010, S. 981f.) vermuten aufgrund der Unterschiedlichkeit der Bewertung des Unterrichts der Lehrkräfte der IG Plus im Vergleich zu den verbleibenden ebenfalls fortgebildeten Lehrkräften, dass ursächlich die Orientierung an den Basismodellen für die Ausprägung der Ergebnisse der Studie von Wackermann (2008) verantwortlich ist. Auf die Studie von Zander et al. (2013) ist diese Aussage dabei nicht übertragbar, da keine entsprechende Gruppeneinteilung der teilnehmenden Lehrkräfte vorgenommen wird.

Grundlegend für die Dissertation von Draxler (2005) ist die Modifikation der Basismodelle nach Reyer. Damit verfügt Draxler (2005) über ein Repertoire von zehn Basismodellen zur Ausgestaltung und Durchführung des Unterrichts der Interventionsgruppe. Die Planung des Unterrichts der Kontrollgruppe gestaltet sich inhaltlich und zeitlich analog, die Sequenzierung des Ablaufs und dessen Ausführung erfolgt durch eine betreuende Lehrkraft unabhängig und ohne Kenntnis der BMT. Bei den eingesetzten Wissenstests (Prä- und Posttest) zur Elektrizitätslehre und der Mechanik weist die Interventionsgruppe in beiden Fällen einen etwas höheren Mittelwert als die Kontrollgruppe auf. Signifikante Unterschiede der Gruppenmittelwerte ergibt die Auswertung der Posttests allerdings nicht. Der Autor vermutet die Ursache der fehlenden Signifikanz der

Ergebnisse in der geringen Probandenzahl von insgesamt 27 Personen eines Physikkurses der Stufe 11. Dass Draxler (2005) die vorab formulierte Hypothese: „Basismodellorientierter Physikunterricht wirkt sich positiv auf die Leistung von Schülerinnen und Schülern aus“ (Draxler, 2005, S. 208) dennoch als bestätigt erachtet, liegt an Gruppenunterschieden im Vortest. Im Vortest weisen die Lernenden der Interventionsgruppe zu Beginn der Untersuchung im Mittel einen signifikant schlechteren Kennwert als die Kontrollgruppenschüler auf. Der Lernzuwachs ist folglich in der Interventionsgruppe größer. Eine Generalisierung der Ergebnisse wird aufgrund der geringen Stichprobengröße aber nur als bedingt möglich erachtet und eine umfangreichere Interventionsstudie angeregt. Unklar verbleibt auch der Einfluss der unterrichtenden Personen in den Teilgruppen der Studien. Eine erfahrene Lehrkraft unterrichtet die Kontrollgruppe, Draxler als zum Zeitpunkt der Studie unterrichtlich unerfahrene Person (Draxler, 2005, S.110) die Interventionsgruppe. Letztlich deutet sich eine lernförderliche Wirkung der Strukturierung von Physikunterricht mit Hilfe der Basismodelle aber auch in dieser Studie an.

Diese lernförderliche Wirkung deutet sich auch in einer Studie von Ohle, Fischer und Kauertz (2011) in der Primarstufe an. Die Untersuchung im Rahmen der sogenannten PLUS-Studie (Professionswissen von Lehrkräften, naturwissenschaftlicher Unterricht und Zielerreichung seitens der Schüler im Übergang von der Primar- zur Sekundarstufe) erfolgte zur Erfassung des Einflusses des physikalischen Fachwissens von Primarstufenlehrkräften auf die Unterrichtsgestaltung und die Schülerleistung. Ohle et al. (2011, S. 357) konnten als Ergebnis der videobasierten Analyse von Unterrichtsstunden von 110 Lehrkräften feststellen, dass das Fachwissen (CK) einen positiven Einfluss auf die Leistungen von Schülern hat. Ohle et al. (2011, S. 357) konnten nachweisen, dass dieser Einfluss des Fachwissens (CK) moderiert wird durch das Ausmaß, zu welchem die jeweils beobachtete Unterrichtssequenz mit einer basismodellkonformen Lernprozessesequenzierung übereinstimmt. Etwaige Übereinstimmungen entstammen dabei keiner expliziten Schulung teilnehmender Lehrkräfte, sondern können nur in einer „intuitiven und zufälligen Umsetzung“ der Basismodelltheorie gefunden werden (Ohle et al., 2011, S. 364). Allgemein steuert eine Moderatorvariable die Richtung bzw. die Stärke eines Zusammenhangs zwischen abhängiger und unabhängiger Variable (Frazier et al., 2004, S. 116). Ohle et al. (2011, S. 357) formulieren deshalb folgende Aussage als Fazit: „Die Sequenzierung, die das Einhalten Lehrzielspezifischer Schrittfolgen im Unterricht fordert (Oser & Baeriswyl, 2001) ist neben dem Lehrer- CK ein positiver Prädiktor für Schülerleistung“.

Wagner (1999) führt im Rahmen ihrer Dissertation eine empirische Überprüfung der Wirksamkeit von Unterricht nach den Vorgaben der BMT außerhalb des Physikunterrichts durch. Wagner (1999) untersuchte Gymnasialunterricht zum Fach Deutsch. Ihre Untersuchung erfolgte innerhalb des Gesamtprojektes zu der von Oser et al. (1997) betreuten mehrjährigen Studie an schweizerischen

2. Strukturierung

Schulen. Wagner (1999) legt dabei für die Konzeption der Lernsequenzen für ihre Interventionsstudie die von Oser und Baeriswyl formulierten zwölf Basismodelle zugrunde. Die von ihr verwendete Kontrollgruppe wird lediglich als „herkömmlicher Unterricht“ (Wagner, 1999, S. 3) beschrieben und nicht näher präzisiert. „Der Basismodell-Unterricht hat den Schüler/n/innen, die unter dieser Bedingung unterrichtet wurden, nicht nur zu einer erleichterten Wahrnehmung der Handlungsschritte des Unterrichts und ihrer eigenen Lernschritte verholfen [...], sondern auch dazu geführt, dass sie einen signifikant größeren Lernerfolg hatten, als diejenigen der Vergleichsgruppe“ (Wagner, 1999, S. 218). Letztgenannte Aussage kann streng genommen nur für eine der zwei Teilstudien getroffen werden. In ihrer ersten Studie erhalten drei Klassen Unterricht auf Grundlage der BMT (Gruppe 1), drei weitere Klassen werden von anderen Lehrkräften ohne Kenntnis der BMT unterrichtet (*herkömmlicher Unterricht*, Gruppe 2). Die Schüler der Gruppe 1 weisen einen höchstsignifikant größeren Lernerfolg als diejenigen der Gruppe 2 auf. Eine Stärke dieses Effekts ist nicht angegeben, über die tatsächliche Bedeutsamkeit für die Unterrichtsqualität kann demnach keine Aussage getroffen werden. Zudem wird der dargebotene Deutschunterricht in der Interventionsstudie von den Schülern als nachvollziehbar und gut strukturiert empfunden. In einer nachfolgenden Replikationsstudie mit anderen Sachinhalten werden Schüler der Gruppe 2 nach der BMT geschult, über die Sequenzierung des Unterrichts von Gruppe 1 werden von Seiten der Autorin keine Forderungen gestellt. Mit dem Wechsel der Lernenden zwischen Interventions- und Kontrollgruppe geht auch ein Wechsel der Lehrkräfte zwischen den Gruppen einher. Im Gegensatz zur ersten Studie stehen bei der Replikation Lehrkräften beider Gruppen die theoretischen Vorgaben der BMT zur Verfügung. Eine Orientierung an verschiedenen Zieltypen der BMT erfolgte nach Aussage der beteiligten Lehrkräfte nun auch in der Kontrollgruppe, eine detaillierte Planung mit Hilfe der Handlungskettenschritte der Basismodelle jedoch nicht. Wagner (1999, S. 157ff.) selbst stuft die Lernbedingungen in beiden Teilstudien als so unterschiedlich ein, dass Ergebnisse der zweiten Studie nur bedingt zur Replikation verwendet werden können. In der Nachfolgestudie zeigen sich auch keine signifikanten Unterschiede hinsichtlich des Lernerfolgs. Berücksichtigt man die nur eingeschränkt aussagekräftige Replikationsstudie nicht, so zeigt sich eine lernförderliche Wirkung der BMT auch im Fach Deutsch, wobei diese Wirkung eindeutig nicht zufällig, in ihrer Ausprägung aber unbekannt ist. Die untersuchten affektiven Komponenten (Lernsicherheit, Lernzufriedenheit und Unterrichtsklima) blieben hingegen weitgehend unbeeinflusst (Wagner, 1999, S. 224).

Insgesamt zeigt sich in den dargestellten Studien eine einheitliche Einschätzung hinsichtlich des positiven Einflusses der Strukturvorgabe der Basismodelltheorie auf den Lernerfolg der unterrichteten Schüler. Diese beurteilen basismodellkonformen Unterricht weiterhin positiv in mehreren qualitativ relevanten Merkmalen wie der Strukturierung. Die Ebene der affektiven oder motivationa-

len Belange der Lernenden scheint hingegen nicht oder nur in geringem Maße beeinflusst durch eine entsprechende Sequenzierung des Lernangebots zu sein.

2.3.5. Beziehungen zwischen den Strukturvorgaben

Verschiedene Strukturvorgaben zur Sequenzierung von Unterricht und deren Auswirkungen sind im vorangegangenen Abschnitt dargestellt. Diese Auflistung ist dabei keineswegs vollständig. Eine Auswahl erfolgte hinsichtlich der Relevanz für diese Arbeit und dem Grad der Verbreitung, Dokumentation und empirischen Erforschung der Strukturvorgaben. Deren theoretische Grundkonzeptionen und die daraus resultierenden Unterrichtsverläufe variieren in unterschiedlichem Maße, weisen aber auch gemeinsame Merkmale und Prozesse auf. Zusammenhänge zwischen den einzelnen Strukturvorgaben werden im folgenden Abschnitt näher beleuchtet. Ausgangspunkt der Bezugnahme bilden das FeU und die BMT aufgrund derer Verwendung als theoretische Grundlage zur Konzeption der Unterrichtsstunden zur Durchführung der Interventionsstudie.

Beziehungen zwischen dem FeU und den aufgeführten Vorgaben aus dem angloamerikanischen Sprachraum

Das Grundkonzept des FeU wird geprägt durch eine Orientierung am Prozess der naturwissenschaftlichen Erkenntnisgewinnung und die damit verbundene Prägung des Unterrichtsablaufs durch Experimente. Eine ähnliche Ausrichtung weisen auch die drei Strukturvorgaben zum Inquiry-Learning auf. Welche sich von dieser Grundausrichtung ebenso wie das FeU eine Steigerung der Motivation, eine hohe Eigenaktivität und Möglichkeiten für Erfolgserlebnisse der Lernenden versprechen. Das FeU stellt dabei ebenso wie das Problem Based Learning das Lernen an Problemsituationen in den Vordergrund. Auf Basis dieser Situationen sollen beim FeU Erkenntnisse auf Grundlage des Vorwissens durch die Lernenden selbst entdeckt werden. Für Bruner (1961) stellt dieser Aspekt wiederum auch den Mittelpunkt des Discovery Learnings dar. In Grundzügen lässt sich dieser Fokus auf das Entdecken der Lerninhalte auch beim Problem Based Learning und dem 5E-Modell wieder finden. In der nachfolgenden Unterrichtsphase soll beim FeU aus den experimentell gewonnenen Erkenntnissen die zu lernenden Konzepte und Inhalte Schritt für Schritt in einer Abstraktionsphase entwickelt und dadurch hervorgehoben werden. Das Herausarbeiten der Theorie aus den experimentellen Daten findet beim 5E-Modell in der Phase der *explanation* funktionell sehr ähnlich statt. Der kleinschrittige Ablauf der Konstruktion der Theorie erinnert dabei aber an die Präsentation der Konzepte bei der Direct Instruction, welche in dieser Phase auch hinsichtlich der Rolle der Lehrkraft vergleichbar zum FeU erscheint. Die Aufgabe der Lehrkraft ist beim FeU variabel bezüglich der einzelnen Unterrichtsphasen. Während der Überlegungen zur Problemlösung und der Durchführung der

2. Strukturierung

Lösungsvorschläge begleitet und betreut die Lehrkraft die Lernenden auf deren Erkenntnisweg, entsprechend der Lehrerrolle beim Problem Based Learning oder Discovery Learning. Die Phasen der Abstraktion und Wissenssicherung erfordern wiederum eine stärkere Lenkung, welche nicht wie beim Problem Based Learning oder Discovery Learning nur im Bedarfsfall, sondern beim FeU im Normalfall so erfolgen soll. Eine ähnliche Bandbreite der Lehrerrolle findet sich ebenfalls beim 5E-Modell.

Dieses Modell weist auch grundsätzliche Ähnlichkeiten bei der Vorgabe der unterrichtlichen Struktur zum induktiven Zweig des FeU auf. Sowohl das FeU als auch das 5E-Modell beginnen mit einer Stufe der Präsentation des Kontexts des Lerninhaltes. Ziel dieser Stufe ist bei beiden Verfahren, die Lernenden mit der Lerngelegenheit vertraut zu machen und so eine Basis für die nachfolgende experimentelle Auseinandersetzung zu schaffen. Ebenfalls vergleichbar werden anschließend aus den experimentell resultierenden Ergebnissen die zu lernenden Konzepte und Inhalte herausgearbeitet. Diese werden nachfolgend bei beiden Strukturvorgaben in neuen Situationen angewendet, worauf abschließend eine Stufe der Rückmeldung erfolgt. Beim FeU erfährt die Lehrkraft während der Lernzielkontrolle den Wissensstand der Lernenden, wohingegen beim 5E-Modell der Fokus auf Hinweisen über die Korrektheit der erworbenen Kenntnis von Seiten der Lehrkraft an die Lernenden liegt. Die Richtung der in den Vorgaben beschriebenen Rückmeldung ist damit unterschiedlich. Unterschiede existieren auch bei der Anwendung des Erlernten. Legt das 5E-Modell bei der Anwendung Wert auf einen Austausch über das Erlernte zwischen den Lernenden, fokussiert das FeU auf eine fächerübergreifende Vernetzung. Zudem liegt beim FeU der didaktische Schwerpunkt auch auf der Planung und Vorbesprechung der durchzuführenden Handlungen sowie einer Problematisierung des Sachgegenstandes. Ersteres findet beim 5E-Modell nach Bybee (1997) keine explizite Erwähnung, letzteres ist ebenfalls möglich, stellt einen alternativen Beginn der Sequenzierung, aber im Gegensatz zum FeU keine grundlegende Komponente des Verfahrens dar. Die Ähnlichkeiten bei der unterrichtlichen Strukturierung zwischen den beiden Vorgaben sind insgesamt erkennbar, die beiden Verfahren differieren allerdings insbesondere bei der zeitlichen Gewichtung und Ausgestaltung einzelner Phasen.

Größere Unterschiede zum induktiven Zweig des FeU zeigen sich in den Abläufen zum Problem Based Learning. Zwar verlaufen das FeU und das Problem Based Learning anfänglich in guter Passung, die Verarbeitung der aus den Experimenten oder Erfahrungen der Lernenden gewonnenen Ergebnisse erfolgt aber unterschiedlich. Wird im FeU gemeinsam mit der Lehrkraft das Wesentliche abstrahiert, folgt zumindest im Problem Based Learning Cycle nach Hmelo-Silver (2004) eine Phase des selbstgesteuerten Lernens, in welcher die Lernenden die erkannten theoretischen Lücken eigenständig schließen sollen. Als gemeinsamen Abschnitt weisen beide Verfahren eine an die Erkenntnisgewinnung anschließende Anwendung des neuen Wissens auf. Zentrales Element beider

Verfahren ist die Problemorientierung, wobei die Problemsituation im FeU im Gegensatz zum Problem Based Learning generell mehr Informationen enthalten darf als zur Lösung minimal notwendig.

Die Denkstufen der Überlegungen zur Problemlösung und der Durchführung der Lösungsvorschläge des induktiven Zweigs des FeU stimmen ebenfalls in wesentlichen Zügen mit dem ersten beiden Phasen des dreiteiligen Zyklus des Discovery Learnings überein, allerdings erscheinen die Vorgaben des Discovery Learnings weniger umfassend für die Gestaltung des Unterrichts zu sein. So fehlt in der Grundstruktur des Discovery Learnings eine Phase der Motivation oder Problematisierung des Inhalts ebenso wie eine anschließende Anwendung oder Wissenssicherung der entdeckten Inhalte und Zusammenhänge.

Eine Entdeckung von Wissenselementen durch die Lernenden ist nicht Ziel eines Ablaufs nach Vorgaben der Direct Instruction. Folglich ist der Verlauf der Stunde auch nicht geprägt durch lange Phasen der Eigentätigkeit der Lernenden beim Aufbau des Wissens und unterscheidet sich damit wesentlich von der Strukturvorgabe des induktiven Zweigs des FeU. Einzig die gestufte Abstraktion zur Ausschärfung der Konzepte und deren anschließende Anwendung stellen gemeinsame Phasenverläufe der beiden Vorgaben dar.

Der induktive Zweig des FeU stimmt damit in wesentlichen Grundzügen und Abläufen mit dem 5E-Modell überein. Integriert in seine Struktur aber auch eine starke Problemfokussierung und damit Elemente des Problem Based Learnings, sowie eine entdeckend-entwickelnde Gewinnung der Erkenntnisse als Element des Discovery Learnings. Die schrittweise Abstraktion erinnert an den Wissensaufbau der Direct Instruction. Insgesamt bündelt das FeU wesentliche Elemente verschiedener Strukturvorgaben in einem Ansatz, weist aber auch charakteristische Eigenschaften auf. Gerade die Gestaltung des Stufen- und Phasenverlaufs in Entsprechung des formalen Ablaufs der Denkprozesse der Schüler, die sehr detaillierte Strukturierung mit insgesamt 5 Denkstufen und 15 Denkphasen, sowie Stellenwert und Einübung der Abstraktion grenzen das FeU von Direct Instruction, Discovery Learning, Problem Based Learning und dem 5E-Modell ab. Ohne die Denkstufe der Abstraktion gestaltet sich der deduktive Zweig des FeU in großer Passung zum Ablauf des Problem Based Learnings, ohne aber Methoden oder Sozialformen wie das selbstgesteuerte Lernen im Problem-based Learning Cycle festzulegen.

Beziehungen zwischen dem FeU und der BMT

Vergleicht man die Grundlagen der beiden Strukturvorgaben, so fordern weder das FeU noch die BMT bestimmte Methoden oder Sozialformen bei der Umsetzung der jeweiligen strukturellen Vorgaben ein. Beide Strukturvorgaben geben feste Schritte im Lernprozess vor, gewähren aber Spielraum in der methodischen Ausgestaltung des Unterrichts. Die BMT ist dabei noch stärker durch das Zusammenspiel aus vorgegebenen Lernschrittfolgen und Methodenfreiheit geprägt. Die Methoden dienen als Werkzeug, um die vorgeschriebenen

2. Strukturierung

Handlungskettenschritte bestmöglich umzusetzen. Keinesfalls soll eine methodische Entscheidung den schrittweisen Ablauf der Lernvorgänge verändern. Das FeU gibt keine Regelung dieses Zusammenspiels vor, legt aber ebenfalls den Ablauf durch die Vorgabe der Denkstufen fest. Diese sind wie bei der BMT an jeweils vermuteten Denk- und Lernschritten der Lernenden orientiert. Obwohl von beiden Verfahren die Lernvorgänge als Basis zur Formulierung der Lehrvorgänge verwendet werden, erfolgt deren Ausgestaltung durch die Strukturvorgaben nicht konform, sondern weist nur in Teilen gleiche Phasierungen auf. Diese Verschiedenartigkeit der vorgegebenen Lehrschritte resultiert dabei auch aus der Unterscheidung mehrerer Lehrzieltypen bei der BMT. Für jeden der vier differenzierten Zieltypen sind unterschiedliche Abläufe vorgegeben, um den angenommenen Lernvorgängen im Lernenden optimal zu entsprechen. Auch beim FeU findet eine Anpassung des Lehrwegs an unterschiedliche unterrichtliche Voraussetzungen statt (*Induktiver* bzw. *Deduktiver Zweig*), die unterrichtlichen Abläufe differieren dabei aber in geringerem Ausmaß und auch die Anzahl an berücksichtigten unterschiedlichen Ausgangssituationen ist geringer. Die BMT ermöglicht zusätzlich eine im Vergleich zum FeU flexiblere Verwendung von Experimenten im Unterricht. Experimente können in einer Vielzahl an verschiedenen unterrichtlichen Phasen mit wiederum verschiedenen Funktionen eingesetzt werden (vgl. sechs kognitive Funktionen bei Wackermann & Priemer, 2012, S. 218f.), da es sich bei der Ausgestaltung der Lehrprozesse nicht primär am Prozess der wissenschaftlichen Erkenntnisgewinnung orientiert. Dies ist aber ein Grundprinzip des FeU, welches eine Ähnlichkeit der Erkenntnisgewinnung bei Lernenden und Wissenschaftlern annimmt. Die daraus resultierende nahezu feste Positionierung und je nach Zweig auch feste Funktion des Hauptexperiments (induktiver Zweig: weiterführendes Experiment; deduktiver Zweig: Bestätigungsexperiment) im FeU wird in der BMT grundsätzlich variabler umgesetzt, obgleich verschiedenen Basismodellen auch spezifische Stellungen und Funktionen des Experiments zugeschrieben werden können (Wackermann & Priemer, 2012). Die im FeU ebenso fest geforderte Problematisierung des Sachgegenstandes, tritt bei der BMT nur im Basismodell *Problemlösen* auf.

Vergleicht man die einzelnen strukturellen Vorgaben der FeU bzw. der BMT zum unterrichtlichen Ablauf (Zweige im FeU vs. Basismodelle bei der BMT), so bedingt der *deduktive Zweig* des FeU vergleichbare Lehrabläufe wie das Basismodell *Problemlösen* der BMT. Allerdings unterscheiden sich die Vorgaben über den Verlauf im Anschluss an die abgeschlossene Problemlösung. Kann diese beim FeU weitere Anwendungsbeispiele, eine Wiederholung der Lösungsergebnisse und gegebenenfalls eine Lernzielkontrolle erhalten, kann diese aber auch ganz entfallen (Schmidkunz & Lindemann, 1992, S. 34). Die BMT fordert eine anschließende Phase explizit ein. Diese dient der Vernetzung und des Transfers auf andere Problemklassen, durch die Anwendung des Lösungsweges (nicht des Lösungsergebnisses) bei anderen aber ähnlichen Problemen (Wackermann, 2008,

S. 144). Macht die Problemsituation aber den Erwerb neuen Wissens notwendig, findet der *induktive Zweig* des FeU Anwendung. Neben weiteren Unterschieden zwischen den beiden Zweigen im FeU beinhaltet der induktive Zweig zusätzlich die Diskussion und die Abstraktion der gewonnenen Ergebnisse. Diese Denkstufen des induktiven Zweiges des FeU ähneln Handlungskettenschritten im Basismodell *Lernen durch Eigenerfahrung*, welches auch insgesamt Parallelen zum induktiven Zweig des FeU aufweist. Die Einführung des Kontextes bei Lernen durch Eigenerfahrung fordert zwar keine Problematisierung der Lern Gelegenheit, soll aber vergleichbar zur Denkstufe der Problemgewinnung beim FeU die Grundlagen für den anschließenden Erfahrungsprozess bereit stellen. Beide Verfahren geben weiterhin eine Planungsphase der Handlungen vor deren Durchführung und der anschließender Diskussion der erhaltenen Ergebnisse und Erfahrungswege vor. Der gestufte Aufbau der Abstraktion im FeU gestaltet die Phase der Generalisierung der Ergebnisse deutlich detaillierter als dies in der BMT angegeben wird. Beim FeU werden in dieser Phase nicht nur experimentelle Ergebnisse generalisiert, sondern zusätzlich auch weiterführende Inhalte aus den vorliegenden Ergebnissen abstrahiert. Zum Aufbau weiterführenden Inhalte, welche die Lernenden nicht direkt aus der Erfahrung aufbauen können, sieht die BMT hingegen andere unterrichtliche Abläufe vor (Basismodell Konzeptaufbau). Beide Vorgaben fordern in der jeweils fünften Stufe eine Anwendung und Transfer des Gelernten. Bei Lernen durch Eigenerfahrung soll eine Dekontextualisierung der Lernkonsequenz erfolgen, indem geprüft wird, inwiefern das Lernergebnis auch in anderen meist physikalischen Kontexten nachweislich richtige Vorhersagen erlaubt (Wackermann, 2008, S. 142), wohingegen beim FeU die Vernetzung überfachlich durch den Aufbau von Querverbindungen zu anderen Fächern oder Phänomenen des Alltags und eine Phase der Wiederholung erfolgen soll (Schmidkunz & Lindemann, 1992, S. 31). Schematisch betrachtet kommen viele Phasen des induktiven Zweiges mit denen des Basismodells Lernen durch Eigenerfahrung zur Deckung, unterscheiden sich dabei aber in deren Binnenstruktur.

Für bestimmte Inhalte können damit die Vorgaben der BMT bzw. des FeU zu einer Konzeption von Unterricht führen, welche sich in vielen Phasen ähnlich gestaltet. Die Vorgaben der Basismodelle Lernen durch Eigenerfahrung bzw. Problemlösen decken sich in weiten Teilen mit denen des induktiven bzw. deduktiven Zweiges. Die BMT gibt allerdings abhängig von den zu behandelnden Inhalten noch weitere, verschiedene unterrichtliche Strukturen neben Problemlösen und Lernen durch Eigenerfahrung an. Lernen durch Eigenerfahrung gestaltet den Lehrverlauf zu Inhalten, welche explizit von den Lernenden *gefunden* werden können. Hier geht es um das Finden von Regeln, Zusammenhängen oder Gesetzmäßigkeiten, nicht jedoch um das (Er-)Finden von Konzepten. Diese Einschränkung wird im FeU nicht getroffen und ein gemeinsames Nacherfinden oder Nachentdecken ist in der Abstraktionsphase auch von Konzepten möglich. Konzepte werden in der BMT hingegen anhand

2. Strukturierung

eines anderen Basismodells unterrichtet. Die Instruktionsvorgabe gestaltet sich beim Basismodell des Konzeptaufbaus im Sinne einer Darstellung eines Konzepts. Dies erfordert allerdings andere unterrichtliche Abläufe aus Sicht der BMT, als diese im induktiven Zweig des FeU vorgegeben werden. Das FeU liefert auch keine Vorgaben, wie Unterricht gestaltet werden kann, der auf einen Konzeptwechsel oder eine Überarbeitung bestehender Konzepte zielt. Das Basismodell Konzeptwechsel fokussiert aber eben auf diese Zielsetzung und weist dabei kaum Gemeinsamkeiten zu einem der Zweige des FeU bezüglich der Gestaltung von Unterricht auf.

Werden also ausschließlich findbare Inhalte oder Problemsituationen zu bereits bekannten Inhalten thematisiert, erscheinen die Abweichungen in den Vorgaben zwischen beiden Verfahren gering. Sind im Lerngegenstand aber auch oder nur erfundene Aspekte (Konzepte) enthalten oder sollen Konzepte gewechselt oder überarbeitet werden, führen beide Strukturvorgaben zu unterschiedlichen Strukturierungen von Unterricht.

Beziehungen zwischen der BMT und den aufgeführten Vorgaben aus dem angloamerikanischen Sprachraum

Vergleicht man die Vorgaben der BMT mit denen der Direct Instruction, des Problem Based Learnings, des Discovery Learnings und auch des 5E-Modells treten erneut einzelne Überschneidungen auf. Diese Überschneidungen treten dabei nicht auf genereller, die ganze BMT betreffender Basis auf, vielmehr sind es einzelne Basismodelle, welche Ähnlichkeiten zu den Strukturvorgaben aus dem angloamerikanischen Sprachraum aufweisen. Weswegen im Folgenden Parallelen aber auch Unterschiede einzelner Basismodelle mit den zu vergleichenden Strukturvorgaben dargestellt werden.

Bei *Lernen durch Eigenerfahrung* unterscheiden Oser und Baeriswyl (2001) ursprünglich ein Basismodell 1A (*Learning through experience*) und ein Basismodell 1B (*Learning through discovery*). Letzteres führt beispielsweise zu einer Möglichkeit des Wiederentdeckens von Zusammenhängen, Theorien oder Objekten, wobei das Ziel des Lernprozesses von Beginn an bekannt ist. Der Lernprozess beim erfahrungsbasiertem Lernen (Basismodell 1A) bezieht sich ganz allgemein auf ein ziel-orientiertes Lernen im Kontext und das entdeckende Lernen (Basismodell 1B) stellt einen Spezialfall von 1A dar (Oser & Baeriswyl, 2001, S. 1051f.). Einzelne Abläufe beim entdeckenden Lernen können sich in einem Kreisprozess wiederholen, insgesamt ergeben beide Varianten des ersten Basismodells aber „fast identische Handlungsketten“ (Reyer, 2004, S. 33), weshalb Reyser im Rahmen dessen Modifikation die Basismodelle 1A und 1B zu einem einheitlichen Lernskript zusammenfasst. In diesem Lernskript sind damit aber wesentliche Elemente des Discovery Learnings als *entdeckender* Lernaspekt enthalten. Betrachtet man zudem den Fokus auf dem erfahrungsbasierten Lernen vor physikalischem Hintergrund, scheint die Arbeit mit Experimenten und mit den daraus resultierenden Erfahrungen zentraler Aspekt von Lernen durch Ei-

generierung zu sein. Das gilt auch für das 5E-Modell. Dieses fünfstufige Modell fordert zudem ähnliche unterrichtliche Abläufe wie das Basismodell Lernen durch Eigenerfahrung (Version der Modifikation von Wackermann, 2008; Vorlage der Konzeption der Interventionen in dieser Studie). Sowohl das 5E-Modell als auch Lernen durch Eigenerfahrung geben eine Einführung des Kontextes zu Beginn des Lernprozesses vor, welche den Sinn der nachfolgenden Auseinandersetzung mit dem Sachgegenstand erkenntlich werden lässt. In beiden Modellen sollen die Lernenden den Lerngegenstand mit Experimenten explorieren und Erfahrungen damit sammeln (Wackermann & Priemer, 2012; Bybee, 1997). Die Ergebnisse dieser Experimente werden übereinstimmend zwischen beiden Vorgaben zunächst von den Lernenden präsentiert und anschließend zu Regeln oder Zusammenhängen generalisiert. Das Basismodell fokussiert dabei stärker auf der Arbeit mit den gewonnenen Daten, das 5E-Modell ermöglicht an dieser Stelle auch eine Präsentation der wissenschaftlich korrekten Erklärungen für die Erfahrungen der Lernenden. Sind für diese die erarbeiteten Regelmäßigkeiten oder Erklärungen verfügbar, soll in beiden Verfahren eine Anwendung auf weiterführende oder neue Kontexte erfolgen. Das 5E-Modell gibt nachfolgend zusätzlich eine Phase der Evaluation vor, wohingegen *Lernen durch Eigenerfahrung* zwischen der Einführung des Kontextes und der Exploration großen Wert auf eine sorgfältige Planung vor der Durchführung der meist experimentellen Handlungen legt. Überschneidungen in den Vorgaben zu unterrichtlichen Abläufen des Basismodells mit dem Discovery Learning finden sich vor allem in der Planung der Experimente und deren Durchführung, reichen aber weniger weit als die Parallelen zum 5E-Modell. Ergänzt man die Abläufe des 5E-Modells um die Planungsphase beim Discovery Learning ergibt sich insgesamt eine noch größere Deckung der Struktur des *Lernens durch Eigenerfahrung* zu dieser Kombination aus Discovery Learning und 5E-Modell.

Unklar erscheint aber, ob die innerhalb der BMT theoretisch vorgesehene Anwendung des Basismodells *Lernens durch Eigenerfahrung* nur auf findbare Lerninhalte mit den Anwendungsgebieten des Discovery Learnings oder des 5E-Modells übereinstimmen. Sowohl das Discovery Learning als auch das 5E-Modell weisen zwar grundsätzlich eine lernförderliche Wirkung auf, allerdings deuten die Aussagen zu deren Wirksamkeit (vgl. S. 37 und S. 38) ebenfalls nicht auf eine generelle Eignung für alle Sachinhalte hin. Ob allerdings genau die inhaltliche Einschränkung, welche in der Adaption der BMT für *Lernen durch Eigenerfahrung* getroffen wird, auch auf die Strukturvorgaben des Discovery Learnings und des 5E-Modells getroffen werden könnte, bleibt offen. Dabei ist allerdings auch generell offen, ob sich die inhaltlichen Einschränkungen bei Wackermann (2008) oder Wackermann et al. (2012) zur Verwendung verschiedener Basismodelle auch empirisch nachweisen lassen.

Das Basismodell *Konzeptwechsel* in seiner gezielten Auseinandersetzung, Gegenüberstellung und Ausschärfung von alten und neuen Konzepten und der Verwendung von kognitiven Konflikten findet im Feld der hier dargestellten

2. Strukturierung

Strukturvorgaben kein Analogon. Kognitive Konflikte werden in den Vorgaben zwar beim Problem Based Learning oder dem 5E-Modell in der Anfangsphase berücksichtigt, allerdings im Sinne einer Motivation oder eines Überraschungseffektes. Nicht jedoch im Sinne einer Konfrontation bestehender Konzepte, um diese aufzulösen und dazu konkurrierende Konzepte anzunehmen, wie dies beim *Konzeptwechsel* explizit vorgesehen ist (Wackermann et al., 2012, S. 7f.). Implizit könnte beim Problem Based Learning, Discovery Learning oder dem 5E-Modell aber eine eigenständige Auseinandersetzung der Lernenden mit kognitiven Konflikten während der Experimentierphase auftreten.

Problemlösen als Basismodell und das Lösen von Problemen im Problem Based Learning weisen erwartungskonform im Stellenwert der Problematisierung der Lerninhalte, aber auch hinsichtlich der unterrichtlichen Vorgaben zur Strukturierung wesentliche Übereinstimmungen auf. Erwartbar deswegen, weil Oser und Baeriswyl (2001, S. 1053) die Handlungkettenschritte zum Basismodell *Problemlösen* aus den Vorgaben einer Vielzahl grundsätzlich ähnlicher, aber individuell verschiedener Problemlöseansätze abgeleitet haben. Nicht unbedingt erwartbar ist hingegen die Kongruenz der Aussagen der empirischen Forschung mit den theoretischen Vorgaben der BMT über das Einsatzgebiet des *Problemlösens*. Theoriekonform (vgl. Trendel et al., 2007, S. 14 oder Wackermann et al., 2012) erfolgt in der BMT der Einsatz dieses Basismodells, wenn zum problematisierten Sachgegenstand bereits das zur Lösung notwendige Wissen vorhanden und das Ziel des Lernvorgangs eine Anwendung oder Verbindung der Wissens Elemente ist. Vergleicht man diese Regelung der BMT mit den zusammenfassenden Aussagen zur Wirkung des Problem Based Learnings von Hattie (2009), ist die Deckungsgleichheit bemerkenswert (vgl. S. 40).

Deutliche Entsprechungen des Basismodells *Konzeptaufbau* finden sich bei einem Vergleich mit den Vorgaben der Direct Instruction. Ziel beider Vorgaben ist der Aufbau von Wissen unter Berücksichtigung bestehender Kenntnisse zum Lerninhalt. Die Aktivierung dieser Kenntnisse bildet im Verlauf des Konzeptaufbaus und der Direct Instruction den ersten unterrichtlichen Schritt. Anschließend erfolgt die Präsentation der wesentlichen Elemente und Merkmale durch oder mit geeigneten Beispielen und eine nachfolgende Anwendung des Gelernten in ähnlicher Form bei beiden Verfahren. Fordert die Direct Instruction nun explizit eine Phase der Rückmeldung an die Lernenden ein, sieht der Konzeptaufbau eine inner- und idealerweise auch überfachliche Vernetzung zu anschlussfähigen oder abzugrenzenden Konzepten vor. Übereinstimmend beinhalten sowohl das Basismodell als auch die Direct Instruction in ihren abschließenden Phasen eine Dekontextualisierung des Erlernten, wobei die Direct Instruction zusätzlich vorgibt, dass die Lernenden diese Anwendung in anderen Kontexten eigenständig durchführen (*independent practice* Hattie, 2009, S. 205f.). In der Grobstruktur und teilweise auch in einigen Details führen beide Vorgaben damit zu ähnlichen unterrichtlichen Ausgestaltungen. Die Inhalte dieser Ausgestaltungen wiederum sind in der BMT vorgegeben. So thematisiert werden *erfundene* Aspekte

2.4. Zusammenfassung des Forschungsstandes und Formulierung der Forschungsfragen

wie Konzepte oder Begriffe. Die Wirksamkeit der Direct Instruction ist zwar nachweislich zwischen einzelnen Fächern unterschiedlich (Hattie, 2009, S. 204f.), hinsichtlich der Eignung der Direct Instruction für einzelne Teilbereiche innerhalb der Physik wurden keine Aussagen gefunden.

Vergleicht man die BMT mit weiteren Strukturvorgaben so findet man keine einzelne Vorgabe, welche der BMT in ihrer Bandbreite der unterrichtlichen Möglichkeiten entspricht. Einzelne Entsprechungen können aber für drei der vier wesentlichen Basismodelle gefunden werden. Entspricht das Basismodell des *Problemlösens* weitestgehend dem Problem Based Learning, so finden sich Parallelen zwischen *Lernen durch Eigenerfahrung* dem 5E-Modell kombiniert mit Elementen des Discovery Learnings und Überschneidungen zwischen dem Basismodell *Konzeptaufbau* und der Direct Instruction. Im theoretischen Rahmen der BMT sind damit zumindest wesentliche Elemente des Discovery Learnings, des Problem Based Learnings, der Direct Instruction und des 5E-Modells zusammen mit dem davon unabhängigen Modell des *Konzeptwechsels* angeordnet. Geordnet ist dabei auch, wann welches Modell eingesetzt wird. Ob sich diese theoretische Vorgabe allerdings für die Basismodelle *Lernen durch Eigenerfahrung*, *Konzeptwechsel* und *Konzeptaufbau* empirisch ebenso klar nachweisen lässt wie für das *Problemlösen*, ist nach jetzigem Forschungsstand nicht beantwortbar und damit offen.

2.4. Zusammenfassung des Forschungsstandes und Formulierung der Forschungsfragen

Forschungsstand

Obwohl eine eindeutige Festlegung des Begriffs der Strukturierung fehlt, besteht Einigkeit in deren Kennzeichnung als wesentliches Qualitätsmerkmal von Unterricht. Welche Teilaspekte, welches Ausmaß oder welche Form der Strukturierung wirklich guten Unterricht bedingen, ist dabei nicht eindeutig zu beantworten. „Forschung zu Kriterien guten Unterrichts beschäftigte sich bisher vornehmlich mit allgemeinen Unterrichtsmerkmalen und nahm dabei wenig Bezug auf Fachstrukturen, Fachinhalte und fachbezogene Lernprozesse“ (Trendel et al., 2007, S. 10), welche zur Beantwortung vieler offener Aspekte notwendig wäre. Insgesamt existiert „überraschend wenig gesichertes Wissen über die Qualität intendierter und implementierter Lernprozesse“ (Trendel et al., 2007, S. 10), obwohl zur Anregung und Steuerung dieser Prozesse für der Physikunterricht konzeptionelle Modelle vorhanden sind. Die Darstellung einer Auswahl verfügbarer Modelle ist in diesem Kapitel erfolgt. Die Wirkungen der Vorgaben der Modelle aus dem angloamerikanischen Raum (Direct Instruction, Discovery Learning, Problem Based Learning und 5E-Modell) sind zumindest im Hinblick auf deren Auswirkungen auf den Lernerfolg empirisch ausführlich erforscht, dennoch basiert die Entscheidung für oder gegen eine

2. Strukturierung

der Strukturvorgaben oft auf unhinterfragten Vorlieben und weniger auf empirisch gut gesicherten Erkenntnissen (vgl. Carnine, 2000). Nur in Ansätzen gesichert ist dabei das Wissen, für welche Themen sich welche Vorgaben in welchen Fächern besonders eignen. Eine generelle Unterschiedlichkeit in den Wirkungen bei Teilbereichen wird konstatiert, kann dabei aber nicht erklärt werden. Es fehlen Studien, welche den Unterricht auf einer Detailebene erfassen und ergründen worin sich diese Unterschiedlichkeit einzelner Vorgaben in der Wirkung begründet. Zusammenhänge zwischen den Vorgaben bleiben damit ebenso unerforscht. Im deutschsprachigen Raum ist die Auswahl eines konzeptionellen Modells vorrangig durch unhinterfragten Vorlieben oder Traditionen beeinflusst. Eine systematische Erforschung einzelner Strukturvorgaben scheint jetzt nicht erfolgt zu sein. Das FeU zeigt sich besonders im naturwissenschaftlichen Unterricht weit verbreitet und akzeptiert (di Fuccia & Ralle, 2010, S. 297), Studien, die dessen Lernwirksamkeit statistisch signifikant im Vergleich zu Kontrollgruppen nachweisen, sind aber nicht existent. Ein Vergleich zu anderen Arten von Strukturvorgaben fehlt ebenso wie eine Untersuchung über mögliche Einflüsse auf Motivation, Interesse oder weitere allgemeine Merkmale der Unterrichtsqualität. Zur BMT und deren Adaption auf den Physikunterricht liegen hingegen mehrere Studien vor, welche die Annahme eines positiven Einflusses auf verschiedene Aspekte des Unterrichts plausibel erscheinen lassen. Gesichert und im Ausmaß festgelegt sind diese Erkenntnisse dabei noch nicht. Offen bleibt, ob die nachgewiesene Lernwirksamkeit durch andere Einflussfaktoren als die Art der Sequenzierung nach der BMT bewirkt wird. Unklar bleibt auch, ob sich eine so geartete Sequenzierung als generell lernwirksamer als andere didaktisch gut begründete Vorgaben erweist und ob sich bei einem entsprechenden Vergleich Unterricht nach den Vorgaben der BMT erneut als besonders wirksam bei im Vortest schwächeren Schülerinnen oder Schülern zeigt (vgl. Zander et al., 2015). Weitere grundsätzliche Fragen zur BMT sind noch nicht beantwortet. Das Wissen über die Anzahl, die Gestaltung oder die Kombination der Basismodelle bedarf ebenso zusätzlicher Untersuchungen wie die Klärung der Frage, wie genau die vollständige bzw. teilweise Umsetzung der Handlungskettenschritte oder deren Variation in der Reihenfolge eines Basismodells das Lernen beeinflusst.

Forschungsfragen

Ziel dieser Untersuchung ist es, zu einer fachspezifischen systematischen Erforschung konzeptioneller Modelle zur Gestaltung von Lehrprozessen im Unterricht beizutragen. Der Fokus liegt damit auf Elementen der didaktischen Strukturierung, spezieller auf bestimmten Arten von Strukturvorgaben. Insbesondere deren unterrichtliche Auswirkungen sollen in einem direkten Vergleich zweier didaktisch gut begründeter Vorgaben ermittelt werden. Die Untersuchung der Auswirkungen von nach den Vorgaben zweier Modelle gestalteten Unterrichts führt dazu, dass die Intervention in beiden Untersuchungsgruppen sehr genau definiert ist. Dies ermöglicht einen Blick in die Detailebene des

Unterrichts.

Vor diesem Hintergrund werden folgende Forschungsfragen formuliert:

- F 1: Ist eine Sequenzierung nach den Vorgaben des FeU umsetzbar und lernwirksam im Physikunterricht?

Das FeU ist primär in der Praxis des Chemieunterrichts in Deutschland verankert (di Fuccia & Ralle, 2010, S. 297f.), nach Schmidkunz und Lindemann (1992) aber für die Planung naturwissenschaftlichen Unterrichts im Allgemeinen geeignet. Eine konkrete Umsetzung in einer Studie in der Fachdidaktik Physik scheint bislang nicht zu existieren, weshalb diese hier so grundlegend formulierte Forschungsfrage noch nicht beantwortet ist.

- F 2: Führen unterschiedliche Sequenzierungen zu einem unterschiedlichen Maß der kognitiven Aktivierung oder der grundlegenden Bedürfnisse?

Für die Auswirkungen von Elementen der didaktischen Strukturierung auf affektiv-motivationale Faktoren liegen generell wenige und gleichsam wenig einheitliche Ergebnisse vor. Einflüsse einer Instruktion nach den Vorgaben der BMT auf motivationale Aspekte oder das Fachinteresse deuten sich bei Wackermann (2008) beispielsweise an, können in einer Nachfolgestudie bei Zander (2013) aber nicht belegt werden. Lindemann spricht im Kontext des FeU vom Prinzip der Motivation und dem Ziel einer hohen eigenen kognitiven Aktivierung (Lindemann, 1989, S. 21), ein empirischer Nachweis allerdings fehlt. Rakoczy et al. (2007, S. 115) wiederum liefern einen empirischen Nachweis eines Zusammenhangs der kognitiven Aktivierung und Elementen der Motivation (vgl. Deci & Ryan, 1993) insbesondere mit dem Konstrukt der *strukturierten Präsentation der Lerninhalte*. In der Operationalisierung dieses Konstrukts lassen sich auch Elemente der didaktischen Strukturierung wiederfinden. Ein Einfluss von Elementen der didaktischen Strukturierung eben auf Elemente der Motivation oder die kognitive Aktivierung scheint demnach plausibel und soll im Rahmen dieser Studie untersucht werden. Eine Untersuchung anderer oder weiterer affektiver oder motivationaler Aspekte wäre dabei möglich, eine Einschränkung auf diese beiden Teilaspekte erfolgte auch aus testökonomischen Gesichtspunkten.

- F 3: Zeigen sich die unterschiedlichen Sequenzierungen auch unterschiedlich lernwirksam?

In vielen Studien entscheidet das Ausmaß an Strukturierung über die Qualität des Unterrichts (Seidel et al., 2006; Rakoczy et al., 2007; Drollinger-Vetter, 2011). Drollinger-Vetter (2011, S. 322f.) stellt dabei zumindest für das von ihr untersuchte Merkmal der strukturellen Klarheit fest, dass unterschiedliche Unterrichtsmuster zu strukturell klarem Unterricht führen. Eindeutige Vorteile einzelner Strukturierungsmuster kann Drollinger-

2. Strukturierung

Vetter (2011, S. 322f.) auf Basis der ihr vorliegenden Daten nicht erkennen, kann die Existenz entsprechender Vorteile (beispielsweise beim Lernerfolg) aber auch nicht ausschließen. Hattie (2009, S. 297f.) hingegen erkennt in seiner Meta-meta-Studie wesentliche Unterschiede in den Auswirkungen von verschiedenen Arten der Sequenzierung und damit der didaktischen Strukturierung auf den Lernerfolg. Für die im Rahmen dieser Studie untersuchten Arten der Sequenzierung liegen allerdings keine vergleichbaren empirischen Daten zu deren Lernwirksamkeit vor und sollen deswegen erhoben werden.

- F 4: Zeigen sich die unterschiedlichen Sequenzierungen auch hinsichtlich einzelner Teilgruppen der Lernenden unterschiedlich lernwirksam?

Helmke (2009, S. 200) betrachtet gute Strukturen als um so wichtiger, je geringer die Vorkenntnisse der Schüler sind. Soll mit einem geeigneten Unterricht einer zu starken Differenzierung hinsichtlich der Leistung der Schüler entgegen gewirkt werden, muss in besonderem Maße auf einen klaren und strukturierten Unterricht geachtet werden (Helmke, 2009, S. 200). Diese Aussage ist dabei allgemein auf die Strukturierung bezogen. Speziell für Elemente der didaktischen Strukturierung konnte Zander (2015) ebenfalls unterschiedlich starke Wirkungen auf Teilgruppen von Schulklassen nachweisen. Die Teilgruppe der in einem Vortest schwächeren Schüler profitierte dabei besonders von einer Sequenzierung des Unterrichts nach den Vorgaben der Basismodelltheorie (bezogen auf eine Vergleichsgruppe mit „herkömmlichen“ Unterricht). Ob sich nun aber zwei didaktisch gut begründete, aber unterschiedliche Arten der strukturierten Sequenzierung in diesem Effekt unterscheiden, ist unklar. Ebenso sind unterschiedliche unterrichtliche Auswirkungen hinsichtlich weiterer Teilgruppen wie dem Geschlecht der teilnehmenden Schüler möglich.

- F 5: Führen unterschiedliche Sequenzierungen auch zu einem unterschiedlichen Maß der von den Schülern wahrgenommenen Strukturierung?

„Die Wahrnehmung der Schüler hat für die Entwicklung der Schüler die größte Bedeutung“ (Clausen, 2002, S. 188). Sie ist dabei insbesondere relevant, sollen Zusammenhänge von Aspekten zur Leistung oder dem Interesse gefunden werden (Clausen, 2002, S. 188). Untersucht man also einen Einfluss der Art der Sequenzierung u.a. auf den Lernerfolg ist die Wahrnehmung der Strukturierung des Unterrichts durch die Schüler entscheidend. Die Schüler sind dabei auch grundsätzlich zu einer differenzierten Bewertung fähig, wenngleich diese oft durch ein Generalurteil zur unterrichtenden Lehrkraft überlagert ist (Clausen, 2002, S. 188).

- F 6: Zeigt sich eine zu favorisierende Sequenzierung der Lernprozesse hinsichtlich der untersuchten Merkmale?

2.4. Zusammenfassung des Forschungsstandes und Formulierung der Forschungsfragen

Die Frage nach der Existenz oder Form einer möglichen optimalen Sequenzierung von Unterricht kann im Rahmen dieser Arbeit nicht umfassend beantwortet werden. Ein zusammenfassender Vergleich beider für die Studie zugrunde gelegter Vorgaben zur Gestaltung von Unterricht kann aber einen kleinen Teil zur Erörterung dieser Frage beitragen.

2. Strukturierung

3. Konzeption und Durchführung der Studie

Die Konzeption der Studie zielt darauf ab, die Forschungsfragen möglichst erschöpfend beantworten zu können. Die Darstellung der Konzeption der Studie beginnt mit dem grundsätzlichen Studiendesign auf Basis des vorliegenden Forschungsanliegens und des theoretischen Hintergrundes der Arbeit. Die Auswahl der verwendeten Strukturvorgaben wird im direkt nachfolgenden Unterkapitel begründet dargestellt. Anschließend folgen Überlegungen zur Konzeption und Ausgestaltung der in der Studie eingesetzten Unterrichtsstunden und es werden Gemeinsamkeiten und Unterschiede zwischen diesen Stunden aufgeführt. Nachfolgend werden die Methoden zur Datenerhebung erläutert.

Zur Beschreibung der Durchführung der Studie wird daran anschließend vom Aufbau der Stichprobe der Hauptstudie berichtet. Der letzte Abschnitt dieses Kapitels behandelt die bei der Auswertung eingesetzten statistischen Verfahren.

3.1. Studiendesign

Die Studie wurde als explorative, quasiexperimentelle Laborstudie konzipiert. Die Entscheidung zur Durchführung einer Laboruntersuchung ist motiviert durch das Bestreben Störgrößen zu kontrollieren oder zu eliminieren. Von besonderem Interesse waren dabei Störgrößen, für welche ein Einfluss primär auf den Lernerfolg vermutet wurde.

Verschiedenen Aspekten der Strukturierung wird ein entsprechender Einfluss auf den Lernerfolg zugeschrieben (vgl. S. 26). Einflüsse von Aspekten der Strukturierung, welche nicht im Fokus der Untersuchung liegen, sollten weitestgehend eliminiert werden. Der Fokus liegt auf Elementen der didaktischen Strukturierung. Nicht von Interesse sind unterrichtliche Auswirkungen spezifischer Methoden oder Maßnahmen zur Verarbeitung der Sachinhalte oder zur Steuerung des Verhaltens der Lernenden. Um diese Aspekte in den Interventionsgruppen möglichst konstant zu halten ist eine Labor- im Vergleich zu einer Feldstudie geeigneter. Die vorliegende Studie zielt zudem darauf ab, die Wirkung der unterrichtlichen Umsetzungen der zugrunde gelegten theoretischen Vorgaben zu untersuchen. Werden diese Vorgaben in den durchgeführten Unterrichtssequenzen nicht theoriekonform umgesetzt, ist anzunehmen, dass davon die Wirkung auf den Lernerfolg beeinflusst wird. Eine theoriekonforme Umsetzung solcher Vorgaben hat sich in Feldstudien allerdings als schwierig

3. Konzeption und Durchführung der Studie

erwiesen (Reyer, 2004; Ohle et al., 2011). Speziell bei der auch in dieser Arbeit als Vorgabe verwendeten BMT, werden die Basismodelle durch die an einer Studie beteiligten Lehrkräfte oft nicht vollständig durchgeführt (Reyer, 2004, S. 278). Eine Kontrolle dieses Effekts erscheint dabei schwierig. Zum einen ist die Effektstärke des Grades der Umsetzung der theoretischen Vorgaben auf den Lernerfolg nicht eindeutig für die verwendeten theoretischen Vorgaben erforscht, zum anderen ist der Grad der Umsetzung zumindest für das FeU nicht eindeutig operationalisiert und damit erfassbar. Für die BMT liefert Wackermann (2008 und 2012) mit den sogenannten Umsetzungsstufen einen Vorschlag zur Operationalisierung, doch auch dieser „enthält nicht notwendigerweise eine Wertung“ (Wackermann, 2008, S. 14) über die unterrichtlichen Wirkungen der formulierten Umsetzungsstufen. Eine Kontrolle des Einflusses einer nicht vollständig theoriekonformen Umsetzung der BMT bzw. des FeU ist somit kaum möglich, ein Einfluss der Umsetzung auf den Lernerfolg ist hingegen anzunehmen. Für eine Untersuchung der Wirkungen unterrichtlicher Umsetzungen der der BMT bzw. des FeU ist demnach eine möglichst theoriekonforme Umsetzung der Vorgaben in allen in der Studie untersuchten Unterrichtseinheiten für die Qualität der zu erhebenden Daten förderlich. Da zudem ja Reyner (2004) und Ohle et al. (2011) berichten, dass eine theoriekonforme Umsetzung in Feldstudien schwer zu erreichen ist, ist damit ein weiteres Argument für die Durchführung einer Laborstudie gegeben. Im Rahmen einer Feldstudie wäre zudem eine Fortbildung oder Schulung teilnehmender Lehrkräfte notwendig. Wackermann et al. (2010, S. 980f.) und Zander et al. (2013, S. 504) berichten diesbezüglich auch von Limitationen ihrer Feldstudien. Der Einfluss einer Fortbildung oder Schulung teilnehmender Lehrkräfte und damit auf den Unterricht welcher der Datenerhebung zugrunde liegt, ist nicht immer klar abzugrenzen und stellt damit eine mögliche Störgröße dar. Die letztlich in den Studien von Wackermann (2008) und Zander et al. (2013) empirisch gefundenen Effekte lassen sich folglich nicht eindeutig nur auf die Variation in der Gestaltung des Unterrichts zurückführen. Dies führt im Rahmen dieser Studie zur Entscheidung, dass Konzeption und Umsetzung der Unterrichtseinheiten vom Versuchsleiter und Autor des vorliegenden Textes durchgeführt wurden. Der Versuchsleiter ist dabei eine ausgebildete Physiklehrkraft, welche von Februar 2006 bis August 2011 im aktiven Lehrdienst an drei bayerischen Gymnasien stand.

Die Entscheidung auch die Umsetzung des Unterrichts durch den Versuchsleiter durchführen zu lassen, bedingt allerdings das Hinzukommen einer Störvariablen. Mögliche Erwartungen des Versuchsleiters über Ergebnisse seiner Untersuchung *können* den Umgang mit den Versuchsteilnehmern und damit die erhobenen Daten auch unbeabsichtigt beeinflussen. Man nennt dies den Versuchsleiter-Erwartungseffekt oder auch Rosenthaleffekt. Dieser Effekt ist dabei nicht immer gleich stark ausgeprägt und kann zudem kontrolliert, oder zumindest im Ausmaß seiner Auswirkung deutlich begrenzt werden (Huber,

2009, S. 184). „Die größte Gefährdung einer gleichmäßigen Wirkung störender Untersuchungsbedingungen auf alle Versuchspersonen in allen Untersuchungsgruppen besteht in der Kenntnis der Untersuchungshypothese“ (Bortz & Döring, 2006, S. 84). Eine solche kann allerdings aufgrund der Beschaffenheit der Literaturlage für einen Vergleich der Wirkung von Unterrichtseinheiten nach der BMT oder konform zum FeU nicht formuliert werden. Die Ausrichtung der Studie ist explorativ. Deshalb wird im Rahmen dieser Arbeit das Ausmaß der Störung durch den Versuchsleitereffekt als eher gering eingeschätzt. In geringem Maße aber können die empirischen Daten durchaus durch den Versuchsleitereffekt beeinflusst sein, da mögliche unbewusste Erwartungen über die Wirkung des FeU oder der BMT auf Seiten des Versuchsleiters nicht auszuschließen sind. Folglich wurden zusätzliche Maßnahmen zur Begrenzung des Effekts dieser Störvariablen für die Umsetzung der Studie eingeplant. Für die Begrenzung des Effekts stellen Bortz und Döring (2006, S. 84f.) bzw. Huber (2009, S. 187) verschiedene Maßnahmen zur Verfügung, welche allerdings nicht bei jeder Studie Anwendung finden können. Einzelne dieser empfohlenen Maßnahmen wurden in der vorliegenden Studie angewendet. So wurde die Instruktion möglichst standardisiert (begleitende Power-Point-Präsentation für beide Interventionsgruppen), sowie der Ablauf aller Teilschritte der Intervention schriftlich fixiert, konstant gehalten und videographiert. Zudem wurde eine Nachbefragung der Untersuchungsteilnehmer über eine mögliche Unterschiedlichkeit in der Wirkung des Versuchsleiters zwischen beiden Interventionsgruppen durchgeführt.

Die Wahl der quasiexperimentellen Vorgehensweise wurde vorwiegend aus organisatorischen und pragmatischen Gründen getroffen. Es wurde angenommen, dass die Anreise von Schülergruppen zur Intervention an der Universität Regensburg im Klassenverbund erfolgt und auf eine zufällige Zuordnung der Lernenden auf die Interventionsgruppen verzichtet. Die Gesamtstichprobe setzt sich demnach aus natürlichen Gruppen mit deren jeweiligen Besonderheiten zusammen. Eine Unterschiedlichkeit innerhalb aber auch zwischen den Gruppen muss folglich bei der Auswertung der Daten berücksichtigt werden. Für diesbezügliche Auswertungsmethoden (z.B. Mehrebenenanalyse, vgl. Abschnitt 3.5) sind oftmals große Stichprobenumfänge erforderlich. Vor diesem Hintergrund erfolgte auch die Festlegung auf der Untersuchung von Wirkungen genau zweier Strukturvorgaben. Eine Untersuchung einer größeren Zahl an Vorgaben würde letztlich eine Verkleinerung der Stichproben pro Interventionsgruppe bedingen und eine nachfolgende Auswertung möglicherweise erschweren.

3.2. Auswahl der Strukturvorgaben

Primäres Ziel dieser Studie ist es, unterschiedliche unterrichtliche Auswirkungen von Lehr-Lern-Sequenzen in der Physik zu erfassen. Diese Sequenzen sollen

3. Konzeption und Durchführung der Studie

dabei konform zu den Vorgaben unterschiedlicher konzeptioneller Modelle gestaltet sein. Ziel dieser Studie ist es nicht, die Wirkung einzelner Maßnahmen im Rahmen einer kognitionspsychologisch verstandenen Strukturierung zu untersuchen. Spezifische Methoden oder auch Handlungen der Lehrkraft insbesondere zur Verbindung von neuem Wissen mit bereits Gelerntem liegen also nicht im Fokus der Studie, sollen nicht untersucht werden und sollten sich demnach auch nicht in den beiden konzipierten Interventionsstunden unterscheiden. Für die Auswahl der Strukturvorgaben bedeutet dies, dass diese eben keine Vorgabe über die Verwendung einzelner Methoden oder spezieller Lehrerhandlungen beinhalten sollten und damit eine Konzeption der Unterrichtsstunden mit vergleichbarer methodischer Gestaltung ohne Einschränkungen möglich ist. Für die Realisierung methodisch gleichgearteter Unterrichtsstunden ist es ebenfalls möglich, dass diesbezügliche Vorgaben in beiden untersuchten konzeptionellen Modellen identisch sind, oder sich die Vorgaben eines Modells auf das zu vergleichende Modell übertragen lassen. In den beiden letztgenannten Fällen ist aber die methodische Planungs- und Gestaltungsfreiheit für die Konzeption des Unterrichts bereits eingeschränkt.

Weiterhin entscheidend ist eine gute Umsetzbarkeit der Strukturvorgaben für eine Studie im Rahmenkontext der Physikdidaktik. Für diese Umsetzbarkeit ist eine detaillierte Dokumentation der vorgegebenen unterrichtlichen Phasen und Phasenverläufe notwendig, um eine zu den Vorgaben konforme Umsetzung der gewählten konzeptionellen Modelle zu ermöglichen. Zusätzlich ist auch eine generelle Eignung der Strukturvorgaben für den Physikunterricht maßgeblich.

Letztlich sollen beide Vorgaben auch vergleichbar sein. Eine Vergleichbarkeit im Grad der didaktischen Strukturierung (nicht in deren Art) und der Zielsetzung der einzelnen Strukturvorgaben sollte gegeben sein. Eine Strukturvorgabe mit dem Ziel des Aufbaus von neuem Wissen ist hinsichtlich des Lernerfolgs nicht vergleichbar mit einer Vorgabe deren Fokus auf der Anwendung bereits vorhandenem Wissens liegt.

Somit sind die Vorgaben der einzelnen konzeptionellen Modelle zu den Kriterien Methoden und Lehrerhandlungen, Umsetzbarkeit und Vergleichbarkeit für die Auswahl der im Rahmen der Studie verwendeten spezifischen Strukturvorgaben ausschlaggebend. Die Entscheidung für die Auswahl des FeU und der BMT als theoretische Vorgaben zur Konzeption der Interventionen wird im folgenden anhand dieser drei Kriterien dargestellt.

Methoden und Lehrerhandlungen

Bei der BMT bildet die fest vorgegebene Unterrichtsstruktur die Grundlage für Entscheidungen über die nachfolgende freie Ausgestaltung der Sichtstrukturen (Oser et al., 1997, S. 21). Als Elemente der Sichtstruktur sind damit auch die Planungen zu Methoden und Lehrerhandlungen flexibel, solange damit der vorgegebene strukturelle Ablauf des Unterrichts unterstützt wird. Das Zusammenspiel der Methodenfreiheit einerseits und einer nahezu gesetzmäßig

vorgegebenen Struktur durch die Basismodelle andererseits, ist dabei sogar einer der Kernpunkte der BMT. Das FeU hingegen gibt einzelne Aspekte der Methoden oder der Lehrerhandlungen bzw. der Lehreraktivität vor, wenngleich dieser Vorgabe nicht der Status einer Gesetzmäßigkeit zuzuschreiben ist. So werden einzelne Phasen mit höherer oder geringerer Aktivität bei den Lernenden und Lehrenden gekennzeichnet, allerdings darauf verwiesen, dass bei situativ erkennbarem Bedarf davon abgewichen werden kann (Schmidkunz & Lindemann, 1992, S. 44f.). Eine Abweichung ist zudem auch von der Empfehlung im FeU über die Verwendung des Experiments in der Phase der *praktischen Durchführung eines Lösevorhabens* möglich. Selbst in dieser Schlüsselstelle des FeU könnten andere Medien (Filmausschnitt, Dia, Realobjekt, Schilderung usw.) an Stelle des Experiments durchgeführt werden (Schmidkunz & Lindemann, 1992, S. 26). Grundsätzlich legt das FeU aber für diese unterrichtliche Phase die Verwendung eines Experiments mit klar definierter didaktischer Funktion nahe. Alle anderen Phasen des Verfahrens können aber methodisch frei gestaltet werden. Somit steuert das FeU die Stellung und Funktion des zentralen Experiments im Unterricht, erlaubt von dieser Ausnahme abgesehen aber eine große Flexibilität bei der Wahl geeigneter Methoden und Lehrerhandlungen. Bezüglich der Methodenwahl und der Vorgabe zu den Lehrerhandlungen unterscheiden sich das Discovery Learning und das 5E-Modell nicht wesentlich vom FeU. Deutlichere Unterschiede finden sich in den diesbezüglich detaillierteren Vorgaben des Problem Based Learnings. Aus den Kerncharakteristiken des Problem Based Learnings nach Borrows (1996) folgen klare Vorgaben zu den Sozialformen (Kleingruppen) im Unterricht, zur Rolle der Lehrkraft (Moderator) und einzelner Methoden (z.B. selbstgesteuertes Lernen) (Gijbels et al., 2005, S. 29f.). Kontrollierbare und leistungsrelevante Aspekte wie Methoden und Lehrerhandlungen will die Direct Instruction grundsätzlich vorgeben (Engelmann, 1980, S. 4). Um diesem Anspruch gerecht zu werden, gibt die Direct Instruction nicht nur die Sequenzierung des Unterrichts vor, sondern zusätzlich auch die spezifische Ausgestaltung einzelner Phasen hinsichtlich zu wählender Methoden oder Lehrerhandlungen.

Insgesamt zeigt sich für das hier betrachtete Kriterium eine besonders deutliche Eignung der BMT zur Verwendung als theoretische Vorlage im Rahmen der Studie. Geeignet erscheinen zudem das FeU, das Discovery Learning und das 5E-Modell. Wohingegen das Problem Based Learning und die Direct Instruction durch spezifische Vorgaben zu Elementen der kognitionspsychologisch verstandenen Strukturierung eine in diesem Bereich einheitliche Konzeption der in der Studie eingesetzten Unterrichtsstunden erschweren.

Umsetzbarkeit

Sehr genaue Vorgaben über die Ausgestaltung der unterrichtlichen Sequenzierung geben sowohl die BMT als auch das FeU und die Direct Instruction (Oser & Baeriswyl, 2001; Wackermann, 2008; Schmidkunz & Lindemann, 1992; Engelmann, 1980, vgl.) vor. Nur etwas weniger detailliert erscheinen die Vorgaben

3. Konzeption und Durchführung der Studie

im 5E-Modell nach Bybee (1997), wohingegen das Problem Based Learning und das Discovery Learning einen größeren Interpretationsspielraum bei der Konzeption der Unterrichtsstunden lassen (Hmelo-Silver, 2004; Jong & Joolingen, 1998, vgl.). Bezüglich des Detailgrades der Vorgaben eignen sich demnach die Direct Instruction, das FeU, die BMT und mit leichten Einschränkungen das 5E-Modell für eine möglichst theoriekonforme Konzeption der Interventionen.

Das 5E-Modell, das Discovery Learning und das Problem Based Learning sind mit dem »Inquiry-Ansatz« und damit mit dem naturwissenschaftlichen Unterricht im allgemeinen und dem Physikunterricht im speziellen verbunden, wobei das Problem Based Learning seinen Schwerpunkt in der medizinischen Ausbildung aufweist (Hmelo-Silver et al., 2007, S. 102f.). Auch das FeU weist einen besonderen Fokus auf. Grundsätzlich ist das FeU für die naturwissenschaftlichen Fächer Chemie, Physik, teilweise auch für Biologie oder Technologie konzipiert (Schmidkunz & Lindemann, 1992, S. 11f.), das FeU weist aber den größten Verbreitungsgrad im Unterricht des Faches Chemie auf (di Fuccia & Ralle, 2010, S. 297). Die Direct Instruction und die BMT nach Oser und Baeriswyl (2001) sind hingegen fachunspezifische Strukturvorgaben, wobei letztere durch Reyer (2004), Trendel (2007) und Wackermann (2008) noch spezifisch für den Physikunterricht modifiziert wurde. Somit zeichnet sich bei diesem Teil des Kriteriums der Umsetzbarkeit eine generelle Eignung aller betrachteten Vorgaben ab.

Beide Teilbereiche des Kriteriums der Umsetzbarkeit werden demnach von der Direct Instruction, der BMT, dem FeU und etwas weniger gut vom 5E-Modell erfüllt.

Vergleichbarkeit

Ziel der Studie ist unter anderem zu explorieren, ob für den Lernerfolg oder affektiv-motivationale Aspekte die Art der Strukturierung unterrichtlicher Phasen entscheidend ist. Das Ausmaß an bzw. der Grad der Strukturierung ist dabei nicht im Fokus des Forschungsvorhabens, weshalb sich die für die Intervention entwickelten Unterrichtsstunden diesbezüglich möglichst wenig unterscheiden sollten. Sowohl das FeU als auch die BMT sehen eine Orientierung der Strukturierung des Unterrichts auf der Basis von Denk- oder Lernprozessen der Lernenden vor (Fries & Rosenberger, 1967; Oser et al., 1997) und untergliedern beide den Ablauf des Unterrichts im wesentlichen in fünf Teilschritte. Diese Teilschritte werden im FeU zudem in jeweils drei Phasen untergliedert. Vergleichbares findet sich in der BMT nicht, allerdings werden insbesondere in der Modifikation der BMT für den Physikunterricht die ursprünglich formulierten fünf Schritte nochmals detaillierter ausgearbeitet, so dass sich zumindest auf einer subjektiven Ebene der Grad an Strukturierung als vergleichbar darstellt. Auch die Direct Instruction und das 5E-Modell weisen eine fünfstufige Untergliederung auf. Sind diese bei der Direct Instruction recht genau vorgegeben, ist die Ausführlichkeit der Darstellung der Vorgaben zur Ausgestaltung der ein-

zelenen Unterstufen im 5E-Modell nach Bybee als etwas geringer einzuschätzen. Die Vorgabe des Discovery Learnings gibt grundlegend einen dreistufigen und oft weniger exakt beschriebenen Verlauf vor, die Vorgabe des Problem Based Learnings weist im *Problem-based- Learning Cycle* nach Hmelo-Silver (2004) sechs Phasen auf, deren Dokumentation aber weniger umfassend erscheint als beispielsweise im FeU oder in der BMT. Das Ausmaß an Strukturierung zeigt sich zumindest subjektiv beim FeU der BMT und der Direct Instruction vergleichbar.

Nachdem die Lernwirksamkeit der Sequenzierung im Fokus der Untersuchung liegt, sollten die als theoretische Vorlage verwendeten Strukturvorgaben ebenfalls diesen Fokus und damit als Hauptziel den Aufbau neuen Wissens verfolgen. Dieses Hauptziel weisen sowohl die Direct Instruction, das 5E-Modell als auch das Discovery Learning, nicht jedoch das Problem Based Learning auf. Für Letzteres konnte Hattie (2009, S. 210f.) nachweisen, dass sich Problem Based Learning für die Anwendung von bereits bekanntem Wissen eignet, nicht jedoch für den Aufbau neuen Wissens. Die Festlegung auf dieses Ziel bedeutet für das FeU bzw. die BMT eine Einschränkung auf einzelne Zweige bzw. Basismodelle. Aus dem FeU kann demnach nur der *induktive Zweig* Anwendung im Rahmen der Studie finden, da dieser im Gegensatz zum *deduktiven Zweig* für Sachverhalte geeignet ist, zu denen die Lernenden noch nicht die zur Lösung des Problems erforderlichen Kenntnisse aufweisen, sondern erst im Laufe des Unterrichts erwerben (Schmidkunz & Lindemann, 1992, S. 35ff.). Das Erwerben neuen Wissens ist auch Ziel der Basismodelle *Lernen durch Eigenführung* und *Konzeptaufbau* (Wackermann, 2008, S. 141-145), welche sich demnach ebenfalls für eine Verwendung in der Untersuchung eignen. Die Basismodelle *Problemlösen* zur Anwendung von Wissen und *Konzeptwechsel* zur Überarbeitung von bestehendem, physikalisch korrektem Wissen oder Alltagsvorstellungen (Wackermann et al., 2012, S. 7ff.) werden bei der Konzeption der Unterrichtsstunden nicht eingesetzt.

Insgesamt erfüllen die BMT und das FeU die drei dargestellten Kriterien nahezu ohne Einschränkungen, weshalb diese beiden Strukturvorgaben auch als theoretische Vorlage bei der Konzeption der Unterrichtseinheiten verwendet wurden. Denkbar wäre auch eine Untersuchung, welche das 5E-Modell oder die Direct Instruction als Grundlagen verwendet, dies ginge aber mit etwas weitergehenden Begrenzungen bei der Planung der Intervention einher.

3.3. Konzeption der Unterrichtseinheiten

Festlegung des fachlichen Inhalts der Unterrichtseinheiten

Der fachliche Inhalt für die zu konzipierenden Unterrichtseinheiten wurde auf den Impuls festgelegt. Diese Entscheidung wird nachfolgend umfassender dargestellt und basierte einerseits auf den organisatorischen Rahmenbedingungen der Studie und andererseits sollte gewährleistet sein, dass sich der gewählte

3. Konzeption und Durchführung der Studie

Inhalt für die beiden ausgewählten Strukturvorgaben gleichermaßen eignet.

Im Rahmen der Studie werden die Wirkungen von Unterrichtsstunden auf Basis zweier didaktisch gut begründeter Strukturvorgaben exploriert. Feste Hypothesen über womöglich auftretende Unterschiede zwischen beiden Gruppen konnten auf Basis der Literaturanalyse nicht formuliert werden. Die im Vorfeld der Studie durchgeführte Stichprobenumfangsplanung mit Hilfe der Software GPower erfolgte auf Basis der Annahme eher geringer Mittelwertunterschiede zwischen den untersuchten Interventionsgruppen, da auch kleinere Effekte in der geplanten Untersuchung statistisch signifikant nachzuweisen sein sollten. Diese Planung ergab eine notwendige Stichprobengröße von ca. 300 bis 1300 Testteilnehmern (für $d = 0.2$ bis $d = 0.1$). Bei der Annahme einer Klassengröße von 25 Schülern würde dies die Teilnahme von 12 bis 52 Schulklassen erfordern. Die Intervention fand im sogenannten Lernforschungslabor der Universität Regensburg statt, wobei dieser Raum an zwei Tagen pro Woche zur Verfügung stand. Demzufolge war bei Belegung aller zur Verfügung stehender Termine für die Intervention ein Zeitraum von 6 bis 26 Wochen notwendig. Zusätzlich wurde angenommen, dass sich mehr Lehrkräfte zu einer Beteiligung an der Studie entscheiden würden, wenn sich der in der Intervention thematisierte Inhalt konform zum bayerischen Lehrplan (als Grundlage des Unterrichts der Schulen in der Umgebung von Regensburg) darstellt. Um nun auch kleine Effekte ($d = 0.1$) mit der Studie statistisch signifikant nachweisen zu können, sind folglich Lehrplaninhalte besonders geeignet, welche von den Lehrkräften an verschiedenen Zeitpunkten im Schuljahr in ihren Klassen unterrichtet werden. Nach der eigenen Unterrichtserfahrung an bayerischen Gymnasien (2006 bis 2011) trifft dies insbesondere auf das Thema des Impulses zu. Als Teil der Inhalte des Lehrplans der zehnten Jahrgangsstufe für Physik an Gymnasien in Bayern kann der Impuls an verschiedenen Zeitpunkten im Schuljahr thematisiert werden. So ist es möglich dieses Thema bereits Anfang Oktober oder auch erst Ende März zu unterrichten. Damit steht für die Intervention pro Schuljahr ein Zeitraum von fast sechs Monaten zur Verfügung. Deshalb wurde der Impuls als fachlicher Inhalt der Unterrichtseinheiten festgelegt, obwohl damit Schulklassen der Realschulen nicht teilnehmen können, da im Realschullehrplan in Bayern der Impuls nicht enthalten ist.

Um den Impuls in der vorliegenden Untersuchung als fachlichen Rahmen verwenden zu können, muss sich der Impuls aber auch für eine unterrichtliche Gestaltung auf Basis beider eingesetzter Strukturvorgaben eignen. „Eine wichtige Voraussetzung, um problemlösenden Unterricht zu gestalten, ist das Schaffen von Problemsituationen für den Lernenden“ (Schmidkunz & Lindemann, 1992, S. 16). Speziell der für die Konzeption der Studie relevante induktive Zweig des FeU geht von einer Problemsituation aus, welche für einen Lernenden immer dann entsteht, „wenn er mit Gegenständen, Situationen, Geschehnissen oder Vorgängen konfrontiert wird, die er mit seinem Vorwissen und seinen bisherigen Erfahrungen nicht erklären [...] kann“ (Schmidkunz & Lindemann, 1992,

S. 16f.). Für die Gestaltung einer entsprechenden Problemsituation zum Thema Impuls würde sich demnach eine Vielzahl verschiedener Stoßprozesse eignen. Damit ist eine wesentliche Voraussetzung für eine unterrichtliche Gestaltung zum Impuls auf Basis des FeU bereits erfüllt. Schmidkunz und Lindemann (1992) stellen neben der Forderung nach der Schaffung einer Problemsituation auch keine expliziten weiteren Forderungen an das zu unterrichtende Thema. Das festgelegte Thema des Impulses bereitete bei der Konzeption der Interventionsstunde zum FeU zumindest subjektiv auch keine Schwierigkeiten.

Aus der BMT stehen für die Studie die Basismodelle *Lernen durch Eigenerfahrung* und *Konzeptaufbau* zur Verfügung. Mit *Lernen durch Eigenerfahrung* können Inhalte thematisiert werden, welche von den Schülern selbst auf der Basis eigener Erfahrungen gefunden werden können, wie beispielsweise die Abhängigkeiten des Impulses von Masse und Geschwindigkeit. Erfahrungen aber „führen keineswegs auf direktem Wege zu neuen Konzepten, weil Konzepte im Allgemeinen nicht entdeckt, sondern erfunden werden“ (Trendel et al., 2007, S. 15). Das Konzept des Impulses macht damit auch zusätzliche Überlegungen notwendig, welche in der BMT nach dem Basismodell *Konzeptaufbau* unterrichtet werden müssen. Das Thema Impuls als ganzes erfordert demnach eine Kombination beider Basismodelle. Die Erschließung eines Themenbereichs mit unterschiedlichen Lehrzieltypen kann dabei durch eine Kombination von Basismodellen im Rahmen der BMT durchgeführt werden (Reyer, 2004, S. 30f.) und demnach auch eine Konzeption einer zur BMT konformen Unterrichtseinheit zum Thema Impuls.

Analyse des Sachinhalts

Die Darstellung der fachphysikalischen Aspekte zum Impuls basiert im Rahmen dieser Arbeit auf den Ausführungen von Halliday, Resnick und Walker (2009). Der Impuls ist dort definiert als:

$$\vec{p} = m\vec{v} \quad (3.1)$$

mit m als Masse und \vec{v} als Geschwindigkeit des Teilchens. Als Abgrenzung zum Drehimpuls wird der so definierte Impuls auch gelegentlich als *linearer* Impuls bezeichnet. Für die zeitliche Änderung des Impulses für ein Teilchen gilt zudem (ursprüngliche Formulierung des zweiten Newtonschen Gesetzes):

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

In Betrag und Richtung stimmt also die zeitliche Änderung des Impulses eines Teilchens mit der insgesamt auf dieses Teilchen wirkenden Kraft \vec{F} überein. Der Impuls eines Teilchensystems aus n Teilchen mit verschiedenen Massen, Geschwindigkeiten und Impulsen ist unter Berücksichtigung von möglichen Wechselwirkungen der Teilchen untereinander und dem möglichen Einwirken äußerer Kräfte auf das System definiert als die Vektorsumme der Impulse der

3. Konzeption und Durchführung der Studie

einzelnen Teilchen:

$$\vec{P} = \vec{p}_1 + \vec{p}_2 + \vec{p}_3 + \dots + \vec{p}_n \quad (3.3)$$

Die zeitliche Änderung dieser Vektorsumme entspricht gerade der insgesamt auf das System wirkenden effektiven äußeren Kraft:

$$\vec{F}_{eff} = \frac{d\vec{P}}{dt} \quad (3.4)$$

Dies stellt eine Verallgemeinerung der für ein Teilchen formulierten Gleichung (3.2) auf ein Vielteilchensystem dar.

Betrachtet man ein System mit einer konstanten Teilchenzahl auf welches keine effektiven äußeren Kräfte wirken (sog. geschlossenes und isoliertes System), so ergibt sich $d\vec{P}/dt = 0$. Solange diese Bedingungen an das System erfüllt sind, ist der Gesamtimpuls \vec{P} des Systems konstant und der Impuls eine Erhaltungsgröße. Alternativ kann dieser Impulserhaltungssatz auch so formuliert werden, dass der Gesamtimpuls eines geschlossenen und isolierten Systems zu einem anfänglichen Zeitpunkt t_0 gerade dem Gesamtimpuls zu einem späteren Zeitpunkt t_1 entspricht. Durch den Vektorcharakter des Gesamtimpulses ergeben sich bei Betrachtung dreier aufeinander senkrecht stehender Richtungen drei unabhängige Gleichungen für die Impulserhaltung.

Entscheidende Bedeutung hat die Impulserhaltung neben der Energieerhaltung insbesondere bei Stoßprozessen. Unter einem Stoß versteht man „ein isoliertes Ereignis, während dessen zwei oder mehrere Körper für relativ kurze Zeit relativ starke Kräfte aufeinander ausüben“ (Halliday et al., 2009, S. 262). Während eines Stoßes können sich die Impulse einzelner Teilchen des Systems durch einen Kraftstoß \vec{J} ändern.

$$\vec{J} = \int_{t_0}^{t_1} \vec{F}(t) dt = \int_{p_0}^{p_1} d\vec{p} \quad (3.5)$$

Die Impulsänderung eines am Stoß beteiligten Teilchens entspricht gerade dem Kraftstoß, welcher auf das Teilchen wirkt.

$$\Delta\vec{p} = \vec{J} \quad (3.6)$$

Zu einem Stoß zweier Körper kann es dabei nur kommen, wenn sich mindestens eines der beiden Teilchen bewegt. Damit besitzt das System eine von null verschiedene kinetische Energie. Bleibt diese kinetische Energie eines Systems beim Stoß erhalten, so spricht man von einem *elastischen* Stoß, bleibt diese nicht erhalten, spricht man vom *inelastischen* Stoß. Man spricht von einem *vollkommen inelastischen* Stoß, wenn die Stoßpartner nach dem Stoß aneinander haften. Im Gegensatz zur kinetischen Energie kann sich der Gesamtimpuls eines geschlossenen isolierten Systems während eines Stoßprozesses nicht ändern. Dazu wären äußere Kräfte notwendig, bei einem Stoß treten aber nur innere Kräfte auf. Innere Kräfte sind dabei Kräfte, welche von einem Teil des betrachteten Systems auf einen anderen Teil des Systems wirken.

Die Darstellung des Impulses im Unterricht der zehnten Jahrgangsstufe an bayerischen Gymnasien erfolgt in der Regel im Anschluss an die Behandlung der Newtonschen Gesetze. Die Newtonschen Gesetze werden als Grundlage zur Beschreibung von Bewegungsabläufen, sowie zur Einführung und Anwendung eines numerischen Verfahrens zum Lösen von Bewegungsgleichungen verwendet. Neben der Definition und Veranschaulichung des Begriffes des Impulses soll der unterrichtliche Fokus auf dem Impulserhaltungssatz, sowie der Anwendung von Impuls- und Energieerhaltungssatz auf einfache Fälle liegen. Bei der Anwendung der Erhaltungssätze empfiehlt der bayerische Lehrplan eine Beschränkung auf eindimensionale Fälle (Bayerisches Staatsministerium für Unterricht und Kultus, 2004, Jahrgangsstufe 10 Physik). Der Vektorcharakter des Impulses wird so reduziert auf die Frage, ob sich Teilchen in die gleiche Richtung oder gerade genau entgegengesetzt gerichtet bewegen. In Aufgaben zum Lösen von Bewegungsgleichungen resultiert diese Frage in der Überlegung des Vorzeichens des Impulses und damit letztlich der Geschwindigkeit. Geschwindigkeiten, bzw. Impulse mit gleicher Richtung weisen gleiche Vorzeichen auf, entgegengesetzt gerichtete werden mit unterschiedlichen Vorzeichen gekennzeichnet. Eine Berechnung mit Hilfe von Vektoren erfolgt nicht, man beschränkt sich auf die Skalarmultiplikation von Masse und den mit dem entsprechenden Vorzeichen versehenem Geschwindigkeitsbetrag zur Berechnung des Impulses. Insgesamt weist das Konzept des Impulses im Unterricht der zehnten Jahrgangsstufe damit sowohl Stoffmengen- als auch Vektorcharakter auf, wenngleich letzterer eher schwach ausgeprägt ist. Im schulfachlichen Kontext zum Impuls stehen insbesondere die Kraft und die kinetische Energie. Die Kraft wird dabei als Grundlage für die Erklärung einzelner Bewegungsabläufe verwendet. Für die Erklärung weiterer Bewegungsabläufe wird dann als Erweiterung das Konzept des Impulses eingeführt. Bei Anwendungsaufgaben zum Thema des Impulses müssen Überlegungen insbesondere zur kinetischen Energie und deren Erhaltung bei meist eindimensionalen Stoßprozessen mit berücksichtigt werden.

Vor diesem Hintergrund wurde als inhaltlicher Rahmen für beide zu entwickelnden Unterrichtseinheiten die Behandlung des Begriffs des Impulses und dessen Erhaltung festgelegt. Der Begriff des Impulses umfasst dabei dessen Definition sowie Stoffmengen- und Vektorcharakter. Beim Vektorcharakter des Impulses wird der Empfehlung des bayerischen Lehrplans zu einer Einschränkung auf eindimensionale Fälle entsprochen. Für die Gestaltung der Problemsituation im FeU bzw. dem Kontext in der BMT wurde ein Frontalzusammenstoß (vollkommen inelastischer Stoß) eines PKW mit einem SUV vorgegeben. Durch die inhaltliche Festlegung den Begriff des Impulses *und* die Impulserhaltung zu thematisieren und dabei Phasen der experimentellen Tätigkeit zu beinhalten, wurde für die Konzeption der Unterrichtseinheiten eine Dauer von circa 90 Minuten veranschlagt. Diese hier formulierten Rahmenbedingungen sollten bei der Gestaltung der Abläufe der beiden Interventionen berücksichtigt werden.

3. Konzeption und Durchführung der Studie

Gestaltung des Ablaufs im FeU

Für die Gestaltung des Ablaufs der Unterrichtseinheit entsprechend der Vorgaben des FeU sind die Vorgaben zum *induktiven Zweig* des FeU bestimmend. Dieser Zweig wird gewählt, wenn die zu behandelnde Problemsituation noch nicht mit bereits bekanntem Wissen gelöst werden kann (Schmidkunz & Lindemann, 1992, S. 35). Aus fachphysikalischer Sicht ist die für die Konzeption vorgegebene Problemsituation ein vollkommen inelastischer Stoß, für welchen den Lernenden die zur Lösung der Situation notwendigen Kenntnisse eben noch nicht zur Verfügung stehen. Die Vorgaben zur Struktur des induktiven Zweiges sind im vorangegangenen Kapitel dargestellt (Siehe: 2.3.3). Die inhaltliche Ausgestaltung dieser Struktur wird im Folgenden kurz erläutert, eine ausführlichere Dokumentation der konzipierten Unterrichtseinheiten findet sich im Anhang (A.1 bis A.3).

Die Denkstufe der Problemgewinnung beginnt mit der Präsentation der Problemsituation. Es sollen Aussagen über die Wucht (umgangssprachlicher Ausdruck für den Impuls) eines Aufpralls mit Fahrzeugen unterschiedlicher Masse und Tempo formuliert werden. Anhand dieser Situation werden die Probleme der Lernenden erfasst und anschließend aufgezeigt. Die Probleme der Lernenden gestalten sich hier in der Art, dass diese nicht über die notwendigen Kenntnisse verfügen. So ist den Lernenden nicht bekannt, wie genau die Wucht (Impuls) von der Masse und dem Tempo der Fahrzeuge abhängt. Die Art der Übertragung der Wucht bei einem Aufprall und Auswirkungen der Wucht auf die Unfallgefahr bei einem Aufprall sind ebenfalls unbekannt.

Die zweite Denkstufe umfasst die Überlegungen zur Problemlösung. Das dafür notwendige Vorwissen (u.a. Variablenkontrolle) wird bereitgestellt. Die Lösung der Probleme soll mit Hilfe von zwei Experimenten erfolgen. Aus einem Experiment soll auf die Abhängigkeiten des Impulses von Masse und Geschwindigkeit geschlossen werden, das zweite Experiment thematisiert eine mögliche Übertragung der Wucht bei einem Zusammenprall.

Die Durchführung des Lösungsvorschlags beinhaltet als erste Phase die Planung des Experimentes. Die zur Verfügung stehenden Materialien werden präsentiert und die Art der Durchführung und die Dokumentation der Experimente besprochen. Anschließend werden beide Experimente durchgeführt und die Ergebnisse der Gruppen im Plenum präsentiert und diskutiert.

Eine ikonische Abstraktion wird nicht explizit eingeplant, sondern ist implizit Teil der veranschaulichenden Darstellung der Ergebnisse der Experimente bei deren Präsentation im Klassenplenum. Die sprachliche Abstraktion der experimentellen Ergebnisse des ersten Experiments liefert die Abhängigkeiten des Impulses von Masse und Geschwindigkeit und die Definition des Impulses. Aus dem zweiten Teilerperiment wird die Impulserhaltung über eine symbolische Auswertung unter Berücksichtigung beim Stoß auftretender Kräfte, sowie der Vektorcharakter des Impulses über eine Detailbetrachtung eines Beispiels

abstrahiert.

Die Sicherung des Wissens über Anwendungsbeispiele erfolgt durch die Bearbeitung eines Musterbeispiels im Kontext der anfänglichen Problemsituation. Zudem suchen die Lernenden nach Beispielen für die Rolle des Impulses und der Impulserhaltung in anderen Fächern, im Alltag oder aus Bereichen der Technik. Die gewonnenen Erkenntnisse und der Weg zu diesen wird anschließend im Plenum wiederholt. Als Lernzielkontrolle ist die Vorhersage des Ausgangs eines Experiments mit Hilfe der Impulserhaltung vorgesehen.

Gestaltung des Ablaufs in der BMT

Die inhaltliche Festlegung für die Konzeption der Unterrichtseinheiten umfassen das Konzept des Impulses und der Impulserhaltung sowie die Abhängigkeiten des Impulses eines Körpers von dessen Masse und Geschwindigkeit. Diese Abhängigkeiten des Impulses sind dabei direkt durch Experimente erfahrbar. Damit kann dieses Wissen aus einer Erfahrung der Lernenden heraus aufgebaut werden und somit entsprechend der Vorgaben des Basismodells *Lernen durch Eigenerfahrung* unterrichtet werden (Wackermann, 2008, S. 141). Dieses Basismodell eignet sich allerdings nicht für die unterrichtliche Gestaltung zur Begrifflichkeit des Impulses und auch nur eingeschränkt für die Impulserhaltung. Für diese beiden inhaltlichen Teilaspekte wurde das Basismodell *Konzeptaufbau* verwendet. Konzepte entsprechen dabei nach Wackermann „Begriffen, die ihre mehr oder weniger komplexe Bedeutung in einem fachlichen Kontext erhalten“ (Wackermann, 2008, S. 145). Dies trifft auf die Begrifflichkeit des Impulses zu. Aber auch die Impulserhaltung erhält seine komplexe Bedeutung eigentlich erst im Hinblick auf die fachlichen Kontexte der Definition des Impulses eines Teilchensystems und vor allem der Definition und Betrachtung eines geschlossenen, isolierten Systems. Dadurch entspricht zumindest in einem weiteren Sinne auch die Impulserhaltung der für das Basismodell *Konzeptaufbau* zugrunde gelegten Definition eines Konzeptes. Zur Umsetzung der inhaltlichen Festlegung sind damit zwei Basismodelle notwendig, welche in einer Unterrichtseinheit kombiniert werden sollen. Eine Kombination von Basismodellen ist grundsätzlich möglich, diese können dabei hintereinander ausgeführt werden, oder ein Basismodell komplett in ein zweites Basismodell eingefügt werden (Oser et al., 1997, S. 22ff.).

Ausgehend von der festgelegten Kontextsituation eines Frontalzusammenstoßes wird die eigene Erfahrung der Lernenden zum Impuls motiviert und erlaubt eine Umsetzung der Handlungskettenschritte eins bis vier des Basismodells *Lernen durch Eigenerfahrung*. Im fünften Schritt dieses Basismodells ist die Anwendung und auch ein Transfer des Lernergebnisses vorgesehen. Das Lernergebnis kann beispielsweise dazu eingesetzt werden auch in anderen Kontexten sinnvolle Erklärungen für Beobachtetes zu generieren. Sehr oft spielt im Zuge anderen Kontexte aber auch die Impulserhaltung eine Rolle. Weitere Kenntnisse zum Impuls und der Impulserhaltung können an dieser Stelle deshalb das weitere

3. Konzeption und Durchführung der Studie

Lernen erleichtern, bzw. erst ermöglichen. Eine Erleichterung des Lernvorgangs beim Schüler ist eine Möglichkeit im Rahmen der BMT, das Einfügen eines Basismodells in ein anderes zu begründen (Oser & Baeriswyl, 2001, S. 1049). In der Interventionsstunde zur BMT erfolgt der Einschub des Basismodells *Konzeptaufbau* deshalb zwischen Handlungskettenschritt vier und fünf des Basismodells Lernen durch Eigenerfahrung (Siehe Abb. 3.1). Erst nach Beendigung des eingefügten Basismodells erfolgt Handlungskettenschritt fünf des Basismodells *Lernen durch Eigenerfahrung*. Die einzelnen Handlungskettenschritte der Basismodelle sind in Tabelle 2.2 dargestellt.

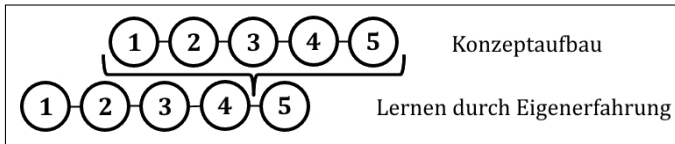


Abbildung 3.1.: Verkettung der Basismodelle für die Konzeption der Unterrichtseinheit entsprechend der BMT

Damit ist die Strukturierung für die Konzeption der Unterrichtseinheit nach den Vorgaben der BMT festgelegt. Die inhaltliche Ausgestaltung dieser Strukturierung wird wie im Falle des FeU kurz erläutert, eine detailliertere Dokumentation des unterrichtlichen Ablaufs kann wiederum dem Anhang (A.4 bis A.6) entnommen werden.

Die Unterrichtseinheit beginnt gemäß den Vorgaben des Basismodells *Lernen durch Eigenerfahrung* mit der Einführung des für die Intervention festgelegten Kontextes. Auch bei der BMT-Unterrichtseinheit sollen demnach Aussagen über die Wucht eines Aufpralls mit Fahrzeugen unterschiedlicher Masse und Tempo formuliert werden. Unklar bei diesen Aussagen bleibt die genaue Abhängigkeit der Wucht eines Körpers von dessen Masse und Tempo, sowie die Art der Übertragung der Wucht bei einem Aufprall. Der Sinn der bevorstehenden Eigenerfahrung besteht in der experimentellen Bearbeitung dieser Fragen. Anschließend erfolgt die Planung der experimentellen Handlungen vor Beginn der Experimente. Es werden die zur Verfügung stehenden Materialien präsentiert, sowie die Ziele der Eigenerfahrung expliziert. Nach der zielbezogenen Durchführung der Experimente werden in der ersten Ausdifferenzierung die experimentellen Ergebnisse durch einen Repräsentanten jeder der Experimentiergruppen im Plenum vorgestellt. In der nachfolgenden Generalisierung der experimentellen Ergebnisse der Gruppen wird festgehalten, dass die Werte für Masse und Tempo von Körpern mit gleicher Wucht indirekt proportional zueinander sind. Das Produkt aus m und v hat dann jeweils den gleichen Wert. Zudem kann sich die Wucht übertragen oder aufteilen bei einem Zusammenprall mit einem zweiten Körper. Dies stellt zugleich einen Teil des Bewusstmachen

des Vorwissens als ersten Schritt im hier eingefügten Basismodell *Konzeptaufbau* dar. Als zweiten Schritt gibt dieses Modell das Vorstellen und Durcharbeiten eines prototypischen Musters vor. Am Beispiel eines Zusammenstoßes eines SUV mit einem PKW werden die Definition, und die Erhaltung des Impulses thematisiert und eine Musteraufgabe durchgearbeitet. Im Rahmen dieser Aufgabe wird anschließend der Vektorcharakter des Impulses und dessen Rolle bei der Impulserhaltung als Merkmale des Konzeptes des Impulses erarbeitet. Die aktive Anwendung dieses Konzeptes erfolgt durch die Bearbeitung von Musterbeispielen, an welchen die Lernenden primär die neu erlernten Sachinhalte festigen und zusätzlich die Rolle des Impulses und der Impulserhaltung in anderen Fächern, im Alltag oder aus Bereichen der Technik erkennen sollen. Als Transfer und Vernetzung folgt ein kurzer Vergleich des Impulses mit der kinetischen Energie im Kontext des vorher verwendeten prototypischen Musters. Damit ist das Basismodell *Konzeptaufbau* abgeschlossen. Zum Abschluss des Basismodells *Lernen durch Eigenerfahrung* erfolgt als letzter Teilschritt der Unterrichtseinheit die Erprobung und Festigung des neuen Konzeptes, indem die Vorhersage des Ausgangs eines Experiments (als neuer Kontext) mit Hilfe der Impulserhaltung formuliert und nachfolgend experimentell überprüft wird.

3.4. Analyse der Unterrichtseinheiten

Eine Analyse der Unterrichtseinheiten wurde im Rahmen der schriftlichen Hausarbeit (Zulassungsarbeit) von Hahn (2014) durchgeführt. Hierzu standen Hahn (2014) je drei Unterrichtsvideos zu beiden konzipierten Einheiten zur Verfügung. Die Auswahl dieser Videos aus dem ersten Teil der Hauptstudie erfolgte per Losverfahren. Für die deskriptive Auswertung mittels einer kategorienbasierten Videoanalyse wurden die Kodiermanuale von Wackermann et al. (2012) für die Interventionsstunde zur BMT und das Kodiermanual von Hansen (2014) für Interventionsstunde zum FeU verwendet. Letztgenanntes Manual entstammt ebenfalls einer Qualifikationsarbeit, die an der Universität Regensburg durchgeführt wurde. Die Arbeit von Hahn (2014) zielte unter anderem darauf ab, die zur Verfügung gestellten Videos eindeutig einer der beiden bei der Konzeption zugrunde gelegten theoretischen Vorgaben zuzuordnen. Als zentrale Funktion der Analyse der Unterrichtseinheiten sollte die Arbeit von Hahn (2014) die Umsetzung der theoretischen Vorgaben untersuchen. Ziel dieser Untersuchung war es, zu ermitteln, zu welchem Grad die videographierten Unterrichtsstunden jeweils als theoriekonform (zur BMT bzw. zum FeU) bezeichnet werden können. Zusätzlich sollten Gemeinsamkeiten und Unterschiede zwischen den beiden Interventionsstunden identifiziert werden. Zunächst werden die Ergebnisse der Arbeit von Hahn (2014) dargestellt und in einzelnen Punkten ergänzt, anschließend folgt ein abschließender Vergleich der beiden konzipierten Unterrichtseinheiten.

3. Konzeption und Durchführung der Studie

Zuordnung und Theoriekonformität

Grundsätzlich konnte Hahn (2014, S. 31) die theoretische Grundlage jeder der sechs Unterrichtsvideos identifizieren und somit eine eindeutige Zuordnung der Videos zu den theoretischen Vorgaben vornehmen. „Im Unterrichtsgeschehen lassen sich die einzelnen Handlungskettenschritte bzw. Denkphasen deutlich beobachten und in die entsprechende Kategorie des Schemas einordnen“ (Hahn, 2014, S. 25). Hahn (2014, S. 25) erkennt bei der Sequenz zum FeU bis auf die ikonische Abstraktion und die Lernzielkontrolle alle Denkphasen in der Unterrichtseinheit. Damit wird die von mir als implizit in der Diskussion der Ergebnisse eingeplante ikonische Abstraktion ebenso wenig erkannt, wie der Einsatz des abschließenden Experiments primär als Lernzielkontrolle und nicht vorrangig zur Anwendung des neuen Wissens. Zwischen der Konzeption der Einheit zum FeU und der nachfolgenden Analyse durch Hahn (2014), bestehen somit Unterschiede in der Zuordnung einzelner Teilphasen zu den theoretischen Vorgaben, nicht jedoch in deren Existenz. Die fünf umfassenderen Denkstufen werden in Übereinstimmung zur ursprünglichen Konzeption von Hahn (2014, S. 22ff.) beschrieben.

Der Unterrichtsverlauf nach der BMT entspricht zunächst „exakt den Operationalisierungen der ersten vier Handlungskettenschritte“ (Hahn, 2014, S. 25) des Basismodells *Lernen durch Eigenerfahrung*, der Übergang zum Basismodell *Konzeptaufbau* hingegen bezeichnet Hahn (2014, S. 26) als schwer zu erkennen und es „bedurfte einer vielfach wiederholten Wahrnehmung der Szene“, um die unterrichtlichen Vorgänge den entsprechenden Handlungskettenschritten zuzuordnen. Insgesamt konnten aber alle theoretisch vorgegebenen Schritte auch in den der BMT zugeordneten Unterrichtsvideos identifiziert werden (Hahn, 2014, S. 29). Zwischen der Konzeption der Unterrichtseinheit zur BMT und deren Analyse durch Hahn treten damit keine Unterschiede in der Auffassung der Phasierung des Ablaufs auf.

„Gemessen an den theoretisch formulierten Ansprüchen bzw. Zielen agiert die Lehrkraft im beobachteten Unterricht auf einem hohen Anforderungsniveau und sehr theoriekonform“ (Hahn, 2014, S. 30), Unterschiede im Grad der Umsetzung der beiden theoretischen Vorgaben wurden nicht festgestellt.

Gemeinsamkeiten und Unterschiede

Hahn (2014, S. 26f.) führt an, dass die Einführung des Kontextes, die Vorbereitung des Experiments, dessen Durchführung und der Diskussion der experimentellen Ergebnisse in beiden Unterrichtseinheiten bis auf vernachlässigbare Unterschiede inhalts- und methodengleich verlaufen. Lediglich hinsichtlich der „beiden Formulierungen ‚Unklar‘ und ‚Probleme‘ unterscheiden sich“ (Hahn, 2014, S. 26) die analysierten Videos bezüglich der Formulierung der noch unbekannten Fachaspekte zum Eingangsbeispiel des Frontalzusammenstoßes. Erst bei der Erarbeitung und Anwendung der Sachinhalte erkennt Hahn (2014) Unterschiede in den unterrichtlichen Umsetzungen der beiden Strukturvorgaben.

„Beim forschenden Unterricht werden alle Elemente aus dem Experiment abstrahiert und der Impuls über das Wechselwirkungsprinzip hergeleitet. Der Inhalt wird fächerübergreifend vernetzt. Im Gegensatz dazu werden bei der Verschränkung aus entdeckendem Lernen und Konzeptaufbau theoretische Zusatzinformationen gegeben und das Konzept innerhalb des Faches angewandt“ (Hahn, 2014, S. 31). Zudem ist nach Hahn (2014) nur beim Unterricht nach dem FeU eine kurze Wiederholung der wesentlichen Ergebnisse zu erkennen, wohingegen die abschließenden Unterrichtsphasen beider Einheiten wiederum identisch bei beiden Interventionsstunden verlaufen. Für Aspekte, welche nicht die Gestaltung der didaktischen Strukturierung betreffen, werden bei Hahn keine Unterschiede zwischen den entwickelten Unterrichtseinheiten aufgeführt. Insgesamt wurde nach Hahn hinsichtlich Inhalt, Dauer und Methodik „der Unterricht nach den beiden didaktischen Konzepten weitgehend gleich abgehalten“ (Hahn, 2014, S. 28). Der Aspekt der Methodik umfasst dabei in der Analyse durch Hahn (2014) die in den Einheiten eingesetzten Experimente und Methoden, sowie die Rolle der Lehrkraft in den jeweiligen unterrichtlichen Phasen (Hahn, 2014, S. 26-29). Der Aspekt der Dauer konnte in der Arbeit von Frau Hahn wiederum nur auf Basis von sechs ihr zur Verfügung stehenden Unterrichtsvideos formuliert werden und kann im Gegensatz zu den Aussagen bezüglich Inhalt und Methodik nicht auf die Gesamtstudie übertragen werden.

Vergleich der Unterrichtseinheiten

Ein Vergleich aller im Rahmen der Studie durchgeführter Unterrichtseinheiten ergibt hinsichtlich der Dauer einen durchschnittlichen Wert von 100 Minuten (Standardabweichung 4.2) für die Einheiten zum FeU und für diejenigen konform zur BMT eine durchschnittliche Dauer von 99.5 Minuten (Standardabweichung 4.2). Die Einheiten gestalten sich demnach in der gesamten Studie vergleichbar auch bezüglich dieses Aspektes. Betrachtet man die beiden Unterrichtseinheiten nicht vor deren jeweiligem theoretischen Hintergrund und den daraus resultierenden vorgegebenen Phasierungen und beschreibt die unterrichtlichen Handlungen aus einer allgemeineren, theorieungebundenen Perspektive, kann eine (nahezu) identische Darstellung der didaktischen Strukturierung für beide Einheiten angegeben werden (vgl. Tabelle 3.1). Zusammen mit der von Hahn (2014) attestierten Vergleichbarkeit beider Einheiten hinsichtlich Inhalt, Dauer und Methodik, erscheint damit eine empirische Untersuchung der Wirkung dieser Einheiten auf den ersten Blick wenig sinnvoll.

Nicht vergleichbar hingegen ist die Ausgestaltung der didaktischen Struktur der einzelnen Phasen. Diese unterscheidet sich insbesondere bei der Auswertung und Analyse der Ergebnisse, dem Durcharbeiten eines Beispiels, sowie der Phase des Transfers und weiterer Anwendungsbeispiele. Die theoretischen Vorgaben des FeU und der BMT unterscheiden sich an diesen Stellen. Sie unterscheiden sich im wesentlichen bei der *Erarbeitung* und der *Verarbeitung* der Sachinhalte (Hahn, 2014, vgl. S. 31). Erfolgt die *Erarbeitung* dieser beim FeU über eine Abstraktion aller Erkenntnisse aus den durchgeführten Experimenten,

3. Konzeption und Durchführung der Studie

FeU		BMT	
<i>allgemein formulierter Ablauf</i>	<i>Medien Methoden</i>	<i>Medien Methoden</i>	<i>allgemein formulierter Ablauf</i>
Hinführen auf Thematik	PP	PP	Hinführen auf Thematik
Planung der Experimente	PP	PP	Planung der Experimente
Durchführung der Experimente	EXP	EXP	Durchführung der Experimente
Präsentation der Ergebnisse	Folien Tafel	Folien Tafel	Präsentation der Ergebnisse
Auswertung u. Analyse der Ergebnisse	PP	PP	Auswertung u. Analyse der Ergebnisse
Durcharbeiten eines Beispiels	PP	PP	Durcharbeiten eines Beispiels
Transfer und weitere Anwendungsbeispiele	PP GA	PP GA	Transfer und weitere Anwendungsbeispiele
Vorhersage eines Experiments	Tafel EXP	Tafel EXP	Vorhersage eines Experiments

Tabelle 3.1.: Allgemein formulierter Ablauf der Unterrichtseinheiten (PP: Power-Point Präsentation; EXP: Experiment; GA: Gruppenarbeit)

sieht die BMT im vorliegenden Beispiel eine Generalisierung von Teilaspekten aus den Experimenten (Basismodell *Lernen durch Eigenerfahrung*), sowie eine Darstellung weiterführender Inhalte (Basismodell *Konzeptaufbau*) vor. Die Verwendung zweier Basismodelle wurde durch die Festlegung des Sachinhaltes notwendig (vgl. Abschnitt 3.3). Bei der *Verarbeitung* der Sachinhalte erfolgen die Anwendungen im FeU nicht „um zu üben und zu gewissen Fertigkeiten zu gelangen, sondern es soll ein beziehungsreiches Wissen mit vielen Querverbindungen aufgebaut werden“ (Schmidkunz & Lindemann, 1992, S. 31). Die fächerübergreifende Vernetzung steht neben der Wiederholung der Erkenntnisse und deren Entstehungsprozess im Fokus dieses Abschnitts. Die Vorgaben zur BMT durch Wackermann et al. (2012, S. 6-12) sehen hingegen vor, dass die neuen Inhalte explizit an Beispielen geübt und angewendet werden sollen. Transfer und Vernetzung sollen auch innerfachlich erfolgen und zusätzlich eine Dekontextualisierung des Sachinhalts erreicht werden.

Zwischen den auf Basis der verwendeten konzeptionellen Modelle entwickelten Unterrichtseinheiten existieren damit Unterschiede, welche eine empirische Untersuchung möglicher Auswirkungen rechtfertigen. Unterschiede allerdings,

die bei einer rein oberflächlichen Betrachtung der unterrichtlichen Struktur nicht in Erscheinung treten (vgl. Tabelle 3.1). Hierfür ist ein Blick in die Tiefenstruktur dieser Einheiten notwendig, denn die untersuchten Unterrichtseinheiten unterscheiden sich nur in Details ihrer Strukturierung.

3.5. Methoden der Datenerhebung

Der vorliegende Abschnitt beginnt mit der Beschreibung der Variablen dieser Untersuchung. Der Ablauf dieser Untersuchung wird daran anschließend dargestellt. Zusätzlich werden die Konzeption der einzelnen Methoden zur Datenerhebung und statistische Kennwerte angegeben, welche die Güte der eingesetzten Methoden kennzeichnen. Grundsätzlich lässt sich die Qualität oder Güte eines eingesetzten Verfahrens anhand von drei zentralen Kriterien festmachen: Objektivität, Reliabilität und Validität (Bortz & Döring, 2006, S. 195).

Das Gütekriterium der Reliabilität (Zuverlässigkeit) beschreibt wie genau, bzw. fehlerfrei das entsprechende Merkmal gemessen wird, wohingegen ein Test als valide bezeichnet werden kann, „wenn er das Merkmal, das er messen soll, auch wirklich misst und nicht irgendein anderes“ (Moosbrugger & Kelava, 2012, S. 13). Sowohl Aspekte zur Reliabilität als auch der Validität werden in den jeweiligen Ausführungen zu den einzelnen eingesetzten Methoden näher beleuchtet. Für die Objektivität der eingesetzten Methoden kann hingegen eine globale Aussage formuliert werden. „Die Objektivität eines Tests gibt an, in welchem Ausmaß die Testergebnisse vom Testanwender unabhängig sind“ (Bortz & Döring, 2006, S. 195). Bortz und Döring (2006, S. 195) unterscheiden innerhalb der Objektivität drei Facetten. Sie sprechen von Durchführungsobjektivität, wenn das Testergebnis vom Untersuchungsleiter unbeeinflusst ist. Ist das Testergebnis zusätzlich von der auswertenden Person (kann auch ungleich dem Untersuchungsleiter sein) unabhängig, liegt sogenannte Durchführungsobjektivität vor. Obliegen die Testwerte weiterhin keiner individuellen Deutung für die Interpretation des Testwertes, so ist auch die dritte Facette der Interpretationsobjektivität gegeben. „Bei standardisierten quantitativen Verfahren, die von ausgebildeten Psychologen oder geschulten Testanweisern unter kontrollierten Bedingungen eingesetzt und ausgewertet werden, ist davon auszugehen, dass perfekte Objektivität vorliegt. In der Tat ist die Objektivität meist ein recht unproblematisches Testgütekriterium“ (Bortz & Döring, 2006, S. 195).

Bei allen eingesetzten Verfahren handelt es sich um standardisierte, quantitative Verfahren die im Rahmen einer Laborstudie (kontrollierbare Bedingungen) stets vom Versuchsleiter durchgeführt und ausgewertet wurden. Die von Bortz und Döring (2006, S. 195f.) formulierten Objektivitätsanforderungen eben an Methoden der quantitativen Datenerhebung sind damit erfüllt. Aus diesem Grund wird im Rahmen der Studie davon ausgegangen, dass die Objektivität

3. Konzeption und Durchführung der Studie

bei allen eingesetzten Verfahren gegeben ist, auf eine empirische Überprüfung der Objektivität wird deshalb verzichtet.

3.5.1. Variablen der Untersuchung

unabhängige Variable

Die Gruppenzugehörigkeit zur Interventionsgruppe FeU oder zur Interventionsgruppe BMT stellt die unabhängige Variable der Untersuchung dar, welche damit durch das Studiendesign vorgegeben ist.

abhängige Variablen

Die abhängigen Variablen hingegen resultieren direkt aus den Forschungsfragen (vgl. Abschnitt 2.4), ihre Betrachtung erfolgt hinsichtlich zweier Ebenen. Die erste Ebene bildet die Einschätzungen der Lernenden über Aspekte der eingesetzten Unterrichtseinheiten. Diese Ebene umfasst die Aspekte der kognitiven Aktivierung, der grundlegenden Bedürfnisse und der wahrgenommenen Qualität der Strukturierung des Unterrichts. Aspekte der ersten Ebene werden mittels eines Schülerfragebogens erhoben. Die zweite Ebene ist die Untersuchung der Lernwirksamkeit der Unterrichtseinheiten in beiden Interventionsgruppen. Hierzu wird ein für diese Studie entwickelter Wissenstest zum Thema Impuls (WTI) eingesetzt.

Kontrollvariablen

Es wird zudem angenommen, dass das Ausmaß des Lernzuwachses dabei von einer möglichen Unterschiedlichkeit in den organisatorischen Rahmenbedingungen zwischen den Interventionsgruppen, den kognitiven Grundfähigkeiten der Lernenden, oder deren Kompetenz zur Erkenntnisgewinnung beeinflusst werden kann. Diese Aspekte stellen somit potentielle Störvariablen dar und sollen in dieser Untersuchung kontrolliert werden. Die Kompetenz zur Erkenntnisgewinnung beschreibt im schulischen Kontext allgemein die Nutzung experimenteller oder anderer Untersuchungsmethoden, sowie den Einsatz von Modellen (Kultusministerkonferenz, 2005, S. 9), im Kontext dieser Arbeit allerdings nur die Nutzung von Experimenten zur Gewinnung von Erkenntnissen, da im wesentlichen nur dieser Teilaspekt für die Lernprozesse der Schüler in den konzipierten Unterrichtseinheiten relevant erscheint. Aufgrund der Stellung der unterrichtlichen Experimente und deren Funktion im Lernprozess in beiden Interventionen kann ein Einfluss der Kompetenz der experimentellen Erkenntnisgewinnung auch auf den Lernzuwachs vermutet werden, wodurch sich die Verwendung dieser Kompetenz als Kontrollvariable innerhalb der Untersuchung rechtfertigt. Die Ausprägung dieser Kompetenz soll bei den Lernenden mit Hilfe eines Tests erhoben werden. Die Verarbeitung und Anwendung von erworbenem Wissen kann zudem von den kognitiven Grundfähigkeiten der Lernenden abhängen, weshalb diese Fähigkeiten ebenfalls kontrolliert und

wiederum durch einen Test bestimmt werden sollen.

Als mögliche Einflussgrößen auf Seiten der organisatorischen Rahmenbedingungen werden die Dauer der Intervention, deren zeitlicher Beginn, das Geschlecht der Lernenden und deren Wahl des schulischen Zweiges (naturwissenschaftlicher Schwerpunkt oder andere) in der Untersuchung kontrolliert. Schulklassen der zehnten Jahrgangsstufe des naturwissenschaftlichen Zweiges an bayerischen Gymnasien haben in Physik den identischen verpflichtenden Lehrplan wie Klassen eines nicht naturwissenschaftlichen Zweiges, allerdings eine Unterrichtsstunde Physik mehr pro Schulwoche (sogenannter Profilbereich). Die Schüler des Profilbereichs vertiefen dabei einzelne Inhalte aus einer im Lehrplan angegebenen Liste von möglichen Themen bevorzugt in schülerzentrierten Unterrichtsformen (Bayerisches Staatsministerium für Unterricht und Kultus, 2004, vgl.). Eine höhere Routine von Schülern des naturwissenschaftlichen Zweiges bei Schülerexperimenten ist demnach möglich. Dies könnte wiederum die Wirkung der Interventionen beeinflussen, da in diesen eben Schülerexperimente eine entscheidende Rolle spielen. Demzufolge soll dieser Aspekt in der Untersuchung kontrolliert werden.

Zusätzlich wurde im zweiten Teil der Hauptstudie auf Schülerseite die Einschätzung zur Lehrkraft erfragt. Nach dem ersten Teil der Hauptstudie waren dem Versuchsleiter und damit der unterrichtenden Person in beiden Interventionsgruppen erste Ergebnisse der Untersuchung bekannt. Dies kann das Verhalten des Versuchsleiters unbewusst beeinflussen. Um etwaige Unterschiede im Verhalten der Lehrkraft (Versuchsleiter) zwischen den Interventionsgruppen zu identifizieren, wurde demnach zur Erfassung des Verhaltens des Versuchsleiters ein Fragebogen eingesetzt, welcher die Wahrnehmung der Lernenden hinsichtlich des Interesses des Lehrenden und dessen Beitrag zu einem lernförderlichem Klima und der Motivation der Lernenden erhebt. Die Zeiträume für den ersten bzw. der zweiten Teil der Hauptstudie, sowie der zeitliche Ablauf der Interventionen und der Datenerhebung ist im nachfolgenden Abschnitt dargestellt.

3.5.2. Ablauf der Untersuchung

Vor Beginn der Untersuchung wurde eine Pilotierung der eingesetzten Verfahren und des geplanten Ablaufs der Interventionen und der Datenerhebung an drei repräsentativen Schulklassen (zehnte Jahrgangsstufe Gymnasium Bayern) vorgenommen. Im Anschluss daran, wurde der Ablauf der Untersuchung, der zeitliche Bedarf einzelner Abschnitte, sowie der Wissenstest zum Thema Impuls (siehe Abschnitt 3.5.3) für die Hauptstudie angepasst. Für die Durchführung der Hauptstudie wurde ein Zeitraum von Oktober 2013 bis März 2015 vorgegeben. Dieser Zeitraum umfasst die Schuljahre 2013/2014 (erster Teil der Hauptstudie) sowie 2014/2015 (zweiter Teil der Hauptstudie). Der Ablauf der Interventionen

3. Konzeption und Durchführung der Studie

wurde in beiden Teilstudien konstant gehalten, der Ablauf der Datenerhebung leicht verändert. Der gesamte Untersuchungsplan beider Teilstudien ist in Abbildung 3.2 dargestellt.

1. Teil der Hauptstudie (Oktober 2013 - März 2014)							
Ablauf	Begrüßung	WTE	WTI	Intervention FeU oder BMT	Schülerfragebogen	WTE	WTI
Dauer	3 min	6 min	18 min	90 min	5 min	6 min	15 min

2. Teil der Hauptstudie (Oktober 2014 - März 2015)						
Ablauf	Begrüßung	KFTk	WTI	Intervention FeU oder BMT	Schülerfragebogen	WTI
Dauer	3 min	7 min	18 min	90 min	5 min	15 min

WTE := Wissenstest zur Kompetenz der Erkenntnisgewinnung
 WTI := Wissenstest zum Impuls
 KFTk := Kurztest zur Subfacette N2 des KFT

Abbildung 3.2.: Ablauf der Datenerhebung und der Intervention in den Teilstudien

Nach einer standardisierten Begrüßung begann die erste Phase der Datenerhebung (WTE bzw. KFTk und WTI). Direkt im Anschluss begann die Intervention, auf welche wiederum direkt nachfolgend der Fragebogen von den Lernenden beantwortet wurde. Nach einer kurzen Pause von ca. zehn Minuten folgte der letzte Teil der Datenerhebung (WTE und WTI oder nur WTI). Die Notwendigkeit einer kurzen Pause zeigte sich in den Erfahrungen aus der Pilotierung.

3.5.3. Wissenstest Impuls (WTI)

Um das Fachwissen zum Impuls vor und nach der Intervention zu erfassen, wurden zwei parallele Wissenstests entwickelt. Paralleltest A sollte den Wissensstand vor und Paralleltest B den Wissensstand nach der Unterrichtseinheit messen. Das Fachwissen sollte dabei möglichst inhaltsvalide erfasst werden. Für eine inhaltsvalide Messung ist es erforderlich, „das zu messende Konstrukt in seinen wichtigsten Aspekten erschöpfend“ (Bortz & Döring, 2006, S. 200) zu erfassen. Auf Basis der Analyse des Sachinhaltes (vgl. Abschnitt 3.3) wurde für die Sachstruktur des Paralleltests festgelegt, dass dieser Items zur Begrifflichkeit und Erhaltung des Impulses, Items zu dessen Stoffmengen- oder Vektorcharakter, Items zu Stoffmengen- und Vektorcharakter, sowie Items zu Impuls- und Energieerhaltung beinhalten sollte. Aufgaben zum elastischen und unelastischen Stoß sollten in etwa zu gleichen Teilen vertreten sein. Die Konstruktion

der Paralleltests erfolgte analog zu den vier Schritten, welche Bortz und Döring (2006, S. 197f.) für eine entsprechende Testkonstruktion empfehlen. In einem ersten Schritt wurde ein Itempool von 60 Aufgaben entworfen, wobei eine Planung von möglichen parallelen Aufgaben und die festgelegte Sachstruktur berücksichtigt wurden. Als zweiter Schritt erfolgte die Itemanalyse zuerst auf Basis einer studentischen Stichprobe ($N=39$, Lehramt Physik) und im Anschluss auf Basis einer Schülerstichprobe ($N=27$, eine Klasse, zehnte Jahrgangsstufe, Pindl Gymnasium Regensburg). Alle Items wurden so mit ihren jeweiligen Schwierigkeitsindizes und ihren Trennschärfekoeffizienten gekennzeichnet. Im dritten Schritt erfolgte die Erstellung der Paralleltests, indem je zwei Items mit jeweils vergleichbaren Kennwerten ein Zwillingsspaar bilden. Aus insgesamt 22 Zwillingsspaaren wurden abschließend (vierter Schritt) per Los je ein Item jedes Paares entweder Paralleltest A oder B zugeordnet. Die Reliabilität der so gebildeten Paralleltests wurden nachfolgend über die Paralleltestmethode (Bortz & Döring, 2006, S. 197f.) an einer dritten Stichprobe ($N=77$, drei Klassen, zehnte Jahrgangsstufe, Pindl Gymnasium Regensburg) bestimmt, blieb aber unter dem angestrebten Wert von 0.8. Dieser Wert wurde als Ziel der Reliabilität der Paralleltests angestrebt, da Rost (2013, S. 178) diesen Wert als untere Schranke für sehr gute Reliabilitäten von Leistungstests angibt und das Ziel des Wissenstests eben eine sehr wenig fehlerbehaftete Messung des Wissensstandes sein sollte. Demzufolge erfolgte eine Überarbeitung der Paralleltests und die Reliabilität dieser überarbeiteten Versionen wurde an einer vierten Stichprobe ($N=46$, zwei Klassen, zehnte Jahrgangsstufe, Ostendorfer Gymnasium Neumarkt) erneut ermittelt. Die erneute Berechnung der Reliabilität anhand der Paralleltestmethode ergab für die in der Studie geplanten Testformen eine Reliabilität von 0.83.

Für die Pilotierung der Interventionen anhand dreier Schulklassen wurden die so konzipierten Paralleltests wie geplant eingesetzt. Die Auswertung der Tests lieferte jedoch widersprüchliche Ergebnisse. So erreichten die teilnehmenden Schulklassen teilweise im Test nach der Intervention einen niedrigeren Gruppenmittelwert als vor der Intervention. Eine Befragung einzelner teilnehmender Schüler deutete einerseits auf eine Verunsicherung aufgrund der Ähnlichkeit der Aufgaben und andererseits auf ein verändertes Antwortverhalten hin. So gaben einige Befragte an, Aufgaben im Nachtest rein *taktisch* beantwortet zu haben, ohne dabei die Aufgabe erneut kritisch zu beurteilen (Wenn sich in Test A der schwerere Wagen nach links bewegt, dann muss es in Test B bestimmt der leichtere sein). Um solche möglichen Verzerrungen bei der Datenerhebung zu vermeiden, wurde vor und nach der Intervention von jedem Lernenden der gleiche Wissenstest bearbeitet. Zusätzlich wurde dieser um vier Aufgaben erweitert. Die Auswertung der Besuche von weiteren Schulklassen ergab nach diesen Änderungen hinsichtlich des Lernzuwachses keine Widersprüchlichkeiten, weshalb diese Form des Wissenstests im weiteren Verlauf der Studie beibehalten wurde.

3. Konzeption und Durchführung der Studie

Die Reliabilität der Endform des Wissenstests zum Thema Impuls wurde nicht mehr über die Paralleltestmethode errechnet. Dies hätte eine erneute Pilotierung und damit eine zeitliche Verzögerung der Hauptstudie und damit eine Verringerung der Stichprobe der Hauptstudie bedingt. Die Reliabilität wurde nun über die Testhalbierungsmethode (Split-Half-Reliabilität) zu 0.72 (unter Berücksichtigung der Spearman-Brown-Korrektur) ermittelt. Rost (2013, S. 178) kennzeichnet die Reliabilität bei (Schul-)Leistungstests als „gut“, wenn diese einen Wert von 0.75 übersteigt, als „sehr gut“ bei einem Wert über 0.8. „Bei Forschungsprojekten, in denen man nur an einem Gruppenvergleich interessiert ist, kann die Reliabilität niedriger liegen“ (Rost, 2013, S. 179), soll aber immer mindestens 0.55 betragen. Rost legt dabei die Paralleltest- bzw. die Retest-Reliabilitätskoeffizienten zugrunde und nicht die in dieser Studie durchgeführte Split-Half-Reliabilität. Allerdings ist nach Bortz die „Testhalbierungsmethode nur eine Sonderform der Paralleltestmethode“ (Bortz & Döring, 2006, S. 198) und demnach sollten die Koeffizienten grundsätzlich vergleichbar sein. Insgesamt betrachtet, kann die Split-Half-Reliabilität der Endform des Wissenstests von 0.72 als zufriedenstellend für die geplante Untersuchung (u. a. zum Vergleich der Mittelwerte zweier Gruppen) betrachtet werden. Eine zur Kontrolle durchgeführte Berechnung einer Reliabilität des Nachtests über Cronbachs α ergab ebenfalls einen Wert von 0.72.

Die endgültige Form des WTI umfasst insgesamt 22 Items, davon 2 Items zum Begriff des Impulses, 6 Items zum Stoffmengencharakter, 4 Items zum Vektorcharakter, 7 Items zum Stoffmengen- und Vektorcharakter und 3 Items zu Impuls- und Energieerhaltung. Von diesen 22 Items lassen sich zudem jeweils 9 Aufgaben dem elastischen bzw. dem inelastischen Stoß zuordnen. Insgesamt wird angenommen, dass die Items des eingesetzten Tests auch „tatsächlich die Menge möglicher relevanter Items“ (Moosbrugger & Kelava, 2012, S. 170) für den Inhaltsbereich des Impulses repräsentieren und der Test damit nach Moosbrugger und Kelava (2012) als inhaltsvalide betrachtet werden kann. Eine externe Validierung, oder eine Überprüfung einer Konstruktvalidität des WTI wird nicht vorgenommen. Für die Überprüfung der Konstruktvalidität fehlt das grundlegende theoretische Konstrukt, auf dessen Basis überprüfbare Vorhersagen formuliert werden könnten. Auf eine externe Validierung wurde verzichtet, da das Ziel des WTI nicht in der Formulierung einer diagnostischen Entscheidung auf Basis der ermittelten Testwerte über das Fachwissen zum Impuls der Lernenden liegt und Moosbrugger und Kelava (2012, S. 161f.) eine entsprechende externe Validierung damit als nicht notwendig einstufen.

Auf eine mögliche Auswertung von Subskalen des WTI (Begriff des Impulses, Stoffmengencharakter, Vektorcharakter, Stoffmengen- und Vektorcharakter, Impuls- und Energieerhaltung, elastische Stöße und inelastische Stöße) wird aufgrund zu geringer Reliabilitäten (Werte von 0.13 bis 0.59; Cronbachs α) verzichtet.

Der WTI stellt insgesamt eine Mischform aus Speed- und Powertest dar, da

Aufgabentypen verschiedenen Schwierigkeitsgrades enthalten sind und zudem eine feste Zeitspanne zur Bearbeitung der Items festgelegt werden sollte, um zu einem einheitlichen Ablauf der Untersuchung beizutragen. Für eine Mischform soll dabei die Testzeit gerade so bemessen werden, dass auch die Testpersonen mit der größten Leistungsfähigkeit nicht alle Items korrekt beantworten (Lienert & Raatz, 1998, S. 34). Nach den Erfahrungswerten aus der Pilotstudie wurde demnach die Zeitvorgaben von 18 Minuten vor und 15 Minuten nach der Intervention festgelegt.

3.5.4. Test zur Kompetenz zur Erkenntnisgewinnung (WTE)

Wie im Abschnitt 3.5.1 dargelegt, sollte die Kompetenz zur experimentellen Erkenntnisgewinnung im Rahmen der Studie als Kontrollvariable auf Seiten der Lernenden mit Hilfe eines Tests erhoben werden. Als Vorlage zur Konzeption dieses Tests wurde der prozessbezogenen Kompetenzen-Test (PK-Test) von Klos (2008) verwendet. Der Schwerpunkt dieses Tests, wurde „auf solche Kompetenzen gelegt, die für die Planung, Durchführung und Auswertung von Experimenten häufig benötigt werden“ (Klos, 2008, S. 54). Diese Kompetenzen im Umgang mit Experimenten finden sich auch in den Bildungsstandards im Fach Physik für den mittleren Schulabschluss für den Bereich der Erkenntnisgewinnung wieder (Kultusministerkonferenz, 2005, S. 11). Die Zielsetzung des zu entwickelnden Tests zur Messung der Kompetenz der Erkenntnisgewinnung stimmt damit im wesentlichen mit dem Schwerpunkt der PK-Tests überein. Die Reliabilität des PK-Tests wird über (Cronbachs α) zu 0.63 berechnet, er besteht insgesamt aus 24 Items, für deren Beantwortung eine Dauer von 40 Minuten vorgesehen sind (Klos, 2008, S. 65). Aus Gründen der Testökonomie werden nur einzelne Items aus dem PK-Test entnommen. Teilweise wurden Aufgaben aus dem PK-Test auch auf die spezifische Zielgruppe und Verwendungszweck angepasst. Der Test zur Messung der Kompetenz der Erkenntnisgewinnung sollte die drei von der Kultusministerkonferenz (2005) angegebenen Anforderungsbereiche mit jeweils zwei Items abdecken.

Aus dem PK-Test wurden hierzu zwei Aufgaben zum Anforderungsbereich *Fachmethoden beschreiben* übernommen (Item 3 und 18 des PK-Tests), zwei Aufgaben zum Bereich *Fachmethoden nutzen* angepasst (Item 4 und 23) und zum Bereich *Fachmethoden problembezogen auswählen und anwenden* wiederum zwei Aufgaben übernommen (Item 8 und 24) (Klos, 2008; Kultusministerkonferenz, 2005, vgl.). Bei der Anpassung der Items 4 und 23 wurde die Grundidee und das Design der Aufgabe weitestgehend übernommen, jedoch der inhaltliche Fokus auf die indirekte Proportionalität gelegt (vorher: direkte Proportionalität), da hiermit eine größere Passung zu den experimentellen Aufgaben der Lernenden während der Unterrichtseinheit erreicht wurde. Der in der Studie eingesetzte Test ist im Anhang dargestellt (B.7 und B.8). Die Testdauer wurde aufgrund der Erfahrungswerte der Pilotstudie auf sechs Minuten festgelegt.

3. Konzeption und Durchführung der Studie

Die Reliabilität des in der Studie eingesetzten Tests wurde ebenfalls über Cronbachs α bestimmt und zu 0.52 berechnet. Dieser Wert liegt unter dem entsprechenden Wert von Klos (2008) und auch unter der nach Rost (2013) minimalen Schranke von mindestens 0.55 für die Verwendung eines Tests in einer Studie. Diese Reliabilität ist damit nicht zufriedenstellend, die Ergebnisse aus dem Test zur Messung der Kompetenz der Erkenntnisgewinnung sind damit auch nur eingeschränkt belastbar, weshalb dieser Test nur in der ersten Phase der Hauptstudie eingesetzt und nachfolgend durch einen Test zu den kognitiven Fähigkeiten der Lernenden ersetzt wurde.

Durch die Abdeckung der drei Anforderungsbereiche mit jeweils zwei Aufgaben wurde versucht das Kriterium der Inhaltsvalidität grob abzudecken. Der hier eingesetzte Test umfasst dabei nur sechs Items. Mit dieser Anzahl an Items gelang es nicht, die Ausprägung der Kompetenz zur experimentellen Erkenntnisgewinnung repräsentativ zu erfassen. Auf eine größere Anzahl an Testitems wurde aber aufgrund von testökonomischen Aspekten verzichtet, da der WTE keinen Schwerpunkt dieser Arbeit darstellt. Das Kriterium der Inhaltsvalidität ist demnach kritisch zu beurteilen. Wie beim WTI wird auf die Überprüfung der Konstruktvalidität und eine externe Validierung verzichtet.

3.5.5. Kognitiver Fähigkeitstest (KFT)

Um die kognitiven Grundfähigkeiten der Lernenden zu kontrollieren, wird die KFT-Subskala N2-Figurenanalogien erhoben. Heller und Perleth (2000, S. 45) geben an, dass mit Hilfe dieser Subskala „gleichermaßen abstraktlogische und anschauungsgebundene bzw. konstruktive Aspekte intelligenten Verhaltens“ erfasst werden können. Ein Teil der erhobenen kognitiven Grundfähigkeiten bildet damit das Abstraktionsvermögen der Lernenden ab. Dieses Abstraktionsvermögen liegt dabei im Fokus des FeU, soll dadurch gefördert werden (Lindemann, 1989, S. 20f.), kann aber auch eine Grundvoraussetzung für den erfolgreichen Einsatz dieses Verfahrens darstellen.

Die kombinierte Re- und Parallel-Test-Reliabilität der Subskala N2 führen Heller und Perleth (2000, S. 21) für die zehnte Jahrgangsstufe mit 0.9 an. Für diese Jahrgangsstufe beinhaltet die Skala der Figurenanalogien insgesamt 25 Items, bedarf einer Instruktion der Länge 2.5 Minuten und soll innerhalb von 8 Minuten durch die Lernenden beantwortet werden (Heller & Perleth, 2000, S. 9-21). Für die Durchführung des Tests stehen zwei verschiedene Testformen zur Verfügung. Diese weisen bei der in der Analyse der Verteilungsmerkmale der KFT-Subtests verwendeten Stichprobe aber unterschiedliche Mittelwerte für die Anzahl richtiger Antworten auf. Bei einem Einsatz in einer zehnten Jahrgangsstufe an Gymnasien berichten Heller und Perleth (2000, S. 82) von einer Mittelwertdifferenz von 1.47. Eine mögliche Mittelwertdifferenz muss damit auch für die in dieser Studie verwendeten Stichprobe angenommen werden.

Die Validität des KFT-Tests wurde hinsichtlich der Kriteriums-, Konstrukt und Faktoriellen Validität überprüft. Die entsprechenden Kennwerte und Ergebnisse sind dem Manual zum KFT zu entnehmen (Heller & Perleth, 2000, S. 28-47)

Aufgrund der zumindest teilweisen Erfassung auch abstraktlogischer Aspekte, welche im Kontext zum FeU relevant sein könnten, sowie der im Vergleich zu anderen Subskalen des KFT geringen Zeitbedarfs, wird diese Subskala des KFT in der Studie eingesetzt. Um den Zeitbedarf aus testökonomischen Gesichtspunkten noch weiter zu reduzieren, wurde wie bei Seidel, Rimmele und Dalehefte (2003, S.340f) ein Kurztest der Subskala eingesetzt. Dieser Kurztest umfasst bei Seidel et al. (2003) sieben Items und weist eine Reliabilität von 0.6 (Cronbachs α) auf. Um diese Reliabilität des Kurztests etwas zu steigern, wurde der Umfang des Kurztests für die vorliegende Untersuchung auf elf Items angehoben. Als Reliabilität der in der Studie eingesetzten Kurzform der Subskala N2 des KFT ergab sich ein Wert von 0.65 (Cronbachs α , bei $N = 503$). Dieser Wert liegt damit über dem Mindestwert von 0.55 und kann damit in der Studie eingesetzt werden, liegt aber deutlich unter dem als gut gekennzeichneten Bereich von über 0.75 (Rost, 2013, S. 178f.), weshalb die Reliabilität von 0.65 insgesamt als akzeptabel für die Erhebung einer Kontrollvariablen betrachtet wird. Für die Testdauer des Kurztests mit elf Items wurde ein Wert von 3.5 Minuten festgelegt (ganze Subskala N2: 25 Items; 8 Minuten).

Die beiden in der Untersuchung verwendeten Testformen A und B weisen auch für die Stichprobe dieser Studie unterschiedliche Mittelwerte ($M_A = 8.3$; $M_B = 9.1$) für die Anzahl richtiger Antworten auf. Dieser Unterschied ist höchstsignifikant ($p < 0.001$; $t(500) = -4.23$) bei einer Mittelwertdifferenz von 0.8. Um die Testwerte miteinander vergleichen zu können und um diese Unterschiedlichkeit der Mittelwerte der beiden Testformen zu eliminieren, wurden die Mittelwerte für die Anzahl richtiger Antworten innerhalb der beiden Testformgruppen z-standardisiert (Rasch et al., 2006a, S. 24). Dies bedingt, dass nun die Mittelwerte der z-standardisierten Daten beider Testformen identisch Null sind und die Daten jeweils eine Streuung von $\sigma = 1$ aufweisen.

3.5.6. Schülerfragebogen

Zur Erfassung der *Grundlegenden Bedürfnisse*, der *kognitiven Aktivierung*, der *Wahrnehmung der Strukturierung der Unterrichtseinheiten* und der *Einschätzung des Lehrenden* (nur zweiter Teil der Hauptstudie) wurde ein Schülerfragebogen eingesetzt. Dieser Schülerfragebogen ist im Anhang vollständig abgebildet.

Die Wahrnehmung von Aspekten der Unterrichtsqualität (hier: Erfüllung der grundlegenden Bedürfnisse, kognitive Aktivierung und Strukturierung des Unterrichts) hängt dabei stark von der Beurteilerperspektive ab (Clausen, 2002, S. 186). Clausen (2002) unterscheidet hierbei zwischen der Schüler- und Lehrerperspektive sowie derjenigen von externen Beobachtern. Generell liege dabei

3. Konzeption und Durchführung der Studie

keine der Sichtweisen „näher an einer »Unterrichtswirklichkeit« als die der anderen Perspektiven“ (Clausen, 2002, S. 186). „Eine Erhebung der Schülersicht gewinnt ihre Bedeutung durch die vergleichsweise engen Zusammenhänge zu Leistung und Interesse bzw. deren Entwicklung“ (Clausen, 2002, S. 188). Im Fokus der geplanten Untersuchung liegt dabei gerade die Entwicklung der Leistung der Lernenden und hierfür ordnet Clausen (2002) der Wahrnehmung der Schüler die größte Bedeutung zu. Aus diesem Grund werden obige Aspekte der Unterrichtsqualität mittels eines Schülerfragebogens erhoben. Auf die Erhebung der Wahrnehmung dieser Aspekte aus weiteren Perspektiven wird im Rahmen dieser Arbeit verzichtet. Es könnte aber im Anschluss auf Basis einer Auswertung der videographierten Unterrichtseinheiten die Erhebung der Wahrnehmung aus Sicht externer Beobachter erfolgen. Bei der Erfassung der Schülersicht weist Clausen (2002, S. 188) zusätzlich darauf hin, dass die Schüler zwar grundlegend zu einer differenzierten Bewertung fähig sind, diese aber von einer globalen Beurteilung der Lehrkraft überlagert sein könnte. Da in beiden Interventionen aber die gleiche Lehrkraft unterrichtet, sollte bei einem vergleichbarem Verhalten der Lehrkraft in beiden Gruppen auch ein vergleichbares globales Urteil in die erhobenen Daten beider Gruppen eingehen, welches sich demnach nicht entscheidend auf mögliche Gruppenmittelwertunterschiede auswirken sollte.

Die Skalen der *Grundlegenden Bedürfnisse* und der *kognitiven Aktivierung* wurden von Berger (2007, S. 84f.) übernommen. Zur Erhebung der *Grundlegenden Bedürfnisse* sind bei Berger (2007) dazu je zwei Items zur sozialen Eingebundenheit ($\alpha = 0.65$), dem Kompetenzerleben ($\alpha = 0.70$) und dem Autonomieerleben ($\alpha = 0.70$) formuliert, die entsprechenden Reliabilitäten sind in der Klammer angegeben. Die Skala der *kognitiven Aktivierung* umfasst drei Items zu tiefen Lernstrategien, sowie zwei Items zum kognitiven Engagement, die Reliabilität gibt Berger (2007, S. 84f.) mit $\alpha = 0.64$ an. Eine Neuberechnung der Reliabilitäten (Cronbachs α) der Skalen anhand der Untersuchungss Stichprobe ergab dabei teilweise abweichende Werte. So lag die Reliabilität der Skala der sozialen Eingebundenheit bei $\alpha = 0.34$ und für die Skala des Autonomieerlebens wurde $\alpha = 0.44$ berechnet. Beide Reliabilitäten liegen dabei unter der von Rost (2013) formulierten Mindestgrenze von 0.55. Ein möglicher Grund für die beobachteten Abweichungen könnte in der Unterschiedlichkeit der verwendeten Skalen liegen. Berger setzt eine fünfstufige Skala ein, wohingegen in dieser Studie eine vierstufige Skala für die Schülerfragebögen verwendet wurde. Die Entscheidung zum Einsatz einer vierstufigen Skala erfolgte, da einer Tendenz zur Mitte im Antwortverhalten der Lernenden entgegengewirkt werden sollte (Moosbrugger & Kelava, 2012, S. 60f.). Für die Skalen zum Kompetenzerleben und zur kognitiven Aktivierung mit $\alpha = 0.68$ bzw. $\alpha = 0.64$ ergaben sich hingegen akzeptable und zur Originalstudie vergleichbare Kennwerte.

Die Items zur *Wahrnehmung der Strukturierung* wurden einem Beobachtungsfragebogen für Lehrkräfte aus der IPN-Studie Lehr-Lern-Prozesse im Physikunter-

richt entnommen und für den in dieser Studie eingesetzten Schülerfragebogen modifiziert. Dieser Beobachtungsfragebogen zur Zielorientierung im Physikunterricht beinhaltet dabei die vier Subfacetten *Orientierung am Ziel*, *Klärung der Anforderungen*, *Organisation des Lernstoffs*, sowie *Schlüssigkeit im Unterrichtsverlauf* (Trepke et al., 2003, S. 201ff.). Die *Zielorientierung* (z.B. Zielbekanntgabe) und die *Organisation des Lernstoffs* (z.B. Zusammenfassung wichtiger Fakten oder Verknüpfung mit Vorwissen) betreffen primär Elemente der kognitionspsychologisch verstandenen Strukturierung. Elemente der didaktischen Strukturierung finden sich in der *Schlüssigkeit im Unterrichtsverlauf* (z.B. Schlüssige Aneinanderknüpfung der Unterrichtssequenzen). Die Skala der *Klärung der Anforderungen* bezieht sich hingegen stärker auf das Konstrukt der Klarheit und wird nicht in den Schülerfragebogen übernommen. Die Skalen zur kognitionspsychologisch verstandenen Strukturierung werden in den Schülerfragebogen integriert, um zu überprüfen ob die Lernenden die Unterrichtseinheiten hinsichtlich dieser Subfacette der Strukturierung unterschiedlich bewerten, obwohl diese Facette nicht im direkten Fokus der Untersuchung steht. Das Vorgehen bei der Modifikation der Items soll folgendes Beispiel verdeutlichen:

- Original: „Wichtige Fakten werden durch die Lehrkraft explizit hervorgehoben und zusammengefasst.“ (Trepke et al., 2003, S. 217)
- Modifikation: „Ich finde, dass die Lehrkraft wichtige Fakten deutlich hervorgehoben und zusammengefasst hat.“

Jedes Item im modifizierten Schülerfragebogen beginnt mit *Ich* oder *Mir*, je nachdem ob eine Aktiv- oder Passivsatzgestaltung für die Formulierung gewählt werden musste. Fremdwörter oder ungebräuchliche Begriffe des Beobachtungsfragebogens wie *salient* wurden im Schülerfragebogen ersetzt oder umschrieben. Die Reliabilitäten (Cronbachs α) der modifizierten Skalen an der Untersuchungsstichprobe liegen bei 0.64 für die Skala der Organisation des Lernstoffs, bei 0.58 für die Schlüssigkeit im Unterrichtsverlauf und bei 0.65 für die Orientierung am Ziel. Sämtliche Reliabilitäten liegen damit über dem von Rost (2013) formulierten Mindestwert für Forschungsprojekte mit dem Ziel eine Gruppenmittelwertvergleichs, allerdings auch unter einem guten (0.75) oder zufriedenstellenden (0.70) Wert für die Reliabilität (bei Leistungstests) (Rost, 2013, S. 178f.).

Im zweiten Teil der Hauptstudie wurde der Schülerfragebogen zur Kontrolle des Versuchsleitereffektes um eine Nachbefragung der Untersuchungsteilnehmer hinsichtlich Interesse und Wirkung des Untersuchungsleiter ergänzt. Die Skala der *Erfassung der Wahrnehmung des Interesses des Lehrenden* stammt aus der Teilfacette Unterstützende Lehr-Lern Bedingungen aus dem Schülerfragebogen der IPN-Videostudie Lehr-Lern-Prozesse im Physikunterricht von Trepke et al. (2003, S. 366). Diese Skala besteht aus drei Items. Trepke et al. (2003, S. 366) geben die Reliabilität dieser Skala mit 0.73 (Cronbachs α) an. Für die Untersuchungsstichprobe errechnet sich ein immer noch akzeptabler Wert von 0.63. Die

3. Konzeption und Durchführung der Studie

Wirkung des Untersuchungsleiters sollte zusätzlich über dessen Einfluss auf das Unterrichtsklima und dessen motivierende Wirkung auf die Lernenden erfasst werden. Hierzu wurde die Skala *Lernförderliches Klima und Motivation* aus dem KMK-Projekt EMU (Evidenzbasierte Methoden der Unterrichtsdiagnostik und Unterrichtsentwicklung) entnommen. Diese Skala umfasst insgesamt sechs Items. Auf eine Berechnung etwaiger Gütekriterien wurde in diesem Projekt verzichtet, da dieser Fragebogen eine Reflexion über Unterricht ausschließlich auf der Ebene konkreter Items ermöglichen sollte (Helmke, 2015, vgl.). Die Berechnung der Reliabilität dieser Skala an der Untersuchungsstichprobe ergab einen akzeptablen Wert von 0.65 (Cronbachs α). Fasst man beide Teilskalen zu einer Skala der *Einschätzung der Lehrkraft* zusammen, ergibt sich eine gute Reliabilität von 0.75.

3.6. Stichprobe

Am ersten Teil der Hauptstudie nahmen $N = 291$ Schüler (FeU: $N = 148$; BMT: $N = 143$) aus 12 Schulklassen teil, am zweiten Teil beteiligten sich $N = 505$ Schüler (FeU: $N = 243$; BMT: $N = 262$) aus 20 Schulklassen. Insgesamt wirkten damit $N = 796$ Schüler (davon 425 Mädchen) aus 32 Schulklassen der zehnten Jahrgangsstufe an bayerischen Gymnasien aus dem Umkreis von Regensburg an der Hauptstudie mit. Diese 32 Schulklassen entstammen dabei 12 verschiedenen Gymnasien und wurden von 24 verschiedenen Lehrkräften an ihren jeweiligen Schulen unterrichtet. Grundsätzlich wurde die Zugehörigkeit zur Interventionsgruppe ausgelost, besuchten allerdings mehrere Klassen einer Schule das Lernforschungslabor der Universität oder nahmen mehrere Klassen der gleichen Lehrkraft an der Studie teil, erfolgte die Zuweisung dieser Klassen auf die Interventionsgruppen im Sinne einer Gleichverteilung hinsichtlich dieser Faktoren. Die Verteilung der Gesamtstichprobe auf die beiden Interventionsgruppen ist in Tabelle 3.2 ersichtlich.

Betrachteter Aspekt	FeU	BMT
Anzahl aller Teilnehmer	391	404
Anzahl Schülerinnen	204	221
Anzahl naturwissenschaftlicher Klassen	9	9
Durchschnittliche Klassengröße	24.4	25.3
Durchschnittlicher Interventionsstart	10:20 Uhr	10:19 Uhr
Durchschnittliche Interventionsdauer	100 Minuten	99.5 Minuten

Tabelle 3.2.: Verteilung der Gesamtstichprobe auf die Interventionsgruppen und Kennzeichnung einzelner Rahmenbedingungen der Durchführung

Bei der Berechnung der in Tabelle 3.2 angegebenen durchschnittlichen Werte für Klassengröße, Interventionsstart und Interventionsdauer wurden die unterschiedlichen Größen der einzelnen Schulklassen berücksichtigt. Zwischen den Interventionsgruppen zeigen sich weder auf Klassen- noch auf Schülerebene signifikante Unterschiede hinsichtlich des Geschlechts der Teilnehmer, der Klassengröße, der Interventionszeit, der Interventionsdauer oder der Zugehörigkeit zu naturwissenschaftlichen Klassen.

3.7. Methoden der Auswertung

Sowohl für das Einpflegen, als auch das Auswerten aller erhobener Daten wurde die Statistiksoftware R eingesetzt.

t-Tests

Um bei allgemeinen Gruppenvergleichen ohne Berücksichtigung bestimmter Kontrollvariablen die statistische Signifikanz möglicher Unterschiede in den Gruppenmittelwerten zu bestimmen, wurden entsprechende *t*-Tests berechnet. Die Voraussetzungen (untersuchtes Merkmal ist intervallskaliert und in der Population normalverteilt, sowie Varianzhomogenität der Populationsvarianzen) für die korrekte Anwendung eines *t*-Tests sind dabei nicht bei allen untersuchten Aspekten gegeben. Allerdings wird das Ergebnis eines *t*-Tests erst dann fehlerhaft, wenn die Größe der zu vergleichenden Gruppen kleiner als 30 oder deutlich unterschiedlich zwischen den Gruppen ist (Rasch et al., 2006a, S. 59f.). Keine der Bedingungen für fehlerhafte Ergebnisse ist für die im Rahmen der Arbeit geplanten *t*-Tests erfüllt, weshalb von einer nicht fehlerbehafteten Anwendbarkeit des *t*-Tests ausgegangen wird. Abhängig vom Ergebnis des Tests auf Varianzhomogenität (*F*-Test) wird in R entweder ein klassischer two sample *t*.test (Kriterium der Varianzhomogenität erfüllt) oder ein Welch two Sample *t*-test (Kriterium der Varianzhomogenität nicht erfüllt) gerechnet.

Die praktische Bedeutsamkeit möglicher durch den *t*-Test (für unabhängige Stichproben) erkannter signifikanter Unterschiede in den Gruppenmittelwerten wurde mit dem Effektstärkemaß *d* als standardisierte Distanz zwischen zwei Mittelwerten gekennzeichnet (Rasch et al., 2006a, S. 66). Bei *t*-Tests für abhängige Stichproben (Vergleich Vor- und Nachtest innerhalb einer Interventionsgruppe) empfehlen Rasch et al. (2006a, S. 92) die Verwendung des Effektstärkemaßes partielles Eta-Quadrat. Im Falle von abhängigen Stichproben spielt bei der Berechnung der Effektstärke auch die Stärke der Abhängigkeit der Gruppen eine Rolle. „Die ermittelte Effektgröße variiert also nicht nur in Abhängigkeit vom Erfolg der experimentellen Manipulation, sondern auch als Funktion der Stärke der Abhängigkeit der Daten. Diese Abhängigkeit kann sich von Studie zu Studie stark unterscheiden“ (Rasch et al., 2006a, S. 92). Eine Einordnung in klein, mittel, groß, sowie ein Vergleich mit Effektstärken aus Studien mit unabhängigen Gruppen kann bei einer Abhängigkeit der Gruppen deshalb

3. Konzeption und Durchführung der Studie

nicht vorgenommen werden. Das oben angegebene Effektstärkemaß d eignet sich im Fall abhängiger Stichproben nicht für die Einschätzung der Größe eines beobachteten Effekts (Rasch et al., 2006a, S. 91f.). In der vorliegenden Studie wurden aber auch Untersuchungen an abhängigen Stichproben durchgeführt (Lernzuwachs innerhalb der Interventionsgruppen). Um eine Bewertung der Effektstärke zwischen den einzelnen Untersuchungen an abhängigen Stichproben angeben zu können, wird demzufolge das von Rasch et al. (2006a, S. 92) vorgeschlagene Maß des partiellen Eta-Quadrats eingesetzt. Insgesamt konnten mittels t-Tests die Forschungsfragen F1, F2 und F5 untersucht werden. Zusätzlich wurden mittels dieser Methode der Auswertung Unterschiede zwischen den Interventionsgruppen hinsichtlich der Ausprägung einzelner Kontrollvariablen (insb. WTE und KFTk) oder den vorgegebenen Rahmenbedingungen auf Signifikanz überprüft.

Kovarianzanalysen

Zur Beantwortung der Forschungsfrage F3 wurde nach signifikanten Interventionseffekten in Bezug auf die Lernwirksamkeit gesucht. Dies betrifft damit mögliche Zusammenhänge zwischen der Unterrichtsart und der Ergebnisse des Nachtests im WTI. Notwendig dabei war zusätzlich eine Berücksichtigung der erhobenen Kontrollvariablen, sowie des Vortestergebnisses des WTI. Damit gingen in die geplante statistische Auswertung sowohl nominale als auch metrische Variablen ein. Petzoldt (2014) empfiehlt für dementsprechende Untersuchungsanliegen eine Auswertung der Daten mit Hilfe einer mehrfaktoriellen ANCOVA (Kovarianzanalyse), welche allerdings in R exakt wie eine ANOVA (Varianzanalyse) durchgeführt wird, lediglich der Datentyp der verwendeten Variablen unterscheidet sich (Petzoldt, 2014, S. 104). Für eine fehlerfreie Analyse der Daten durch eine Kovarianzanalyse müssen die Daten ähnlich wie beim t-Test gewisse Voraussetzungen erfüllen. Die Voraussetzungen einer Kovarianzanalyse sind dabei mit denen einer Varianzanalyse vergleichbar. Insgesamt sollten im Idealfall drei Kriterien an die Daten erfüllt sein. Das entscheidendste Kriterium ist dabei die Unabhängigkeit der Daten zwischen den untersuchten Stichproben, welche durch das Studiendesign gegeben sein sollte, allerdings nicht durch einen Test geprüft werden kann (Petzoldt, 2014, S. 97f.). Das zusätzlich angegebene Kriterium der Varianzhomogenität der Daten (Nachtestergebnis des WTI) ist bei der vorliegenden Studie hingegen verletzt (Bartlett Test, $p = 0.01$), wohingegen die dritte Anforderung der Normalverteiltheit der Residuen wieder erfüllt ist (Ermittelt mit dem Shapiro Test). Somit sind nicht alle bei Petzoldt (2014, S. 97f.) angegebenen Voraussetzungen gegeben. Allerdings „handelt es sich bei der Kovarianzanalyse um ein ausgesprochen robustes Verfahren“ (Bortz & Schuster, 2010, S. 311), nur eine gleichzeitige Verletzung mehrerer Voraussetzungen führt zu einer nennenswerten Verzerrung der Ergebnisse der statistischen Auswertung. Dies ist bei den Daten der vorliegenden Studie nicht der Fall, weshalb die Auswertung der Daten mit Hilfe von Kovarianzanalysen nicht zu fehlerhaften Ergebnisse führt. Bei der Berechnung

mehrfaktorieller Kovarianzanalysen mit R muss zusätzlich noch berücksichtigt werden, dass bei R als Standardeinstellung die sog. Typ I Quadratsumme bei der Berechnung der ANCOVA verwendet wird. Field, Miles und Field (2012, S. 474ff.) weisen in ihrem Buch *Discovering Statistics Using R* darauf hin, dass mit dieser Standardeinstellung nur dann die wahre Ausprägung der Haupteffekte bestimmt werden kann, wenn alle in der Analyse integrierten Variablen voneinander unabhängig sind. Varianzen, welche auf *mehrere* Faktoren gleichzeitig zurückzuführen sind, werden bei Verwendung von Berechnungstyp I jeweils nur *einem* Faktor zugeordnet. Hier spielt die Reihenfolge der Auflistung der Faktoren im R-Befehl eine entscheidende Rolle für das Ergebnis der Auswertung, da Varianzen immer nur dem zuerst aufgeführten Faktor zugerechnet werden. Betrachtet man wie in dieser Studie hauptsächlich Haupteffekte und keine Interaktionen zwischen den Faktoren, empfehlen Field et al. (2012, S. 476) die Verwendung von Berechnungstyp II (von insgesamt vier verschiedenen möglichen Berechnungstypen in R). Dieser Berechnungstyp wird demnach für alle Kovarianzanalysen in dieser Untersuchung verwendet (Befehl: *Anova(data, Type="II")* im Paket *car*).

Als Maß der Stärke beobachteter signifikanter Effekte wird bei der ANCOVA das Maß ω^2 („klein Omega Quadrat“) verwendet. Es dient als Schätzer für den Populationseffekt Ω^2 . Dieses Maß beschreibt den Anteil der durch einen Faktor aufgeklärten Varianz auf der Ebene der Population (Rasch, Frieese, Hofmann & Naumann, 2006b, S. 37f.). Alternativ verwendet beispielsweise das Statistikprogramm SPSS als Effektstärkemaß bei (mehrfaktoriellen) Varianzanalysen das (partielle) Eta-Quadrat, welches sich als Quotient aus der errechneten Quadratsumme zwischen den Faktorstufen und der totalen Quadratsumme (Quadratsumme zwischen den Faktorstufen plus Quadratsumme innerhalb der Faktorstufen) ergibt (Moosbrugger & Rieß, 2010, S. 440f.). Rasch et al. (2006b, S. 39) geben hierzu allerdings an, dass das (partielle) Eta-Quadrat die Stärke des beobachteten Effekts im Vergleich zum wahren Effekt auf der Ebene der Population überschätzt, weshalb die Berechnung des Effektstärkemaßes ω^2 empfohlen wird und in dieser Studie Anwendung findet.

Mehrebenenanalyse

Bei der Beantwortung von Forschungsfrage F3 ist allerdings zudem zu berücksichtigen, dass die vorliegenden Daten eine hierarchische Struktur aufweisen. Innerhalb der Interventionsgruppen sind die Daten einzelner Schüler (Ebene-1-Einheiten) wiederum in Schulklassen (Ebene-2-Einheiten) gruppiert. Diese den Interventionsgruppen inhärente Struktur kann zu einer Abhängigkeit der Beobachtungen innerhalb der Gruppen und damit zu Fehlern in der Auswertung führen (Hartig & Rakoczy, 2010, S. 539). Als Maß der Abhängigkeit zwischen den Daten kann der ICC-Koeffizient (ICC: intraclass correlation) verwendet werden. Je höher der errechnete Wert, desto größer die Abhängigkeit der Daten und damit die Fehlerwahrscheinlichkeit (Hartig & Rakoczy, 2010, S. 544). Der ICC-Koeffizient für die abhängige Variable (Nachtstestergebnis des WTI) unter

3. Konzeption und Durchführung der Studie

Berücksichtigung der Gruppierung innerhalb der Interventionsgruppen wurde für diese Studie zu 0.14 berechnet (aus dem Nullmodell und über die Maximum-Likelihood-Methode). Das bedeutet, dass 14 Prozent der Varianz im Nachtestergebnis des WTI über die Unterschiedlichkeit der Klassen erklärt werden kann. Ab einem Wert von 10 Prozent für den Koeffizienten betrachten Hartig und Rakoczy (2010) den ICC als substantiell und empfehlen die Berechnung einer Mehrebenenanalyse. Eine Berechnung einer Mehrebenenanalyse bedarf dabei allerdings einer Mindestzahl an Ebene-2-Einheiten. Als Mindestzahl nennen Hartig und Rakoczy (2010) 30 Ebene-2-Einheiten für eine Mehrebenenanalyse zur Untersuchung fester Effekte (Untersuchung der aus den Daten berechneten festen Koeffizienten der Prädiktoren im Mehrebenenmodell). Zur Schätzung von Zufallseffekten (Untersuchung der Residualvarianz auf Ebene-1 oder der Varianz und der Kovarianz der Residuen auf Ebene-2) wären größere Stichproben notwendig (Hartig & Rakoczy, 2010, Vgl S. 541f.). Im Rahmen dieser Studie werden nur feste Effekte untersucht. Die von Hartig und Rakoczy (2010, S. 544) formulierten Voraussetzungen ($ICC > 0.1$ und $N_{Klassen} > 30$) für den Einsatz einer Mehrebenenanalyse für die Gesamtstichprobe sind damit gegeben. Diese besteht aus 32 Klassen, bzw. Ebene-2-Einheiten. Für Berechnungen, welche Ergebnisse der WTE oder des KFTk berücksichtigen sollen, können allerdings nur Kovarianzanalysen eingesetzt werden, da die Daten zum WTE und dem KFTk nur in Teilen der Gesamtstichprobe enthalten sind und damit die Anzahl an Ebene-2-Einheiten zu gering ist. So umfasst die Teilstichprobe zum WTE (erster Teil der Hauptstudie) 12 Klassen, diejenige zum KFTk (zweiter Teil der Hauptstudie) 20 Klassen.

Die Berechnung der Stärke der in der Mehrebenenanalyse empirisch gefundenen Effekte erfolgt auf Basis der Vorgaben von Tymms (2004). In seinem Buchbeitrag *Effect sizes in multilevel models* unterscheidet Tymms (2004, S. 57-62) dabei die Berechnung der Effektstärke für dichotome Variablen (z.B. Art der Sequenzierung des Unterrichts; FeU oder BMT) von der Berechnung für kontinuierliche Variablen. Für dichotome Variablen gibt Tymms (2004, S. 57-62) vor, dass die Effektstärke über den Quotienten des Mittelwertunterschieds des relevanten Konstrukts und der Varianz innerhalb der Interventionsgruppen berechnet wird. Bei der Berechnung des Einflusses kontinuierlicher Messdaten muss zusätzlich die Standardabweichung des betrachteten Aspekts (Prädiktorvariable) berücksichtigt werden (Tymms, 2004, S. 57-62). Als Basis seiner Vorgaben zur Berechnung der Effektstärke von Mehrebenenanalysen führt Tymms (2004, S. 57-62) den Effektstärkebegriff wie bei Hedges und Olkin (1985) an. Hedges und Olkin (1985) berechnen die Effektstärke durch die Differenz der Gruppenmittelwerte geteilt durch die globale Standardabweichung (engl.: pooled standard deviation; zitiert aus: Tymms, 2004, S. 56). Die so berechnete Effektstärke entspricht damit dem sog. Cohens d (Rasch et al., 2006a, S. 67f.), welches auch als Effektstärkemaß für die t-Tests eingesetzt wird. Die Größe von berechneten Effekten aus unterschiedlichen Methoden der Auswertung

wird damit vergleichbar. Alternativ geben Snijders und Bosker (2012, S. 192) an, dass als Maß für die Effektstärke auch der ICC verwendet werden kann. Ein Vergleich der Effektstärken wäre bei dieser Berechnungsvariante hingegen nur innerhalb eines Auswertungsverfahrens möglich, weshalb die Effektstärke in der vorliegenden Studie entsprechend der Vorgaben von Tymms (2004) bestimmt wird.

Moderatorentest

Die Untersuchung der Forschungsfrage F4 wird mit Hilfe eines Moderator-tests entsprechend der Vorgaben zur Durchführung von Frazier et al. (2004) vorgenommen. Untersucht wird ein möglicher moderierender Einfluss des Vortestergebnisses im WTI auf die Stärke des Interventionseffekts hinsichtlich des Lernerfolgs. Um zu überprüfen, ob nun ein entsprechender Moderator-effekt des Vortestergebnisses vorliegt, geben Frazier et al. (2004, S. 121) als ersten Schritt vor, die relevanten Daten zuerst zu z-standardisieren und daraus anschließend den Produktterm aus den z-standardisierten Variablen für den Prädiktor (Art der Sequenzierung) und den möglichen Moderator (Ergebnis im Vortest) zu bilden. Anschließend erfolgt eine Regressionsanalyse zur Überprüfung der Signifikanz des Einflusses des Produktterms. Ist der Produktterm im berechneten Regressionsmodell signifikanter Einflussfaktor, ist auch der untersuchte Moderatoreffekt signifikant. Abschließend erfolgt eine Interpretation der erhaltenen Ergebnisse durch eine graphische Darstellung der Kennwerte repräsentativer Teilgruppen oder weiterführender Analysen (Frazier et al., 2004, S. 121ff.). Im Rahmen dieser Studie wird hierzu auf Basis des Vortestergebnisses eine Einteilung der Gesamtstichprobe in vier Teilgruppen vorgenommen. Eine Gruppeneinteilung mit Hilfe einer latenten Klassenanalyse wurde geprüft, allerdings konnte mit dieser Methode kein Antwortmuster in den Daten identifiziert werden, welches einer Stufung nach den Vortestergebnissen der Schüler entsprochen hätte (mögliche Ursache: eingesetzter Wissenstest nicht eindimensional). Eine Verwendung einer latenten Klassenanalyse zur Beantwortung von Forschungsfrage F4 ist damit nicht sinnvoll.

In die Beantwortung der Forschungsfrage F6 gehen alle in der Studie ermittelten Daten und damit auch alle Methoden der Datenauswertung mit ein.

3. Konzeption und Durchführung der Studie

4. Ergebnisse der Studie

In diesem Kapitel erfolgt zunächst eine Beschreibung der Ergebnisse in Bezug nur auf einzelne Kontrollvariablen oder einzelne abhängige Variablen. Anschließend werden zusammenfassende Ergebnisse zur gesamten Hauptstudie präsentiert, bei welchen die Ausprägungen der Kontrollvariablen in den entsprechenden Analysen (Mehrebenenanalyse und Kovarianzanalyse) mit berücksichtigt werden. Nachfolgend werden die Effekte der Intervention unter Kontrolle der Kompetenz der Erkenntnisgewinnung (Kovarianzanalyse, nur erster Teil der Hauptstudie) und der kognitiven Grundfähigkeiten (Kovarianzanalyse, nur zweiter Teil der Hauptstudie) angegeben. Im Anschluss daran werden Auswertungen zum Einfluss des Ergebnisses im Vortest des Wissenstest zum Thema Impuls (WTI) auf die Lernwirksamkeit der Interventionen aufgeführt (Moderatorentest). Im letzten Abschnitt dieses Kapitels sind Ergebnisse der Analyse von Zusammenhängen zwischen Merkmalen des Schülerfragebogens und dem WTI dargestellt.

Das Niveau möglicher Signifikanzen wird in Tabellen für die angegebenen Ergebnissen wie folgt gekennzeichnet: Nicht signifikante Ergebnisse ($0.10 < p < 1.00$) werden nicht zusätzlich gekennzeichnet. Marginal signifikante Ergebnisse ($0.05 < p < 0.10$) erhalten den Zusatz '.', signifikante ($0.01 < p < 0.05$) werden mit '**', hoch signifikante ($0.001 < p < 0.01$) mit '***' und höchst signifikante ($p < 0.001$) mit '****' gekennzeichnet. Bei den durchgeführten t-Tests kennzeichnen positive t-Werte bzw. positive Effektstärken (Cohens d) eine höhere Ausprägung des untersuchten Merkmals bei der Interventionsgruppe FeU. Dementsprechend beschreiben negative t-Werte und negative Effektstärken (Cohens d) höhere Ausprägung des untersuchten Merkmals bei der Interventionsgruppe BMT.

4.1. Ergebnisse zu den Kontrollvariablen

Wie bereits im Abschnitt 3.5.1 erläutert, werden organisatorische Aspekte (Geschlecht der Teilnehmer, Dauer und Beginn der Intervention, Klassengröße, schulischer Zweig), die Einschätzung der Lehrkraft (nur zweite Hauptstudie), die Kompetenz zur Erkenntnisgewinnung (WTE, nur erste Hauptstudie) und die allgemeinen kognitiven Grundfähigkeiten (KFTk, nur zweite Hauptstudie) als Kontrollvariablen erhoben.

Hinsichtlich der organisatorischen Kontrollvariablen besteht kein signifikanter Unterschied zwischen den Interventionsgruppen (vgl. Abschnitt: 3.6).

4. Ergebnisse der Studie

Die *Einschätzung des Lehrenden* wurde auf Schülerseite ($N_{ges} = 505$) erfragt und umfasst die Teilkategorien *Interesse des Lehrenden* (3 Items; L1 bis L3), sowie die Kategorie *lernförderliches Klima und Motivation* (6 Items; L4 bis L9). Alle Items der beiden Teilkategorien konnten durch die Schüler auf einer vierstufigen Skala bewertet werden. Die Ergebnisse der Befragung sind in Abbildung 4.1 dargestellt, die Ergebnisse der zugehörigen Auswertung auf mögliche signifikante Unterschiede finden sich in Tabelle 4.1 und in nachfolgender Beschreibung.

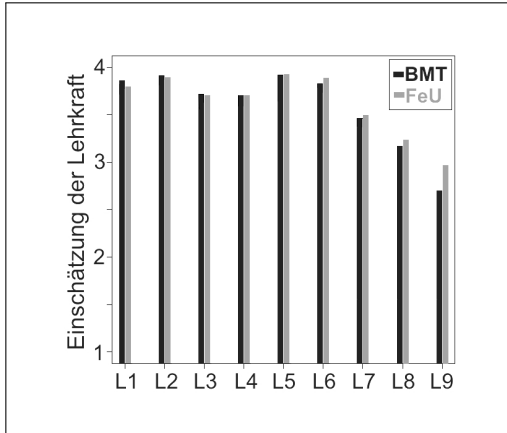


Abbildung 4.1.: Balkendiagramm der Mittelwerte der Einschätzungen der Lehrkraft durch die Schüler der beiden Interventionsgruppen auf einer Skala von 1 bis 4 für die Items L1 bis L9

	FeU	BMT	Signifikanz	Effektstärke
Interesse des Lehrenden	3.79	3.83	$p = 0.143$	$d = -0.13$
Lernförderliches Klima und Motivation	3.53	3.48	$p = 0.131$	$d = 0.13$
Gesamteinschätzung des Lehrenden	3.62	3.60	$p = 0.512$	$ d < 0.1$

Tabelle 4.1.: Mittelwerte der Einschätzung des Lehrenden für beide Teilskalen und die Gesamtskala, sowie Ergebnisse der Auswertung auf signifikante und bedeutsame Unterschiede zwischen den Interventionsgruppen.

Die Gesamteinschätzung des Lehrenden, sowie die Einschätzung hinsichtlich der beiden Teilkategorien unterscheidet sich weder signifikant ($p > 0.05$) noch

bedeutsam ($d < 0.20$) (Rasch et al., 2006a, S. 68) zwischen den beiden Interventionsgruppen. Die Teilskalen weisen jedoch entgegengesetzt gerichtete Tendenzen auf. Wird von Schülern in der Interventionsgruppe BMT ($N_{BMT} = 262$) das Interesse des Lehrenden etwas höher bewertet als von Schülern der Interventionsgruppe FeU ($N_{FeU} = 243$), so bewerten diese wiederum das lernförderliche Klima und die Motivation etwas höher als Schüler der Interventionsgruppe BMT. Signifikant unterschiedlich wird die Lehrkraft nur hinsichtlich einer Einzelfrage (L9) eingeschätzt ($p = 0.007$; $t(467) = 2.715$). Diese Einzelfrage erfasst, inwiefern die Schüler es als zutreffend erachten, dass sie durch die Lehrkraft in der Unterrichtsstunde für eigene Beiträge gelobt wurden. Schüler der Gruppe FeU bewerten die Intervention hinsichtlich dieser Frage damit systematisch höher als Schüler der Gruppe BMT. Alle acht weiterhin erhobenen Einzelfragen unterschieden sich nicht systematisch. Insgesamt liegen damit keine statistisch signifikant unterschiedlichen Einschätzungen der Lehrkraft zwischen der Gesamtheit der Schüler der verschiedenen Interventionsgruppen vor. Die Gesamteinschätzung der Lehrkraft in den beiden Interventionsgruppen ($M_{FeU} = 3.62$, bzw. $M_{BMT} = 3.60$) ist vergleichbar. Somit zeigt sich in den erhobenen Daten kein Indiz für einen substantiellen Versuchsleitereffekt.

Die *Kompetenz zur experimentellen Erkenntnisgewinnung* wird mit 8 Items in der ersten Hauptstudie über den Wissenstest zur Kompetenz der Erkenntnisgewinnung (WTE) vor und nach der Intervention erfasst. Für jedes korrekt bearbeitete Item wird eine Bewertungseinheit (BE) vergeben. Zwischen den Interventionsgruppen ($N_{FeU} = 148$; $N_{BMT} = 143$) zeigt sich ein höchst signifikanter ($p = 0.0008$; $t(289) = -3.40$) Unterschied im Test vor der Intervention bei einer kleinen bis mittleren Effektstärke von $d = -0.40$. Für die Schüler der Interventionsgruppe FeU ergibt sich dabei ein Testmittelwert vor der Intervention von 5.72 von 8 möglichen BE, Schüler der Interventionsgruppe BMT erreichen im Schnitt 6.27 BE vor der Unterrichtseinheit. Auch der WTE nach der Intervention unterscheidet sich signifikant ($p = 0.012$; $t(106) = -2.56$; $d = -0.40$) zwischen den Gruppen, allerdings bei einer kleineren Stichprobe, da aus organisatorischen Gründen bei einigen Klassen auf den Einsatz des WTE nach der Intervention verzichtet werden musste. Die Ergebnisse des Tests vor und nach der Unterrichtseinheit zeigen dabei innerhalb beider Interventionsgruppen keine signifikanten Unterschiede (FeU: $p = 0.87$; $t(124) = -0.168$ und BMT: $p = 0.99$; $t(229) = 0.001$).

Die mit der Kurzskala KFTk erfasste Ausprägung der *kognitiven Grundfähigkeiten* der Lernenden unterscheidet sich nicht signifikant zwischen den Interventionsgruppen ($N_{FeU} = 243$; $N_{BMT} = 262$). Weder die Gruppenmittelwerte (M) der Originaldaten ($M_{FeU} = 8.73$; $M_{BMT} = 8.64$; $p = 0.61$; $t(501) = 0.51$) noch die entsprechenden Mittelwerte der jeweils innerhalb der beiden eingesetzten Testformen des KFTk z-standardisierten Daten ($M_{FeU} = 0.03$; $M_{BMT} = -0.03$; $p = 0.51$; $t(501) = 0.66$) weisen nicht zufällige Unterschiede auf.

Insgesamt zeigt sich damit hinsichtlich der erhobenen Kontrollvariablen nur

4. Ergebnisse der Studie

ein signifikanter Unterschied zwischen den Interventionsgruppen. Dieser Unterschied liegt dabei in den Mittelwerten der Ergebnisse des WTE, welcher die Kompetenz zur experimentellen Erkenntnisgewinnung erfassen soll. Die Erfassung dieser Kompetenz über den WTE erfolgt dabei aber nur wenig zuverlässig $\alpha = 0.52$ (vgl. Abschnitt 3.5.4) und ist somit fehlerbehaftet. Die Frage, ob sich die Kompetenz zur experimentellen Erkenntnisgewinnung in den beiden Interventionsgruppen tatsächlich signifikant unterscheidet, lässt sich demnach mit den in dieser Studie eingesetzten Instrumenten nicht eindeutig beantworten.

4.2. Ergebnisse zu den abhängigen Variablen

In diesem Abschnitt sind Ergebnisse direkt zu den abhängigen Variablen (vgl. Abschnitt 3.5.1) angegeben. Zum einen sind dies die Einschätzungen der Lernenden zu Aspekten der Unterrichtsqualität der in den Interventionen eingesetzten Unterrichtseinheiten, zum anderen Ergebnisse der Lernenden im WTI und damit zum Kenntnisstand der Lernenden vor und nach der Intervention.

Einschätzung der Interventionen durch die Lernenden

Die Mittelwerte der Einschätzungen der Lernenden ($N_{Ges} = 796$; $N_{FeU} = 392$; $N_{BMT} = 404$) über die mit einer vierstufigen Skala erhobenen Aspekte der Unterrichtsqualität sind in Tabelle 4.2 angegeben.

	FeU	BMT	Signifikanz	Effektstärke
Soziale Eingebundenheit (SE)	3.69	3.70	$p = 0.611$	$ d < 0.1$
Kompetenzerleben (KE)	3.25	3.29	$p = 0.287$	$ d < 0.1$
Autonomieerleben (AE)	3.25	3.23	$p = 0.539$	$ d < 0.1$
Kognitive Aktivierung (KOG)	3.00	3.02	$p = 0.506$	$ d < 0.1$
Organisation der Lernumgebung (OL)	3.14	3.17	$p = 0.325$	$ d < 0.1$
Schlüssigkeit im Unterrichtsverlauf (SU)	3.58	3.59	$p = 0.852$	$ d < 0.1$
Orientierung am Ziel (OZ)	3.49	3.53	$p = 0.123$	$d = -0.11$

Tabelle 4.2.: Mittelwerte der Einschätzung der Lernenden bezüglich einzelner Aspekte und der Unterrichtsqualität, sowie Ergebnisse der Auswertung auf signifikante und bedeutsame Unterschiede zwischen den Interventionsgruppen.

Ein signifikanter Unterschied in den Einschätzungen durch die Lernenden findet sich zwischen den Interventionsgruppen bei keinem der untersuchten

Aspekte. Eine Auswertung der Ausprägung der Gesamtskala aller grundlegenden Bedürfnisse (SE, KE und AE; $p = 0.574$; $t(793) = -0.56$) und der Gesamtskala von Elementen der kognitionspsychologisch verstandenen Strukturierung (OZ, OL; $p = 0.127$; $t(793) = -1.53$; $d = -0.11$) weist ebenfalls auf keine signifikanten oder bedeutsamen Unterschiede zwischen den Gruppen hin.

Lernwirksamkeit der Interventionen

Der in der Untersuchung eingesetzte Wissenstest zum Thema Impuls (WTI) umfasst 22 Items und soll die Lernwirksamkeit der Interventionen erfassen. Für jedes korrekt bearbeitete Item wird wie beim WTE eine Bewertungseinheit (BE) vergeben. Vor der Intervention erreichten Schüler der Interventionsgruppe FeU im Mittel $M_{FV} = 11.52$ BE, Schüler der Gruppe BMT $M_{BV} = 11.31$ BE. Der Unterschied von 0.21 BE zwischen den Gruppen ist dabei nicht signifikant ($p = 0.396$; $t(794) = 0.85$), auch die Varianzen der Daten innerhalb der Gruppen unterscheiden sich zwischen den Gruppen nicht signifikant (F-Test, $F = 0.91$; $p = 0.91$), allerdings weist die Interventionsgruppe FeU einen größeren Median auf (vgl. Abb.: 4.2; $Median_{FeU} = 12$ BE; $Median_{BMT} = 11$ BE).

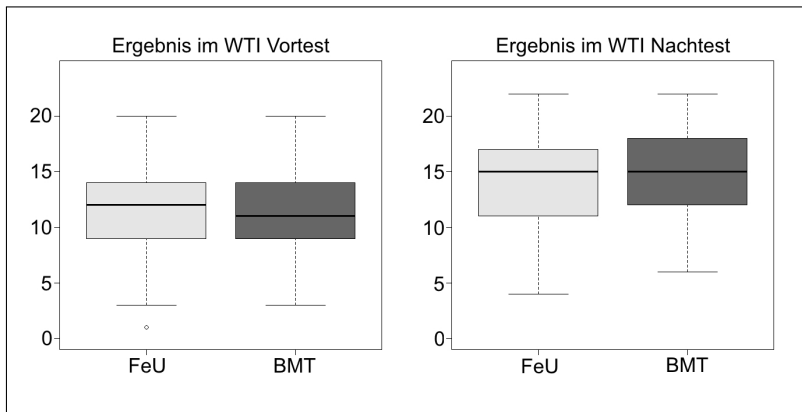


Abbildung 4.2.: Boxplot der Ergebnisse der Schüler im WTI vor und nach der Intervention. Die Mediane der Gruppen sind darin jeweils mit einem Querstrich innerhalb der grau gefärbten Kästen gekennzeichnet.

Nach der Intervention ergeben sich die erreichten Mittelwerte für die BE zu $M_{FN} = 14.25$ BE bzw. zu $M_{BN} = 15.07$ BE für die Interventionsgruppen FeU bzw. BMT. Dieser Unterschied zwischen den Interventionsgruppen ist hoch signifikant ($p = 0.002$; $t(774) = -3.10$) mit einer kleinen Effektstärke $d = -0.22$. Zudem unterscheiden sich im Nachtest des WTI die Varianzen der Daten signifikant (F-Test, $F = 1.29$; $p = 0.012$) zwischen den Gruppen, die Mediane stimmen nun

4. Ergebnisse der Studie

überein ($Median_{FeU} = 15 \text{ BE} = Median_{BMT}$). Die Differenz aus Nachtestwert und Vortestwert unterscheidet sich mit einer kleinen bis mittleren Effektstärke $d = -0.30$ sogar höchst signifikant ($p < 0.001$; $t(794) = -4.290$) zwischen den Gruppen (bei vergleichbarer Varianz; F-Test, $F = 0.86$; $p = 0.12$).

Die Teilgruppen des ersten ($N = 291$) und des zweiten ($N = 505$) Teils der Hauptstudie unterscheiden sich weder im Ergebnis des Vortests ($p = 0.31$; $t(794) = 1.02$) noch des Nachtests ($p = 0.46$; $t(794) = 0.74$) signifikant voneinander. Die Mittelwerte der erreichten Punktzahlen liegen für Schüler der ersten Hauptstudie bei $M_{1VT} = 11.57 \text{ BE}$ im Vortest bzw. $M_{1NT} = 14.79 \text{ BE}$ im Nachtest. Für Schüler der zweiten Hauptstudie ergeben sich die entsprechenden Kennwerte von $M_{2VT} = 11.32 \text{ BE}$ für den Vortest bzw. $M_{2NT} = 14.59 \text{ BE}$ für den Nachtest.

Innerhalb beider Interventionsgruppen sind die Mittelwerte der erreichten BE im WTI bei einem Vergleich von Vor- und Nachtest jeweils höchst signifikant unterschiedlich. Beträgt die Differenz der Mittelwerte in der Interventionsgruppe FeU im Schnitt 2.73 BE und ergibt sich eine Effektstärke von $\eta_p^2 = 0.04$ ($t(390) = 16.7$; $p < 0.001$), so kann der Gruppe BMT ein Mittelwert für die Differenz von 3.76 bei einer Effektstärke von $\eta_p^2 = 0.05$ ($t(404) = 21.6$; $p < 0.001$) zugeordnet werden.

Als Überblick sind einzelne interventionsspezifische Ergebnisse nochmals in Tabelle 4.3 dargestellt.

Aspekt	Relevante Mittelwerte		Signifikanz	Effektstärke
Lernzuwachs FeU	$M_{FNT} = 14.2$	$M_{FVT} = 11.5$	$p < 0.001$	$\eta_p^2 = 0.04$
Lernzuwachs BMT	$M_{BNT} = 15.1$	$M_{BVT} = 11.3$	$p < 0.001$	$\eta_p^2 = 0.05$
FeU - BMT: VT	$M_{FVT} = 11.5$	$M_{BVT} = 11.3$	$p = 0.396$	$ d < 0.10$
FeU - BMT: NT	$M_{FNT} = 14.2$	$M_{BNT} = 15.1$	$p = 0.002$	$d = -0.22$
FeU - BMT: DZ	$M_{FDZ} = 2.73$	$M_{BDZ} = 3.76$	$p < 0.001$	$d = -0.30$

Tabelle 4.3.: Überblick zu den Ergebnissen der Auswertungen des WTI. Die Art des Effektstärkemaßes variiert zwischen t-Tests für gepaarte (η_p^2) bzw. unabhängige Stichproben (d).

Dabei wurden in Tabelle 4.3 folgende Abkürzungen verwendet: Vortest (VT), Nachtest (NT), Differenz aus Nachtestwert und Vortestwert (DZ), Mittelwert (M). In den Indizes der angegebenen Mittelwerte ist die Zugehörigkeit zur Interventionsgruppe zusätzlich durch ein vorangestelltes (B) für Interventionsgruppe BMT oder (F) für die Interventionsgruppe FeU gekennzeichnet.

Insgesamt zeigen die Ergebnisse der durchgeführten t-Tests eine größere Lernwirksamkeit einer Sequenzierung nach den Vorgaben der BMT.

4.3. Zusammenfassende Ergebnisse

Die nachfolgend dargestellten zusammenfassenden Ergebnisse basieren auf Analysen, welche den Einfluss mehrerer Variablen insbesondere auf die Ergebnisse der Lernenden im Nachtest des WTI berücksichtigen.

4.3.1. Analyse der Gesamtstudie

Bei der Analyse der Daten der Gesamtstudie kann dabei der Einfluss der Kompetenz zur experimentellen Erkenntnisgewinnung und der Einfluss der Ausprägung der kognitiven Grundfähigkeiten nicht einbezogen werden, da diese beiden Kontrollvariablen jeweils nur in einem Teil der Hauptstudie erfasst wurden. Zuerst erfolgt eine schrittweise Darstellung der Datenauswertung mittels einer Mehrebenenanalyse. Die Schrittfolge bei der Darstellung folgt dabei weitestgehend entsprechenden Vorgaben von Bliese (2013) zur Durchführung einer Mehrebenenanalyse mit dem Statistikprogramm R. Anschließend werden die Ergebnisse einer Kovarianzanalyse der Gesamtdaten präsentiert. Diese zusätzliche Präsentation erfolgt primär, um die Vergleichbarkeit der Ergebnisse der Analyse der Daten der Gesamtstudie mit den nachfolgenden, mit Kovarianzanalysen ausgewerteten Teildaten gewährleisten zu können.

Eine Mehrebenenanalyse der Daten wurde aufgrund der vorliegenden hierarchischen Datenstruktur durchgeführt. Die innerhalb der Interventionsgruppen inhärente Klassen-Struktur weist einen substantiellen Einfluss auf die abhängige Variable auf (vgl. Abschnitt 3.7). Die Schüler einzelner Klassen unterscheiden sich diesbezüglich also systematisch innerhalb der Interventionsgruppen. Die Unterschiedlichkeit im Nachtestergebnis des WTI zwischen den Klassen ist in Abb. 4.3 dargestellt.

Die Mittelwerte einzelner Klassen innerhalb der Interventionsgruppen streuen dabei in beiden Gruppen vergleichbar stark um den jeweiligen Gruppenmittelwert ($\sigma_{FeU} = 1.7$ bzw. $\sigma_{BMT} = 1.6$). Insgesamt kann 14% der Varianz im Nachtestergebnis des WTI über die Unterschiedlichkeit der Klassen erklärt werden ($ICC = 0.14$). Die Varianz der Klassen im Nachtestergebnis ($\tau_{00} = 2.1$) ist dabei höchstsignifikant von Null verschieden (aus Varianzanalyse eines Modells mit und eines Modells ohne zufälligen Achsenabschnitt). Bliese (2013, S. 56) empfiehlt zur nachfolgenden Auswertung ein Modell mit zufälligem Achsenabschnitt für das Nachtestergebnis der Klassen zu verwenden. Als feste Effekte bei den berechneten Mehrebenenanalysemodellen werden das Ergebnis im Vortest des WTI, die Klassengröße, der Zeitpunkt und die Dauer der Intervention, die Zweigwahl der Lernenden, deren Geschlecht und die Sequenzierungsart (unabhängige Variable) verwendet. Im Modell mit allen festen Effekten ($AIC = 4030$; $BIC = 4077$) weisen weder die Klassengröße ($p = 0.33$), die Zweigzugehörigkeit ($p = 0.74$), noch die Dauer der Intervention ($p = 0.69$) einen signifikanten Effekt

4. Ergebnisse der Studie

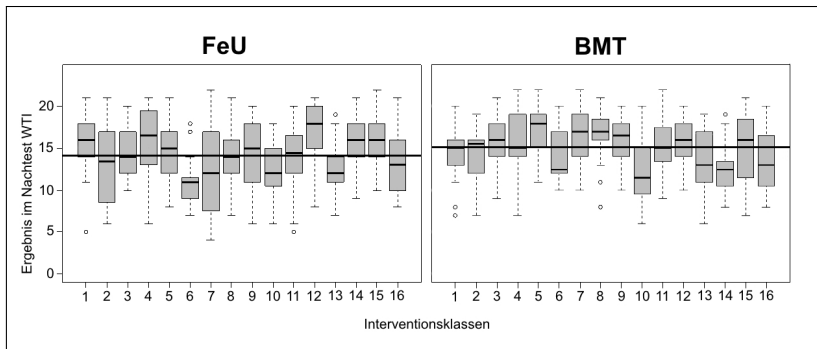


Abbildung 4.3.: Boxplots der Ergebnisse der Schüler einzelner Klassen innerhalb der Interventionsgruppen im Nachtest des WTI. Die Mittelwerte der Nachtestergebnisse der entsprechenden Interventionsgruppe sind mit einem Querstrich innerhalb der Kästen gekennzeichnet.

auf das Nachtestergebnis auf. Die Kennwerte AIC (Akaike Information Criterion) und BIC (Bayesian Information Criterion) beschreiben dabei die Passung des jeweils berechneten Mehrebenenmodells. Die Passung des Modells wird dabei umso besser, je kleiner diese Kennwerte sind (Moosbrugger & Kelava, 2012, S. 309). Den niedrigsten AIC und BIC Wert aller berechneten Modelle weist dabei das Modell auf, welches als feste Effekte nur die signifikanten Einflussfaktoren enthält (AIC = 4016; BIC = 4049). Auf diesem Modell basieren auch die nachfolgend angegebenen p-Werte und die berechneten Effektstärken, welche in Tabelle 4.4 dargestellt sind.

Fester Effekt	Schätzwert	DF	t-Wert	p-Wert	Effektstärke
(Achsenabschnitt)	9.50	762	17.3	0.0000 ***	—
Ergebnis Vortest	0.54	762	16.4	0.0000 ***	$d = 1.24$
Interventionsstart	-0.40	29	-3.50	0.002 **	$d = -0.38$
Geschlecht	-1.40	762	-5.77	0.0000 ***	$d = -0.48$
Sequenzierungsart	0.93	29	2.90	0.007 **	$d = 0.32$

Tabelle 4.4.: Ergebnisse des Mehrebenenmodells zum Einfluss untersuchter Aspekte auf das WTI-Ergebnis mit den niedrigsten AIC und BIC Kennwerten. Bei der Berechnung der Effektstärken wird die Standardabweichungen bei den intervallskalierten Effekten ($sd_{VT} = 3.37$; bzw. $sd_{ST} = 1.39$) berücksichtigt.

Die Standardabweichung der Achsenabschnitte bei diesem Modell beträgt dabei $\sigma_A = 0.69$, die der Residuen $\sigma_R = 2.93$ (entspricht der globalen Standardabweichung). Unter Berücksichtigung möglicher Abhängigkeiten durch die hierarchische Datenstruktur mit Hilfe der Mehrebenenanalyse ergibt sich ein hoch signifikanter Effekt der Sequenzierungsart (SeA, unabhängige Variable) auf das Nachtstestergebnis des WTI. Die Sequenzierungsart des BMT weist hier die größere Lernwirksamkeit auf. Die Stärke dieses Effekts liegt bei $d = 0.32$. Ebenso hoch signifikant zeigt sich der Einfluss des Zeitpunkts des Interventionsstarts (bei $d = -0.38$), lernförderlich ist dabei ein möglichst früher Start der Intervention. Höchst signifikante Effekte wiederum stellen das Geschlecht der Lernenden und das Ergebnis im Vortest des WTI dar. Wobei die teilnehmenden Mädchen mit einer knapp mittleren Effektstärke von $d = -0.48$ ein niedrigeres Nachtstestergebnis erzielen. Das Ergebnis im Vortest stellt mit $d = 1.24$ insgesamt den bedeutendsten Einflussfaktor dar. Positive Ergebnisse im Vortest bewirken dabei auch positive Ergebnisse im Nachttest.

Der Einfluss einzelner Faktoren auf das Nachtstestergebnis soll zudem mit den Ergebnissen einer ANCOVA wiedergegeben werden. Die hierarchische Datenstruktur spielt bei dieser Auswertungsart für die Berechnung der Einflüsse keine Rolle. Die Angabe der Ergebnisse dieser Auswertungsart dient vornehmlich der Vergleichbarkeit mit späteren gleichgearteten Auswertungen von Daten aus Teilgruppen der Gesamtstichprobe. In der ANCOVA werden grundsätzlich die Einflüsse genau der Variablen überprüft, welche auch bei der Auswertung über die Mehrebenenanalyse Berücksichtigung finden. Allerdings werden bei der hier ausgewählten ANCOVA die Einflüsse aller Variablen dargestellt (vgl.: Tabelle 4.5).

Die Ergebnisse der ANCOVA entsprechen sowohl hinsichtlich der Signifikanz als auch der Bedeutsamkeit und der Wirkungsrichtung der untersuchten Aspekte den Ergebnissen der Mehrebenenanalyse. Keinen signifikanten Einfluss auf das Nachtstestergebnis des WTI zeigen die Zweigzugehörigkeit, die Interventionsdauer und die Klassengröße. Den größten und dabei höchst signifikanten Einfluss weist auch hier das Ergebnis im Vortest auf. Die Effektstärke von $\omega^2 = 0.257$ liegt dabei allerdings noch deutlicher im Bereich großer Effekte als bei der Mehrebenenanalyse. Werte von ω^2 über 0.14 kennzeichnen dabei große Effekte, Werte über 0.06 stehen für mittlere und ab einem Wert von $\omega^2 = 0.01$ spricht man von kleinen Effekten (Rasch et al., 2006b, S. 38). Folglich sind durch die berechneten Effektstärken die jeweils höchst signifikanten Einflüsse des Interventionsstarts, der Art der Sequenzierung und des Geschlechts als kleine Effekte zu bezeichnen. Dieser Einfluss des Geschlechts zeigt sich mit etwas größerer Effektstärke ($\omega^2 = 0.089$), wenn nur die Daten von monoedukativen Schulklassen (F-Wert = 18.9; $N = 183$; $N_w = 104$; $N_m = 79$) berücksichtigt werden.

Wendet man die Berechnung (nur für Vergleich zweier Gruppen möglich; Berechnung aus F-Wert und der Größe der Interventionsgruppen) nach Thalheimer und Cook (2002, S. 8) für die Effektstärke Cohens d aus Varianz- oder Kovari-

4. Ergebnisse der Studie

Untersuchter Aspekt	QS	DF	F-Wert	p-Wert	Effektstärke
Ergebnis Vortest (VT)	2482.45	1	275.89	0.0000 ***	$\omega^2 = 0.257$
Zweizugehörigkeit (ZW)	3.22	1	0.36	0.5497	$\omega^2 < 0.010$
Geschlecht (GE)	300.37	1	33.38	0.0000 ***	$\omega^2 = 0.039$
Interventionsstart (ST)	209.61	1	23.29	0.0000 ***	$\omega^2 = 0.027$
Interventionsdauer (DA)	5.37	1	0.60	0.4400	$\omega^2 < 0.010$
Klassengröße (KG)	15.81	1	1.76	0.1853	$\omega^2 < 0.010$
Sequenzierungsart (SeA)	178.78	1	19.87	0.0000 ***	$\omega^2 = 0.023$
Residuen	7090.54	788			

Tabelle 4.5.: Ergebnisse der Kovarianzanalyse der Daten der Gesamtstichprobe zum Einfluss auf das WTI-Ergebnis. Bei der Berechnung der Effektstärken wird das Maß ω^2 („kleines Omegaquadrat“; vgl. Abschnitt 3.7) verwendet. Die Abkürzung QS steht dabei für Quadratsumme, DF für Freiheitsgrad.

anzanalysen an, so ergibt sich für den Effekt der unabhängigen Variable (Art der Sequenzierung) eine Effektstärke von $d = 0.32$ und damit die exakt gleiche Effektstärke wie aus den Daten der bereits dargestellten Mehrebenenanalyse kalkuliert. Für die anderen untersuchten Aspekte lassen sich so ebenfalls die Effektstärken in das Maß d umrechnen: VT: $d = 1.18$; ST: $d = 0.34$; GE: $d = 0.41$. Diese so berechneten Effektstärken sind damit jeweils etwas geringer als aus der Mehrebenenanalyse kalkuliert. Eine Angabe der Richtung der Effekte ist bei dieser Umrechnung im Gegensatz zur Berechnung bei der Mehrebenenanalyse nicht möglich.

4.3.2. Analyse von Teilgruppen der Gesamtstudie

Bei der Analyse der Daten von Teilgruppen der Gesamtstudie werden jeweils Kovarianzanalysen durchgeführt. Die Ergebnisse dieser Analysen werden nachfolgend zu den untersuchten Teilgruppen aufgeführt. Zuerst erfolgt die Präsentation der Ergebnisse einer geschlechtsspezifischen Untersuchung der Gesamtstichprobe, anschließend werden die Resultate aus zwei Kovarianzanalysen für die einzelnen Interventionsgruppen angegeben.

Die Wirkungsrichtung der Effekte wird bei den nachfolgenden Darstellungen nur angeführt, wenn diese von zuvor berichteten Richtungen abweicht, bzw. noch nicht genannt wurde. Durch die Kovarianzanalyse in Abschnitt 4.3.1 sind bereits einige Faktoren identifiziert, welche in der Gesamtstichprobe keinen systematischen Einfluss auf das Nachtestergebnis aufweisen. Grundsätzlich

werden bei mehrfaktoriellen Varianz- oder Kovarianzanalysen aber Faktoren betrachtet, welche einen Einfluss auf die abhängige Variable ausüben (Bortz & Schuster, 2010, S. 237). Werden mehrere Faktoren bei einer Analyse in die Berechnung des Modells integriert, besteht die Möglichkeit, dass mehrere Faktoren gleiche Varianzanteile erklären. Eine eindeutige Zuweisung des entsprechenden Anteils zum jeweiligen Faktor ist nicht möglich (Field et al., 2012, S. 276). Mit einer größeren Anzahl an Faktoren in einem gerechneten Modell kann deshalb eine Verringerung der Qualität der Aussagen dieses Modells einher gehen, weshalb im Rahmen dieser Studie zusätzliche Faktoren nur dann zu den Analysen hinzugefügt werden, wenn diese Faktoren die Varianzaufklärung und damit die Güte des Modells systematisch verbessern. Ob Modelle mit mehr Faktoren nun zu einer signifikant höheren Varianzaufklärung des Modells führen, wird deshalb bei den Analysen der Daten von Teilgruppen jeweils überprüft. In jedem dieser Modelle sind dabei die in Abschnitt 4.3.1 als systematisch für die Gesamtstudie identifizierte Faktoren (VT, GE, ST und SeA) enthalten. Es wird demnach überprüft, ob ein Hinzufügen weiterer Faktoren (ZW, DA; KG) zu einer systematischen Verbesserung des Modells führt. Führt das Hinzufügen zu keiner systematischen Verbesserung, erfolgt keine Darstellung und weitergehende Interpretation des entsprechenden Modells mit zusätzlichen Faktoren. Die Interpretation und die Berechnung der Effektstärken erfolgt dann auf Basis des Modells mit nur den systematischen Faktoren. Eine Prüfung auf Signifikanz weiterer Faktoren wird jedoch jeweils vorgenommen.

Geschlechtsspezifische Teilgruppen

Im Abschnitt 4.3.1 wird ein Einfluss des Geschlechts sowohl in der Mehrebenenanalyse als auch in der ANCOVA der Gesamtstichprobe auf das Nachttestergebnis im WTI aufgeführt. Das Vortestergebnis ist dabei zwischen den Geschlechtergruppen ebenso höchst signifikant unterschiedlich ($p < 0.001$; $t(794) = -8.03$; $d = -0.57$) wie das Nachttestergebnis ($p < 0.001$; $t(794) = -9.51$; $d = -0.68$). Zu beiden Testzeitpunkten weisen die männlichen Testteilnehmer im arithmetischen Mittel die höheren Kennwerte auf ($M_{MVT} = 12.4$ und $M_{MNT} = 16.0$ im Vergleich zu $M_{WVT} = 10.6$ und $M_{WNT} = 13.5$; W im Index für „weiblich“, M für „männlich“). Die Effektstärke des Unterschieds zwischen den Geschlechtergruppen ist dabei als mittel zu kategorisieren. Die Teilstichproben der weiblichen bzw. der männlichen Studienteilnehmer sind damit hinsichtlich des Nachttests (und des Vortests) systematisch verschieden.

Inwiefern diese Verschiedenartigkeit wiederum den Einfluss der Kontrollvariablen bzw. der unabhängigen Variablen auf das Nachttestergebnis beeinflusst, wird in dieser Studie mit zwei Kovarianzanalysen untersucht. Die Prüfung relevanter Faktoren für diese Teilgruppen lieferte bei den weiblichen Teilnehmern die Faktoren VT, ST und SeA, bei den männlichen Teilnehmern zusätzlich KG. Für eine bessere Vergleichbarkeit der Modelle werden für beide Teilgruppen

4. Ergebnisse der Studie

die Faktoren VT, ST, SeA und KG berechnet und deren Ergebnisse in den Tabellen 4.6 und 4.7 aufgelistet. Keine der weiteren Faktoren (ZW und DA) erweist sich in den zusätzlich berechneten Modellen als signifikanter Einflussfaktor.

Untersuchter Aspekt	QS	DF	F-Wert	p-Wert	Effektstärke
Ergebnis Vortest	1147.57	1	130.10	0.0000 ***	$\omega^2 = 0.233$
Interventionsstart	69.28	1	7.85	0.0053 **	$\omega^2 = 0.016$
Sequenzierungsart	91.06	1	10.32	0.0014 **	$\omega^2 = 0.021$
Klassengröße	3.98	1	0.45	0.5023	$\omega^2 < 0.010$
Residuen	3704.78	420			

Tabelle 4.6.: Ergebnisse der Kovarianzanalyse der Daten der Teilstichprobe der weiblichen Studienteilnehmer zum Einfluss auf das WTI-Ergebnis (N= 425).

Untersuchter Aspekt	QS	DF	F-Wert	p-Wert	Effektstärke
Ergebnis Vortest	1261.18	1	138.79	0.0000 ***	$\omega^2 = 0.271$
Interventionsstart	108.06	1	11.89	0.0006 ***	$\omega^2 = 0.029$
Sequenzierungsart	75.85	1	8.35	0.0041 **	$\omega^2 = 0.019$
Klassengröße	74.02	1	8.15	0.0046 **	$\omega^2 = 0.019$
Residuen	3325.93	366			

Tabelle 4.7.: Ergebnisse der Kovarianzanalyse der Daten der Teilstichprobe der männlichen Studienteilnehmer zum Einfluss auf das WTI-Ergebnis (N= 371).

Bedeutendster und zugleich höchst signifikanter Einfluss ist in beiden Teilstichproben das Ergebnis im Vortest. Der Einfluss dieses Faktors ist bei den männlichen Teilnehmern geringfügig stärker ausgeprägt. Vergleichbar zeigen sich die Gruppen hinsichtlich Signifikanz (jeweils hoch signifikant) und Effektstärke ($\omega^2 = 0.021$ bzw. $\omega^2 = 0.019$) der Sequenzierungsart. Der Interventionsstart ist bei den weiblichen Studienteilnehmern in einem etwas geringeren Maße signifikant und effekteffektiv als bei der Gruppe der männlichen Teilnehmer, bei welchen dieser Aspekt höchst signifikant und mit der Effektstärke $\omega^2 = 0.029$ gekennzeichnet ist. Nur bei der männlichen Teilstichprobe signifikant ist dabei die Klassengröße. Dieser Effekt ist als kleiner Effekt $\omega^2 = 0.019$ zu bezeichnen. Die männlichen Teilnehmer weisen bessere durchschnittliche Testwerte auf, je höher der Wert für die Klassengröße ist.

Neben den unterschiedlich starken Einflüssen einzelner Kontrollvariablen wei-

sen die in diesem Abschnitt betrachteten Teilgruppen auch hinsichtlich der Einschätzungen der Qualitätsmerkmale des Unterrichts im Schülerfragebogen Unterschiedlichkeiten auf. Die Mittelwerte dieser Einschätzungen sind in Abbildung 4.4 angegeben.

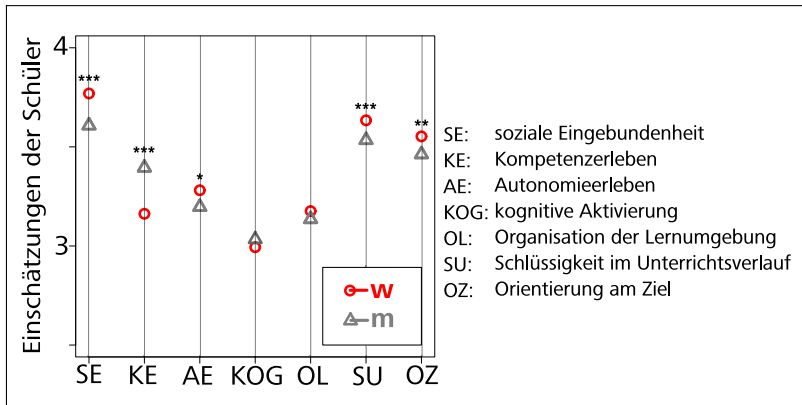


Abbildung 4.4.: Durch Linien verbundene Mittelwerte der Schülereinschätzung der Merkmale des Fragebogens durch weibliche (w) bzw. männliche (m) Schüler. Signifikante Mittelwertunterschiede sind entsprechend gekennzeichnet.

In Abbildung 4.4 ist dabei die Skala der Einschätzung der Schüler verkürzt dargestellt, um die Unterschiedlichkeiten in den Bewertungen durch die weiblichen und männlichen Studienteilnehmer optisch deutlicher werden zu lassen. Die Mittelwerte der Bewertungen der weiblichen Teilnehmer liegen für fünf der erfragten Merkmale über den entsprechenden Bewertungen der männlichen Teilnehmer, für zwei unter deren Werten. Ist diese niedrigere Bewertung der männlichen Schüler des Merkmals des Kompetenzerlebens (KE) ein höchst signifikanter, kleiner bis mittlerer Effekt ($p < 0.001$; $t(794) = -5.73$; $d = -0.41$), erweist diese Bewertungstendenz sich bei der kognitiven Aktivierung (KOG) nicht als systematischer Einfluss ($p = 0.251$; $t(794) = -1.15$; $|d| < 0.10$). Ebenfalls als nicht systematisch unterschiedlich bewertet, zeigt sich das Merkmal der Organisation der Lernumgebung (OL) ($p = 0.195$; $t(794) = 1.30$; $|d| < 0.10$). Die Höherbewertungen der weiteren Merkmale (SE, AE, SU und OZ) durch die weiblichen Teilnehmer sind höchst signifikant für die soziale Eingebundenheit (SE) ($p < 0.001$; $t(666) = 5.45$; $d = 0.39$) und die Schlüssigkeit der Lernumgebung (SU) ($p < 0.001$; $t(712) = 4.04$; $d = 0.29$), hoch signifikant für die Orientierung am Ziel (OZ) ($p = 0.002$; $t(666) = 3.07$; $d = 0.22$), sowie signifikant für das Autonomieerleben (AE) ($p = 0.034$; $t(753) = 2.13$; $d = 0.15$). Letztgenannter empirisch gefundener Unterschied ist dabei aber nicht bedeutsam ($d < 0.20$), die weiteren

4. Ergebnisse der Studie

Effekte weisen kleine Effektstärken ($0.22 < d < 0.39$) auf. Die Interventionen werden damit insgesamt hinsichtlich der Mehrheit der untersuchten Qualitätsmerkmale systematisch und bedeutsam unterschiedlich von weiblichen bzw. männlichen Studienteilnehmern bewertet.

Interventionsspezifische Teilgruppen

Zur Untersuchung der Einflüsse der Kontrollvariablen hinsichtlich der einzelnen Arten der Sequenzierung wird auch für jede der Interventionsteilgruppen (FeU-Gruppe und BMT-Gruppe) eine ANCOVA berechnet. Eine erneute Faktorenprüfung ergibt bei der Teilgruppe FeU keine signifikante Verbesserung der Modelle durch Hinzufügen weiterer Faktoren als die für die Gesamtstichprobe systematischen. Bei der Gruppe BMT hingegen erklärt das Modell mit den Faktoren VT, ZW, GE, SeA, DA und KG signifikant die meiste Varianz. Um wiederum die Analyse beider Teilgruppen gut vergleichen zu können, werden für beide Kovarianzanalysen die Faktoren des BMT-Modells zugrunde gelegt. Die Ergebnisse der Analysen sind in den Tabellen 4.8 sowie 4.9 zusammengefasst.

Untersuchter Aspekt	QS	DF	F-Wert	p-Wert	Effektstärke
Ergebnis Vortest	1754.90	1	196.42	0.0000 ***	$\omega^2 = 0.333$
Zweigzugehörigkeit	13.67	1	1.53	0.2168	$\omega^2 < 0.010$
Geschlecht	128.17	1	14.35	0.0002 ***	$\omega^2 = 0.033$
Interventionsstart	175.94	1	19.69	0.0000 ***	$\omega^2 = 0.046$
Interventionsdauer	2.40	1	0.27	0.6043	$\omega^2 < 0.010$
Klassengröße	7.92	1	0.89	0.3469	$\omega^2 < 0.010$
Residuen	3430.82	384			

Tabelle 4.8.: Ergebnisse der Kovarianzanalyse der Daten der Teilstichprobe der Interventionsgruppe FeU zum Einfluss auf das WTI-Ergebnis (N=391).

In beiden Teilgruppen zeigt sich kein systematischer Effekt der Zweigzugehörigkeit auf das Nachtstergesult. Vergleichbar ist zudem die Signifikanz, Richtung und Effektstärke für den Einfluss des Geschlechts innerhalb der Gruppe des FeU ($\omega^2 = 0.033$) bzw. des BMT ($\omega^2 = 0.038$). Das Ergebnis im Vortest und der Interventionsstart sind für beide Teilgruppen höchst signifikante Einflussgrößen, die Größe der Effekte ist jedoch in der BMT-Gruppe jeweils niedriger. Weist das Ergebnis im Vortest in der FeU-Gruppe einen Wert von $\omega^2 = 0.333$ auf, liegt die Effektstärke bei der BMT-Gruppe zwar ebenso im hohen Einflussbereich, allerdings bei einem niedrigeren Wert von $\omega^2 = 0.190$. Entsprechend verhält es sich auch bei den jeweils mittleren Effektstärken für den Interventionsstart

Untersuchter Aspekt	QS	DF	F-Wert	p-Wert	Effektstärke
Ergebnis Vortest	817.62	1	96.12	0.0000 ***	$\omega^2 = 0.190$
Zweigzugehörigkeit	5.40	1	0.63	0.4261	$\omega^2 < 0.010$
Geschlecht	144.59	1	17.00	0.0000 ***	$\omega^2 = 0.038$
Interventionsstart	106.88	1	12.57	0.0004 ***	$\omega^2 = 0.028$
Interventionsdauer	27.90	1	3.28	0.0709 .	$\omega^2 < 0.010$
Klassengröße	59.44	1	6.99	0.0085 **	$\omega^2 = 0.015$
Residuen	3385.45	398			

Tabelle 4.9.: Ergebnisse der Kovarianzanalyse der Daten der Teilstichprobe der Interventionsgruppe BMT zum Einfluss auf das WTI-Ergebnis (N=405).

(FeU: $\omega^2 = 0.046$; BMT: $\omega^2 = 0.028$). Die Interventionsdauer wird nur in der Teilgruppe BMT marginal signifikant, dabei hängt eine kürzere Interventionsdauer mit einem besseren Nachtestergebnis zusammen. Bedeutsam ist dieser Faktor allerdings in keiner der Teilgruppen (jeweils $\omega^2 < 0.010$). Nur in einer der beiden Teilgruppen bedeutsam und signifikant ist dabei die Klassengröße. So zeigt sich die Klassengröße in der BMT-Gruppe als hoch signifikanter ($p = 0.0085$) Einfluss mit einer kleinen Effektstärke ($\omega^2 = 0.015$), in der FeU-Gruppe jedoch nicht.

In der Bewertung der Qualitätsmerkmale der Interventionsstunden im Schülerfragebogen unterscheidet sich die FeU-Gruppe insgesamt in keinem der untersuchten Merkmale von der BMT-Gruppe (vgl. Tabelle 4.2). Unterschiedliche Bewertungen ergeben sich allerdings bei geschlechterspezifischer Betrachtung der Schülerbewertungen. Betrachtet man geschlechter- und interventionspezifische Teilgruppen, zeigen sich unterschiedliche Muster bei den Einschätzungen der Qualitätsmerkmale durch weibliche bzw. männliche Studienteilnehmer zwischen der FeU- und der BMT-Gruppe. Diese Unterschiedlichkeiten und deren Signifikanzen sind in Abbildung 4.5 dargestellt.

Vergleichbar zwischen diesen Gruppen zeigen sich die geschlechterspezifischen Einschätzungen bezüglich der sozialen Eingebundenheit und dem Kompetenzerleben und stimmen demzufolge auch mit den Bewertungen dieser beiden Merkmale in Abbildung 4.4 für die Gesamtstichprobe überein. Die Einschätzungen der weiblichen Schüler bezüglich der weiteren im Schülerfragebogen erhobenen Merkmale unterscheidet sich zu den Bewertungen durch die männlichen Schüler zwischen den Interventionsgruppen. Liegt in der FeU-Gruppe die Bewertung der weiblichen Teilnehmer signifikant unter (KOG) oder auf gleichem Niveau (AE, OL, SU, OZ) mit den Werten der männlichen Teilnehmer, ergeben sich bei diesem Vergleich signifikant höhere (AE, OL, SU, OZ) oder

4. Ergebnisse der Studie

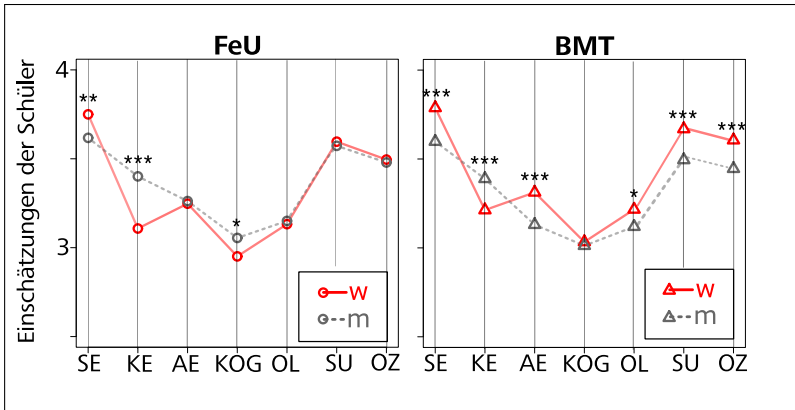


Abbildung 4.5.: Durch Linien verbundene Mittelwerte der Schülereinschätzung der Merkmale des Fragebogens durch weibliche (w) bzw. männliche (m) Schüler der Interventionsgruppen FeU bzw. BMT. Signifikante Mittelwertsunterschiede sind entsprechend gekennzeichnet. Die gestrichelten Linien deuten keine Entwicklung der Werte an, sondern dienen hier nur der Visualisierung der Bewertungsunterschiede. Benennungen vgl. Abb. 4.4.

gleich hohe (KOG) Werte in der BMT-Gruppe. Damit bewerten die weiblichen Teilnehmer die Intervention in der BMT-Gruppe insgesamt positiver als die männlichen Teilnehmer, in der FeU-Gruppe ist dabei keine so eindeutige Tendenz erkennbar. Betrachtet man vor diesem Hintergrund den Lernzuwachs (berechnet aus der Differenz des Nachtest- und des Vortestergebnisses im WTI) der weiblichen und männlichen Schüler innerhalb der Interventionsgruppen liegen die durchschnittlichen Ergebnisse der männlichen Studienteilnehmer jeweils über denjenigen der weiblichen Teilnehmer. Dieser Unterschied wird in der FeU-Gruppe marginal signifikant bei einer Effektstärke von $d = -0.19$ ($p = 0.054$; $t(389) = -1.94$), in der BMT-Gruppe zeigt sich bei einer minimal niedrigeren Effektstärke allerdings kein signifikanter Unterschied im Lernzuwachs ($d = -0.15$; $p = 0.124$; $t(403) = -1.54$). Der Einfluss des Geschlechts auf das Nachtestergebnis unter Kontrolle des Vortestergebnisses innerhalb der Interventionsgruppen errechnet sich aus den Ergebnissen der oben angegebenen mehrfaktoriellen Kovarianzanalysen allerdings für die BMT-Gruppe als etwas stärker ausgeprägt (vgl. Tabellen 4.9 und 4.8).

4.3.3. Zusammenfassung der Ergebnisse der Kovarianzanalysen

Für einen Vergleich der Ergebnisse der Kovarianzanalysen aus den vorangegangenen Abschnitten sind diese in leicht verkürzter Form in Tabelle 4.10 als Überblick zusammengefasst. Die Einflüsse des Geschlechts (GE) und der Sequenzierungsart (SeA) unterscheiden sich dabei kaum zwischen den Analysen verschiedener Teile der Stichprobe. Nur bei einzelnen Teilen (M und BMT) erweist sich die Klassengröße (KG) als signifikanter und kleiner Effekt, bei den weiteren Teilstichproben findet sich hingegen kein Einfluss dieser Variable auf das Nachtestergebnis im WTI. Der hoch bis höchst signifikante Einfluss des Interventionsstarts (ST) wird aus jeder Teilanalyse ersichtlich, jedoch unterscheiden sich die daraus kalkulierten Effektstärken. Diese kennzeichnen den Interventionsstart (ST) dabei aber stets als kleinen Effekt. Keinen Effekt hingegen stellen die Zweigzugehörigkeit (ZW) der einzelnen Schüler und die Dauer der Intervention (DA) bei den betrachteten Teilgruppen dar. Durchgehend einflussreichste Variable ist das Ergebnis im Vortest des WTI (VT), wobei die Effektstärke des VT vor allem zwischen den Interventionsgruppen FeU und BMT schwankt.

	GES		W		M		FeU		BMT	
	ω^2	Sig.	ω^2	Sig.	ω^2	Sig.	ω^2	Sig.	ω^2	Sig.
VT	0.257	***	0.233	***	0.271	***	0.333	***	0.190	***
ZW	<0.01		<0.01		<0.01		<0.01		<0.01	
GE	0.039	***	—	—	—	—	0.033	***	0.038	***
ST	0.027	***	0.016	**	0.029	***	0.046	***	0.028	***
DA	<0.01		<0.01		<0.01		<0.01		<0.01	.
KG	<0.01		<0.01		0.019	**	<0.01		0.015	**
SeA	0.023	***	0.021	**	0.019	**	—	—	—	—

Tabelle 4.10.: Effektstärken und Signifikanzen der untersuchten Variablen bei der Gesamtstichprobe (GES) den weiblichen (W) sowie männlichen (M) Studienteilnehmern und den Interventionsgruppen (FeU und BMT). In den einzelnen Analysen nicht berücksichtigte Variablen sind mit „—“ gekennzeichnet. Signifikante Effekte sind fett gedruckt.

4.3.4. Effekte der Intervention unter Kontrolle der Kompetenz der Erkenntnisgewinnung (WTE)

Die Erfassung der WTE-Ergebnisse zur Messung der Kompetenz der experimentellen Erkenntnisgewinnung der teilnehmenden Schüler erfolgte ausschließlich in der ersten Hauptstudie von Oktober 2013 bis März 2014. Für die Untersuchung des Einflusses dieser Kontrollvariablen auf die Effekte der Intervention stehen demnach nur die Daten von 12 Schulklassen mit $N = 291$ ($N_{FeU} = 148$; bzw. $N_{BMT} = 143$) Schüler zur Verfügung. Die Auswertung dieser Daten erfolgt erneut über eine mehrfaktorielle Kovarianzanalyse. Für diese Analyse wurde neben den in Abschnitt 4.3.1 als systematischen Faktoren bezeichneten noch als zusätzlicher Faktor das Ergebnis im WTE berücksichtigt. Modelle mit weiteren Faktoren führen zu keinem signifikant besseren Modell und weisen zudem keine Signifikanzen für die zusätzlich einzeln geprüften Faktoren auf. Die Ergebnisse der Kovarianzanalyse der Daten der ersten Hauptstudie mit den Faktoren VT, GE, ST, SeA, WTE sind in Tabelle 4.11 aufgeführt.

Untersuchter Aspekt	QS	DF	F-Wert	p-Wert	Effektstärke
Ergebnis Vortest (VT)	740.25	1	87.58	0.0000 ***	$\omega^2 = 0.229$
Geschlecht (GE)	263.65	1	31.19	0.0000 ***	$\omega^2 = 0.094$
Interventionsstart (ST)	184.37	1	21.81	0.0000 ***	$\omega^2 = 0.067$
Sequenzierungsart (SeA)	87.81	1	10.39	0.0014 **	$\omega^2 = 0.031$
Ergebnis im WTE (WTE)	25.43	1	3.01	0.0839 .	$\omega^2 < 0.010$
Residuen	2408.97	285			

Tabelle 4.11.: Ergebnisse der Kovarianzanalyse der Daten der ersten Hauptstudie zum Einfluss des WTE-Ergebnisses auf das WTI-Ergebnis ($N = 291$). WTE: Test zur Erfassung der experimentellen Erkenntniskompetenz.

Das Ergebnis des nur in der ersten Hauptstudie erhobenen WTE hat damit einen marginal signifikanten ($p = 0.0839$), aber nicht bedeutsamen Einfluss ($\omega^2 < 0.010$) auf das Nachtestergebnis im WTI. Sowohl das Nachtestergebnis im WTI als auch das WTE-Ergebnis weisen dabei bei der Interventionsgruppe der BMT einen signifikant höheren Wert auf (vgl.: Abschnitte 4.1 und 4.2) und unterscheiden sich somit systematisch zwischen den Interventionsgruppen. Eine zusätzliche Untersuchung (zweifaktorielle Kovarianzanalyse mit den Faktoren VT und WTE) möglicher, unterschiedlicher Ausprägungen des Einflusses des WTE-Ergebnisses zwischen diesen Gruppen ergibt aber weder bei Schülern der Interventionsgruppe FeU ($N = 148$; $F = 2.1$; $p = 0.15$) noch bei den Schülern aus der BMT-Gruppe ($N = 143$; $F = 0.1$; $p = 0.76$) einen signifikanten Einfluss der

WTE- auf die WTI-Ergebnisse im Nachtest. Innerhalb der Interventionsgruppen stellt damit das Ergebnis im WTE keinen signifikanten oder bedeutsamen Einflussfaktor auf das Nachtestergebnis im WTI dar.

Betrachtet man die weiteren Ergebnisse der Analyse der Daten der Stichprobe der ersten Hauptstudie (vgl. Tabelle 4.11), erweisen sich auch für diesen Teil der Stichprobe das Ergebnis im Vortest, das Geschlecht, der Interventionsstart und die Sequenzierungsart als höchst signifikante und bedeutsame Einflussfaktoren auf das Nachtestergebnis im WTI. Die Effektstärken der Faktoren des Vortestergebnisses (großer Effekt; $\omega^2 = 0.229$) und der Sequenzierungsart (kleiner Effekt; $\omega^2 = 0.031$) gestalten sich dabei vergleichbar mit den entsprechenden Werten der Faktoren aus der Gesamtstichprobe (vgl. Tabelle 4.5), wohingegen die Faktoren Geschlecht und Interventionsstart in der ersten Hauptstudie einen größeren Einfluss (jeweils mittlerer Effekt; GE: $\omega^2 = 0.094$; ST: $\omega^2 = 0.067$) als in der Gesamtstudie (jeweils kleiner Effekt; GE: $\omega^2 = 0.039$; ST: $\omega^2 = 0.027$) aufweisen.

4.3.5. Effekte der Intervention unter Kontrolle der kognitiven Grundfähigkeiten (KFTk)

In diesem Abschnitt werden die Ergebnisse der statistischen Auswertungen zu Effekten der Intervention auf das Nachtestergebnis unter Kontrolle der kognitiven Grundfähigkeiten (KFTk) angegeben. Die statistischen Auswertungen basieren auf den Daten der zweiten Hauptstudie von Oktober 2014 bis März 2015. Während dieses Zeitraums wurden die kognitiven Grundfähigkeiten über den KFTk bei $N=505$ ($N_{FeU} = 243$; bzw. $N_{BMT} = 262$) Schülern aus 20 Klassen ermittelt. Die Analyse der Daten dieser 20 Klassen erfolgt erneut über eine mehrfaktorielle Kovarianzanalyse. Als Faktoren dieser Analyse ergeben sich neben den als systematisch bezeichneten Faktoren und den über den KFTk ermittelten kognitiven Grundfähigkeiten keine weiteren Faktoren, welche die empirischen Daten signifikant besser erklären können. Keine der weiteren Faktoren stellt in zusätzlich berechneten Modellen einen signifikanten Einflussfaktor dar. Die Ergebnisse sind in Tabelle 4.12 angegeben.

Der Einfluss des Ergebnisses des KFTk ist höchst signifikant und kann als kleiner bis knapp mittlerer Effekt ($\omega^2 = 0.038$) bezeichnet werden. Sowohl beim WTE als auch beim KFTk hängen positive Testergebnisse auch mit positiven Ergebnissen der Schüler im Nachtest zusammen. Die Stärke dieses Zusammenhangs erweist sich dabei aber als unterschiedlich stark zwischen den Interventionsgruppen. Die Ergebnisse der Untersuchung (zweifaktorielle Kovarianzanalysen mit den Faktoren VT und KFTk) des Einflusses des KFTk innerhalb der beiden Interventionsgruppen lassen bei der FeU-Gruppe einen signifikanten, aber geringeren Einfluss ($N=243$; $F=5.4$; $p=0.0214$; $\omega^2 = 0.018$) auf das Nachtestergebnis im WTI erkennen, als in der BMT-Gruppe ($N=262$; $F=15.5$; $p=0.0001$; $\omega^2 = 0.052$).

4. Ergebnisse der Studie

Untersuchter Aspekt	QS	DF	F-Wert	p-Wert	Effektstärke
Ergebnis Vortest	1349.08	1	151.15	0.0000	$\omega^2 = 0.229$
Geschlecht	181.93	1	20.38	0.0000	$\omega^2 = 0.037$
Interventionsstart	37.85	1	4.24	0.0400	$\omega^2 < 0.010$
Sequenzierungsart	91.92	1	10.30	0.0014	$\omega^2 = 0.020$
KFTk	185.63	1	20.80	0.0000	$\omega^2 = 0.038$
Residuen	4435.88	497			

Tabelle 4.12.: Ergebnisse der Kovarianzanalyse der Daten der zweiten Hauptstudie zum Einfluss des KFTk-Ergebnisses auf das WTI-Ergebnis (N= 505). KFTk: Kurztest zur Erfassung der kognitiven Grundfähigkeiten.

In dieser Teilgruppe ergibt sich sogar ein höchst signifikanter Einfluss des KFTk.

In Bezug auf die Ergebnisse aus den Auswertungen der Daten der Gesamtstichprobe weisen die in Tabelle 4.12 neben dem KFTk angegebenen, weiteren Einflussfaktoren nur bei einem der betrachteten Faktoren wesentliche Unterschiede in den Effektstärken auf. Der Interventionsstart ist in der zweiten Hauptstudie zwar ein signifikanter aber im Gegensatz zur Gesamtstichprobe (vgl. Tabelle 4.5) kein bedeutsamer Effekt. In Übereinstimmung zu den Ergebnissen der Gesamtstichprobe (vgl. Tabelle 4.5) stellen das Ergebnis im Vortest einen großen Effekt ($\omega^2 = 0.229$), das Geschlecht ($\omega^2 = 0.037$) und die Sequenzierungsart ($\omega^2 = 0.020$) jeweils einen kleinen Effekt für das Nachtestergebnis im WTI dar. Die Effektstärken unterscheiden sich dabei nur geringfügig zwischen der Gesamt- bzw. der zweiten Hauptstudie, auch wenn sich der Einfluss der Sequenzierungsart für die hier betrachtete Teilstichprobe nicht als höchst signifikant wie in der Gesamtstudie erweist. Das Signifikanzniveau der Faktoren des Vortestergebnisses und des Geschlechts differieren hingegen nicht vom Ergebnis aus den Daten der gesamten Studie.

Beim Vergleich beider Teile der Hauptstudie erweist sich jeweils der Einfluss des Vortestergebnisses als höchst signifikant und stark ausgeprägt (jeweils: $\omega^2 = 0.229$). Bezüglich der anderen in beiden Analysen untersuchten Faktoren bestehen allerdings Unterschiede. Die untersuchten Aspekte des Geschlechts und der Art der Sequenzierung sind in beiden Gruppen höchst bzw. hoch signifikante Einflussfaktoren, weisen aber jeweils in der Teilgruppe der zweiten Hauptstudie geringere Effektstärken auf. Stellt das Geschlecht der Teilnehmer in der ersten Hauptstudie mit $\omega^2 = 0.094$ einen mittleren Effekt dar, repräsentiert dieser für die Teilnehmer der zweiten Hauptstudie nur einen kleinen Effekt $\omega^2 = 0.037$. Als kleiner Effekt erweist sich auch die Sequenzierungsart in beiden Teilgruppen, wie oben bereits angeführt unterscheiden sich allerdings die Werte

für die Effektstärken (1.Teil: $\omega^2 = 0.031$ bzw. 2.Teil: $\omega^2 = 0.020$). Der Einfluss des Interventionsstarts unterscheidet sich zudem hinsichtlich der Signifikanz zwischen den Teilgruppen. Zeigt sich der Einfluss dieses Aspekts im ersten Teil höchst signifikant und stellt einen mittleren Effekt ($\omega^2 = 0.067$) dar, so ist der Effekt des Interventionsstarts im zweiten Teil zwar noch signifikant aber nicht mehr bedeutsam ($\omega^2 < 0.010$).

4.3.6. Einfluss des Ergebnisses im Vortest des WTI

In diesem Abschnitt erfolgt die Präsentation der Ergebnisse aus der Untersuchung eines möglichen Moderatoreffektes des Vortestergebnisses auf die Stärke des Interventionseffektes hinsichtlich des Nachtestergebnisses im WTI. Das Vortestergebnis ist dabei genau dann eine Moderatorenvariable, wenn es die Stärke oder die Richtung eines Zusammenhangs zwischen einer Prädiktorvariable und der abhängigen Variable abändert. „Thus, a moderator effect is nothing more than an interaction whereby the effect of one variable depends on the level of another“ (Frazier et al., 2004, S. 116). Die Ergebnisse aus der Überprüfung der Signifikanz des untersuchten Moderatoreffektes sind in Tabelle 4.13 dargestellt.

Variable	Schätzwert	t-Wert	p-Wert
(Achsenabschnitt)	-0.003	-0.108	0.9136
Ergebnis Vortest (VT)	0.551	18.879	0.0000 ***
SeA (SeA)	0.126	4.325	0.0000 ***
Produktterm	-0.105	-3.598	0.0003 ***

Tabelle 4.13.: Ergebnisse der Regressionsanalyse zum möglichen Einfluss des untersuchten Moderatoreffektes des VT-Ergebnisses. Die angegebenen Variablen wurden vorher z-standardisiert

Der Produktterm der in Tabelle 4.13 angegebenen Regressionsanalyse ist ein höchst signifikanter Einflussfaktor und damit nach Frazier et al. (2004) das Vortestergebnis ein höchst signifikanter Moderator des Interventionseffekts auf das Nachtestergebnis im WTI. Zur Interpretation und Visualisierung dieses Moderatoreffekts erfolgt die Bildung von vier repräsentativen Teilgruppen der Gesamtstichprobe. Die Zuteilung einzelner Schüler zu diesen Gruppen wird ausgehend von deren Vortestergebnis vollzogen. Lag dieses um mehr als eine Standardabweichung unter dem Gesamtmittelwert (Ergebnis $VT < 9$; $N = 158$), erfolgt eine Zuordnung in Gruppe 1, lag das Vortestergebnis eines Schülers um mehr als eine Standardabweichung über dem Mittelwert, wurde dieser Schüler der Gruppe 3 zugewiesen (Ergebnis $VT > 13$, $N = 221$). Alle Schüler deren Vortestergebnis um weniger als eine Standardabweichung vom Gesamtmittel-

4. Ergebnisse der Studie

wert abweicht, bilden Gruppe 2 ($9 < \text{Ergebnis VT} < 13$; $N = 417$). Diese Gruppe wurde zusätzlich in Gruppe 2_1 ($9 < \text{Ergebnis VT} < 11$; $N = 243$) und Gruppe 2_2 ($11 < \text{Ergebnis VT} < 13$; $N = 174$) aufgeteilt, um vier nahezu gleich große Teilgruppen zu erhalten. Die Mittelwerte der Vor- und Nachtestergebnisse im WTI der so gebildeten Teilgruppen sind in Abbildung 4.6 dargestellt.

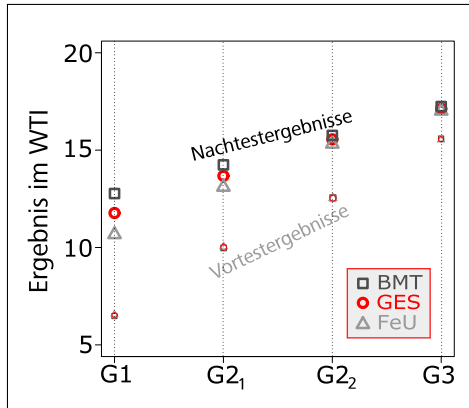


Abbildung 4.6.: Mittelwerte der Vor- und Nachtestergebnisse im WTI von Schülern der Teilgruppen G1, G2₁, G2₂ und G3. Die mittleren Vortestergebnisse der Schüler dieser Gruppen sind durch die unteren, kleineren Symbole dargestellt. „GES“ kennzeichnet die Mittelwerte aller Schüler in den Teilgruppen; „FeU“ bzw. „BMT“ die Mittelwerte von Schülern der Teilgruppen, welche nach den Vorgaben des FeU bzw. der BMT unterrichtet wurden.

Die Vortestergebnisse unterscheiden sich dabei zwischen den Interventionsgruppen nur unwesentlich, wohingegen bei den Nachtestergebnissen größere Mittelwertsunterschiede zu erkennen sind. Der Einfluss der Vortestergebnisse auf den Nachtest zeigt sich in Abbildung 4.6 in zweierlei Hinsicht. Zum Einen wird durch die geringer werdenden Abstände zwischen den Gesamtmittelwerten von Vor- und Nachtest deutlich, dass mit zunehmendem Wert im Vortest ein generell geringerer Lernzuwachs verbunden ist. Mit zunehmenden Vortestwerten sinkt zum anderen auch die Unterschiedlichkeit in den Gruppenmittelwerten zwischen den Interventionsgruppen. Dies spiegelt sich auch im negativen Vorzeichen des t-Werts des Produktterms in Tabelle 4.13 wieder. Der Interventionseffekt nimmt also bei steigendem Vortestwert ab. Dementsprechend hat die Intervention in den hier betrachteten Teilgruppen einen unterschiedlich starken Effekt und die Effektstärke des Interventionseffekts variiert. Um die entsprechenden Effektstärken angeben zu können und zu untersuchen, ob einzelne der Kontrollvariablen innerhalb dieser Teilgruppen ebenfalls

unterschiedlichen Einfluss aufweisen, erfolgen erneut Kovarianzanalysen der entsprechenden Daten. Bei der zugehörigen Faktorenprüfung für alle Teilgruppen wird jeweils kein signifikant besseres Modell gefunden, als das durch die systematischen Einflussfaktoren (VT, GE, ST, SeA) vorgegebene. Die einzelnen Ergebnisse der durchgeführten Kovarianzanalysen (Berücksichtigung nur der systematischen Faktoren) befinden sich im Anhang C, ein zusammenfassender Überblick der Ergebnisse ist in Tabelle 4.14 gegeben.

	Vortest		Geschlecht		Interventionsstart		Sequenzierungsart	
	ω^2	Sig.	ω^2	Sig.	ω^2	Sig.	ω^2	Sig.
G1	0.023	*	0.029	*	< 0.010	—	0.094	***
G2 ₁	0.014	*	0.017	*	0.058	***	0.024	**
G2 ₂	< 0.010	—	0.105	***	0.023	*	< 0.010	—
G3	0.114	***	0.078	***	< 0.010	—	< 0.010	—

Tabelle 4.14.: Effektstärken ω^2 der Faktoren VT, GE, ST und SeA aus den Kovarianzanalysen der Daten der Teilgruppen G1, G2₁, G2₂ und G3. Nicht signifikante Effekte sind hier zusätzlich mit „—“ gekennzeichnet. Die exakten p-Werte können dem Anhang C entnommen werden.

Die Varianzen der Variable VT (Vortestergebnis) innerhalb der betrachteten Teilgruppen sind dabei durch die Zusammensetzung der untersuchten Gruppen stark beeinflusst. Weisen Schüler der Gruppe G2₂ nur Vortestwerte von 12 oder 13 auf, treten bei Schülern der Gruppe G3 hingegen Vortestwerte zwischen 14 und 20 auf. Die Varianzen des Faktors VT unterscheiden sich paarweise signifikant zwischen den Gruppen. Auf einen Vergleich der Effektstärken aus den Teilgruppenanalysen mit den Werten für die Auswertung der Daten der Gesamtstichprobe wird deshalb verzichtet. Bei der Betrachtung der Effektstärken des Vortestergebnisses zwischen den einzelnen Teilgruppen zeigt sich der größte und zugleich höchst signifikante Einfluss bei der Gruppe G3. Der Einfluss in den restlichen Teilgruppen ist nur teilweise signifikant und weniger bedeutsam. Maximal zeigen sich hier kleine Effekte des Faktors VT. Der Faktor des Geschlechts ist als einziger der untersuchten Einzelaspekte in jeder Teilgruppe ein signifikanter und bedeutsamer Einflussfaktor. Allerdings erweist sich das Geschlecht bei den im Vortest unterdurchschnittlichen Schülern (G1 und G2₁) als kleiner und signifikanter Effekt, wohingegen die Rolle des Geschlechts für überdurchschnittliche Schüler (G2₂ und G3) einen höchst signifikanten und in einem stärkerem Maße bedeutsamen (mittlerer Effektstärke) Einflussfaktor auf das Nachtestergebnis darstellt. Auf dieses wirkt sich der Faktor des Interventionsstarts nur signifikant in den Gruppen G2₁ und G2₂ (die im Vortest als

durchschnittlich zu bezeichnenden Schüler) aus. Dabei ergibt sich bei Schülern der Gruppe G2₁ ein höchst signifikanter Effekt bei einer knapp mittleren Effektstärke und bei den der Gruppe G2₂ zugeordneten Schülern ein signifikanter und kleiner Effekt. In den Gruppen G1 und G3 zeigt sich kein signifikanter oder bedeutsamer Einfluss der Variable ST (Interventionsstart). Für den Einfluss der Art der Sequenzierung lässt sich eine abnehmende Effektstärke bei steigenden Vortestwerten erkennen. Zeigt sich die Art der Sequenzierung in G1 noch als deutlich mittlerer ($\omega^2 = 0.094$) und höchst signifikanter Effekt, so ist der Einfluss für Schüler der Gruppe G2₁ hoch signifikant und klein ($\omega^2 = 0.024$) und für die Schüler in den Gruppen G2₂ und G3 als nicht mehr systematisch zu bezeichnen. Die Richtung des Einflusses der Art der Sequenzierung ist dabei bei den untersuchten Teilgruppen gleich (vgl. Abbildung: 4.6). Die Schüler der BMT weisen in allen Gruppen im Durchschnitt bessere Testwerte auf, der Abstand zwischen den Interventionsmittelwerten für das Nachtestergebnis im WTI nimmt dabei von G1 bis G3 ab.

4.4. Zusammenhang der Merkmale des Schülerfragebogens mit dem WTI

Die Skalen des eingesetzten Schülerfragebogens sollen über die Schülereinschätzung die Ausprägung verschiedener Qualitätsmerkmale in den Interventionsstunden erfassen. Insgesamt werden Aspekte der Motivation im Rahmen der grundlegenden Bedürfnisse (SE, KE, AE), der kognitiven Aktivierung (KOG) und der Strukturierung des Unterrichts erhoben (OL, SU; OZ). Die Ergebnisse einer Untersuchung möglicher Zusammenhänge (Korrelationen) dieser Qualitätsmerkmale mit dem Nachtestergebnis im WTI wird in diesem Abschnitt für die Gesamtstichprobe und den interventionsspezifischen Teilgruppen angegeben. Diese Teilgruppen unterscheiden sich hinsichtlich der Bewertungen der mit dem Schülerfragebogen ermittelten Qualitätsmerkmale nicht signifikant (vgl. Tabelle 4.2), hinsichtlich der Korrelationen mit dem WTI-Nachtestergebnis sind in Tabelle 4.15 allerdings Unterschiede zu erkennen.

Auch für die Einordnung von Korrelationskoeffizienten r existieren Konventionen zur Einschätzung derer Bedeutsamkeit. Cohen (1988) liefert entsprechende Konventionen zur Beschreibung der Größe des Korrelationskoeffizienten r . Werte von $r = 0.10$ kennzeichnen einen kleinen, Werte von $r = 0.30$ einen mittleren und Koeffizienten ab $r = 0.50$ kennzeichnen große Effekte (zitiert nach Rasch et al., 2006a, S. 133). Die Beschreibung der Ergebnisse aus den Korrelationsanalysen erfolgt zunächst für die Gesamtstudie, anschließend werden Unterschiedlichkeiten zwischen den Interventionsgruppen thematisiert.

In der Gesamtstudie zeigen sich mehrere signifikante Korrelationen der Qualitätsmerkmale zum Nachtestergebnis im WTI. Als signifikanter, aber nicht bedeutsamer Zusammenhang erweist sich die Orientierung am Ziel (OZ). Kleine

4.4. Zusammenhang der Merkmale des Schülerfragebogens mit dem WTI

Merkmal	FeU		BMT		Gesamtstudie	
	cor	p-Wert	cor	p-Wert	cor	p-Wert
SE	0.01	0.878	−0.10	0.040 *	−0.05	0.149
KE	0.37	0.000 ***	0.24	0.000 ***	0.31	0.000 ***
AE	0.02	0.724	−0.08	0.123	−0.03	0.412
KOG	0.22	0.000 ***	0.07	0.149	0.15	0.000 ***
OL	0.13	0.008 **	0.08	0.115	0.11	0.002 **
SU	0.19	0.000 ***	0.01	0.775	0.10	0.003 **
OZ	0.12	0.014 *	−0.02	0.703	0.06	0.008 **

Tabelle 4.15.: Koeffizienten der Korrelationen (cor) zwischen mit dem Schülerfragebogen erhobenen Merkmale und dem Nachtestergebnis, sowie deren p-Wert aus der Überprüfung der Signifikanz der jeweiligen Korrelation. Signifikante Korrelationen sind fett gedruckt.

Korrelationseffekte finden sich bei Bewertung der Ausprägung der kognitiven Aktivierung (KOG), der Organisation der Lernumgebung (OL) und der Schlüssigkeit im Unterrichtsverlauf (SU). Wohingegen die Korrelation der Selbsteinschätzung der Schüler bezüglich ihres Kompetenzerlebens (KE) mit dem Nachtestergebnis im WTI ein mittlerer Effekt ist. Weder signifikante noch bedeutsame Korrelationen stellen die Einschätzungen der Schüler hinsichtlich der sozialen Eingebundenheit (SE) und des Autonomieerlebens (AE) dar.

Die für die Gesamtstichprobe ermittelten Werte für die Signifikanz und die Bedeutsamkeit dieser Korrelationen entsprechen dabei weitestgehend den Ergebnissen aus der Analyse der Korrelationen der Daten der Interventionsgruppe FeU, wenngleich für diese Teilgruppe die jeweiligen Koeffizienten der Korrelationen einen etwas höheren Wert annehmen. Die aus den Daten der Interventionsgruppe BMT berechneten Kennwerte weichen hingegen deutlich von den Werten der beiden anderen untersuchten Gruppen ab. Einzig die Korrelation des Kompetenzerlebens erweist sich als signifikant in allen untersuchten Gruppen, wobei der Korrelationskoeffizient in der BMT-Gruppe nur einen kleinen Effekt kennzeichnet. Die restlichen aus den Daten der Gesamtstudie oder der Interventionsgruppe FeU gefundenen signifikanten Korrelationen ergeben sich für die BMT-Gruppe nicht. Allerdings erweist sich in dieser Gruppe die Einschätzung der sozialen Eingebundenheit als kleiner negativer Zusammenhang mit dem Nachtestergebnis im WTI. Zeigen sich für alle Aspekte der Strukturierung (OL, SU; OZ) für die FeU-Gruppe kleine signifikante Korrelationen, finden sich diese für die BMT-Gruppe nicht. Auch der Zusammenhang der kognitiven Aktivierung mit dem WTI-Nachtestergebnis scheint ein anderer zu sein. Insgesamt zeichnet sich für die Schüler der Interventionsgruppe BMT eine

4. Ergebnisse der Studie

vergleichsweise große Unabhängigkeit derer Einschätzungen der Qualität der Interventionsstunden mit dem erreichten Nachtestergebnis ab. Die Art der Sequenzierung ist damit für den Zusammenhang der Unterrichtseinschätzungen der Schüler mit dem Nachtestergebnis von Belang.

Das Nachtestergebnis ist dabei einerseits von der Sequenzierungsart und andererseits auch stark vom Vortestergebnis abhängig (vgl. Abschnitt 4.3.1). Das Vortestergebnis moderiert dabei zusätzlich die Stärke des Einflusses der Art der Sequenzierung (vgl. Abschnitt 4.3.6). Dass damit abhängig vom Vortestergebnis und der Art der Sequenzierung Unterschiedlichkeiten zwischen Teilgruppen der Gesamtstudie bei der Einschätzungen der Lernenden über die Qualitätsmerkmale der Intervention auftreten könnten, erscheint vor diesem Hintergrund plausibel. Die Rolle des Vortestergebnisses und der Sequenzierungsart hinsichtlich dieser Einschätzung wird deshalb nachfolgend näher beleuchtet. Hierzu wird die Gesamtstichprobe wie in Abschnitt 4.3.6 je nach Vortestergebnis im WTI und Interventionsgruppenzugehörigkeit in acht Gruppen unterteilt (G1, G2₁, G2₂ und G3; jeweils noch in FeU und BMT unterteilt). Die durchschnittliche Einschätzung der Qualitätsmerkmale der Schüler der einzelnen Gruppen sind in Abbildung 4.7 dargestellt.

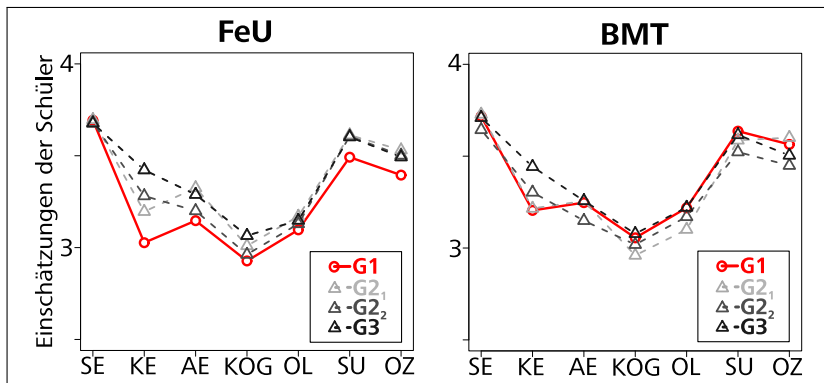


Abbildung 4.7.: Durch Linien verbundene Mittelwerte der Schülereinschätzung der Merkmale des Fragebogens durch die Gruppen G1, G2₁, G2₂ und G3 für die FeU- und die BMT-Gruppe. Die Linien deuten keine Entwicklung der Werte an, sondern dienen hier der Visualisierung einer Bewertungstendenz. Benennungen vgl. Abb. 4.4.

Die Skala der Einschätzung der Schüler ist in Abbildung 4.7 dabei nicht vollständig (1 bis 4) wiedergegeben, um etwaige Unterschiede in den Einschätzung der Schüler der einzelnen Gruppen hervortreten zu lassen. Grundsätzlich weichen die Mittelwerte der Schülereinschätzung um nicht mehr als 0.4 voneinander ab.

Die Mittelwerte für die soziale Eingebundenheit unterscheiden sich zwischen allen betrachteten Gruppen nur unwesentlich und nicht signifikant. Ein wesentlicher Unterschied zwischen den Interventionsgruppen zeigt sich bei den Bewertungen der Schüler aus G1 verglichen mit den Bewertung der Schüler der anderen Teilgruppen. Weist die mittlere Einschätzung von Schülern der G1 bei der Interventionsgruppe der BMT nur für das Kompetenzerleben den niedrigsten Wert auf, bewerten Schüler der G1 bei der Interventionsgruppe FeU alle Qualitätsmerkmale bis auf die soziale Eingebundenheit am niedrigsten. Diese geringere Bewertung der Merkmale durch die FeU-G1 Schüler ist dabei teilweise signifikant und bedeutsam. Die Unterschiede bei der Einschätzung des Kompetenzerleben sind beispielsweise marginal (Im Vergleich zu G₂₁), normal (Im Vergleich zu G₂₂) bzw. sogar höchst signifikant (Im Vergleich zu G3), bei Effektstärken zwischen $d = -0.31$ und $d = -0.81$. Kleine Effekte mit Effektstärken bei ca. $d = -0.3$ finden sich bei der Bewertung der Schlüssigkeit im Unterrichtsverlauf zwischen FeU-Schülern aus G1 und den restlichen Teilgruppen. Diese Effekte sind dabei mindestens marginal signifikant. Marginal signifikante Unterschiede treten auch in den Einschätzungen der G1 FeU-Schüler bei der Orientierung am Ziel (im Vergleich zu G₂₁) und beim Autonomieerleben (im Vergleich zu G3) auf. Einige dieser Gruppenunterschiede sind damit auch (marginal) systematisch. Systematisch unterschiedlich schätzen BMT-Schüler der G1 nur das Kompetenzerleben im Vergleich zu BMT-Schülern der G3 ein ($d = -0.43$). Die Bewertungen der Schüler aus der Interventionsgruppe BMT hängen damit kaum von deren Ergebnis im Vortest des WTI ab. Die im Vortest unterdurchschnittlichen (G1) FeU Schüler hingegen, weichen in ihren Bewertungen im Vergleich zu den anderen Teilgruppen hinsichtlich einiger Merkmale systematisch nach unten ab.

Ob sich nun auch die Einschätzungen der Schüler der Gruppe G1 systematisch voneinander unterscheiden, je nachdem welche Art der Intervention sie erfahren, ist Tabelle 4.16 zu entnehmen.

In allen untersuchten Merkmalen bis auf die soziale Eingebundenheit lässt sich eine einheitliche Tendenz in der Bewertungen der Qualitätsmerkmale erkennen. Dabei bewerten die BMT-Schüler der hier betrachteten Teilgruppe G1 die Merkmale Kompetenzerleben, Autonomieerleben, kognitive Aktivierung, Organisation der Lernumgebung, Schlüssigkeit im Unterrichtsverlauf, und Orientierung am Ziel im Schülerfragebogen jeweils höher und schätzen diese damit als stärker ausgeprägt ein. Diese Unterschiede im Ausprägungsgrad sind dabei unterschiedlich groß, signifikant und bedeutsam. Lassen sich für das Autonomieerleben und die kognitive Aktivierung keine signifikanten Effekte mittels t-Test nachweisen, ergeben sich marginal signifikante Effekte beim Kompetenzerleben und der Organisation der Lernumgebung, sowie signifikante Mittelwertunterschiede für die Schlüssigkeit im Unterrichtsverlauf und die Orientierung am Ziel. Die Effektstärken zwischen $d = -0.22$ und $d = -0.40$ deuten aber durchweg auf bedeutsame Unterschiede für Schüler der

4. Ergebnisse der Studie

Merkmal	M_{FeU}	M_{BMT}	p-Wert	Effektstärke
SE	3.69	3.72	$p = 0.710$	$ d < 0.10$
KE	3.03	3.20	$p = 0.054$.	$d = -0.30$
AE	3.15	3.25	$p = 0.144$	$d = -0.22$
KOG	2.93	3.06	$p = 0.101$	$d = -0.26$
OL	3.10	3.22	$p = 0.073$.	$d = -0.29$
SU	3.49	3.64	$p = 0.018$ *	$d = -0.38$
OZ	3.39	3.56	$p = 0.015$ *	$d = -0.40$

Tabelle 4.16.: Mittelwerte der Einschätzungen der Qualitätsmerkmale der FeU-Schüler (N= 75) und der BMT-Schüler (N= 83) aus der Gruppe G1. Zusätzlich sind die p-Werte und die Effektstärken für die Unterschiedlichkeit der jeweiligen empirisch gefundenen Mittelwerte zwischen den Interventionsgruppen angegeben.

Gruppe G1 zwischen den Interventionsgruppen hin. Die Effektstärken für die Unterschiedlichkeit in den Einschätzungen der Schüler liegen dabei im Bereich kleiner Effekte. Für Schüler der Gruppe G1 ist damit die Zugehörigkeit zur Interventionsgruppe FeU bzw. BMT für ihre Wahrnehmung über die Qualität der Interventionsstunden zumindest hinsichtlich einzelner Qualitätsmerkmale nachweislich als wesentlich einzuschätzen.

5. Interpretation und Einordnung der Ergebnisse

In Kapitel 4 wurden die wesentlichen Ergebnisse aus dieser explorativen Studie angegeben. In diesem Kapitel soll nun eine Einschätzung dieser Erkenntnisse primär vor dem Hintergrund bereits vorhandener Erkenntnisse erfolgen. Hierzu werden im ersten Abschnitt 5.1 die vorab formulierten Forschungsfragen der Studie anhand der einzelnen interventionsabhängigen Ergebnisse näher beleuchtet. Dabei werden auch die Ergebnisse weiterer Studien berücksichtigt. In einem zweiten Abschnitt 5.2 werden neue Erkenntnisse der Studie interpretiert, welche sich nicht direkt auf die Intervention und damit die unabhängige Variable der Untersuchung beziehen. Die in diesem Abschnitt thematisierten Erkenntnisse stehen auch nicht in direktem Zusammenhang zu den Forschungsfragen dieser Studie, sondern sind vielmehr Resultat der explorativen Ausrichtung der Untersuchung.

5.1. Wirkungen der Interventionen hinsichtlich der Forschungsfragen

Ziel dieses Abschnitts ist die Einschätzung derjenigen neuen Erkenntnisse, welche aus den Unterschieden zwischen den Interventionen erwachsen. Die Beschaffenheit dieser Unterschiede ist für die Einschätzung relevant und wird deshalb nachfolgend nochmals kurz angegeben.

Dabei ist von Belang, dass sich die Interventionen bzw. die zugehörigen Interventionsstichproben hinsichtlich vieler Aspekte eben nicht unterscheiden. Keine Unterschiede zwischen den Interventionsstichproben zeigen sich hinsichtlich der organisatorischen Kontrollvariablen (vgl. Abschnitt 3.5.1) und der kognitiven Grundfähigkeiten der Schüler. Signifikant unterschiedlich sind die Gruppen lediglich bezüglich des Ergebnisses im wenig reliablen Wissenstest zur Erkenntnisgewinnung (WTE). Kaum Unterschiede zwischen den Interventionseinheiten bestehen zudem hinsichtlich Lehrerverhalten, sowie Inhalt und Methodik (vgl. Abschnitt 3.4). Die Unterschiede liegen auf einer tieferen Ebene des Unterrichts, nur in Details der jeweiligen Strukturierung der Interventionseinheiten. Sie liegen dabei im wesentlichen in der Er- und Verarbeitung der Sachinhalte. Eine Abstraktion aller Erkenntnisse aus den experimentell gewonnenen Daten und keine reine Übung des Inhalts sondern primär dessen

5. Interpretation und Einordnung der Ergebnisse

fächerübergreifende Vernetzung im FeU stehen in der BMT einer Generalisierung nur von Teilaspekten aus den experimentellen Daten, verbunden mit einer Darstellung weiterführender Inhalte, deren Anwendung gerade um zu üben und einer auch innerfachlichen Vernetzung der erlernten Sachinhalte gegenüber. Diese Unterschiedlichkeiten zwischen den Interventionsgruppen sind es, welche für die Interpretation der Ergebnisse aus den Analysen der Werte für die abhängigen bzw. kontrollierten Variablen zugrunde gelegt werden.

5.1.1. Umsetzbarkeit und Lernwirksamkeit des FeU

F1: Ist eine Sequenzierung nach den Vorgaben des FeU umsetzbar und lernwirksam im Physikunterricht?

Die Frage nach der Lernwirksamkeit ist im Hinblick auf die Vor- und Nachtestergebnisse im Wissenstest zum Impuls (WTI; vgl. Abschnitt 4.2) eindeutig zu bejahen. Die Schüler erlernen zusätzliches Fachwissen zum Impuls. Eine Einordnung der Lernwirksamkeit des FeU im Vergleich zu anderen Strukturvorgaben im Sinne einer Rangfolge wie dies im Rahmen der Studie von Hattie (2009, S. 297ff.) vorgenommen wurde, kann hier nur sehr eingeschränkt vorgenommen werden. Dies liegt zum einen an der fehlenden Literaturbasis vor allem zum FeU und zum anderen am Design dieser Studie. Nur ein Vergleich zur Lernwirksamkeit der Unterrichtseinheit zum Thema Impuls nach den Vorgaben der BMT kann angegeben werden. Dies wiederum ist Schwerpunkt des Abschnitts 5.1.2. Grundsätzlich erweist sich die Lernwirksamkeit des FeU-Unterrichts dabei als im Vergleich zum BMT-Unterricht geringer. Alle im Rahmen dieser Arbeit durchgeführten Datenauswertungen kennzeichnen diese Unterschiedlichkeit im Lernerfolg als systematischen und kleinen Effekt. Über die Nachhaltigkeit des Lernzuwachses in der FeU-Gruppe können mit dieser Untersuchung keine zusätzlichen Aussagen formuliert werden. Ob sich also wie in der (etwas eingeschränkt belastbaren) Studie von Herbers (1990) angedeutet die im FeU erlernten Sachinhalte als besonders zeitstabil erweisen, bleibt offen.

Für den in dieser Untersuchung verwendeten Sachkontext des Impulses gestalten sich die Vorgaben des FeU zur Sequenzierung ebenso gut umsetzbar wie die entsprechenden Richtlinien der BMT. Die Unterrichtseinheiten die konform zu den theoretischen Vorgaben konzipiert wurden, werden von den Studienteilnehmern hinsichtlich der erhobenen Qualitätsmerkmale auch vergleichbar positiv bewertet (vgl. Tabelle 4.2). Diese Positivbewertung (alle Merkmale werden als überdurchschnittlich gut eingeschätzt) kann dabei auch als Maß für eine gelungene Umsetzung des FeU angesehen werden. Die Umsetzung der Vorgaben der FeU führt also zu einem von den Schülern als gut empfundenen und nachweislich lernwirksamen Unterricht.

Trotzdem sehe ich die generelle Umsetzbarkeit der Vorgaben des FeU als etwas eingeschränkt im Vergleich zu den Vorgaben der BMT an. Das FeU benötigt

zwingend einen Sachverhalt, dessen wesentliche Eigenschaften direkt ableitbar aus den Ergebnissen von idealerweise von Schülern selbst durchgeführten Experimenten sind. Im Rahmen der Konzeption der Studie wurden mehrere, ebenfalls als Grundlage für die Interventionsstunden mögliche Sachverhalte (z.B. Einführung des Konzepts der Kraft) verworfen, da sich diese nach meiner Einschätzung eben nicht gut aus Experimentiererergebnissen (allein) ableiten lassen.

Insgesamt schätze ich eine Sequenzierung nach den Vorgaben des FeU für eine Vielzahl an physikalischen Sachverhalten als gut umsetzbar und lernwirksam im Physikunterricht ein, auch wenn möglicherweise lernwirksamere Alternativen verfügbar wären.

5.1.2. Lernwirksamkeit der Interventionsstunden

F3: Zeigen sich die unterschiedlichen Sequenzierungen auch unterschiedlich lernwirksam?

Die unterschiedlichen Sequenzierungen erweisen sich eindeutig, systematisch und unabhängig von der Art der durchgeführten Datenauswertung auch als unterschiedlich lernwirksam. Diese Aussage hat ihre Gültigkeit für die hier durchgeführte Intervention im Rahmen eines Physik(labor)unterrichts zum Thema Impuls. Eine Generalisierung auf reale schulische Situationen oder andere Sachinhalte innerhalb der Physik ist nicht zulässig und lag auch nicht im Fokus dieser Untersuchung.

Bemerkenswert eindeutig sind insgesamt die Ergebnisse und Effektstärkeberechnungen aus den verschiedenen Analysen. Sowohl die durchgeführten t-Tests zum Unterschied im Nachtestergebnis oder dem Lernzuwachs, die mehrfaktorielle Kovarianzanalyse (Berücksichtigung der Kontrollvariablen) oder die Mehrebenenanalyse (Berücksichtigung der hierarchischen Datenstruktur) kennzeichnen die Sequenzierungsart als einen hoch bis höchst signifikanten kleinen Effekt. Repräsentativ für die Stärke des Effekts kann dabei der Wert $d = -0.3$ betrachtet werden. Dieser Wert ergibt sich aus dem t-Wert zum Test der Unterschiedlichkeit des Lernzuwachses, aus der Mehrebenenanalyse und der Umrechnung der Effektstärke $\omega^2 = 0.023$ aus den Kovarianzanalysen nach Thalheimer und Cook (2002). Was bedeutet dieser Wert von $d = -0.3$? Nach den Konventionen von Cohen (1988) kennzeichnet eine Effektgröße von $0.20 \leq d < 0.50$ einen kleinen Effekt (zitiert nach Rasch et al., 2006a, S. 68). Allein nach dieser Konvention beurteilt, liegt damit ein Effekt vor, dessen Einfluss auf den Unterricht vorhanden, wenngleich nicht besonders stark ausgeprägt ist. Die durchschnittliche Effektgröße für unterrichtliche Einflussfaktoren gibt Hattie (2009, S. 16) mit $d = 0.4$ an, bezeichnet diesen Kennwert als *hinge-point* und wertet alle Einflussfaktoren mit größeren Effektstärken als definitiv erstrebenswert. Nun liegt der in dieser Untersuchung gefundene Kennwert unter diesem

5. Interpretation und Einordnung der Ergebnisse

hinge-point und damit im unterdurchschnittlichen Einflussbereich aller bei Hattie berücksichtigter unterrichtlicher Faktoren. Oberflächlich betrachtet, läge es damit durchaus nahe, das Ergebnis dieser Studie zur Kenntnis zu nehmen, den eigenen Fokus aber gleich wieder auf eben überdurchschnittliche Einflussfaktoren zu legen. Eine tiefer gehende Betrachtung des Effekts und seiner Stärke liefert nach meiner Ansicht allerdings Argumente für einen Fokus auf den Einfluss der Sequenzierungsart.

Bei der Festlegung des *hinge-points* schränkt Hattie selbst dahingehend ein, dass Effekte mit Stärken unter $d = 0.4$ ebenso erstrebenswert sein könnten. Dies hänge dabei von zwei verschiedenen Bedingungsfaktoren ab. Als ersten Bedingungsfaktor nennt Hattie (2009, S. 16) den mit dem Effekt verbundenem finanziellen und organisatorischen Aufwand. Der zusätzliche Aufwand einer Berücksichtigung der BMT bei der Strukturierung von Unterrichtseinheiten ist dabei offensichtlich nicht finanziell hoch, da kein zusätzliches Material oder Personal benötigt wird. Aus organisatorischer Perspektive ist der mit der BMT oder dem FeU verbundene zeitliche Aufwand ebenfalls als nicht besonders hoch einzuschätzen. Der Umgang mit der BMT oder dem FeU muss zwar erlernt und geübt werden, kann aber den Planungsaufwand bei der Gestaltung von Unterricht sogar verringern. Zahlreiche Fragen im Kontext unterrichtlicher Planungen werden speziell durch die Vorgaben der BMT automatisch beantwortet. Ein Zusätzlicher Aufwand ist bei der Arbeit mit der BMT oder dem FeU damit nicht verbunden. Als zweiten Bedingungsfaktor um eigentlich unterdurchschnittliche Effekte doch als relevant zu bewerten, stellt Hattie (2009, S. 16) die Forderung, dass diese Effekte im direkten Zusammenhang mit anderen Einflussfaktoren stehen. Die Art der Sequenzierung als Grundlage des hier betrachteten Effekts ist ein wesentlicher Teilaspekt der didaktischen Strukturierung. Für diese Bedeutungsfacette der Strukturierung vermutet Lipowsky (2009, S. 86), „dass ein Mindestmaß an didaktischer Strukturierung eine notwendige Voraussetzung für eine wirksame Klassenführung darstellt, die wiederum als wichtige Voraussetzung dafür angesehen werden kann, dass inhaltsbezogene Strukturierungen und Hinweise Wirkungen entfalten können“. Ein Zusammenhang von Aspekten der Sequenzierung mit weiteren Einflussfaktoren kann folglich ebenfalls angenommen werden. Trotz einer für unterrichtliche Bedingungsfaktoren unterdurchschnittlichen Effektstärke deutet sich damit aus zweierlei Perspektive eine Berechtigung für eine Fokussierung auf bestimmte Sequenzierungen bzw. Strukturvorgaben an.

Legt man den Fokus auf die Effektstärken verschiedener Strukturvorgaben in der Studie von Hattie (2009, S. 297ff.) (Direct Instruction $d = 0.59$; Problem-based learning $d = 0.15$; Inquiry-based teaching $d = 0.31$) oder der Metastudie von Alfieri et al. (2011, S. 9) (Discovery Learning in den Naturwissenschaften $d = 0.11$) zeigt sich, dass diese Effektstärken selten Hatties *hinge-point* übersteigen. Die Aussagen dieser Metastudien basieren dabei auf Untersuchungen in denen die Wirkungen in der Interventionsgruppe (Unterricht nach den theoreti-

schen Vorgaben der untersuchten Strukturvorgaben) mit denen einer Vergleichs- oder Kontrollgruppe (meist ein Konglomerat verschiedener Unterrichtsabläufe, oft ohne theoretische Fundierung) gegenübergestellt wurden. In meiner Studie wurde hingegen eine möglicherweise unterschiedliche Wirkung zweier verschiedener, dabei didaktisch gut begründeter theoretischer Strukturvorgaben untersucht. Die dabei angegebene Effektstärke kennzeichnet damit die Unterschiedlichkeit der Wirkung zweier sorgsam geplanter, theoretisch fundierter Unterrichtseinheiten untereinander und nicht im Vergleich zu einem eher undefiniertem Unterrichtsmix, welcher gerne als *herkömmlicher Unterricht* bezeichnet wird (Wagner, 1999, S. 3). Die empirisch gefundene Effektstärke von $d = -0.3$ für die Art der Sequenzierung beschreibt also das Ausmaß der Unterschiedlichkeit der Wirkung zweier Strukturvorgaben. Vergleichbar wäre beispielsweise eine Untersuchung, welche die Wirkung des Inquiry-based teachings $d = 0.31$ mit derjenigen der Direct Instruction $d = 0.59$ vergleichen würde. Dieses Beispiel ist dabei nicht zufällig gewählt. So teilt sich das FeU wesentliche Grundzüge mit dem 5E-Modell, welches wiederum zur Kategorie des Inquiry-based teachings gezählt werden kann und die im Rahmen dieser Studie eingesetzte Verkettung zweier Basismodelle enthält zusätzlich wiederum wesentliche Aspekte, welche sich auch in der Direct Instruction wieder finden. Die gefundene Effektstärke für die Sequenzierungsart von $d = -0.3$ kennzeichnet damit in gewisser Weise auch die Unterschiedlichkeit in der Wirkung des Inquiry-based teachings im Vergleich zur Direct Instruction am gewählten Sachbeispiel und steht mit der von Hattie gefundenen Unterschiedlichkeit in den einzelnen Effektstärken in Einklang. Die Annahme einer größeren Effektstärke als $|d| = 0.3$ und ein damit verbundenes Überschreiten von Hatties *hinge-point* bei einem entsprechendem Vergleich der Wirkung der BMT mit der Wirkung von *herkömmlicher Unterricht* erscheint damit zumindest plausibel, kann mit dieser Studie aber nicht nachgewiesen werden. Eine Einordnung in die von Hattie vorgenommene Effektstärkenrangfolge verschiedener Einflussfaktoren erscheint nur begrenzt möglich. Dazu fehlt eine Angabe der Lernwirksamkeit des FeU bzw. der BMT gegenüber einer klassischen Kontroll- oder Vergleichsgruppe. Im Falle der BMT liegt durch die Studie von Zander et al. (2013) zwar ein entsprechender Kennwert vor ($d = 0.32$), allerdings unterscheidet sich diese Studie in ihrem Design (Feldstudie, Lehrerfortbildung, Vergleichsgruppe) und der teilweise damit verbundenen geringeren Theoriekonformität der Intervention (Wackermann, 2008, S. 63f. und 77) wesentlich von der Konzeption meiner Untersuchung, so dass ich diesen Wert zumindest nicht direkt auf die Wirkung der von mir durchgeführten Intervention übertragen würde.

Grundsätzlich deutet sich mit der Effektstärke von $d = -0.3$ zwischen den Wirkungen der Strukturvorgaben eine bedeutsame Distanz in deren Lernwirksamkeit an, welche für die Verwendung der BMT spricht. Der Einsatz der BMT ist dabei nicht mit einem dauerhaft größeren Aufwand verbunden, könnte aber die Wirkungen zusätzlicher Strukturierungsmaßnahmen positiv beeinflussen,

5. Interpretation und Einordnung der Ergebnisse

weshalb die Verwendung der BMT bei der Gestaltung von Lehr-Lern-Sequenzen zu einer bemerkbaren und leistbaren Verbesserung des Physikunterrichts führen könnte.

Damit soll auch die Bewertung der Effektstärke des Unterschieds abgeschlossen und die Frage nach dessen Ursache in den Vordergrund gerückt werden. Es wird angenommen, dass die Ursache für die Unterschiedlichkeit der Lernwirksamkeit in den oben angeführten wesentlichen aber im Detail liegenden Unterschieden der Sequenzierung der Unterrichtseinheiten liegt. Diese Unterschiede lagen dabei in der Er- und Verarbeitung der Sachinhalte. Eine differenzierte Betrachtung welcher Teilaspekte sich nun genau für die Unterschiedlichkeit der Lernwirksamkeit verantwortlich zeigt, kann dabei allerdings nicht vorgenommen werden. Die Abstraktion der neuen Sachinhalte aus den experimentell gewonnenen Daten in Kombination mit einer fächerübergreifenden Vernetzung ohne expliziter reiner Übungsphasen kann aber als weniger lernwirksam als die alternative Umsetzung nach den Vorgaben der BMT bezeichnet werden. Bei dieser alternativen Umsetzung werden einzelne Sachinhalte dargestellt und müssen dabei eben nicht abstrahiert oder hergeleitet werden, werden dafür aber explizit eingeübt und sollen auch innerfachlich vernetzt werden. Wobei an dieser Stelle darauf hingewiesen werden sollte, dass auch die Intervention nach den Vorgaben des FeU Anwendungsphasen der neuen Sachinhalte aufweist, allerdings zumindest teilweise mit einer anderen übergeordneten Zielsetzung (Fächerübergreifende Vernetzung). In beiden Gruppen erfolgt aber die Bearbeitung eines identischen Musterbeispiels nach der Erarbeitung der wesentlichen neuen Sachinhalte. Welche neuen Inhalte dabei in Abhängigkeit von deren Art (also ob diese nun *gefunden* werden können, oder ob sie *erfunden* sind), wie erarbeitet oder eingeführt werden, gibt die BMT in ihren theoretischen Vorgaben zum Unterricht an. Das FeU differenziert bei der Erarbeitung neuer Inhalte nicht. Dies unterscheidet die BMT wesentlich von dem FeU und erlaubt ein unterrichtlich flexibleres Vorgehen. Diese Flexibilität ermöglicht anscheinend eine Unterrichtsgestaltung, welche eine größere Passung zu den tatsächlich notwendigen Lernprozessen der Schüler erreicht, als dies mit den Vorgaben des FeU gelingt. Auch für eine explizite Übungsphase und die Betonung der innerfachlichen Vernetzung der Inhalte deutet sich in dieser Studie eine größere Lernwirksamkeit an, als für eine implizite Anwendungsphase mit rein fächerübergreifender Vernetzung.

Im Kontext zu diesen Einschätzungen stehen auch die Ergebnisse zweier videobasierter Vergleichsstudien von finnischem, deutschem und schweizerischem Unterricht innerhalb des QuIP-Projekts (Quality of Instruction in Physics). In der ersten Videostudie konnten Helaakoski und Viiri (2014, S. 101) den positiven Einfluss der innerfachlichen Vernetzung auf die Schülerleistung nachweisen. Aspekte der innerfachlichen Vernetzung (Frequency of connections between physics concepts, Connection density) zeigen in dieser Untersuchung signifikante, kleine bis mittlere positive Korrelationen mit den durchschnittlichen Lernzu-

wachsen von Schülern in N= 98 Klassen. Ein Vergleich zum Zusammenhang von Aspekten der fächerübergreifenden Vernetzung mit der Schülerleistung kann hier nicht vorgenommen werden, da diese in der Studie von Helaakoski und Viiri (2014) nicht erhoben wurden. Die durchschnittliche Häufigkeit von Phasen innerfachlicher Vernetzung fällt im Ländervergleich in Deutschland dabei am geringsten aus (Helaakoski & Viiri, 2014, S. 102). In der zweiten QuIP-Studie untersuchten Geller, Neumann und Fischer (2014) das Vorkommen und die Häufigkeit von bestimmten Lernprozessorientierungen zur Instruktion von Konzepten. Geller et al. (2014, S. 86) unterschieden dabei zwischen *mainly concept building*, *concept building as introducing part*, *concept building as continuing part* und *rarely concept building*. Die Instruktion von Konzepten erfolgte dabei in Deutschland meist in Form einer Lernprozessorientierung, welche dem *rarely concept building* zugeordnet werden kann. Diese Orientierung vermischt dabei Aspekte verschiedener Basismodelle mit Phasen in denen keinerlei Bezug zu den Basismodellen ersichtlich ist. Die Übereinstimmung zwischen theoretischen Vorgaben durch die BMT und dem beobachteten Unterricht sind bei dieser Lernprozessorientierung gering. Die Lernprozessorientierung des *rarely concept buildings* weist dabei in den N= 99 teilnehmenden Klassen einen negativen Zusammenhang zur Schülerleistung auf. Positive Zusammenhänge ergaben sich für die Orientierung *mainly concept building*, welche in Finnland und der Schweiz das am häufigsten beobachtete Unterrichtsmuster darstellt (Geller et al., 2014, S. 87ff.) und im wesentlichen mit den Vorgaben der BMT zur Deckung kommt. Etwas vereinfacht formuliert, zeigt also die innerfachliche Vernetzung positive Auswirkungen, eine Abweichung von der Struktur der Basismodelle wiederum negative Auswirkungen auf die Schülerleistung.

Die Befunde beider Videostudien innerhalb des QuIP-Projekts stehen damit in Übereinstimmung zu den Einschätzung für die Ursachen der unterschiedlichen Lernwirksamkeit der beiden Strukturvorgaben in meiner Studie. Abweichungen vom BMT-Muster zum Aufbau eines physikalischen Konzepts und das Fehlen von Phasen der innerfachlichen Vernetzung sind demnach mit einer geringeren positiven Wirkung auf die Schülerleistung verbunden (Geller et al., 2014; Helaakoski & Viiri, 2014).

5.1.3. Lernwirksamkeit der Intervention bei Teilgruppen

Zeigen sich die unterschiedlichen Sequenzierungen auch hinsichtlich einzelner Teilgruppen der Lernenden unterschiedlich lernwirksam?

Die systematisch unterschiedliche Lernwirksamkeit des FeU und der BMT wurde zusätzlich noch auf weitere Unterschiedlichkeiten hinsichtlich einzelner Teilgruppen untersucht. Zunächst erfolgt eine Einschätzung der Lernwirksamkeit bezüglich geschlechterspezifischer Teilgruppen und anschließend bezüglich vier Teilgruppen (G1, G2₁, G2₂ und G3), welche ex post auf Basis des Vortester-

5. Interpretation und Einordnung der Ergebnisse

gebnisses im WTI gebildet wurden.

Geschlechterspezifische Teilgruppen

Der Lernzuwachs (berechnet aus Differenz des Nachtest- und des Vortestergebnisses im WTI) der weiblichen Schüler der FeU-Gruppe ($M_{WF} = 2.43$) und der BMT-Gruppe ($M_{WB} = 3.51$) ist dabei jeweils geringer als der Zuwachs bei den männlichen Schülern ($M_{MF} = 3.06$; bzw. $M_{MB} = 4.05$). Auf diese interventionsunabhängige Unterschiedlichkeit wird in Abschnitt 5.2.1 näher eingegangen. An dieser Stelle soll der Fokus zunächst auf den geschlechterspezifischen Lernzuwachsunterschieden zwischen und innerhalb der Interventionsgruppen liegen. Innerhalb der Interventionsgruppen zeigt sich bei den FeU-Schülern ein systematischer aber nicht besonders bedeutsamer Unterschied ($d = -0.19$) zu Ungunsten der weiblichen Teilnehmer. Nur unwesentlich schwächer ($d = -0.15$) aber nicht mehr systematisch zeigt sich der Einfluss des Geschlechts auf den Lernzuwachs bei den BMT-Schülern. Auf Basis der t-Tests deutet sich folglich eine leichte Unterschiedlichkeit der Wirkung der Sequenzierungen hinsichtlich der hier betrachteten Teilgruppen an, wobei die FeU-Gruppe einen etwas größeren Geschlechtereffekt aufweist. Betrachtet man aber die Ergebnisse der Kovarianzanalyse der interventionsspezifischen Teilgruppen (Vergleiche Tabellen 4.8 und 4.9) deuten diese wiederum bei der BMT-Gruppe einen leicht größeren Geschlechtereffekt an. Damit liegen zunächst einmal widersprüchliche Aussagen vor, welche eine sichere Einschätzung des Einflusses der Sequenzierungsart auf die geschlechterspezifischen WTI-Ergebnisunterschiede erschweren. Eine mögliche Ursache für diese Widersprüchlichkeit könnte in der stark unterschiedlichen Einflussnahme des Vortestergebnisses in den entsprechenden mehrfaktoriellen Kovarianzanalysen liegen. So erklärt das Vortestergebnis in der FeU-Gruppe wesentlich mehr Varianz der Daten. Dies kann zur Folge haben, dass sich weitere Faktoren wie das Geschlecht der Schüler verstärkt gleiche Varianzen mit dem Vortestergebnis teilen und folglich der Einfluss dieser weiteren Faktoren unterschätzt wird. Eine Betrachtung einer einfaktoriellen Kovarianzanalyse hinsichtlich des Einflusses des Geschlechts auf den Lernzuwachs ist dabei nicht zielführend, da diese Analyse identisch mit oben angegebenen t-Tests wäre. Ebenso wenig zielführend wäre eine Kovarianzanalyse ohne den bedeutendsten Prädiktor (Vortestergebnis) durchzuführen.

Somit verbleibt dieser Einfluss der Intervention auf die Unterschiedlichkeit im Lernzuwachs bei weiblichen und männlichen Studienteilnehmern schwierig einzuschätzen. Zumindest erscheint dieser Einfluss als zu klein um diesen aus den vorhandenen Daten als systematisch zu identifizieren. Die Existenz des Effekts ist aber möglich und würde auch im Einklang zu Ergebnissen einer ebenfalls von mir durchgeführten Studie stehen, bei welcher sich weibliche Schüler als wesentlich sensibler für Reihenfolgeaspekte von Experimenten erwiesen haben (Maurer & Rincke, 2013, S. 121). Ein Befund einer höheren Sensitivität für die erwiesenermaßen leistungsrelevante Art und Weise der Reihenfolge und Verbindung von unterrichtlichen Phasen wäre vor diesem

Hintergrund nicht überraschend gewesen.

Teilgruppen nach Vortestergebnis

Die Unterschiede im Nachtestergebnis des WTI zwischen Schülern der FeU- bzw. BMT-Gruppe hängen eindeutig vom Ergebnis im Vortest ab (vgl. Abschnitt 4.3.6). Im Vortest schwächere Schüler profitieren überdurchschnittlich stark von der Strukturierung der Intervention nach den Vorgaben der BMT, ohne dass stärkere Schüler durch diese Strukturierung benachteiligt würden. Je geringer die Vorkenntnisse der Schüler also sind, desto mehr profitieren diese von der Strukturvorgabe der BMT. Vergleicht man diese Aussage mit der Einschätzung von Helmke (2009, S. 200) über die Personenspezifität der Strukturierung: „Je geringer die Vorkenntnisse der Schüler, desto wichtiger sind gute Strukturen“, könnte man folgern, dass nun die BMT die bessere Strukturvorgabe als das FeU sei. Zumindest für das in dieser Studie gewählte Thema des Impulses scheint diese Aussage auch zulässig zu sein. Bezieht man zusätzlich die Ergebnisse der Untersuchung von Zander et al. (2015) über guten Physikunterricht für schwächere Schülerinnen und Schüler mit ein, so gewinnt obige Aussage zusätzlich an Plausibilität und deutet auch eine Zulässigkeit für weitere Sachinhalte an. Zander et al. (2015, S. 391f.) konnten im Rahmen des Mechanikunterrichts in der achten Jahrgangsstufe an Gymnasien in Nordrhein-Westfalen eine besonders positive Wirkung von basismodellkonformen Unterricht für schwächere Schüler nachweisen. Zander et al. (2015, S. 391f.) konnten dabei ebenfalls keinen negativen Einfluss auf den Lernprozess der als stark eingestuften Schüler feststellen. Diese Ergebnisse von Zander et al. (2015) stehen damit in sehr großer Übereinstimmung zu den entsprechenden Befunden meiner Studie und untermauern damit die Einschätzung der Vorgaben der BMT als nachweislich gute Strukturvorgabe zur unterrichtlichen Thematisierung mehrerer Inhalte aus der Physik.

Bei der Betrachtung der für diesen Teilgruppeneffekt ursächlichen Unterschiede zwischen den Interventionen kann weitgehend eine ähnliche Einschätzung wie im vorangegangenen Abschnitt 5.1.2 getroffen werden. Mit der zusätzlichen Berücksichtigung dass die Wirkung der Unterschiede in der unterrichtlichen Gestaltung besonders deutlich bei den im Vortest schwächeren Schülern werden, können auch die Ursachen für die unterschiedliche Lernwirksamkeit noch genauer eingeschätzt werden.

Die im FeU verankerte Abstraktion kann kognitiv sehr fordernd sein. Fordernd zum einen, weil hierfür ein gewisses Maß an Abstraktionsvermögen vorhanden sein sollte, viel wichtiger in diesem Kontext aber ist der fordernde Charakter bezüglich des für die Abstraktion benötigten Vorwissens. Im Falle des Impulses werden verschiedene physikalische Definitionen (u.a. der Kraft) und Konzepte (z.B. Wechselwirkungsprinzip) benötigt, um neue Sachinhalte aus den experimentellen Ergebnisse abzuleiten. Schüler ohne entsprechendes Vorwissen müssen an dieser Stelle im Unterricht ein größeres Maß an neuer Information

5. Interpretation und Einordnung der Ergebnisse

aufnehmen und deren Anwendung nachvollziehen, als dies bei Schülern eben mit den entsprechenden Vorkenntnissen der Fall ist. Die Aufnahmefähigkeit für die abstrahierten, neuen Inhalte könnte demnach bei den Schülern mit geringerem Vorwissen eingeschränkt sein und eine geringere Lernwirksamkeit für diese Schüler in der FeU-Gruppe zumindest teilweise erklären. Zum anderen Teil könnte diese verringerte Wirksamkeit des FeU auch durch das Fehlen einer expliziten Übungsphase bedingt sein. Übungsphasen dürften für alle Schülergruppen von Bedeutung sein, dennoch erscheint es mir gerade für Schüler mit weniger Vorwissen entscheidend zu sein, während der Übung nicht vordergründig bereits andere Ziele zu verfolgen. Wird aber nun wie im FeU explizit gefordert, angewendet um Verbindungen zu suchen, erfordert dies einerseits bereits die Anwendung der Inhalte aber gleichzeitig auch die Suche nach Bezügen in Sport oder anderen Fächern. Diese fächerübergreifende Vernetzung kann dabei auch Aspekte positiv beeinflussen, welche nicht in dieser Studie erhoben wurden. Ein positiver Einfluss auf Teilaspekte des Interesses oder der Motivation ist durchaus anzunehmen. Für die Fähigkeit Aufgaben korrekt mit den neuen Inhalten zu beantworten (entspricht der Messung im WTI) vermute ich diesen Zusammenhang aber eher nicht. Mit der Suche nach Querverbindungen zu den neuen Inhalten, ohne dass sich durch schlichtes Einüben der Sachinhalte der Umgang mit diesen automatisiert hat, werden zwei Lernprozesse gleichzeitig angestrebt. Noch bevor der Lernprozess der Anwendung und Sicherung des Wissens abgeschlossen und ein klares Bild des neuen Sachinhaltes abgespeichert ist, beginnt bereits dessen Vernetzung zu anderen Gedächtnisinhalten. Die Verankerung des neuen Wissens auf einer noch unsicheren Basis kann so die Fähigkeit im Umgang mit den Inhalten der Schüler im FeU negativ beeinträchtigen. Besonders die Lernprozesse von Schüler mit einer weniger gut gesicherten Vorwissensbasis könnten dadurch gehemmt werden. Ähnliche Überlegungen könnten auch über die Wirkung der innerfachlichen Vernetzungsphase in der BMT-Interventionsstunde angeführt werden. Die innerfachliche Vernetzung in der BMT-Gruppe kann gerade bei Schülern mit einer weniger stark ausdifferenzierten Wissensbasis von Vorteil sein. Indem Zusammenhänge zwischen bestehenden physikalischen Konzepten und den neu erlernten hergestellt werden, werden gleichzeitig auch die verschiedenen Konzepte gegeneinander abgegrenzt. Dies erleichtert die Anwendung der alten aber auch der neuen Konzepte und kann durchaus eine bessere Beantwortung im WTI bewirken.

Insgesamt zeigt sich, dass Schüler mit geringerem Vorwissen bei verschiedenen Themen im Physikunterricht von einer basismodellkonformen Strukturierung von Unterricht besonders stark profitieren. Dies könnte durch die in der BMT ebenfalls besonders große Flexibilität bei der Erarbeitung der Sachinhalte bedingt werden, oder aus dem Vorkommen einer expliziten Übungsphase und der zusätzlichen innerfachlichen Vernetzung resultieren. Speziell das Fehlen einer oft stark vorwissensbasierten Herleitung oder Abstraktion könnte das Lernen

gerade derjenigen Schüler positiv beeinflussen, welche über keine elaborierte Vorwissensbasis verfügen, da dies neben der oben angedeuteten Vorteile zudem zusätzliche zeitliche Spielräume in der unterrichtlichen Gestaltung eröffnet. Diese Spielräume könnten einerseits natürlich für eine Herleitung bzw. Abstraktion der neuen Inhalte aber andererseits eben auch für eine explizite Anwendung und eine innerfachliche Vernetzung verwendet werden. Gerade Schüler mit geringem Vorwissen scheinen aber deutlich mehr von der zweiten Alternative zu profitieren. Von Nachteil ist diese Variante der Gestaltung dabei für keine der betrachteten Schülergruppen.

5.1.4. Schülerwahrnehmung der Interventionsstunden

F2: Führen unterschiedliche Sequenzierungen zu einem unterschiedlichen Maß der kognitiven Aktivierung oder der grundlegenden Bedürfnisse?

F5: Führen unterschiedliche Sequenzierungen auch zu einem unterschiedlichen Maß der von den Schülern wahrgenommenen Strukturierung?

Die Ergebnisse zu den Forschungsfragen F2 und F5 werden in diesem Abschnitt zusammengefasst interpretiert. Der Abschnitt beginnt mit der Interpretation der Bewertung der einzelnen Qualitätsmerkmale durch die Schüler der Gesamtstichprobe. Im weiteren Verlauf des Abschnitts werden auch die Bewertungen von den in der Untersuchung betrachteten Teilgruppen der Gesamtstichprobe interpretiert.

Schülerwahrnehmungen in der Gesamtstichprobe

Grundsätzlich unterscheiden sich die Bewertungen der Schüler innerhalb der FeU-Gruppe nicht systematisch von denjenigen der BMT-Schüler. Die Qualität der beiden Interventionsstunden wird hinsichtlich der erhobenen Merkmale also einheitlich eingeschätzt. Vor dem Hintergrund der tatsächlich relativ geringen Unterschiede zwischen den Ausgestaltungen der Interventionsstunden ist dies vielleicht auch nicht verwunderlich, gerade da sich die wenigen Unterschiede zudem primär in dem weniger leicht zu erfassenden Bereich der Tiefenstruktur des Unterrichts wiederfinden. Etwas verwunderlich aber schon, wenn man berücksichtigt, dass die beiden Interventionsstunden zu einem systematisch und bedeutsam unterschiedlichen Lernzuwachs führen und Clausen (2002, S. 188) der Schülersicht von Qualitätsmerkmalen für die Entwicklung der Schüler die größte Bedeutung einräumt und damit ein Zusammenhang dieser Größen existiert. Zudem bescheinigt Clausen (2002, S. 188) den Schülern die Fähigkeit zu einer differenzierten Wahrnehmung dieser Merkmale. Diese Wahrnehmung ist dabei allerdings oft durch eine globale Beurteilung des Unterrichtenden flankiert, welche „die Einschätzung der eher effektivitätsbezogenen Unterrichtsmerkmale überlagert“ (Clausen, 2002, S. 188). Mögliche systematische Unterschiede in den Bewertungen können so auch im (statistisch) Verborgenen bleiben. Die tatsächlich vorhandenen Unterschiede zwi-

5. Interpretation und Einordnung der Ergebnisse

schen den Interventionsstunden führen bei der Gesamtstichprobe zu keiner Unterschiedlichkeit in der Bewertung der Stunden. Dies kann daran liegen, dass diese hinsichtlich der erfassten Qualitätsmerkmale wirklich gleich stark einzuschätzen sind, oder dass die tatsächlichen Unterschiede nicht ausgeprägt oder prägnant genug sind, um von den Schülern erkannt zu werden. Zusätzlich erfolgte die Messung dieser Qualitätsmerkmale teilweise nicht reliabel genug. Stark fehlerbehaftet war dabei die Erfassung der sozialen Eingebundenheit und des Autonomieerlebens. Die Reliabilitäten der Skalen zu Erhebung der restlichen Merkmale liegen zumindest über einer minimal notwendigen Schwelle, erreichen aber keine als gut zu bezeichnenden Kennwerte (vgl. Abschnitt 3.5.6). Diese Ungenauigkeit in den Messungen könnte zusätzlich das Hervortreten etwaiger Unterschiede in den Bewertungen zwischen den Interventionsgruppen erschweren.

Schülerwahrnehmung in den Teilgruppen nach Vortestergebnis

Betrachtet man die Teilgruppen G1, G2₁, G2₂ und G3 treten insbesondere in der Teilgruppe G1 hinsichtlich der Lernwirksamkeit der Interventionen Unterschiede auf (vgl. Abschnitt 4.3.6). Diese Unterschiede zeigen sich auch in einer zumindest teilweise systematisch verschiedenen Bewertung durch FeU bzw. BMT-Schüler innerhalb der Teilgruppe G1 (vgl. Tabelle 4.14). Im Gegensatz zu den FeU-Schülern dieser Teilgruppe bewerten die BMT-Schüler die Intervention nicht grundsätzlich niedriger als andere Vortest-Teilgruppen, sondern tendenziell höher. Gerade für Schüler mit einem geringeren Vortestergebnis scheinen sich die Interventionen hinsichtlich der Qualitätsmerkmale also echt zu unterscheiden. Diese Teilgruppe der Schüler nimmt offenbar Defizite in der Ausgestaltung der Erarbeitung und Verarbeitung der Sachinhalte entsprechend der Vorgaben des FeU, aber nicht hinsichtlich derer der BMT wahr. Schüler mit geringeren Vorkenntnissen sind demnach nicht nur besonders abhängig von Unterricht mit guten Strukturen (Helmke, 2009, S. 200), sondern scheinen diesbezüglich auch besonders sensitiv zu sein. Sie nehmen strukturelle Unterschiede anscheinend eher wahr und ihre Einschätzungen werden nicht mehr überlagert von einem Globalurteil der Lehrkraft. Dieses Globalurteil ist dabei sicherlich weiter existent, verdeckt aber den nun für diese Teilgruppe größeren Effekt der Unterschiedlichkeit der Stunden nicht. Diese vom Vortestergebnis unabhängige Bewertung der Qualitätsmerkmale in der BMT-Gruppe erklärt dabei auch die empirisch gefundenen interventionsspezifischen Unterschiede in den Zusammenhängen der einzelnen Qualitätsmerkmale mit dem Nachtestergebnis (vgl. Tabelle 4.15). Da das Nachtestergebnis stark mit dem Vortestergebnis korreliert und bei der FeU ein geringes Vortestergebnis wiederum zu einer geringeren Einschätzung der Qualitätsmerkmale führt, hängen auch diese Merkmale in dieser Interventionsgruppe erkennbar mit dem Nachtestergebnis zusammen. Die Schüler der gesamten BMT-Gruppe führen mit Ausnahme des Kompetenzerlebens eine vortestunabhängige Einschätzung der Merkmale durch, diese Merkmale weisen dann auch in der gesamten BMT-Gruppe keine Korrelation

zum Nachtestergebnis auf.

Die systematische Bewertungsunterschiedlichkeit bei der wohl struktur-sensitiveren Gruppe G1 macht dabei die Annahme plausibel, dass die Schüler der Gesamtstichprobe einen tatsächlich vorliegenden, wenngleich wohl nicht besonders stark ausgeprägten aber echten Unterschied hinsichtlich der Qualitätsmerkmale schlichtweg nicht erfassen konnten. Die Existenz eines echten Unterschieds hinsichtlich der Qualitätsmerkmale der Interventionsstunden schätze ich auf dieser Basis insgesamt als eher wahrscheinlich ein.

Schülerwahrnehmung in den geschlechterspezifischen Teilgruppen

Die geschlechterspezifische Unterrichtswahrnehmung ist zwischen den Interventionsgruppen systematisch unterschiedlich. Bewerten die weiblichen Teilnehmer die Unterrichtseinheit nach den Vorgaben des FeU in allen erhobenen Merkmalen außer der sozialen Eingebundenheit niedriger oder gleich hoch, liegen die Mittelwerte der Einschätzungen der weiblichen Teilnehmer in der BMT-Gruppe bis auf die Einschätzung des Kompetenzerlebens stets über dem Referenzwert der männlichen Teilnehmer. Der basismodellkonforme Unterricht wird also von den weiblichen Schülern positiver bewertet als von den männlichen. Besonders deutlich wird dies bei der höchst signifikant höheren Bewertung der Merkmale des Autonomieerlebens (AE), der Schlüssigkeit im Unterrichtsverlauf (SU) und der Orientierung am Ziel (OZ). Zudem sind die Unterschiede in der Bewertung des Kompetenzerlebens in der BMT-Gruppe weniger stark ausgeprägt als in der FeU-Gruppe. Nun wird die Intervention nach den Vorgaben der BMT ja grundsätzlich von den im Vortest schwächeren Schülern positiver bewertet. In dieser Schülergruppe befinden sich überproportional viele Mädchen, welche ja im arithmetischen Mittel einen systematisch niedrigeren Kennwert im Vortest des WTI aufweisen. So befinden sich unter den 158 Schülern in der Gruppe G1 insgesamt 116 Mädchen, welche sich zu fast gleichen Teilen auf die BMT-G1-Gruppe (N= 60) und die FeU-G1-Gruppe (N= 56) aufteilen. Die Schnittmenge der Gruppe G1 mit der Gruppe der weiblichen Teilnehmer ist also relativ groß und es wäre möglich, dass hier zwei mal der gleiche Effekt betrachtet wird. Allerdings bewerten innerhalb der Gruppe 1 sowohl die Teilgruppe der weiblichen als auch die der männlichen Lernenden jeweils die BMT-Intervention in der Tendenz positiver bezüglich der erhobenen Qualitätsmerkmale. Dieser Effekt hängt in der Gruppe G1 damit nicht vom Geschlecht der Teilnehmer ab. Der Effekt der positiveren Einschätzung der Merkmale für die BMT-Intervention durch die weiblichen Studienteilnehmer zeigt sich unabhängig vom Vortestergebnis für alle Teilgruppen G1 bis G3, welche auf Basis des Vortestergebnisses gebildet wurden. Dieser Effekt ist damit zumindest grundsätzlich vom Vortestergebnis unabhängig. Aus diesem Grund wird die angegebene Unterschiedlichkeit in der geschlechterspezifischen Unterrichtswahrnehmung zwischen den Interventionsstunden als eigenständiger Effekt eingeschätzt.

Insgesamt führen in der Gesamtstichprobe unterschiedliche Sequenzierungen bei dieser Untersuchung zu keinem unterschiedlichen Maß der kognitiven Aktivierung, der grundlegenden Bedürfnisse oder der von den Schülern wahrgenommenen Strukturierung. Dabei kann nicht ausgeschlossen werden, dass echte, aber kleine Unterschiede vorhanden sind, diese aber von einem Globalurteil über die Lehrkraft überlagert sind oder aufgrund einer etwas fehlerbehafteten Messung nicht statistisch erfasst wurden. Als sicher ist hingegen die Unterschiedlichkeit in den Bewertungen dieser Merkmale bei einzelnen Teilgruppen einzuschätzen. Sowohl die Gruppe der im Vortest schwächeren Schüler, als auch die weiblichen Studienteilnehmer bewerten die Qualitätsmerkmale der Intervention nach den Vorgaben der BMT hinsichtlich einiger Merkmale signifikant höher. Die unterschiedlichen Sequenzierungen führen demnach in Teilgruppen zu systematisch unterschiedlichen Maßen obiger Merkmale.

5.1.5. Abschließender Vergleich der Strukturvorgaben

F6: Zeigt sich eine zu favorisierende Sequenzierung der Lernprozesse hinsichtlich der untersuchten Merkmale?

Die Existenz einer idealen Strukturvorgabe ist strittig, könnte allein mit dieser Untersuchung auch nicht geklärt werden und ist vielleicht sogar als unwahrscheinlich einzuschätzen. Unwahrscheinlich deshalb, weil eine große Bandbreite an Anforderungen an diese eine Strukturvorgabe zu stellen wären. Diese soll möglichst alle Qualitätsmerkmale von Unterricht erreichen, alle Lernenden gleichermaßen ansprechen und auch verschiedenste mit dem Lernen verbundene Zielsetzungen abdecken. Vielleicht liegt hier gerade der entscheidende Vorteil der BMT, ihre Flexibilität eben Unterricht entsprechend verschiedener Lern- oder Lehrziele angepasst zu gestalten. Diese Flexibilität erlaubt anscheinend eine engere Passung zu den tatsächlichen Lernschritten, welche bei den Lernenden durchlaufen werden sollten. Dies zeigt sich zumindest in dieser Untersuchung in der systematisch höheren Lernwirksamkeit insgesamt, insbesondere bei der Teilgruppe an Schülern, welche noch kein besonders fundiertes Vorwissen aufweisen. Mit erkennbaren Nachteilen bei der Betrachtung weiterer Merkmale ist die BMT dabei nicht verbunden.

Eine andere Perspektive neben der Lernwirksamkeit eröffnet sich bei Betrachtung der Varianzen der Schülerergebnisse im Nachtest des WTI. Es ist eine alternative Perspektive der Begegnung von Leistungsheterogenität. Aus Perspektive der Lernwirksamkeit ist Leistungsheterogenität weder besonders förderlich noch hemmend für zukünftigen Lernerfolg (Hattie, 2009; Gröhllich, Scharenberg & Bos, 2009). Die Perspektive ist aber eine andere, will man „Physik als einen Modus der Weltbegegnung“ (Rincke, 2015, S. 17) vermitteln. „Physik ist Teil unserer kulturellen Identität“ (Rincke, 2015, S. 17). Vor diesem Hintergrund ist es relevant, ob sich Schüler im Physikunterricht abgehängt fühlen,

5.1. Wirkungen der Interventionen hinsichtlich der Forschungsfragen

vom Fach abwenden und sich vielleicht auf Basis dessen auch der physikalischen Weltbegegnung verschließen. Für diese Teilgruppe kann dieser Teil des Bildungsauftrages des Faches Physik dann nur schwerlich erreicht werden. Verringert sich allerdings die Leistungsheterogenität innerhalb einer Klasse, ist die Annahme durchaus berechtigt, dass sich auch weniger Schüler abgehängt fühlen und so offen gegenüber der Physik als Modus der Weltbegegnung verbleiben. Der Leistungsheterogenität wirkt dabei die BMT im besonderen Maße entgegen. Dies zeigt sich zum einen in einem im Vergleich zum FeU geringeren Einfluss des Vorwissens auf den Lernerfolg der Schüler und bestätigt dabei gleichgeartete Ergebnisse der Studie von Zander et al. (2013, S. 504), bei welcher das Vorwissen in der Gruppe mit basismodellkonformen Unterricht ebenfalls weniger stark vom Vorwissen der Schüler als in einer Vergleichsgruppe abhängt. Die BMT scheint dabei zu helfen, eine Überforderung der Schüler zu vermeiden. Ein weiteres Indiz für die heterogenitätsmindernde Wirkung der BMT findet sich in den Varianzen der Nachtestergebnisse im WTI der Interventionsgruppen. Die Varianz in der BMT-Gruppe ist systematisch geringer als in der FeU-Gruppe (vgl. Abbildung 4.2). Die Varianz ist geringer, weil in der BMT-Gruppe merklich seltener WTI-Ergebnisse unter 50% der maximalen Punktzahl erzielt werden. Die BMT wirkt also der Leistungsheterogenität entgegen. Dieser Befund findet sich in einer zweiten Veröffentlichung zur Studie von Zander et al. (2015, S. 392) wieder.

Die Beantwortung der Forschungsfrage F6 kann demnach eindeutig formuliert werden. Es zeigt sich eine zu favorisierende Sequenzierung und es ist die BMT. Diese Aussage kann dabei nur im Rahmen dieser Studie formuliert werden, eine Übertragung auf andere Situationen ist grundsätzlich nicht zulässig, wenngleich eine Berücksichtigung der Befunde weiterer Studien aus der Fachdidaktik dies durchaus plausibel erscheinen lässt.

Nimmt man zusätzlich noch die Ergebnisse aus Abschnitt 2.3.4 in Betracht, so zeigt sich ein einheitliches Bild. Die Vergleiche der Wirkungen der BMT mit anderen Unterrichtsformen zeigt jeweils einen Vorteil zu Gunsten der BMT. Dieser Vorteil zeigt sich unabhängig von Schulfach oder Schulart in den angegebenen Studien. Berichte über nachteilige Wirkungen der BMT wurden im Rahmen meiner Literaturrecherche nicht gefunden. Die BMT deshalb als ideale Strukturvorgabe auszuwählen ist dennoch voreilig. Dazu bedürfte es einer Vielzahl weiterer Untersuchungen mit entsprechendem Ergebnis, welche die BMT mit »fairen« Kandidaten vergleichen würden. Nur so erscheint eine Einordnung der kognitiven Wirkungen der BMT und eine detailliertere Abschätzung der affektiven Wirkungen möglich.

5.2. Interventionsgruppenunabhängige Studienergebnisse

Zur Beantwortung der Forschungsfragen standen bei dieser Untersuchung interventionsabhängige Erkenntnisse im besonderen Fokus. Neben diesen traten aber auch weitere aus meiner Sicht berichtenswerte Erkenntnisse zu Tage. Dies sind die interventionsgruppenunabhängigen Ergebnisse bezüglich der geschlechterspezifischen Unterschiedlichkeit im generellen Lernzuwachs, des Einflusses der erhobenen Kontrollvariablen, sowie bezüglich der Ausprägung verschiedener Qualitätsmerkmale und deren Zusammenhang mit dem Nachtestergebnis im WTI.

5.2.1. Geschlechterspezifische Unterschiedlichkeit des generellen Lernzuwachses

Aus den in der Studie erhobenen Daten ergibt sich eine systematische geschlechterspezifische Unterschiedlichkeit des generellen Lernzuwachses. Die Ergebnisse im Nachtest des WTI zeigen einen recht deutlichen Einfluss des Geschlechts. Die weiblichen Schüler weisen höchst signifikant und mit einer mittleren Effektstärke ($d = -0.68$) gekennzeichnet, negativere Ergebnisse im physikalischen Fachwissen zum Impuls auf. Fächer- und Länderübergreifend allerdings ist kein so deutlicher Unterschied erwartbar. Hattie (2009, S. 55ff.) gibt einen wenig bedeutsamen Geschlechtereffekt von $d = 0.12$ zu Gunsten der männlichen Schüler an. Rückt man das Betrachtungsfeld wieder in einen engeren Kontext zu meiner Studie, so zeigen sich hingegen mit meinen Ergebnissen übereinstimmende Literaturangaben. So wird in einer Zusammenfassung zur Pisa Studie von 2006 für fünfzehnjährige Schüler in Deutschland ebenfalls von signifikant höheren Kompetenzwerten der Jungen im Bereich der »physikalischen Systeme« berichtet (Prenzel et al., 2007, S. 94f.), weshalb von einem nicht generellen aber physiktypischen und nicht studienbedingten Geschlechtereffekt ausgegangen wird.

Der deutlichste und bedeutsamste Unterschied zwischen den männlichen und weiblichen Teilnehmern außerhalb der Leistungen in den Wissenstests zeigt sich im Kompetenzerleben. Der Geschlechtereffekt beim Kompetenzerleben zu Gunsten der männlichen Schüler ist höchst signifikant bei einer kleinen bis mittleren Effektstärke. Dieser Effekt ist dabei besonders relevant, da er die höchste Korrelation aller erhobener Qualitätsmerkmale mit dem Nachtestergebnis im WTI aufweist. Die Mädchen weisen also im Schnitt geringere Kompetenzen in den Tests auf und bewerten ihr Kompetenzerleben auch entsprechend niedriger. Ob das geringere Kompetenzerleben während der Intervention (Hinweis: Das Kompetenzerleben wurde zeitlich vor dem Nachtest-WTI erhoben) verantwortlich ist für das schlechtere Abschneiden im WTI oder den etwas geringeren Lernzuwachs, ist möglich, kann aber aufgrund der vorliegenden Daten nicht eindeutig geklärt werden. Ebenso wenig kann ausgeschlossen wer-

den, dass aufgrund der Vorerfahrungen im Schulunterricht zum Fach Physik das Kompetenzerleben niedriger eingeschätzt wurde. Die Unterschiede in den physikalischen Kompetenzen bestehen bereits vor der Intervention und werden von dieser in ihrer Unterschiedlichkeit auch nicht wesentlich beeinflusst.

Grundsätzlich sind beide Geschlechter dabei als psychologisch gesehen weitestgehend ähnlich einzuschätzen (The Gender Similarities Hypothesis; Hyde, 2005). Zu diesem Urteil gelangt Hyde (2005) auf Basis eines Review Artikels zu 46 Metaanalysen. Eine Erklärung für den beobachteten Kompetenzunterschied findet sich damit nicht direkt in einer psychologisch oft vermuteten aber empirisch nicht gegebenen Unterschiedlichkeit der Geschlechter (Hyde, 2005, S. 581). Kein Unterschied zwischen den Geschlechtern zeigt sich zudem bei den kognitiven Grundfähigkeiten der Studienteilnehmer (KFTk). Ein tatsächlicher Geschlechterunterschied findet sich aber in der gesellschaftlichen Sichtweise der Passung der Jungen oder Mädchen zur Physik. Es zeigt sich „sowohl auf der Ebene ‚expliziter‘ als auch ‚impliziter‘ Stereotype, dass Mathematik und Physik im Durchschnitt als weniger gut zu Mädchen passend charakterisiert werden“ (Kessels, 2015, S. 22). Gerade diese Passung ist aber ein entscheidender Faktor für die Art und Weise der Auseinandersetzung mit oder der Erfolgserwartung in dem entsprechenden Fach. Beides ist dabei bei den Mädchen negativ beeinflusst, da sie „eine größere Distanz zwischen dem eigenen Selbst und diesen Fächern empfinden“ (Kessels, 2015, S. 22; Anmerkung: Kessels bezieht sich hier auf die Fächer Mathematik und Physik). Diese empfundene Distanz auf Seiten der weiblichen Schüler kann durchaus einen wesentlichen Beitrag zum Geschlechtereffekt in den Leistungen im Fach Physik liefern.

Der empirisch gefundene Geschlechtereffekt in den erhobenen physikalischen Kompetenzen ist insgesamt kein reiner Studieneffekt, sondern stellt einen erwartbaren Leistungsunterschied dar, welcher wohl erheblich von sozialen und gesellschaftlichen Gegebenheiten beeinflusst ist.

5.2.2. Einfluss verschiedener Kontrollvariablen auf das Nachtestergebnis im WTI

Der Einfluss der neben dem Geschlecht erhobenen Kontrollvariablen auf das Nachtestergebnis im WTI ist zwischen den einzelnen Variablen unterschiedlich und wird nachfolgend für jede der Variablen kurz eingeschätzt. Eine Beschreibung der Kontrollvariablen findet sich in Abschnitt 3.5.1.

Zweigezugehörigkeit

Die Zugehörigkeit zum naturwissenschaftlichen Zweig an bayrischen Gymnasien zeigt weder in den Kovarianzanalysen der Gesamtstichprobe noch bei den entsprechenden Analysen von Daten von Teilgruppen der Studie einen systematischen oder bedeutsamen Effekt auf das Nachtestergebnis im WTI. Auch im Vortestergebnis unterscheiden sich die naturwissenschaftlichen Klas-

5. Interpretation und Einordnung der Ergebnisse

sen nicht signifikant von den restlichen Klassen ($p = 0.43$; $t(794) = 0.79$). Die reinen Nachtestergebnisse hingegen erweisen sich in einem t-Test als hoch signifikant unterschiedlich ($p = 0.002$; $t(733) = 3.12$). Dieser Unterschied ist mit einer Effektstärke von $d = 0.22$ verbunden, dabei weisen die Schüler aus naturwissenschaftlichen Klassen (NW-Gruppe) die etwas höheren Testergebnisse auf. Die Zusammensetzung der NW-Gruppe unterscheidet sich dabei wiederum von der Zusammensetzung der nicht naturwissenschaftlichen Klassen (AL-Gruppe; AL steht dabei für alternative Zweigzugehörigkeit) insbesondere hinsichtlich der Verteilung von Jungen und Mädchen. Befinden sich in der NW-Gruppe 237 Jungen und 125 Mädchen, sind es in der AL-Gruppe 134 Jungs und 300 Mädchen. Eine zusätzliche geschlechterspezifische Auflistung der Vor- und Nachtestwerte ist in Tabelle 5.1 dargestellt.

Teilgruppe	N	VT	NT
NW - Mädchen	125	10.2	13.5
NW - Jungen	237	12.2	16.0
NW - GES	362	11.5	15.1
AL - Mädchen	300	10.7	13.6
AL - Jungen	134	12.8	15.9
AL - GES	434	11.3	14.3

Tabelle 5.1.: Gruppengrößen und Mittelwerte der WTI-Ergebnisse der Mädchen, Jungen und der Gesamtgruppen der Klassen mit naturwissenschaftlicher oder alternativer Zweigzugehörigkeit. VT: Vortestergebnis im WTI, NT: Nachtestergebnis im WTI

Betrachtet man nun die Kennwerte im Nachtest der NW-Gruppe und der AL-Gruppe spezifisch nach Geschlecht sind keine Unterschiede mehr zu erkennen. Der höhere Kennwert resultiert anscheinend allein aus der Unterschiedlichkeit der Zusammensetzung der Gruppe und ist damit primär auf den Einfluss der Kontrollvariable des Geschlechts zurückzuführen. Bemerkenswert und zumindest kontraintuitiv sind die jeweils schlechteren Testergebnisse im Vortest für Jungen und Mädchen in der NW-Gruppe. Vor allem vor dem Hintergrund, dass diese Schüler seit über zwei Schuljahren eine Stunde Physikunterricht pro Schulwoche zusätzlich erhalten. Statistisch signifikant sind diese Unterschiede allerdings nicht. Zudem handelt es sich beim Vortest um noch nicht im Unterricht thematisierte Sachinhalte, weshalb diese geringeren Vortestwerte in der NW-Gruppe an dieser Stelle auch nicht weitergehend interpretiert werden.

Zusammenfassend erscheint letztlich das Ergebnis der Kovarianzanalysen belastbarer, da hier der Einfluss mehrerer Kontrollvariablen berücksichtigt werden kann und die Gruppenzusammensetzung der NW- und der AL-Gruppe deut-

lich von diesen abhängt. Folglich wird der Einfluss der Zweigzugehörigkeit insgesamt als nicht systematisch und nicht bedeutsam eingeschätzt.

Interventionsstart

Der Startzeitpunkt der Intervention erweist sich oftmals als ein hoch bis höchst signifikanter kleiner Effekt bei den durchgeführten Analysen. Dabei erreichen grundsätzlich Klassen höhere mittlere Ergebnisse im WTI, je früher die Intervention beginnt. Die Intervention inklusive der Testphasen weist insgesamt eine durchschnittliche Dauer von ca. 150 Minuten auf. Der Startzeitpunkt des Nachtests im WTI liegt in einem Zeitbereich von 11 bis 17 Uhr. In diesem Zeitbereich unterliegt die Leistungsbereitschaft der Schüler durchaus Schwankungen. Diese Schwankungen sind in Abbildung 5.1 in der sogenannten REFA Normkurve dargestellt.

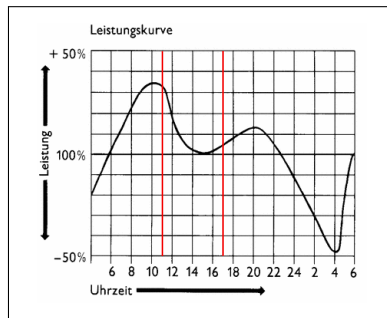


Abbildung 5.1.: Leistungsbereitschaft als prozentuale Abweichung vom Normwert in Abhängigkeit der Tageszeit. Der relevante Zeitraum ist zusätzlich rot gekennzeichnet (Seiwert, 1984, S. 32)

Diese Normkurve weist dabei im relevanten Zeitraum ein tendenzielles Absinken der Leistungsbereitschaft von ca. 135% um 11 Uhr bis zu ca. 105% um 17 Uhr auf. Der Einfluss der Kontrollvariable des Interventionsstarts kann nach meiner Einschätzung auf dieses Absinken der Leistungsbereitschaft im relevanten Zeitraum zurückgeführt werden.

Interventionsdauer

Die Dauer der Intervention erweist sich hingegen in keiner der durchgeführten Analysen als ein systematischer bzw. bedeutsamer Einflussfaktor auf das Nachtestergebnis. Für einen Großteil der Varianz dieser Kontrollvariable ist die Dauer der Experimentierphase verantwortlich. Die Dauer der Experimentierphase war dabei in den Interventionsstunden primär von der benötigten Zeitdauer zum Aufbau der Experimentiermaterialien beeinflusst. Das nun die Aufbauschnelligkeit keinen systematischen Einfluss auf das Nachtestergebnis

5. Interpretation und Einordnung der Ergebnisse

aufweist, erscheint mir durchaus plausibel. Da in den Interventionseinheiten jeweils erst zur nächsten Phase übergegangen wurde, wenn fast alle Schüler die Experimentiertätigkeit abgeschlossen hatten, prägt nun aber die Aufbauschnelligkeit in besonderem Maße die Interventionsdauer. Weshalb deren geringer Einfluss ebenfalls als nicht überraschend eingeschätzt wird.

Klassengröße

Hattie (2009, S. 85f.) kennzeichnet die Klassengröße als systematischen Faktor mit eher geringem unterrichtlichem Einfluss ($d = 0.22$). Kleinere Klassen führen also nicht zu deutlich größeren Lernzuwächsen. Hattie (2009, S. 88) führt dies unter anderem darauf zurück, dass Lehrer ihre Unterrichtsstrategien nicht maßgeblich der Klassengröße anpassen. Auch in meiner Studie erfolgte keine Anpassung der Unterrichtsstrategien an die jeweilige Klassengröße, weshalb ein mittlerer oder großer Einfluss dieser Kontrollvariable in meiner Untersuchung auch nicht erwartbar ist. Ein Einfluss auf das Nachtestergebnis im WTI zeigt sich entweder als nicht systematischer Effekt (Gesamtstichprobe, weibliche Studienteilnehmer, FeU-Gruppe) oder als hoch signifikanter aber kleiner Effekt in den Teilgruppe der männlichen Studienteilnehmer und der BMT-Gruppe. Das ist soweit als erwartungskonform zu bezeichnen. Nicht zu erwarten war dabei allerdings die Richtung des Effektes in letztgenannten Teilgruppen. Größere Klassen weisen im Mittel auch höhere Kennwerte im Nachtest auf. Die Richtung des Effekts ist damit zu dem in der Literatur berichteten genau entgegengesetzt gerichtet. Eine Erklärung dieser Tendenz findet sich nach meiner Einschätzung weder in der Ausgestaltung der BMT-Intervention im Vergleich zur FeU-Intervention noch in etwaigen psychologischen Unterschiedlichkeiten der Geschlechter, welche aus empirischer Sicht ja auch gar nicht vorhanden sind (Hyde, 2005). Letztlich vermute ich die Ursachen für diesen Effekt als in der Zusammensetzung der (Teil-)Stichproben gelagert, ohne diese Ursachen explizit benennen zu können. Sicher aber ist, dass sich die Schüler einzelner Klassen innerhalb der Interventionsgruppen systematisch unterscheiden ($ICC = 0.14$). Unterschiede in der Zusammensetzung können dabei insbesondere bei Teilgruppen mit entsprechend geringeren Stichprobengrößen zu stärkeren Effekten in den durchgeführten Analysen führen. Passend erscheint vor diesem Kontext, dass die Klassengröße weder in der Kovarianzanalyse der Gesamtstichprobe noch in der Gruppenunterschiedlichkeiten berücksichtigenden Mehrebenenanalyse einen systematischen Effekt darstellt. Weshalb insgesamt davon ausgegangen wird, dass die Klassengröße zusammenfassend keinen systematischen Effekt in der Untersuchung darstellt.

Kompetenz zur experimentellen Erkenntnisgewinnung (WTE)

Die Messung der Kompetenz zur experimentellen Erkenntnisgewinnung über den WTE ist leider fehlerbehaftet (Reliabilität $\alpha = 0.52$). Dies muss bei der Interpretation der Ergebnissen zu dieser Variable mit berücksichtigt werden. In den Daten der Untersuchung zeigt sich ein unterschiedlich hoher Testwert

im WTE zwischen den Interventionsgruppen, wobei sich innerhalb beider Interventionsgruppen kein systematischer Einfluss des Ergebnisses im WTE auf das Nachtestergebnis im WTI ergibt. Der in der Gesamtstudie beobachtete marginal signifikante Effekt resultiert damit voraussichtlich aus einer leicht unterschiedlichen Zusammensetzung der Interventionsgruppen, aber nicht aus einer Bedeutung des mit dem WTE erfassten Merkmals auf das Nachtestergebnis.

Insgesamt wird deshalb der Einfluss des Ergebnisses im WTE als nicht systematisch und nicht bedeutsam gewertet. Über einen möglichen Einfluss der tatsächlichen Kompetenz der experimentellen Erkenntnisgewinnung kann keine Aussage formuliert werden, dazu hätte diese Kompetenz reliabler erfasst werden müssen.

Kognitive Grundfähigkeiten (KFTk)

Die Erfassung der kognitiven Grundfähigkeiten der Schüler über die Kurzska KFTk erfolgt hingegen ausreichend reliabel (Reliabilität $\alpha = 0.65$). Die mit der Kurzform gemessenen Ausprägungen der kognitiven Grundfähigkeiten der an der Studie teilnehmenden Schüler stimmen auch weitestgehend mit den im Handbuch des KFT angegebenen Werten überein. So erreichen die Gymnasialschüler der 10. Jahrgangsstufe im KFTk eine vergleichbare mittlere Lösungswahrscheinlichkeit ($Lw = 0.789$) als aus den im Handbuch des KFT angegebenen Daten berechnet ($Lw = 0.795$; ebenfalls für Gymnasialschüler) (Heller & Perleth, 2000, S. 82). Die Ausprägung der Testwerte im KFTk und im KFT sollen „gleichermaßen abstraktlogische und anschauungsgebundene bzw. konstruktive Aspekte intelligenten Verhaltens“ (Heller & Perleth, 2000, S. 45) erfassen. Dass diese Grundfähigkeiten auch einen systematischen und bedeutsamen Einfluss auf das Nachtestergebnis im WTI aufweisen (vgl. Tabelle 4.12) erscheint mir wenig überraschend, dass allerdings der Einfluss des Ergebnisses im KFTk in den Interventionsgruppen unterschiedlich stark ist (FeU: $\omega^2 = 0.018$; BMT: $\omega^2 = 0.052$) hingegen schon. Die stärkere Ausprägung des Einflusses des KFTk-Ergebnisses in der BMT-Gruppe steht dabei im Widerspruch zu den Ergebnissen der Studie von Zander et al. (2013), in welcher ein eben geringerer Einfluss des KFT-Ergebnisses in der Gruppe mit basismodellkonformen Unterricht als der Vergleichsgruppe nachgewiesen werden konnte (Zander et al., 2013, S. 504). Dabei ist allerdings nicht angegeben, ob eine Teilfacette oder der Gesamttest des KFT bei Zander et al. (2013) eingesetzt wurde.

Die Unterschiedlichkeit in der Ausprägung des Einflusses der kognitiven Grundfähigkeiten zwischen den Interventionsgruppen ist damit schwer einschätzbar, da aus meiner Sicht auch keine Gestaltungsmerkmale der Interventionen zu einer belastbaren Einschätzung beitragen. Der grundsätzliche positive Einfluss der kognitiven Grundfähigkeiten auf das Nachtestergebnis hingegen erscheint schlicht plausibel.

5.2.3. Ausprägung der allgemeinen Qualitätsmerkmale und deren Zusammenhang mit dem Nachtestergebnis im WTI

Ausprägung der allgemeinen Qualitätsmerkmale

Beide Sequenzierungen werden grundsätzlich hinsichtlich der erhobenen Qualitätsmerkmale überdurchschnittlich hoch bewertet. Beide Sequenzierungen können zudem als theoriegeleiteter Unterricht kategorisiert werden. In dieser Studie werden damit Formen theoriegeleiteten Unterrichts überdurchschnittlich positiv von den Schülern bewertet und damit vermeintlich besser als der von den Schülern in der Unterrichtspraxis erlebte. In ihrem Blick auf den Physikunterricht in der IPN-Videostudie stellen Seidel et al. (2006, S. 804) dazu passend fest, dass hinsichtlich vieler Qualitätsmerkmale eine durchaus mögliche theoretische Orientierung selten in der Unterrichtspraxis Einzug findet. Die vorherrschende Unterrichtspraxis scheint damit eben nicht theoretisch fundiert zu sein. Die überdurchschnittlich positive Bewertung von theoriegeleitetem Unterricht hinsichtlich der erhobenen Qualitätsmerkmalen in meiner Untersuchung kann als Indiz dafür gewertet werden, dass Schüler auch wirklich von nach theoretischen Vorgaben gestalteten Unterricht profitieren würden. Die Einschätzung der Merkmale durch die Schüler ist ja für deren kognitive Entwicklung besonders bedeutsam (Clausen, 2002, S. 188).

Unterricht speziell gestaltet nach den Vorgaben der in dieser Studie untersuchten Strukturvorgaben könnte zudem einem weiteren von Seidel et al. (2006) identifizierten negativen Einflussfaktor eliminieren. Dieser negative Einfluss entsteht nach Seidel et al. (2006, S. 812) dadurch, „dass Lernprozesse in Experimentalphasen oft nur unzureichend und wenig systematisch unterstützt wurden“. Dabei konnten Seidel et al. (2006) auf Basis der ihnen vorliegenden Daten belegen, „dass die Art und Weise des Experimentierens wichtig für die kognitive Lernentwicklung ist“ (Seidel et al., 2006, S. 812). Beide untersuchte Strukturvorgaben geben Hinweise auf Funktion und Stellung der Experimente im Unterricht. Damit liefern beide Vorgaben auch Richtlinien die sowohl eine systematischen Einbettung der Experimente in die Lernprozesse der Schüler, als auch eine Festlegung der Art und Weise der Durchführung dieser Experimente ermöglichen. Die Schüler nehmen dies auch so wahr. Sie finden, dass die Experimente an sinnvollen Stellen im Unterrichtsverlauf durchgeführt wurden (Fragebogenitem 19; $M=3.75$; Skala von 1 bis 4), zudem konnten sie die Bedeutung der Lerninhalte durch die Experimente besser erfassen (Fragebogenitem 20; $M=3.61$). Vor diesem Hintergrund kann eine systematische Unterstützung der Lernprozesse der Schüler durch die sinnvolle Einbettung und zielführende Realisierung der Experimente in den Interventionseinheiten angenommen werden.

Die aus Schülersicht insgesamt überdurchschnittliche hohe Ausprägung der Qualitätsmerkmale und zusätzlich die als gut eingeschätzte Implementierung der Experimente in den Interventionsstunden kennzeichnen beide Strukturvor-

gaben generell als empfehlenswert für den unterrichtlichen Einsatz. Mit deren Einsatz bei der Gestaltung von Unterricht könnte dabei zusätzlich den oben genannten und von Seidel aufgedeckten Schwachstellen in der Unterrichtspraxis entgegen gewirkt werden.

Zusammenhänge der allgemeinen Qualitätsmerkmale mit dem Nachtestergebnis

Unabhängig von der jeweils betrachteten Teilgruppe (FeU, BMT, GES) ergibt sich nur ein einheitlicher statistisch signifikanter und bedeutsamer Zusammenhang zwischen den erhobenen Qualitätsmerkmalen und dem Nachtestergebnis im WTI (vgl. Tabelle 4.15). Dieser positive Zusammenhang des Kompetenzerlebens mit dem WTI-Nachtestergebnis ist dabei als mittel stark ausgeprägt zu kategorisieren ($r = 0.31$) und erweist sich als höchst signifikante Korrelation. In der Gesamtstichprobe (und der FeU-Gruppe) zeigen sich auch Korrelationen der kognitiven Aktivierung (KOG) und Elementen der kognitionspsychologisch verstandenen (OL) aber auch der didaktischen Strukturierung (SU) mit dem WTI-Nachtestergebnis als höchst bzw. hoch signifikant, erweisen sich dabei aber als weniger stark bedeutsam. ($0.10 < r < 0.15$).

Entsprechende Elemente der Strukturierung weisen auch in der Studie im Rahmen der Fachdidaktik Mathematik von Rakoczy et al. (2007) einen positiven aber etwas höheren Zusammenhang (hier $r = 0.26$) mit den Schülerleistungen auf. Dabei muss allerdings berücksichtigt werden, dass bei Rakoczy et al. (2007) diese Elemente nicht durch eine Schülereinschätzung erhoben, sondern auf Basis von Videoanalysen aus einer externen Sichtweise eingeschätzt wurden. Die Einschätzungen von unterrichtlichen Qualitätsmerkmalen aus verschiedenen Perspektiven (Schüler, Lehrer, Externe) ist dabei nur von niedriger Übereinstimmung gekennzeichnet. Diese niedrige Übereinstimmung hängt allerdings von dem zu beurteilenden Merkmal ab (Clausen, 2002, S. 186), so dass auch ein gewisser gemeinsamer Kern zwischen den Einschätzungen besteht. Grundsätzlich stimmen damit die Aussagen meiner Studie mit den Angaben in der fachdidaktischen Literatur zur Korrelation struktureller Merkmale mit der Schülerleistung überein. Die Abweichung in der Stärke des Zusammenhangs ist dabei vermutlich durch die auffallend positive Beurteilung dieser Merkmale von im Vortest (und damit auch meist im Nachtest) schwächeren Schülern in der BMT-Gruppe beeinflusst.

Grundsätzlich weist die positive Korrelation der Schülereinschätzung der Unterrichtsstruktur mit dem Testergebnis im WTI damit literaturkonforme und erwartbare Kennwerte auf.

Diese Aussage kann zudem für die Ergebnisse zum Zusammenhang der kognitiven Aktivierung übernommen werden. Dabei besteht zwischen der Art der Erhebung dieses Merkmals kein Unterschied zwischen der Studie von Rakoczy et al. (2007) und meiner Studie. In beiden Fällen erfolgt eine Schülereinschätzung dieses Merkmals und in beiden Fällen zeigt sich ein systematischer, kleiner Zusammenhang mit den Schülerleistungen (Rakoczy et al., 2007, S. 113: $r = 0.13$;

5. Interpretation und Einordnung der Ergebnisse

meine Studie: $r=0.15$). Bemerkenswert bleibt allerdings die Abweichung des entsprechenden Kennwerts in der BMT-Gruppe ($r=0.07$). Anscheinend konnten in dieser Interventionsgruppe auch die schwächeren Schüler kognitiv aktiviert werden.

Auch die grundlegenden Bedürfnisse (soziale Eingebundenheit; Kompetenzerleben; Autonomieerleben) wurden in der Studie von Rakoczy et al. (2007) durch die Schüler eingeschätzt. Rakoczy et al. (2007, S. 113) führen diese drei Merkmale dabei gemeinsam unter dem Begriff *emotional experience* auf. Die Daten ihrer Studie deuten auf keinen Zusammenhang dieses Sammelmerkmals mit der Schülerleistung hin. Die Daten meiner Studie hingegen lassen auf eine höchst signifikante und kleine Korrelation ($r=0.14$) zwischen der Gesamtskala der grundlegenden Bedürfnisse und dem Nachtestergebnis im WTI schließen. Dabei ist diese Korrelation auf der Gesamtskala nur durch die mittel stark ausgeprägte Korrelation des Kompetenzerlebens mit dem WTI bedingt. Beide verbleibenden Teilmerkmale (soziale Eingebundenheit, Autonomieerleben) korrelieren nicht systematisch mit dem Nachtestergebnis. Ein detaillierter Vergleich der Einzelmerkmale zwischen den Studien ist dabei nicht möglich, da in der Publikation von Rakoczy et al. (2007) die Skala der *emotional experience* hinsichtlich der Einzelmerkmale nicht differenziert dargestellt wird. Dennoch kann angenommen werden, dass der Unterschied zwischen den Korrelationen der Merkmale in den beiden Studien beim Merkmal des Kompetenzerlebens liegt. Eine mögliche Ursache dieser Unterschiedlichkeit könnte in den verschiedenen Designs der Interventionen liegen. Sowohl in der FeU- als auch der BMT-Intervention bilden Phasen einen wesentlichen Teil des Unterrichts, welche eine direkte Begegnung der Schüler mit den Inhalten vorsehen. In diesen Phasen können sich die Schüler auch als entsprechend kompetent erleben und folglich einschätzen. Die der Studie von Rakoczy et al. (2007) zugrunde liegenden Unterrichtseinheiten sollten die Einführung des Satzes des Pythagoras leisten. Weitere Richtlinien für die Gestaltung des Unterrichts wurden nicht vorgegeben (Rakoczy et al., 2007, S. 106). Die so gestalteten Mathematikstunden können einen ebenso großen Stellenwert auf Phasen der direkten Begegnung legen, müssen dies aber nicht. Nachdem in den Mathematikstunden dieser Fokus nicht vorgegeben ist, in den Interventionsstunden bei meiner Studie sich aus den theoretischen Vorgaben aber so ergibt, vermute ich diesbezüglich Unterschiede. Ein Mangel an Möglichkeiten zum als kompetent Erleben könnte demnach den fehlenden Zusammenhang des Kompetenzerlebens mit der Schülerleistung in der Studie von Rakoczy et al. (2007) ebenso erklären wie das Vorkommen dieser Möglichkeiten den positiven Zusammenhang in meiner Studie.

Deci und Ryan (1993, S. 233f.) erwarten insgesamt positive Einflüsse aller Faktoren der grundlegenden Bedürfnisse auf die Lernqualität, heben dabei die Autonomieunterstützung als besonders lernrelevant hervor. Empirisch werden aber auch in weiteren Interventionsstudien keine oder nur uneindeutige Zusam-

menhänge der Merkmale der grundlegenden Bedürfnisse auf den Lernerfolg gefunden (Berger, 2007; Göhring, 2010).

Die in meiner Studie gefundene, eindeutig positive Korrelation des Kompetenzerlebens mit dem Lernerfolg differiert damit von den Ergebnissen anderer Interventionsstudien. Der theoretisch von Deci und Ryan erwartete, insgesamt positive Zusammenhang aller drei Merkmale der grundlegenden Bedürfnisse zeigt sich nicht in den Daten zur Studie.

5. *Interpretation und Einordnung der Ergebnisse*

6. Die Studie in der kritischen Rückschau

Bei der nachfolgenden, rückblickenden Betrachtung der Konzeption der Untersuchung sollen nicht alle Einzelaspekte der Studie vor dem Hintergrund neuer Erfahrungen und Erkenntnissen neu beleuchtet werden. Es erfolgt eine Rückschau auf ausgewählte konzeptionelle Entscheidungen. Diese Rückschau ist dabei in drei Teile gegliedert. Sie beginnt mit der Auseinandersetzung mit dem grundsätzlichen Design der Studie, fährt fort mit einer Beurteilung der Gestaltung der Unterrichtskonzepte und endet mit einer Analyse einzelner Untersuchungsmethoden.

Grundsätzliches Design der Studie

Die explorative Grundausrichtung der Studie erscheint nach wie vor gerechtfertigt, da sich auch im Nachhinein die Forschungssituation nicht derart gestaltet, dass die Formulierung von Hypothesen bezüglich der Kernfragen der Untersuchung auf Basis der vorhandenen Erkenntnisse möglich gewesen wäre. Einzelne Studienergebnisse konnten zwar auf Grundlage der Befunde anderer Untersuchungen eingeordnet oder interpretiert werden, es ergaben sich durchaus auch literaturkonforme Ergebnisse, eine Vorhersage der wesentlichen Ergebnisse (vor allem zu den abhängigen Variablen) meiner Studie ermöglichten die Erkenntnisse anderer Studien allerdings nicht.

Ein wesentlicher Unterschied zu anderen Studien in diesem Forschungsfeld besteht auch in der Ausrichtung meiner Studie als Laboruntersuchung. Dies schränkt zunächst einmal die Generalisierbarkeit der mit meiner Arbeit gewonnenen neuen Erkenntnisse ein. Gelten diese doch nur für einen ausgewählten Sachinhalt und sind zudem nicht direkt übertragbar auf die schulische Realsituation. Im Hinblick auf andere Studien im relevanten Forschungsgebiet stellt meine Studie mit dieser Ausrichtung aber eine sinnvolle Ergänzung dar, da sie grundsätzlich andere Stärken und Schwächen aufweist als sonstige Studien in diesem Kontext. Können beispielsweise Interventionsstudien mit eigens fortgebildeten Lehrkräften (vgl. Wackermann 2008 oder Zander 2013 und 2015) zwar Aussagen zur Wirksamkeit beispielsweise der BMT über eine Vielzahl von Sachinhalten und auch für die schulische Realsituation generieren, verbleibt unklar, ob diese Wirksamkeit aus Trainingseffekten in der Fortbildung (Grundlegende Reflektionen zum Unterricht und Expertenunterstützung bei der Stundenplanung), einer Positivauswahl der teilnehmenden Lehrkräfte oder tatsächlich der Intervention stammen. Diese Aspekte können in Feldstudien nur sehr schwer kontrolliert werden. In Laborstudien können hingegen noch zu-

sätzliche Aspekte berücksichtigt und so etwaige empirische Befunde ggf. auch eindeutig der Intervention zugeordnet werden. Ein wesentliches Kriterium für die Entscheidung zu einer Laborstudie war auch das Ziel einer möglichst theoriekonformen Umsetzung der Strukturvorgaben. Hahn (2014) betrachtet dieses Ziel dabei auch als weitestgehend erreicht und die Ausrichtung als Laborstudie erscheint mir vor diesem Hintergrund auch rückblickend als sinnvoll. Nur mit Hilfe dieser Studienausrichtung wurde ein Blick in die Tiefenstruktur der Interventionen möglich. Aufgrund der oberflächlichen Ähnlichkeit der Interventionsstunden war dieser so geartete Blick auch notwendig.

Gestaltung der Unterrichtskonzepte

Bei der Auswahl der Strukturvorgaben wären auch alternative Entscheidungen möglich gewesen. Die Auswahl einer Strukturvorgabe mit bereits bekannter Lernwirksamkeit (z.B. Inquiry-based Teaching oder Direct Instruction) hätte wohl eine nachträgliche Einordnung der Lernwirksamkeit der zweiten in der Studie untersuchten Strukturvorgabe beispielsweise in Hatties (2009) Lernwirksamkeitsrangfolge erleichtert. Eine entsprechende Auswahl hätte aber eine methodisch und inhaltlich vergleichbare Konzeption der Interventionsstunden (vgl. Abschnitt 3.2) erschwert und damit auch eine Beschreibung möglicher Ursachen für empirische Unterschiede. Die Auswahl des FeU und der BMT gerade in Kombination mit der Studienausrichtung als Laborstudie und dem damit verbundenen möglichen Detailblick auf die Ergebnisse der Studie erscheint auch in der Rückschau zielführend. Zielführend deshalb, weil so eine methodisch und inhaltlich vergleichbare Gestaltung der Interventionsstunden gelang (Hahn, 2014) und damit der Blick auf die eigentliche Zielsetzung der Studie (Wirkungen von Unterschieden in der Sequenzierung) frei wurde. Eine entsprechend inhaltlich und methodisch einheitliche Gestaltung auch nach alternativen Strukturvorgaben wäre dabei möglich gewesen, wurde im Vorfeld allerdings als unwahrscheinlicher oder schwieriger eingeschätzt (vgl. Abschnitt 3.2). Mit einer anders gearteten Wahl der Strukturvorgabe hätte sich zwar das Studienergebnis bezüglich der Lernwirksamkeit wohl leichter in bereits vorhandene Ergebnisse einordnen lassen, dabei allerdings das Risiko erhöht nur unspezifische Aussagen über die Ursachen möglicher Unterschiede in den Wirkungen der Strukturvorgaben zuzulassen.

Die konforme Ausgestaltung der Interventionsstunden zu den jeweiligen theoretischen Vorgaben gelang bis auf wenige meist pragmatisch nicht vermeidbare kleine Abweichungen auch aus externer Sicht (Hahn, 2014). Eine externe Sicht im Sinne einer Expertenberatung bereits im Zuge der Konzeption der Unterrichtseinheiten fand nicht statt, stellt für mich rückblickend aber eine ernst zu nehmende Alternative oder Ergänzung zur letztlich gewählten retrospektivischen Überprüfung der Theoriekonformität durch Hahn (2014) dar. Eine Diskussion der so gestalteten Interventionen erfolgte zwar teilweise im Rahmen eines Vortrags beim naturwissenschaftlichen Kolloquium an der Ruhr-Universität Bochum bei Rainer Wackermann im Juni 2014, eine Berücksichtigung anderer

Perspektiven bei der Interpretation der BMT (Hans E. Fischer, Rainer Wacker-
mann, Heiko Krabbe, Simon Zander) oder der FeU (Helmut Lindemann) fand
bei der Erstellung der Einheiten hingegen nicht statt. Gerade die Suche nach
Austauschmöglichkeiten über die Konzeption der FeU-Einheit erscheint mir
nachträglich als erstrebenswert, da hier keine Modifikation für den Physikun-
terricht vorlag und die in der Literatur angegebenen Unterrichtsbeispiele im
Kontext des Chemieunterrichts wiederzufinden waren. Dabei sehe ich hierin
nicht nur die Möglichkeit der Absicherung bei der Konzeption, sondern auch
die Chance zur nachfolgenden Kommunikation und Diskussion der erhaltenen
Ergebnisse insbesondere im Rahmen der Chemiefachdidaktik als verbessert an.

Methoden der Untersuchung

Verbesserungswürdig ist rückblickend auch die Messung der Kompetenz zur
experimentellen Erkenntnisgewinnung. Die Ausprägung dieser Kompetenz
sollte als Kontrollvariable erfasst werden, dies gelang mit den sechs aus einem
Test (insgesamt 24 Items) von Silke Klos (2008) übernommenen und teilweise
überarbeiteten Items nicht ausreichend gut. Eine größere Zahl an Items
sollte der entsprechende Test (WTE) dabei nicht aufweisen, um besonders lan-
ge Testphasen im Sinne der Testökonomie zu vermeiden. Eine Vergrößerung
der Testlänge hätte vermutlich zu einer weniger fehlerbehafteten Messung,
jedoch auch zu einer längeren Testzeit geführt und dadurch womöglich die
Beantwortung des Wissenstests zum Thema Impuls (WTI) negativ beeinflusst.
Da der WTI allerdings von zentraler Bedeutung für die Aussagequalität der
durchgeführten Untersuchung ist, würde ich in der Rückschau auf den WTE
verzichten. Eine größere Testlänge könnte die Qualität der Messung verbessern,
dies muss aber nicht gelingen. Alternativ könnten so in der Gesamtstudie die
kognitiven Grundfähigkeiten der Schüler über den KFTk ermittelt und in den
Auswertungen kontrolliert werden.

Auch die Messung weiterer Variablen erfolgte nicht immer auf einem zufrieden-
stellend reliablen Niveau. Die Skalen der sozialen Eingebundenheit (SE)
und des Autonomiererlebens (AE) erfüllten dabei auch nicht die Mindestan-
forderungen nach Rost (2013). Die entsprechenden Skalen entstammen der
Arbeit von Berger (2007), welcher die Reliabilitäten Berger (2007, S. 84f.) der
entsprechenden Skalen als ausreichend bis zufriedenstellend ($0.6 < \alpha \leq 0.7$)
kennzeichnet. Die Unterschiede in den Reliabilitäten der von Berger (2007)
übernommenen Items könnten an der im Rahmen dieser Arbeit verwendeten
Skala von 1 bis 4 liegen (bei Berger 1 bis 5). Diese wurde eingesetzt um eine
Bewertungstendenz zur Mitte bei den Schülereinschätzungen zu vermeiden.
Diese Schülereinschätzungen sind nun oft durch ein Globalurteil über den
Unterricht oder die Lehrkraft überlagert (Clausen, 2002). Dies kann zur Folge
haben, dass die Messung einzelner Merkmale ungenauer wird und die Reliabili-
tät der Erfassung dieser Merkmale entsprechend sinkt. Nachträglich betrachtet
könnte eine Skala von 1 bis 6 eventuell zu einer differenzierteren Bewertung
der jeweiligen Merkmale führen, ohne dabei die Antworttendenz zur Mitte zu

fördern. Diese Erweiterung der Skala könnte auch ermöglichen, die Messung weiterer Merkmale (Kompetenzerleben oder die Merkmale zur Struktur des Unterrichts) reliabler zu gestalten. Zwar erfüllen die weiteren Skalen zumindest die Mindestanforderungen bei (Rost, 2013), eine fehlerreduzierte Messung der Merkmale würde aber eine exaktere Auswertung der Daten ermöglichen. Obwohl die Items zur Erfassung der Merkmale aus anderen Studien übernommen oder nur leicht abgeändert wurden, würde ich rückblickend eine andere Skalierung, sowie insgesamt eine ausführlichere Pilotierung und Weiterentwicklung des Schülerfragebogens als sinnvoll erachten.

Zudem könnten alternative Zusammensetzungen des Schülerfragebogens in Betracht gezogen werden. Im Vorfeld der Untersuchung wurde diesbezüglich eine Auswahl getroffen. Diese Auswahl erfolgte auf Basis einer Literaturrecherche über mögliche Zusammenhänge der Strukturierung und des Lernerfolgs mit weiteren Qualitätsmerkmalen und führten zu der gewählten Komposition des Fragebogens. Speziell ein Bezug zu der Studie von Rakoczy et al. (2007) schien relevant und führte zur Integration der Skalen der grundlegenden Bedürfnisse und der kognitiven Aktivierung. Möglich wäre zum Beispiel auch die Erhebung von Aspekten der Klarheit gewesen. Auf die Erhebung der Einschätzung der Klarheit des Unterrichts wurde aber verzichtet, da die Klarheit nicht im direkten Fokus der Untersuchung stand und zudem Klarheit und Strukturierung in der „Realität oft eng zusammenhängen“ (Helmke, 2009, S. 191) und vermutet wurde, dass diese Konstrukte sich nur schwerlich als eigenständige Facetten abgrenzen lassen. Unter Berücksichtigung der ohnehin meist nur ausreichenden Reliabilitäten der Skalen zur Strukturierung erscheint diese Vermutung auch gerechtfertigt gewesen zu sein. Nicht im Fokus der Untersuchung war zudem die Erfassung von geschlechterspezifischen Unterschieden in der Schülerleistung oder der Einschätzung der erhobenen Qualitätsmerkmale. Die Ergebnisse der Untersuchung weisen aber auf ebensolche hin. Vor diesem Hintergrund würde ich bei der Zusammensetzung des Schülerfragebogens Items zur Selbstwirksamkeitserwartung integrieren, unter anderem da sich in einer früheren von mir durchgeführten Studie (Maurer 2013) bei diesem Merkmal ein gewisses Erklärungspotential für die dort gefundenen Geschlechterunterschiede abgezeichnet hatte.

Zusätzlich zur Bewertung von Aspekten der Strukturierung im Schülerfragebogen könnte eine videobasierte Analyse und Einschätzung von Strukturierungsmerkmalen in den Interventionsstunden durch externe Beobachter eine neutrale Perspektive ohne eine globale Urteilsneigung auf den Unterricht geben (Clausen, 2002, S. 188f.). Dies ist dabei weniger eine Alternative als eine Ergänzung und kann (und soll) auch noch im Nachgang der Untersuchung durchgeführt werden.

7. Zusammenfassung und Ausblick

Hintergrund der Untersuchung zum vorliegenden Text ist die Strukturierung von Unterricht. Diese weist grundsätzlich drei verschiedene Bedeutungsfacetten auf (Lipowsky, 2009, S. 83). Die Facette der *kognitionspsychologisch verstandenen Strukturiertheit* beschreibt dabei einzelne unterrichtliche Maßnahmen beispielsweise zur Verankerung des Wissens, werden Aspekte zur Regelung des Schülerverhaltens betrachtet, betrifft dies die Facette der *Strukturierung auf der Verhaltensebene*, die Facette der *didaktischen Strukturierung* beinhaltet die inhaltliche Planung und die Phasierung des Unterrichtsverlaufs. Mit der Sequenzierung von Unterricht nach den theoretischen Richtlinien zweier Strukturvorgaben stand damit ein Teilaspekt der *didaktischen Strukturierung* im Fokus meiner Untersuchung. Ganz allgemein betrachtet, stellt die Strukturierung ein sehr entscheidendes Kriterium für den Lernerfolg der Schüler dar (z.B. Meyer, 2004; Helmke, 2009). Zudem konnte in Metaanalysen der Einfluss einzelner Strukturvorgaben (und damit Teilen der Facette der didaktischen Strukturierung) auf den mittleren Lernerfolg recht genau bestimmt werden. Dieser Einfluss ergab sich meist aus einem Vergleich zu einem Sammelsurium verschiedenster Unterrichtsabläufe, ist grundsätzlich positiv und kann sich zwischen den Strukturvorgaben wesentlich unterscheiden (vgl. Alfieri et al., 2011; Gijbels et al., 2005; Hattie, 2009). Wesentlich unterscheidet sich dabei aber auch die spezifische Lernwirksamkeit einzelner Strukturvorgaben für verschiedene Inhalte, Fächer oder unterrichtlichen Zielsetzungen (Hattie, 2009). Die nun spezifisch lernwirksamste Art der Sequenzierung einer zu planenden Unterrichtsstunde zu einem bestimmten Sachinhalt und bei einer ganz bestimmten Zielsetzung zu nennen, ist (noch) nicht möglich. Um solch spezifische Angaben zu ermöglichen, wären zahlreiche detaillierte empirische Untersuchungen notwendig, deren Umfang und Anzahl dabei nur sehr schwer abgeschätzt werden kann. Allerdings existieren bereits einzelne Strukturvorgaben die für unterschiedliche unterrichtliche Zielsetzungen auch unterschiedliche unterrichtliche Abläufe vorsehen und zumindest für eine Vielzahl an Inhaltsbereichen einsetzbar sein sollen. Zwei dieser Vorgaben wurden für die Gestaltung der Interventionen in dieser Untersuchung ausgewählt. Beide didaktisch gut begründeten Verfahren hegen dabei den Anspruch eine ideale Struktur für den Unterricht zu liefern. Das erste dieser Verfahren ist das Forschend-entwickelnde Unterrichtsverfahren (FeU) nach Schmidkunz und Lindemann (1992), welches grundsätzlich differenziert, ob nun neues Wissen erarbeitet (*induktiver Zweig* des Verfahrens) oder bestehendes in Problemsituationen angewendet (*deduktiver Zweig*) werden soll. Entsprechend der jeweiligen Zielsetzung wird der unterrichtliche Ablauf und

7. Zusammenfassung und Ausblick

auch die Funktion der eingesetzten Experimente verändert vorgegeben. Zum anderen ist dies die physikspezifische Modifikation der Basismodelltheorie (BMT) nach Oser und Baeriswyl (2001). In ihrer modifizierten Form (Wackermann, 2008; Wackermann et al., 2012) unterscheidet diese Theorie vier Zielsetzungen im Physikunterricht (Konzeptaufbau, Konzeptwechsel, Problemlösen und Lernen durch Eigenerfahrung). Als Zielsetzung der Interventionsstunden ist im Kontext dieser Arbeit der Aufbau von neuem Wissen vorgegeben. Im Gegensatz zum FeU sieht die BMT für den Aufbau neuen Wissens jedoch zwei verschiedene Abläufe von Unterricht vor, je nachdem ob neues Wissen in Form eines Konzepts aufgebaut werden soll, oder neues Wissen aus Erfahrungen der Lernenden gewonnen werden soll. Der für die Studie festgelegte Sachinhalt des Impulses entspricht nach meiner Interpretation der Vorgaben der BMT zwei verschiedenen Zielsetzungen. Für diese Zielsetzungen gibt die BMT jeweils unterschiedliche resultierende unterrichtliche Abläufe vor, welche aber innerhalb einer Interventionseinheit kombiniert werden können. Diese Differenzierung nach der Art der Zielsetzung beim Aufbau neuer Inhalte stellt neben der unterschiedlichen Art der Vernetzungsphasen (FeU: fächerübergreifend; BMT auch innerfachlich) einen der wesentlichen Unterschiede in den theoretischen Vorgaben beider Verfahren für die Sequenzierung der Interventionsstunden zum gewählten Thema des Impulses dar. Abgesehen von dieser durch die Verfahren vorgegebene Variation in der Sequenzierung der Interventionsstunden wurden diese Stunden gerade so konzipiert, dass hinsichtlich weiterer unterrichtlicher Gesichtspunkte keine weitergehenden wesentlichen Unterschiede bestehen. Damit sollte gewährleistet werden, dass empirisch gefundene Unterschiede in den Auswirkungen der Interventionsstunden auch eindeutig den im Fokus der Untersuchung liegenden, verschiedenen Sequenzierungsarten zuzuordnen sind.

Um auch weitere Einflussfaktoren auf etwaige Unterschiede in den Auswirkungen kontrollieren zu können, ist die Untersuchung als Laborstudie konzipiert und der Unterricht beider Interventionsgruppen wurde im sogenannten Lernforschungslabor der Universität Regensburg durchgeführt. Die Ausrichtung dieser Laborstudie ist dabei explorativ, da der Stand der Forschung zu den für die Studie gewählten Strukturvorgaben eine theoretische begründete Vorformulierung von Hypothesen über den Studienausgang nicht ermöglichte. Insgesamt beteiligten sich 796 Schüler der zehnten Jahrgangsstufe aus 32 Schulklassen bayerischer Gymnasien an der Hauptstudie. In der Hauptstudie wurde über einen Wissenstest vor und nach den Interventionen die Schülerleistungen zum Impuls erfasst und mittels eines Fragebogens die Schülereinschätzung einzelner Qualitätsmerkmale von Unterricht erhoben.

Die Auswertung der so gewonnenen Daten erfolgte bei vielen Teilaspekten der Studie über t-Tests und mehrfaktorielle Kovarianzanalysen. Für die Gesamtstichprobe konnte zudem eine Mehrebenenanalyse durchgeführt werden. Die Ergebnisse dieser Analysen zeigen einen eindeutigen und systematischen Un-

terschied in der Lernwirksamkeit zwischen beiden Interventionsstunden zum Thema Impuls. Schüler aus der BMT-Interventionsgruppe erreichten im Wissenstest nach der Intervention höchst signifikant bessere Ergebnisse als Schüler der FeU-Gruppe. Die Art der Sequenzierung hat in dieser Studie damit einen Einfluss auf den Lernerfolg. Dieser Einfluss ist systematisch und kann mit einer kleinen Effektstärke von $d = 0.3$ gekennzeichnet werden. Dieser Einfluss ist nach meiner Einschätzung auch für den Unterrichtsalltag und die -praxis bedeutsam, da eine Sequenzierung nach den Vorgaben der BMT nicht mit zusätzlichem Aufwand verbunden ist, aber zusätzlich die Wirkung weiterer Strukturierungsmaßnahmen positiv beeinflussen könnte (Lipowsky, 2009, S. 86). Schüler mit unterdurchschnittlichem Vorwissen werden durch eine basismodellkonforme Unterrichtsgestaltung dabei besonders positiv beeinflusst, ohne dass dabei die Lernprozesse der Schüler gutem Vorwissen wiederum negativ beeinflusst würden. Das Vorwissen der Schüler moderiert die Stärke des Interventionseffektes. Dieser Effekt tritt auch in der Studie von Zander et al. (2015) zur Wirkung von basismodellkonformen Unterricht auf und ist damit wohl kein reiner Effekt des Vergleichs mit dem FeU, sondern scheint mit der speziellen Konzeption nach den Vorgaben der BMT verbunden. Mit Unterricht nach diesen Vorgaben kann damit der Leistungsheterogenität in Klassen wirkungsvoll entgegengewirkt werden. Diese Wirkung schreibe ich dabei der größeren Flexibilität der BMT bei der Erarbeitung neuer Inhalte zu. Die BMT unterscheidet, wann die Schüler mit Hilfe eigener Erfahrungen neue Inhalte *finden* können, und wann die Darstellung eben *erfundener* Inhalte primär durch den Lehrenden geleistet werden muss. Eine Überforderung der Schüler wird so wohl an manchen Stellen des Unterrichtsgeschehens vermieden werden können. Eine Überforderung wird vermutlich auch dadurch vermieden, dass die BMT bei *erfundene*n Inhalten erst dann Phasen der Vernetzung vorsieht, wenn diese neuen Inhalte (Konzepte) durch eine Phase der aktiven Anwendung und Einübung auch gefestigt wurden. Das mag banal klingen, das FeU sieht eben dies aber anders vor. Abweichend von den Vorgaben des FeU fordert die BMT zusätzlich auch eine innerfachliche Vernetzung insbesondere der *erfundene*n Inhalte oder Konzepte. Erst kürzlich konnten Helaakoski und Viiri (2014, S. 101) innerhalb einer deutsch-finnisch-schweizerischen Vergleichsstudie (QuIP-Projekt) auch einen positiven Zusammenhang von Phasen der innerfachlichen Vernetzung mit der Schülerleistung nachweisen.

Die Unterschiede in der Erarbeitung und Verarbeitung der Inhalte in den Interventionsstunden sehe ich dabei als ursächlich für die empirische gefundene Unterschiedlichkeit in der Lernwirksamkeit an. Dabei sollte nicht außer Acht gelassen werden, dass zwischen den Interventionsstunden eigentlich deutlich mehr Gemeinsamkeiten als Unterschiede zu finden sind. Diese wenigen Unterschiede resultieren nur aus Details der Gestaltung der unterrichtlichen Struktur. Bereits Details in der Strukturierung von Unterricht zeigen damit einen bedeutsamen Einfluss auf den Lernerfolg der Schüler.

7. Zusammenfassung und Ausblick

Wendet man den Blick weg von diesen Details hin auf Grundsätzliches zum Physikunterricht in deutschen Klassenzimmern, wird offenbar, dass viele aus theoretischer und empirischer Sicht lernförderliche Maßnahmen schlicht keinen Einzug in die Unterrichtspraxis gefunden haben (Seidel et al., 2006). Viele dieser Maßnahmen finden in den in dieser Studie zugrunde gelegten Strukturvorgaben ihre Berücksichtigung. Berücksichtigt man also diese Strukturvorgaben bei der Gestaltung von Unterricht, könnten diese Maßnahmen auch ihren Einzug in die Unterrichtspraxis finden. Die Schüler könnten davon profitieren. Dies zeigt sich auch in den als deutlich überdurchschnittlich zu bezeichnenden Bewertungen der Studienteilnehmer hinsichtlich der erhobenen Qualitätsmerkmale beider Interventionsstunden, welche bezüglich dieser Bewertungen dabei keine systematischen Unterschiede aufweisen. Erst bei einem erneut detaillierteren Blick lassen sich unterschiedliche Einschätzungen bei einer Teilgruppe der Schüler nachweisen. Schüler mit geringerem Vorwissen profitieren nicht nur besonders stark von der BMT, sie bewerten auch basismodellkonformen Unterricht besonders positiv. Nachgewiesen ist auch, dass sich in mittlerweile zahlreichen Studien zur BMT, Unterricht nach deren Vorgaben stets als lernwirksamer als der Unterricht in der Kontroll- oder Vergleichsgruppe erwiesen hat (z.B. Zander et al., 2013; Wagner, 1999). Die Ergebnisse meiner Studie sind damit in Einklang und es deutet sich ein einheitliches Befundbild zur BMT an, welches deren unterrichtlichen Einsatz ertrebenswert erscheinen lässt.

Letztlich abgesichert ist oben erwähntes Befundbild zur BMT dabei noch nicht. Gerade der Vergleich der Wirkungen der BMT mit denen einzelner Strukturvorgaben aus dem angloamerikanischen Sprachraum öffnet ein breites Fenster für empirische Forschungsvorhaben. Der Forschungsstand zu den Vorgaben aus dem angloamerikanischen Sprachraum ist dabei bereits deutlich elaborierter als derjenige zur BMT oder dem FeU, weist dabei gerade im Hinblick auf die Auswirkungen auf affektive Merkmale aber ebenfalls Lücken auf. Lücken, die auch durch die Ergebnisse meiner Studie kaum geschlossen werden konnten. Eine Berücksichtigung auch affektiver Merkmale bei entsprechenden Vergleichen erscheint mir vor diesem Hintergrund interessant zu sein. Erweist sich die BMT auch gegenüber dieser Strukturvorgaben als lernwirksamer und auch hinsichtlich weiterer unterrichtlicher Bedingungsfaktoren als geeignet, so schätze ich Untersuchungen zu einer gelingenden Implementierung der BMT in die Unterrichtspraxis als lohnenswert ein. Zusätzlich könnten in weiteren Studien die Einflüsse der einzelnen zwischen den Interventionseinheiten unterschiedlichen Komponenten (Art der Erarbeitung, innerfachliche Vernetzung, reine Anwendungsphasen) auf den Lernerfolg ermittelt werden. Eine spezifische Zuordnung, welcher dieser Komponenten sich nun zu welchem Anteil für den beobachteten unterschiedlichen Lernerfolg verantwortlich zeigt, kann meine Untersuchung nicht leisten. Erkenntnisse darüber könnten sich aber für die unterrichtliche Gestaltung als wertvoll erweisen.

A. Unterrichtsmaterialien

Auf den folgenden Seiten sind die geplanten Unterrichtsverläufe und zugehörigen Materialien aufgeführt.

Detailplanung für den Ablauf der Doppelstunde gemäß dem FeU			
Denkphase	Zeitverlauf	L-S-Tätigkeiten und Teilergebnisse	Medien
1) Problem-gewinnung	0-6	<p>Lk schreibt Ablauf der Stunde an Tafel</p> <p>Ablauf der Stunde:</p> <ul style="list-style-type: none"> → Die Rolle Wucht in einer Alltagssituation → Experimente zur Wucht → Auswertungen der Experimente → Nähere Betrachtung der Alltagssituation → Anwendungen der Wucht <p>Besprechung der Rollen von Masse und Tempo für die Wucht (Unfallgefahr) von drei unterschiedlich massereichen Fahrzeugen bei einem bevorstehenden Aufprall. Insbesondere bei der Fahrzeuggattung des SUV.</p> <p>Ergebnis: Die Wucht eines Körpers hängt von dessen Masse und Tempo ab.</p> <p>Probleme:</p> <ul style="list-style-type: none"> • Wie genau hängt die Wucht von diesen Größen ab? • Was passiert mit der Wucht zweier Fahrzeuge bei einem Zusammenstoß? • Wie genau wirkt sich die Wucht auf die Unfallgefahr aus? 	Tafel, Power-Point
2) Überlegungen zur Problem-lösung:	6-9	<ul style="list-style-type: none"> → Bereitstellung des Vorwissens und Analyse des Problems → Besprechung der Experimentiermaterialien - Was steht zur Verfügung → Ziele beim Experimentieren - Informationen zu den Problemfragen → Fragestellung: Wie können wir mit den gegebenen Materialien herausfinden, wie die Wucht von der Masse und dem Tempo abhängt? 	Power-Point, Tafel
3) Durchführung eines Lösungs-vorschlags zu 1)	9-55	<ul style="list-style-type: none"> → Vorgabe der Gruppeneinteilung, der Notation und der Zeitgestaltung (30 + 5) → Durchführung der Experimente und Vorbereitung der Folien → Präsentation der Ergebnisse jeder Gruppe durch die SuS mit Hilfe von selbstklebenden Folien 	Experiment, AB, selbstklebende Folien, Tafel,

Abbildung A.1.: Ablauf der Intervention nach den Vorgaben des FeU - Teil 1

A. Unterrichtsmaterialien

		→ Zusammentragung der Teilergebnisse aus dem arbeitsteiligen Experimentierabschnitt 2) → Zusammenfassung und Diskussion der wesentlichen Ergebnisse der Experimente → Problemfragen teilweise beantwortet - erfordern teilweise noch Auswertung	Power-Point
4) Abstraktion	55 - 65	in Worten gemeinsam (sprachliche Abstraktion): → Die Masse und das Tempo eines Körpers sind zueinander indirekt proportional, wenn sich dessen Wucht nicht verändert. Das Produkt aus Masse und Tempo hat dabei stets den gleichen Wert → Die Wucht kann sich z.B. übertragen oder aufteilen bei einem Zusammenprall mit einem zweiten Körper. Lk präsentiert (Beginn symbolische Abstraktion): Was versteht man in der Physik unter dem Begriff Wucht? (1) In der Physik bezeichnet man die Wucht oder den Schwung eines Körpers mit dem Begriff Impuls Wert des Impulses : $\mathbf{p} = \mathbf{m} \cdot \mathbf{v}$ mit m =Masse und v =Tempo Einheit des Impulses : $[p] = \text{kg} \cdot \text{m/s}$ (2) Der Impuls (die Wucht) eines Körpers ist direkt proportional zu dessen Masse und Tempo . Abstraktion weiterer Inhalte (Aspekt der Impulserhaltung und der Richtung) aus leicht verändertem Experimentierkontext und dem Wechselwirkungsprinzip. Ergebnis: Der Impuls p ist eine vektorielle Erhaltungsgröße (Antw. auf: Wie überträgt sich die Wucht?) und ein eigenständiges Konzept der Physik.	selbstklebende Folien, Fenster, Power-Point
5) Wissenssicherung	65-95	Anwendungsbeispiele: → Bearbeitung eines Musterbeispiels (Frontalzusammenstoß SUV - PKW). Bestimmung der Geschwindigkeit zweier Unfallpartner nach einem vollkommen unelastischen Zusammenstoß. → Arbeitsanweisungen für EA/GA (jeder sucht ein Beispiel für eine Anwendungssituation des Impulses (Sport/Technik/Alltag/andere Fächer - in Gruppe Beispiele gegenseitig darstellen - Begleitaufgaben bearbeiten) → Begleitaufgaben : Erläutert inwiefern in den Musterlösungen <ul style="list-style-type: none"> der Impuls als solcher erkennbar ist. 	Power-Point, Tafel, Experiment 2, Notizblatt

Abbildung A.2.: Ablauf der Intervention nach den Vorgaben des FeU - Teil 2

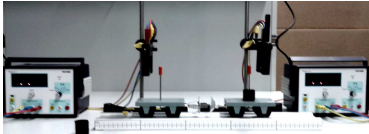
		<ul style="list-style-type: none"> die Impulserhaltung erkennbar ist. → Zeitvorgabe: 12min → kurze Besprechung der Begleitaufgaben Wiederholung: → Ergebnisse und Weg zusammenfassen und kommentieren; Lernzielkontrolle: Lk beschreibt Experiment: zwei verbundene Wagen mit untersch. Gewicht; Anwendung der neuen Inhalte und Überprüfung der erworbenen Kenntnisse. → Ziel: Vorhersage liefern für die Geschwindigkeit der Wagen nach dem Entspannen der Feder (mitschreiben) → wir testen damit sozusagen unser neues Konzept- → Bild:  → Impulserhaltung: Gesamtimpuls bevor die Wagen voneinander wegrollen: $p_{\text{Ges}}=0$ Gesamtimpuls nachher: $p_{\text{Ges}}=0=p_{W1}+p_{W2} \Rightarrow p_{W1} = -p_{W2}$ $m_{W1} v_{W1} = - m_{W2} v_{W2}$ mit $m_{W1} = \frac{1}{2} m_{W2} \Rightarrow v_{W1} = -2v_{W2}$ → Überprüfung durch das Experiment (vorher Test der Apparatur für gleiche Massen der Wagen) → Die Vorhersage konnte bestätigt werden.	
--	--	--	--

Abbildung A.3.: Ablauf der Intervention nach den Vorgaben des FeU - Teil 3

Detailplanung für den Ablauf der Doppelstunde gemäß der Basismodell-Theorie			
Handlungs- kettenschritt	Zeitverlauf	L-5-Tätigkeiten und Teilergebnisse	Medien
E0 – Kontext	0-7	Lk schreibt Ablauf der Stunde an Tafel Ablauf der Stunde: <ul style="list-style-type: none"> Die Rolle Wucht in einer Alltagssituation Experimente zur Wucht Zusatzinformationen Nähere Betrachtung der Alltagssituation Anwendungen der Wucht Besprechung der Rollen von Masse und Tempo für die Wucht (Unfallgefahr) von drei unterschiedlich massereichen Fahrzeugen bei einem bevorstehenden Aufrall. Insbesondere bei der Fahrzeuggattung des SUV. Ergebnis: Die Wucht eines Körpers hängt von dessen Masse und Tempo ab. Unklar: <ul style="list-style-type: none"> Wie genau hängt die Wucht von diesen Größen ab? Was passiert mit der Wucht zweier Fahrzeuge bei einem Zusammenstoß? Wie genau wirkt sich die Wucht auf die Unfallgefahr aus? Ziel: Klärung dieser Unklarheiten mit Hilfe von Experimenten	Power-Point
E1 – Vorentwurf der Handlung	7-10	Besprechung der Experimentiermaterialien - Was steht zur Verfügung Besprechung des Vorgehens beim Experimentieren. Ziele beim Experimentieren - Informationen zu den Unklarheiten Lk gibt die Art der Notation, die Gruppeneinteilung und die Zeitvorgabe für das Experiment: ca. 30 +5 (Erstellung der Folie) min bekannt.	Power-Point
E2 – Handeln im Kon- text	10-45	Durchführung der Experimente und Vorbereitung der Folien	Experiment, AB

Abbildung A.4.: Ablauf der Intervention nach den Vorgaben der BMT - Teil 1

E3 – Erste Ausdifferen- zierung der Er- gebnisse	45 - 55	Präsentation der Ergebnisse jeder Gruppe durch die SuS mit Hilfe von selbstklebenden Folien	selbstklebende Folien
E4 – Generalisierung der Ergebnisse	55-60	→ Zusammenfassung der wesentlichen Ergebnisse der Experimente → Generalisierung von allgemeineren Aussagen aus den Ergebnissen	Fenster, Foli- en, Tafel
K1: Bewusstma- chen des Vorwis- sens	55-57	Lk fasst zusammen: Was wissen wir über die Wucht: → Die Werte für Masse und Tempo von Körpern mit gleicher Wucht sind indirekt proportional zueinander. Das Produkt aus m und v hat dann jeweils den gleichen Wert. → Die Wucht kann sich z.B. übertragen oder aufteilen bei einem Zusammenprall mit einem zweiten Körper.	Power-Point
K2 – Vorstellen und Durcharbeiten ei- nes prototypischen Musters	57-62	Lk präsentiert: Was versteht man in der Physik unter dem Begriff Wucht ? (1) In der Physik bezeichnet man die Wucht oder den Schwung eines Körpers mit dem Begriff Impuls . Wert des Impulses : $p = m \cdot v$ mit m=Masse und v=Tempo Einheit des Impulses : $[p] = \text{kg} \cdot \text{m/s}$ (2) Der Impuls (die Wucht) eines Körpers ist direkt proportional zu dessen Masse und Tempo . (3) Der Impuls ist eine vektorielle Erhaltungsgröße . → Bearbeitung eines Musterbeispiels (Frontalzusammenstoß SUV - PKW) Bestimmung der Geschwindigkeit zweier Unfallpartner nach einem vollkommen unelastischen Zusammenstoß	Power-Point
K3 – Darstellen bzw. Erarbeiten der we- sentlichen Merk- male und Prinzipien	62-65	→ Ausarbeiten der wesentlichen Merkmale des Impulses anhand leicht verändertem Musterbeispiel Teilergebnis: 1) eigenständiges Konzept der Physik 2) Vektorcharakter	Power-Point MB
K4 – Aktiver Umgang mit neuem Kon- zept	65-80	→ Arbeitsanweisungen für EA/GA (jeder ein ausgearbeitetes Beispiel nachvollziehen - in Gruppe Beispiele gegenseitig erklären - Begleitaufgaben bearbeiten) → Begleitaufgaben : Erläutert inwiefern in den Musterlösungen <ul style="list-style-type: none"> der Impuls als solcher erkennbar ist. die Impulserhaltung erkennbar ist. 	Power-Point, Ausgearbeitete Beispiele-

Abbildung A.5.: Ablauf der Intervention nach den Vorgaben der BMT - Teil 2


		<p>-- Zeitvorgabe: 12min</p> <p>-- kurze Besprechung der Begleitaufgaben</p>	
K5 – Kombination verschiedener Konzepte zu größeren Einheiten	80-85	<p>-- Innerfachliche Vernetzung des Konzepts des Impulses mit der (Kinetischen) Energie am Musterbeispiel</p> <p>1) Gültigkeitsbereich, wann gelten beide Erhaltungssätze? Wann ist auch die kin. Energie erhalten? an gleichem Bsp. erkennbar.</p> <p>2) Vorhersagekraft</p>	
E5 – Begegnung mit der Fremderfahrung	85-95	<p>Lk beschreibt Experiment: zwei verbundene Wagen mit untersch. Gewicht; Anwendung der neuen Inhalte auf anderen Kontext.</p> <p>-- Ziel: Vorhersage liefern für die Geschwindigkeit der Wagen nach dem Entspannen der Feder (mitschreiben) --wir testen damit sozusagen unser neues Konzept--</p> <p>-- Bild:</p>  <p>-- Impulserhaltung: Gesamtimpuls bevor die Wagen voneinander weggrollen: $p_{\text{ges}}=0$ Gesamtimpuls nachher: $p_{\text{ges}}=0=p_{k1}+p_{k2} \Rightarrow p_{k1} = -p_{k2}$ $m_{k1} v_{k1} = - m_{k2} v_{k2}$ mit $m_{k1} = \frac{1}{2} m_{k2} \Rightarrow v_{k1} = 2v_{k2}$</p> <p>-- Überprüfung durch das Experiment (vorher Test der Apparatur für gleiche Massen der Wagen)</p> <p>-- Die Vorhersage konnte bestätigt werden.</p>	Tafel, Experiment 2, Notizblatt

Abbildung A.6.: Ablauf der Intervention nach den Vorgaben der BMT - Teil 3

Wovon hängt die Wucht ab?

Worum geht es?

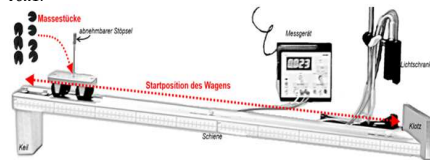
Wir befassen uns mit der **Wucht** von **Körpern**. Mit Hilfe von Experimenten stellt ihr Vermutungen auf, wie genau die Wucht eines Körpers von dessen Masse und Tempo abhängt (Teil 1) und wie und ob sie sich bei einem Zusammenprall übertragen kann (Teil 2).

Was könnt ihr verändern?

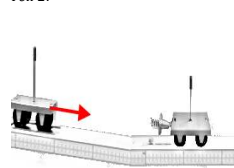
Teil 1 und 2:

- Auflegen von Massestücken (10g oder 50g) auf den Wagen (Masse 50g)
- Verändern der Startposition des Wagens => damit verändert ihr natürlich das Tempo des Wagens.

Teil1:



Teil 2:



Hinweise:

- Achtet darauf, dass der Klotz immer direkt an der Schiene steht.
- Verändert während einer Versuchsreihe nicht die Position der Lichtschranke, oder die Einstellungen am Messgerät.
- Das Tempo des Wagens könnt ihr aus der gemessenen Zeit errechnen. Dazu benötigt ihr den Durchmesser des Stöpsels ($d=8\text{mm}$). In guter Näherung kann dann über $v=d/t$ das Tempo bestimmt werden.

Wie lautet der Erkundungsauftrag?

Teil1: Wie hängt die Wucht des Wagens von dessen Masse und Tempo ab?

- Sucht dazu die niedrigste Startposition für den Wagen so, dass dieser gerade umfällt und notiert euch die jeweilige Masse und das Tempo des Wagens.

Masse in g	60	80	100	120	160	200	240
Wagen 50g + Zusatzgewichte							
Zeit in s							
Tempo in $\frac{m}{s}$							

- Formuliert euer Ergebnis des Erkundungsauftrages in Form von Wenn-Dann-Sätzen.

Teil2: Was passiert mit der Wucht eines Wagens bei einem Zusammenstoß mit einem Zweiten?

- Ersetzt den Klotz aus Teil 1 durch eine dritte Schiene, auf welcher ein zweiter Wagen (mit Feder) steht.
- Lasst den ersten Wagen auf den Zweiten aufprallen und beobachtet was mit der Wucht des ersten Wagens passiert. (Eine Zeitmessung ist nicht mehr erforderlich!)
- Formuliert euer Ergebnis des Erkundungsauftrages in Form eines Wenn-Dann-Satzes.

Abbildung A.7.: Experimentieranleitung für beide Interventionen

MERKBLATT

IMPULS

Experimentelle Ergebnisse (1):

- Die Wucht eines Körpers hängt (proportional) von dessen Masse und Geschwindigkeit ab.
- In der Physik bezeichnet man die Wucht oder den Schwung eines Körpers mit dem Begriff Impuls.
Wert des Impulses: $p = m \cdot v$ mit m =Masse und v =Tempo
Einheit des Impulses $[p] = \text{kg} \cdot \text{m/s}$

Experimentelle Ergebnisse (2):

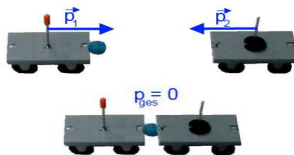
$$p_{N1} + p_{N2} + \dots = p_{V1} + p_{V2} + \dots$$

Die Summe der Impulse
aller Körper **nach** dem
Zusammenstoß

=

Die Summe der Impulse
aller Körper **vor** dem
Zusammenstoß

Experimentelle Ergebnisse (3):



Impulserhaltung!

$$\vec{p}_1 = m_1 \cdot \vec{v}_1 \quad \text{und} \quad \vec{p}_2 = m_2 \cdot \vec{v}_2$$

$$\vec{p}_1 + \vec{p}_2 = 0$$

$$p_{\text{ges}} = 0$$

Der **Impuls** ist eine **vektorielle Erhaltungsgröße**.
Die Richtung der Impulse muss bei der Impulserhaltung berücksichtigt werden.

Anwendungsbeispiele:

Raketenantrieb



Quelle: NASA

Rückstoßpistole



Quelle: NASA

Unfallforschung und Unfallaufklärung



Quelle: Brady Holt

Abbildung A.8.: Merkblatt für die Schüler der Interventionsgruppe FeU

Was versteht man in der Physik unter dem Begriff „Wucht“?

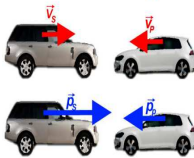
- (1) In der Physik bezeichnet man die Wucht oder den Schwung eines Körpers mit dem Begriff **Impuls**.

Wert des **Impulses**: $\mathbf{p} = m \cdot \mathbf{v}$ mit m =Masse und v =Tempo

Einheit des **Impulses**: $[p] = \text{kg} \frac{\text{m}}{\text{s}}$

- (2) Der **Impuls** ist eine **Erhaltungsgröße**.

Welche Rolle spielt die Richtung?



(1) Multipliziert man die **Geschwindigkeit** eines Körpers mit seiner **Masse** erhält man dessen **Impuls**.

Merkmal 1:

Auch der **Impuls** ist eine **vektorielle** Größe. Zur vollständigen Beschreibung des Impulses eines Körpers muss dessen Stärke und Richtung angegeben sein.

Was bedeutet Erhaltungsgröße?



Gesamtimpuls vor dem Zusammenstoß = Gesamtimpuls nach dem Zusammenstoß

$$\vec{p}_s + \vec{p}_p = \vec{p}_{\text{ges}}$$

Merkmal 2: Der **Impuls** ist eine **Erhaltungsgröße**. Die Richtung der Impulse muss dabei mit berücksichtigt werden.

Und was ist mit der Energie?



Gesamtimpuls: ca. $22000 \text{ kg} \frac{\text{m}}{\text{s}}$

Summe der kinetischen Energien: ca. 1,8 GJ



Gesamtimpuls: ca. $22000 \text{ kg} \frac{\text{m}}{\text{s}}$

Summe der kinetischen Energien: ca. 0,1 GJ

Merke:

Der **Gesamtimpuls** bleibt bei einem Zusammenstoß auch **erhalten**, wenn sich einer oder beide Körper irreversibel **verformen**. Für die Summe der **kinetischen Energien** gilt dies **nicht**.

Wann bleibt auch die kinetische Energie erhalten?



Bedingung:

Keine dauerhafte Verformung der beiden Körper



Dann gilt:

Der **Gesamtimpuls** und die **Summe** der **kinetischen Energien** der Körper des Systems bleiben **erhalten**.

Merke:

- Der **Impuls** ist ein eigenständiges neues Konzept in der Physik.
- Gerade bei Stoßprozessen hat der Impuls eine hohe Vorhersagekraft.

Abbildung A.9.: Merkblatt für die Schüler der Interventionsgruppe BMT

B. Instrumente der Datenerhebung

Auf den folgenden Seiten sind selbst erstellte, oder überarbeitete Instrumente der Datenerhebung aufgeführt.





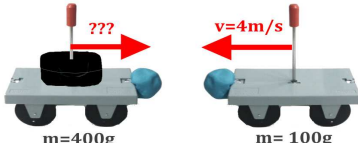

Wissenstest zum Thema Impuls		Datum	FeU/BMT VT/NT
Initialen/Name _____		Geschlecht: <input type="checkbox"/> w: <input type="checkbox"/> m:	
A1) Wucht. Welcher der drei Pkw würde bei einem Unfall mit der größten Wucht auf ein entgegenkommendes Fahrzeug aufprallen?	 PKW1	 PKW2	 PKW3
a) PKW 1 <input type="checkbox"/>	b) PKW2 <input type="checkbox"/>	c) PKW3 <input type="checkbox"/>	d) alle gleich <input type="checkbox"/>
A2) Vor und nachher im Vergleich Zwei Pkw prallen aufeinander. Dabei verkeilen sich die Wagen ineinander und verformen sich. Welche Aussage über die Situation vor und nach dem Zusammenprall ist korrekt?			
			
a) Die Summe der kinetischen Energien vor dem Aufprall ist gleich der kinetischen Energie nach dem Aufprall. Die Wucht ändert sich.		<input type="checkbox"/>	
b) Die Wucht der beiden Wagen vor dem Aufprall ist gleich der Wucht nachher. Die kinetische Energie ändert sich.		<input type="checkbox"/>	
c) Die Summe der kinetischen Energien vor dem Aufprall ist gleich der kinetischen Energie nach dem Aufprall. Die Wucht der beiden Wagen vor dem Aufprall ist gleich der Wucht nachher.		<input type="checkbox"/>	
d) keine der Aussagen ist zutreffend.		<input type="checkbox"/>	
A3) Stillstand? Im Bild rechts ist zu sehen, wie zwei Experimentierwagen unterschiedlicher Masse aufeinander zu rollen. Die Wagen stoßen zusammen. Bei welchem Tempo des linken Wagens kommen sie dabei genau zum Stillstand?		 m=400g m= 100g	
a) $4 \frac{m}{s}$ <input type="checkbox"/>	b) $2 \frac{m}{s}$ <input type="checkbox"/>	c) $1 \frac{m}{s}$ <input type="checkbox"/>	d) beide Wagen können nicht zum Stillstand kommen <input type="checkbox"/>
A4) Tackling Im Bild rechts sieht man Flo beim Rugby Training. Er schafft es dabei das Trainingsgerät gerade so umzustößen. Flo (m=60kg) trifft dabei mit einem Tempo von $6 \frac{m}{s}$ auf das Übungsgerät auf. Seine Teamkollegen Hans und Xaver wollen ebenfalls das Übungsgerät umstoßen. Hans (m=120kg) kommt dabei auf ein Tempo von $3 \frac{m}{s}$. Xaver (m=180kg) dagegen nur auf $2 \frac{m}{s}$. Wer schafft es?			
a) Nur Flo <input type="checkbox"/>	b) Flo und Hans <input type="checkbox"/>		
c) Flo und Xaver <input type="checkbox"/>	d) alle drei <input type="checkbox"/>		

Abbildung B.1.: Wissenstest zum Thema Impuls - Teil 1

A5) Skateboard

Im Bild rechts sieht man Mia auf ihrem Skateboard fahren. Zusammen mit ihrem Skateboard hat Mia eine Masse von 50kg. Als Spitzengeschwindigkeit erreicht Mia ein Tempo von $10 \frac{m}{s}$. Wie hoch ist ihre maximale Bewegungsenergie?

- a) 250J ☐
- b) 500J ☐
- c) 2500J ☐
- d) 5000J ☐



A6) Flummis

Zwei Flummis bewegen sich mit **gleichem Tempo** und gleichen Bewegungsrichtungen. Der kleinere Flummi hat im Vergleich zum größeren Flummi die **halbe Masse**. Welche Aussage über die Wucht der Flummis ist zutreffend?

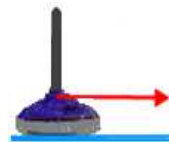
- a) Der kleinere Flummi hat die gleiche Wucht wie der größere. ☐
- b) Der kleinere Flummi hat die viertelte Wucht wie der größere. ☐
- c) Der kleinere Flummi hat die halbe Wucht wie der größere. ☐
- d) keine der Aussagen ist zutreffend. ☐



A7) Eisstockschießen

Im Bild rechts ist ein Eisstock ($m=10kg$) in Bewegung abgebildet. Maximal können erwachsene Eisstockschützen eine Geschwindigkeit von ca. $20 \frac{m}{s}$ erreichen. Jugendliche erreichen mit einem Eisstock der **gleichen Masse** das **halbe Tempo**. Welche Aussage über die maximale Wucht eines Eisstocks trifft beim Vergleich eines jugendlichen mit einem erwachsenen Schützen zu?

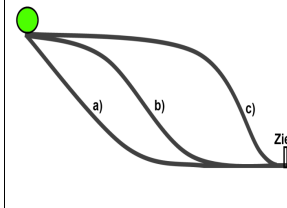
- a) Die maximale Wucht halbiert sich. ☐
- b) Die maximale Wucht bleibt gleich. ☐
- c) Die maximale Wucht sinkt auf ein Viertel. ☐
- d) keine der Aussagen ist zutreffend. ☐

















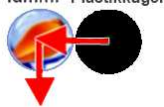
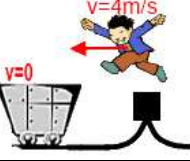
A8) Rollbahn



Im Bild rechts sind eine Kugel und verschiedene Rollbahnen dargestellt. Die Kugel soll jede der Bahnen einmal hinunter rollen (reibungsfrei). Bei welcher Bahn hat die Kugel das höchste Tempo im Ziel?

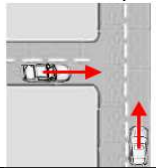
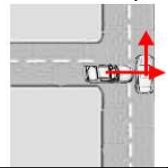




- a) Bei Bahn a) ☐
- b) Bei Bahn b) ☐
- c) Bei Bahn c) ☐
- d) Die Kugel hatte im Ziel immer das gleiche Tempo. ☐

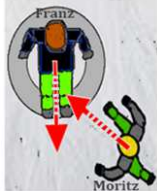






<p>A9) Bullriding Im Bild rechts ist zu sehen, wie ein Rinderbulle auf ein stehendes Fass zu läuft. Die Metallfässer zum Schutz der Clowns sind außen mit einer elastischen Gummischicht überzogen. So sollen weder Bulle noch Clown bei einem Aufprall verletzt werden. Der Bulle ($m=900\text{ kg}$) läuft mit $5\frac{m}{s}$ auf den Clown im Fass (gemeinsam: $m=100\text{ kg}$) auf. Nach dem Stoß hat der Bulle noch ein Tempo von $4\frac{m}{s}$ und läuft in seiner ursprünglichen Richtung weiter. Welches Tempo bekommt der Clown samt Fass dadurch?</p> <p>a) $6\frac{m}{s}$ <input type="checkbox"/> b) $12\frac{m}{s}$ <input type="checkbox"/> c) $9\frac{m}{s}$ <input type="checkbox"/> d) $15\frac{m}{s}$ <input type="checkbox"/></p>	<p>kurz vor dem Stoß:</p>  <p>direkt nach dem Stoß:</p> 
<p>A10) Autoscooter In einem Autoscooter prallt ein Wagen mit einem Tempo von $3\frac{m}{s}$ auf einen zweiten, ruhenden Wagen. Die Wagen sind durch einen elastischen Gummiring um das Fahrzeug geschützt. Der vorher stehende Wagen hat nach dem Aufprall (elastisch) ein Tempo von $5\frac{m}{s}$. Was kann man über die Massen der Wagen aussagen?</p> <p>a) Die Masse des aufprallenden Wagens ist größer als die des vorher ruhenden. <input type="checkbox"/> b) Die Masse des aufprallenden Wagens ist kleiner als die des vorher ruhenden. <input type="checkbox"/> c) Die Masse des aufprallenden Wagens ist gleich der des vorher ruhenden. <input type="checkbox"/> d) Über die Masse der Wagen lässt sich hier nichts aussagen. <input type="checkbox"/></p>	<p>kurz vor dem Stoß:</p>  <p>direkt nach dem Stoß:</p> 
<p>A11) Eisstockschießen. Ein auf der Eisfläche gleitender Eisstock prallt (elastisch) auf einen zweiten baugleichen Eisstock. Was passiert, wenn der geschossene Eisstock den anderen genau mittig (zentral) trifft?</p> <p>a)  <input type="checkbox"/> b)  <input type="checkbox"/> c)  <input type="checkbox"/> d) keine der angebotenen Möglichkeiten ist zutreffend <input type="checkbox"/></p>	<p>kurz vor dem Stoß:</p> 
<p>A12) Münzen schnippen Im Bild rechts ist dargestellt, wie man zwei 1€-Münzen zentral aufeinander schnippt. Wie bewegen sich die 1€-Münzen direkt nach deren elastischem Zusammenstoß?</p> <p>a)  <input type="checkbox"/> b)  <input type="checkbox"/> c)  <input type="checkbox"/> d) keine der angebotenen Möglichkeiten ist zutreffend <input type="checkbox"/></p>	<p>kurz vor dem Stoß:</p> 

<p>A13) Flummis (von oben betrachtet)</p> <p>Im Bild rechts ist zu sehen, wie zwei gleichartige Flummis aufeinander prallen (elastisch). Der von rechts kommende Flummi (2) hat dabei ein höheres Tempo als der andere Flummi (1). Der rechte Flummi kommt dabei zum Stillstand. In welche Richtung bewegt sich der andere Flummi weiter? Kreuze die passende Antwort an.</p>	
<p>a) <input type="checkbox"/> b) <input type="checkbox"/> c) <input type="checkbox"/> d) <input type="checkbox"/></p>	
<p>A14) Eishockey (von oben betrachtet)</p> <p>Die Bilder rechts zeigen wie ein Eishockeyspieler sich einem anderen nähert und mit ihm zusammenstößt. Das Tempo des Spielers mit der Nummer 2 ist dabei größer als das Tempo von Spieler Nummer 8, ihre Massen identisch. Durch die Schutzkleidung sind beide beim Stoß geschützt und prallen (elastisch) voneinander ab. Spieler Nummer 2 kommt dabei gerade zum Stehen. In welche Richtung bewegt sich Spieler 8 weiter?</p>	
<p>a) <input type="checkbox"/> b) <input type="checkbox"/> c) <input type="checkbox"/> d) <input type="checkbox"/></p>	
<p>A15) Kugelstoß (von oben betrachtet)</p> <p>Im Bild rechts ist zu sehen, wie eine Plastikkugel auf einen Flummi prallt (elastisch). Die Plastikkugel hat dabei eine geringere Masse als der Flummi. Beide Kugeln bewegen sich mit dem gleichen Tempo. In welche Richtung bewegen sich die beiden Kugeln direkt nach dem Aufprall weiter? Kreuze die passende Antwort an.</p>	
<p>a) <input type="checkbox"/> b) <input type="checkbox"/> c) <input type="checkbox"/> d) <input type="checkbox"/></p>	
<p>A16) Lorensprung</p> <p>Moritz ($m=60\text{kg}$) springt vom Prallbock des Endes eines Gleises in eine stehende Lore ($m=180\text{kg}$). Er springt dabei mit einem Tempo von $4\frac{\text{m}}{\text{s}}$ ab. Mit welchem Tempo bewegt sich die Lore mit dem Jungen anschließend vom Prallbock weg?</p> <p>a) $4\frac{\text{m}}{\text{s}}$ <input type="checkbox"/> b) $1\frac{\text{m}}{\text{s}}$ <input type="checkbox"/> c) $2\frac{\text{m}}{\text{s}}$ <input type="checkbox"/> d) $0\frac{\text{m}}{\text{s}}$ <input type="checkbox"/></p>	

<p>A17) Aufprall mit Skateboard Im Bild rechts ist zu sehen, wie Mia ($m=40\text{kg}$) auf Kai mit hohem Tempo ($5\frac{\text{m}}{\text{s}}$) zu rast. Kai ($M=100\text{kg}$) steht bereits am Treffpunkt und wartet dort. Mia kann aber nicht mehr bremsen und saust voll in Kai. Dieser hält sie fest und fällt dabei nicht einmal vom Skateboard. Wie bewegen sich die beiden gemeinsam weiter, nachdem Kai Mia aufgefangen hat.</p>	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> $v=5\text{m/s}$  </div> <div style="text-align: center;"> $v=0$  </div> </div>
<p>a) Gar nicht. Kai bleibt samt Skateboard und Mia stehen. <input type="checkbox"/> b) Beide bewegen sich mit dem halben Tempo von Mia nach rechts weiter. <input type="checkbox"/> c) Beide bewegen sich mit dem ursprünglichen Tempo von Mia nach rechts weiter. <input type="checkbox"/> d) keine der Aussagen ist zutreffend <input type="checkbox"/></p>	

<p>A18) Verkehrsunfall (von oben betrachtet) Die Bilder rechts zeigen, wie ein PKW mit einem zweiten kollidiert. Beide haben die gleiche Masse und das gleiche Tempo. Die PKW verhaken sich nach dem Aufprall. In welche Richtung bewegen sie sich nach dem Aufprall gemeinsam weiter? Kreuze an:</p>	kurz vor dem Aufprall: 	direkt vor dem Aufprall: 	
<p>a)  <input type="checkbox"/> b)  <input type="checkbox"/> c)  <input type="checkbox"/> d)  <input type="checkbox"/></p>			

<p>A19) Schneepiste (von oben betrachtet) Das Bild rechts zeigt wie Moritz zu Franz auf dessen Rutschring springt. Beide Jungs haben die gleiche Masse, die Masse des Rutschrings kann man vernachlässigen. Über ihr Tempo ist nichts bekannt. Nach dieser tollkühnen Aktion fahren beide gemeinsam auf dem Rutschring weiter. Wie könnte die gemeinsame Bewegung der Jungs direkt nach dem Sprung sein? Kreuze die passende Antwort an.</p>	
<p>a)  <input type="checkbox"/> b)  <input type="checkbox"/> c)  <input type="checkbox"/> d)  <input type="checkbox"/></p> <p style="text-align: center; margin-top: -10px;">$v=0$</p>	

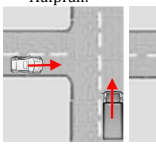
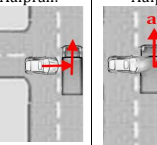
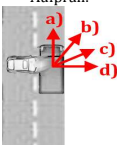

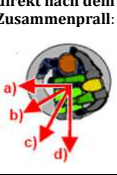
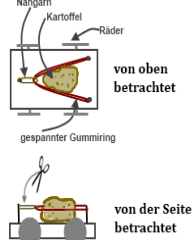










<p>A20) Unfall (von oben betrachtet) Die Bilder rechts zeigen, wie ein LKW mit einem PKW kollidiert. Beide Fahrzeuge haben das gleiche Tempo, der LKW aber eine höhere Masse. Die beiden Fahrzeuge verhaken sich nach dem Aufprall. In welche Richtung bewegen sie sich direkt nach dem Aufprall gemeinsam weiter?</p>	kurz vor dem Aufprall: 	direkt vor dem Aufprall: 	direkt nach dem Aufprall: 
<p>a) <input type="checkbox"/> b) <input type="checkbox"/> c) <input type="checkbox"/> d) <input type="checkbox"/></p>			

Abbildung B.5.: Wissenstest zum Thema Impuls - Teil 5

<p>A21) Schneepiste (von oben betrachtet) Die Bilder rechts zeigen, wie Moritz zu Franz auf dessen Rutschring springt. Beide Jungs haben die gleiche Masse. Moritz hat ein höheres Tempo als Franz. Beide fahren gemeinsam auf dem Rutschring weiter. Wie könnte die gemeinsame Bewegung der Jungs auf der flachen Schneepiste direkt nach dem Sprung sein? Kreuze die passende Antwort an</p>	<p>kurz vor dem Zusammenprall:</p> 	<p>direkt nach dem Zusammenprall:</p> 	
a) <input type="checkbox"/>	b) <input type="checkbox"/>	c) <input type="checkbox"/>	d) <input type="checkbox"/>

<p>A22) Kartoffelkanone Im Bild rechts ist eine mobile Kartoffelkanone abgebildet. Vom beweglichen Wagen aus, wird die Kartoffel nach rechts geschleudert. Die Kartoffel hat eine höhere Masse als der Wagen. Welche der Aussagen über die Bewegung des Wagens trifft zu?</p> <p>a) Kartoffel und Wagen bewegen sich mit gleichem Tempo aber unterschiedlichen Richtungen auseinander. <input type="checkbox"/></p> <p>b) Die Kartoffel bleibt liegen, und der Wagen bewegt sich mit hohem Tempo nach links. <input type="checkbox"/></p> <p>c) Der Wagen rollt nach rechts, Die Kartoffel folgt mit etwas niedrigerem Tempo in die gleiche Richtung. <input type="checkbox"/></p> <p>d) Die Kartoffel fliegt nach rechts, der Wagen mit höherem Tempo als die Kartoffel nach links. <input type="checkbox"/></p>	 <p>von oben betrachtet</p> <p>von der Seite betrachtet</p>
--	--

<p>A23) Münzenkette Im Bild rechts ist dargestellt, wie man eine 1€-Münze auf vier ruhende 1€-Münzen schnippt. Wie bewegen sich die fünf 1€-Münzen direkt nach dem Zusammenstoß?</p> <p> geschnippte Münzen ← Bewegungsrichtung, nicht Betrag</p>	
<p>a)  <input type="checkbox"/></p>	<p>b)  <input type="checkbox"/></p>
<p>c)  <input type="checkbox"/></p>	<p>d) Keine der Aussagen ist zutreffend <input type="checkbox"/></p>

<p>A24) Münzenkette Im Bild rechts ist dargestellt, wie man drei 1€-Münzen auf zwei ruhende 1€-Münzen schnippt. Wie bewegen sich die fünf 1€-Münzen direkt nach dem Zusammenstoß?</p> <p> geschnippte Münzen ← Bewegungsrichtung, nicht Betrag</p>	
<p>a)  <input type="checkbox"/></p>	<p>b)  <input type="checkbox"/></p>
<p>c)  <input type="checkbox"/></p>	<p>d) Keine der Aussagen ist zutreffend <input type="checkbox"/></p>

-6-

Abbildung B.6.: Wissenstest zum Thema Impuls - Teil 6

Wissenstest zur Erkenntnisgewinnung		Datum	FeU/BMT VT/NT
Initialen/Name _____		Geschlecht: <input type="checkbox"/> w: <input type="checkbox"/> m:	

A1) Wenn Wissenschaftler irgendeine Größe mehrere Male sorgfältig messen, erwarten Sie dass...

a) ...alle Werte genau übereinstimmen. ☐

b) ...nur zwei Messwerte genau übereinstimmen. ☐

c) ...alle Messwerte bis auf einen genau übereinstimmen. ☐

d) ...die meisten Messwerte nahe beieinander liegen, jedoch nicht genau übereinstimmen. ☐

A2) Maria hat Gas aufgefangen, das von einem glühenden Stück Holzkohle abgegeben wurde. Das Gas wurde anschließend durch farbloses Kalkwasser abgeleitet. In Marias Bericht steht: „Nachdem das Gas in das Gefäß geleitet wurde, bekam das Kalkwasser allmählich eine milchigweiße Farbe.“ Diese Aussage ist...

a) ...eine Beobachtung. ☐

b) ...eine Schlussfolgerung. ☐

c) ...eine Verallgemeinerung. ☐

d) ...eine Voraussetzung für die Untersuchung. ☐

e) ...eine Annahme ☐

A3) In der Wissenschaft haben Forscher mit einem Gas experimentiert. Sie haben bei diesem Gas das Volumen bei verschiedenen Drücken aber gleichbleibender Temperatur gemessen und folgende Ergebnisse erhalten:

a) Welcher Wert fehlt in der in der vorletzten Zeile?
180 ☐ 150 ☐ 120 ☐ 140 ☐

b) Welcher Wert fehlt in der ersten Zeile?
0,15 ☐ 0,2 ☐ 0,12 ☐ 0,1 ☐

Druck (N/m ²)	Volumen (m ³)
600	
300	0,2
200	0,3
	0,4
100	0,6

A4) An jede Feder aus einem Set Spiralfedern mit verschiedenen **Federhärten D** wird immer das gleiche Massestück angehängt und die **Ausdehnung l** gemessen. Das Diagramm rechts zeigt das Ergebnis der Messungen.

a) Welche Ausdehnung wäre für eine Spiralfeder mit der Federhärte D=20N/m zu erwarten?
0,6m ☐ 0,45m ☐ 0,5m ☐ 0,55m ☐

b) Welche Ausdehnung wäre für eine Spiralfeder mit der Federhärte D=120N/m zu erwarten?
0,15m ☐ 0,2m ☐ 0,1m ☐ 0,05m ☐

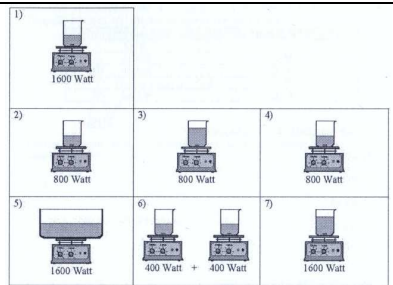
Ausdehnung l (m)	Federhärte D (N/m)
0.2	58
0.3	40
0.4	30
0.8	15
1.0	12

Abbildung B.7.: Test zur Erfassung der Kompetenz Erkenntnisgewinnung - Teil

A5) Die Zeichnungen rechts zeigen Versuche, die Julia durchgeführt hat.

Sie vermutet, dass ein volles Becherglas länger erhitzt werden muss, bis die Flüssigkeit eine Temperatur von 50°C erreicht hat. Welche zwei Versuche muss sie vergleichen, um ihre Vermutung zu überprüfen?

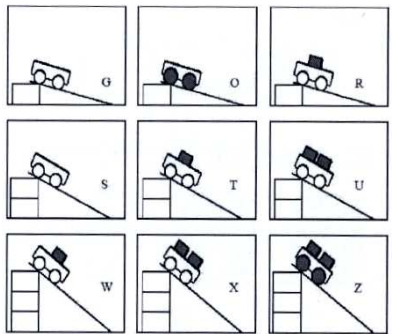
- a) 2 und 5 ☐
- b) 2 und 4 ☐
- c) 2 und 7 ☐
- d) 1 und 6 ☐
- e) 1 und 7 ☐



A6) Die Zeichnungen rechts zeigen mehrere Versuche, die Andrea mit Wagen mit unterschiedlichen Rädern durchgeführt hat. Sie hat sie von unterschiedlichen Höhen hinabrollen lassen. Die Blöcke, die sie hineingelegt hat, haben alle die gleiche Masse.

Sie möchte folgende Vermutung überprüfen. Je schwerer ein Wagen ist, desto größer ist seine Geschwindigkeit am Fuße der Rampe. Welche drei Versuche sollte sie vergleichen?

- a) G, T und X ☐
- b) O, T und Z ☐
- c) R, U und Z ☐
- d) S, T und U ☐
- e) S, W und X ☐



Vielen Dank für deine Mitarbeit 😊

Abbildung B.8.: Test zur Erfassung der Kompetenz Erkenntnisgewinnung - Teil 2

B. Instrumente der Datenerhebung

-Kreuze bitte jeweils die Ziffer an, die für dich zutrifft.
 ① bedeutet dabei, dass die Aussage, so wie sie in der linken Spalte steht gar nicht zutrifft.
 ④ bedeutet dabei, dass die Aussage, so wie sie in der linken Spalte steht völlig zutrifft.

Initialen/ Name: _____ **Geschlecht:** ☐w: ☐m:

Deine Erfahrungen in der heutigen Physikstunde.	trifft gar nicht zu	→	trifft völlig zu
(1) Die Atmosphäre war entspannt.	①	②	③ ④
(2) Ich habe mich in der Gruppe wohlfühlt.	①	②	③ ④
(3) Ich habe gemerkt, dass ich die Dinge verstanden habe.	①	②	③ ④
(4) Ich fühlte mich den Anforderungen gewachsen.	①	②	③ ④
(5) Ich hatte die Möglichkeit, neue Bereiche eigenständig zu erkunden.	①	②	③ ④
(6) Ich hatte das Gefühl, Entscheidungsspielräume zu haben.	①	②	③ ④
(7) Ich fühlte mich konzentriert.	①	②	③ ④
(8) Ich fühlte mich engagiert.	①	②	③ ④
(9) Ich habe versucht, wichtige von unwichtigen Dingen zu unterscheiden.	①	②	③ ④
(10) Ich habe versucht, Stoff mit dem zu verbinden, was ich schon wusste.	①	②	③ ④
(11) Ich habe darüber nachgedacht, ob die Aussagen so stimmen oder nicht.	①	②	③ ④

Deine Einschätzungen über die heutige Physikstunde.	trifft gar nicht zu	→	trifft völlig zu
(12) Ich finde, dass die Lehrkraft wichtige Fakten deutlich hervorgehoben und zusammengefasst hat.	①	②	③ ④
(13) Ich habe die wichtigsten Fakten aus der Stunde, oder einem Experiment selbstständig zusammengefasst.	①	②	③ ④
(14) Ich konnte Zusammenhänge der neuen Lerninhalte mit bereits Gelernten erkennen.	①	②	③ ④
(15) Mir wurde klar gemacht, wie wichtig die Lerninhalte für die Lösung neuer Fragestellungen sind.	①	②	③ ④
(16) Ich wurde aufgefordert, darüber nachzudenken, wie wir zu den neuen Ergebnissen/Inhalten gekommen sind.	①	②	③ ④
(17) Ich finde, dass aufgeworfene Fragen (von mir, Mitschülern oder dem Lehrer) im Verlauf der Stunde geklärt wurden.	①	②	③ ④

Abbildung B.9.: Schülerfragebogen - Teil 1

(18)	Ich finde, dass einzelne Unterrichtsphasen schlüssig aneinander geknüpft waren.	①	②	③	④
(19)	Ich finde, dass die Experimente an sinnvollen Stellen im Unterrichtsverlauf durchgeführt wurden.	①	②	③	④
(20)	Ich konnte die Bedeutung der Lerninhalte durch die Experimente besser erfassen.	①	②	③	④
(21)	Mir wurde erklärt, wie einzelne Phasen im Verlauf der Stunde in den Zusammenhang der gesamten Stunde passen.	①	②	③	④
(22)	Ich konnte dem Unterrichtsverlauf gut folgen.	①	②	③	④
(23)	Mir wurde das Unterrichtsziel zu Beginn der Stunde von der Lehrkraft deutlich gemacht.	①	②	③	④
(24)	Ich finde, dass das Ziel der Stunde immer wieder aufgegriffen wurde.	①	②	③	④
(25)	Ich finde, dass Teilergebnisse des Unterrichts zusammengefasst und mit dem Unterrichtsziel in Verbindung gesetzt wurden.	①	②	③	④
(26)	Ich konnte bei Ergebnissen aus den Experimenten erkennen, dass sie zum Erreichen des Unterrichtsziels beigetragen haben.	①	②	③	④
(27)	Mir war am Ende der Stunde klar, wie alle Erkenntnisse der Stunde zum Erreichen des Unterrichtsziels beigetragen haben.	①	②	③	④

Deine Einschätzung über die Lehrkraft.		trifft gar nicht zu → trifft völlig zu			
(1)	Ich hatte das Gefühl, dass sich der Lehrer für das Thema interessiert.	①	②	③	④
(2)	Ich hatte den Eindruck, dass der Lehrer selbst voll bei der Sache war.	①	②	③	④
(3)	Ich habe bemerkt, dass der Lehrer uns gern unterrichtet.	①	②	③	④
(4)	Mit Schülerbeiträgen ist der Lehrer in dieser Unterrichtsstunde wertschätzend umgegangen.	①	②	③	④
(5)	Der Lehrer war in dieser Unterrichtsstunde freundlich zu mir.	①	②	③	④
(6)	Der Lehrer hat mich in dieser Unterrichtsstunde ausreden lassen, wenn ich dran war.	①	②	③	④
(7)	Wenn der Lehrer in dieser Unterrichtsstunde eine Frage gestellt hat, hatte ich ausreichend Zeit zum Nachdenken.	①	②	③	④
(8)	Der Lehrer hat in dieser Unterrichtsstunde auflockernde Bemerkungen gemacht.	①	②	③	④
(9)	Ich bin in dieser Unterrichtsstunde für eigene Beiträge zum Unterricht gelobt worden.	①	②	③	④

Vielen Dank für Deine Mitarbeit!

Abbildung B.10.: Schülerfragebogen - Teil 2

C. Ergebnisse der Kovarianzanalysen zum Moderatoreneffekt

Auf der folgenden Seite sind die im Hauptdokument nicht explizit aufgeführten Kovarianzanalysen der Teilgruppen nach Vortestergebnis angegeben.

C. Ergebnisse der Kovarianzanalysen zum Moderatoreneffekt

Untersuchter Aspekt	QS	DF	F-Wert	p-Wert	Effektstärke
Ergebnis Vortest	49.76	1	4.79	0.0302 *	$\omega^2 = 0.023$
Geschlecht	58.66	1	5.64	0.0188 *	$\omega^2 = 0.029$
Interventionsstart	23.64	1	2.27	0.1335	$\omega^2 < 0.010$
Sequenzierungsart	180.79	1	17.39	0.0001 ***	$\omega^2 = 0.094$
Residuen	1590.29	153			

Tabelle C.1.: Ergebnisse der Kovarianzanalyse der Daten der Teilgruppe G1

Untersuchter Aspekt	QS	DF	F-Wert	p-Wert	Effektstärke
Ergebnis Vortest	48.83	1	4.57	0.0336 *	$\omega^2 = 0.029$
Geschlecht	54.42	1	5.09	0.0250 *	$\omega^2 = 0.017$
Interventionsstart	171.99	1	16.09	0.0001 ***	$\omega^2 = 0.202$
Sequenzierungsart	74.17	1	6.94	0.0090 **	$\omega^2 = 0.024$
Residuen	2544.51	238			

Tabelle C.2.: Ergebnisse der Kovarianzanalyse der Daten der Teilgruppe G2₁

Untersuchter Aspekt	QS	DF	F-Wert	p-Wert	Effektstärke
Ergebnis Vortest	2.47	1	0.28	0.5973	$\omega^2 < 0.010$
Geschlecht	189.62	1	21.49	0.0000 ***	$\omega^2 = 0.105$
Interventionsstart	44.80	1	5.08	0.0255 *	$\omega^2 = 0.023$
Sequenzierungsart	10.07	1	1.14	0.2869	$\omega^2 < 0.010$
Residuen	1490.89	169			

Tabelle C.3.: Ergebnisse der Kovarianzanalyse der Daten der Teilgruppe G2₂

Untersuchter Aspekt	QS	DF	F-Wert	p-Wert	Effektstärke
Ergebnis Vortest	178.96	1	29.44	0.0000 ***	$\omega^2 = 0.114$
Geschlecht	119.91	1	19.73	0.0000 ***	$\omega^2 = 0.078$
Interventionsstart	12.46	1	2.05	0.1537	$\omega^2 < 0.010$
Sequenzierungsart	4.48	1	0.74	0.3915	$\omega^2 < 0.010$
Residuen	1313.01	216			

Tabelle C.4.: Ergebnisse der Kovarianzanalyse der Daten der Teilgruppe G3

Danksagung

Zu aller erst möchte ich mich bei all denjenigen bedanken, die dazu beigetragen haben, dass ich mich auf dieses Unterfangen eingelassen habe. Karsten Rincke und Josef Reisinger haben mir diese neue Herausforderung zunächst schmackhaft gemacht und sich danach auch dazu entschieden diese gemeinsam mit mir anzugehen. Nach einigen Jahren Berufserfahrung auf meiner Planstelle war dies für mich trotzdem keine leichte Entscheidung, Schule und Schüler in einer gewissen Weise ja zurückzulassen. Eine Entscheidung, die ich womöglich so auch nicht getroffen hätte, hätte man mich nicht auf die Unsinnigkeit des Auslassens dieser Chance (mitunter auch recht deutlich) hingewiesen (Urmel, Alex, Franco, Norbert, Stefan A), oder mir dieses Angebot nicht zugeschickt (Stefan F). Absolut notwendig war dabei auch die Unterstützung dieses Vorhabens durch meine Frau Susanne und dass mir keine Steine in den Weg gelegt wurden (Herr König, stellvertretend für meine alte Stammschule das Gregor-Mendel-Gymnasium in Amberg). Vielen Dank dafür.

Besonderer Dank gebührt natürlich meinem Doktorvater Karsten Rincke - und zwar in vielerlei Hinsicht. Sei es die gemeinsame konzeptionelle Planung und Ausgestaltung der Studie, seine in der Sache stets zutreffende, auch fordernde aber immer fördernde Rückmeldung und Unterstützung bei Texten oder Vorträgen, die familienfreundliche Grundeinstellung (schließlich sind in dieser Zeit auch meine beiden Kinder geboren), oder das Schaffen einer sehr angenehmen Atmosphäre in der Beratung und Begleitung meines Weges und der gesamten Arbeitsgruppe.

Auch den Mitgliedern dieser Arbeitsgruppe will ich an dieser Stelle meinen Dank aussprechen: Hans für seine angenehm ruhige Art und seine Hilfsbereitschaft und den vielen guten Kaffee; Sandra für viele Möglichkeiten der Ablenkung aber auch für den Austausch über die Gestaltung der Unterrichtsstunden; Andreas für die vielen Hinweise unter anderem zu LaTeX oder Grafiken, das sommerliche Eisessen und die damit verbundenen oft notwendigen Kopfpausen und nicht zuletzt Josef für intensive Begriffsreibereien und Gelegenheiten zum Gedankenaustausch die es dann auch oft mir erst ermöglicht haben eine gewisse Strukturiertheit meiner eigenen Gedanken zu erreichen. Als quasi erweitertes Mitglied unserer Arbeitsgruppe möchte ich mich auch noch bei Anja bedanken, dass Sie stets Ausführungen zu meiner Arbeit mit einem offenem Ohr begegnete, gerade wenn es um statistische Belange ging und bei anderen oft ein Ohr zugeing.

Unterstützt wurde ich aber auch im Naturwissenschafts- und mathematikdidaktischem Kolloquium an der Universität Regensburg, in dessen Rahmen ich meine Arbeit in einem recht frühem Stadium schon präsentieren konnte und dabei einige wertvolle Hinweise erhielt. Ebenso wertvolle Hinweise für die Ausgestaltung meiner Untersuchung habe ich auch auf meinen Besuchen der Tagungen und insbesondere der Doktorierendenkolloquien der GDCP erhalten. Vielen Dank damit an alle Teilnehmer dieser Kolloquien und Tagungen.

Ein ausgesprochen großes Dankeschön geht auch an alle beteiligten Lehrkräfte und Schüler, ohne die diese Studie natürlich nicht möglich gewesen wäre.

Vielen herzlichen Dank will ich an dieser Stelle auch meiner Familie aussprechen. Allen voran meiner Frau Susanne und unseren beiden Mädls, aber auch meinen beiden Eltern, der Familie meines Bruders und auch dem „Schwieger-Teil“ unserer Familie. Danke, dass ihr alle so seid wie ihr seid und danke für alles, was ihr mir (mit)gegeben habt.

Literaturverzeichnis

- Adams, G. & Engelmann, S. (1996). *Research on Direct Instruction: 25 Years beyond DISTAR*. Educational Achievement Systems, Seattle, WA.
- Alfieri, L., Brooks, P. J., Aldrich, N. J. & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, 103 No. 1, 1-18. doi: DOI:10.1037/a0021017
- Atkin, J. M. & Karplus, R. (1962). Discovery or invention? *The Science Teacher*, 29 (5), 45-51.
- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COAKTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), (S. S.29-53). Münster u.a.: Waxmann.
- Bayerisches Staatsministerium für Unterricht und Kultus. (2004). *Lehrplan für das Gymnasium in Bayern* (Bd. 1). Kastner.
- Berger, R. (2007). *Das Gruppenpuzzle im Physikunterricht der Sekundarstufe II- eine empirische Untersuchung auf der Grundlage der Selbstbestimmungstheorie der Motivation* (H. Niedderer, H.Fischler & E. Sumfleth, Hrsg.). Logos Verlag Berlin GmbH, Comeniushof, Gubener Str. 47, 10243 Berlin.
- Bliese, P. (2013, März). *Multilevel Modeling in R (2.5) A Brief Introduction to R, the multilevel package and the nlme package*. Zugriff auf http://cran.r-project.org/doc/contrib/Bliese_Multilevel.pdf
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation*. Springer Medizin Verlag, Heidelberg.
- Bortz, J. & Schuster, C. (2010). *Statistik - für Human- und Sozialwissenschaftler*. Springer-Verlag, Berlin, Heidelberg.
- Brophy, J. (2000). Teaching. *International Journal of Education - International Bureau of Education*. Zugriff auf <http://www.ibe.unesco.org/publications/educationalpracticesseriespdf/prac01e.pdf>
- Bruner, J. S. (1961). The Act of Discovery. *Harvard Ed. Rev.*, 31, 21-32.
- Bybee, R. (1997). *Achieving scientific literacy: From purposes to practices*. Portsmouth, NH: Heinemann.
- Bybee, R., Taylor, J. A., Gardner, A., Scotter, P. V., Carlson, J., Westbrook, A. & Landes, N. (2006). *The BSCS 5E Instructional Model: Origins and Effectiveness*. Colorado Springs, CO: BSCS.
- Carnine, D. (2000). *Why education experts resist effective practices (and what it would take to make education more like medicine)*. Thomas B. Fordham Foundation. Washington, DC.

- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive* (D. H. Rost, Hrsg.). Waxmann Münster, New York, München, Berlin.
- Council, N. R. (1996). *National science education standards*. The National Academies Press.
- Council, N. R. (2000). *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning* (S. Olson & S. Loucks-Horsley, Hrsg.). The National Academies Press.
- Deci, E. & Ryan, R. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39 (2), 224-238.
- di Fuccia, D. & Ralle, B. (2010). Forschend-entwickelnd und kontextorientiert -Eine Beziehungsanalyse des forschend-entwickelnden Unterrichtsverfahrens und Chemie im Kontext in fünf Denkstufen-. *MNU*, 63/5, 296-304.
- Draxler, D. (2005). *Aufgabendesign und basismodellorientierter Physikunterricht*. Dissertation. Universität Duisburg-Essen. Zugriff auf <http://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet?id=14098>
- Drollinger-Vetter, B. (2011). *Verstehenselemente und strukturelle Klarheit - Fachdidaktische Qualität der Anleitung von mathematischen Verstehensprozessen im Unterricht* (Bd. 8; G. Krummheuer & A. Heinze, Hrsg.). Waxmann.
- Duit, R., Hepp, R. & Rincke, K. (2013). Guter Frontalunterricht. *Unterricht Physik*, 135/136, S.4-12.
- Eisenkraft, A. (2003). Expanding the 5E Model. A proposed 7E model emphasizes transfer of learning and the importance of eliciting prior understanding. *The Science Teacher*, 70 No.6, 57-59.
- Engelmann, S. (1980). *Direct instruction*. Educational Technology Publications, Inc., Eaglewood Cliffs, New Jersey.
- Field, A., Miles, J. & Field, Z. (2012). *Discovering Statistics Using R*. SAGE Publications. London.
- Frazier, P. A., Barron, K. E. & Tix, A. P. (2004). Testing Moderator and Mediator Effects in Counseling Psychology Research. *Journal of Counseling Psychology*, 51 (1).
- Fries, E. & Rosenberger, R. (1967). *Forschender Unterricht*. Verlag Moritz Diesterweg, Frankfurt am Main, Berlin, Bonn, München.
- Geller, C., Neumann, K. & Fischer, H. E. (2014). Quality of Instruction in Physics- Comparing Finland, Germany and Switzerland. In H. E. Fischer, P. Labudde, K. Neumann & J. Viiri (Hrsg.), (S. 81-92). Waxmann, Münster, New York.
- Göhring, A. (2010). Selbstbestimmtes Lernen im naturwissenschaftlichen Unterricht – eine empirische Interventionsstudie. In *Didaktik der physik - frühjahrstagung - hannover 2010*.
- Gijbels, D., Dochy, F., den Bossche, P. V. & Segers, M. (2005). Effects of Problem-Based Learning: A Meta-Analysis From the Angle of Assessment. *Review of Educational Research*, 75, 27-61. doi: 10.3102/00346543075001027

- Gröhlich, C., Scharenberg, K. & Bos, W. (2009). Wirkt sich Leistungsheterogenität in Schulklassen auf den individuellen Lernerfolg in der Sekundarstufe aus? *Journal for educational research*, 1 (1), 86-105.
- Hahn, E. (2014). *Der Vergleich lernprozessorientierter Strukturmodelle* (Schriftliche Hausarbeit). Universität Regensburg, Fachdidaktik Physik.
- Halliday, D., Resnick, R. & Walker, J. (2009). *Physik* (Bd. 2). Wiley-VCH GmbH & Co KGaA.
- Hammer, D. (1997). Discovery Learning and Discovery Teaching. *Cognition and Instruction*, 15 Nr. 4, 485-529.
- Hansen, D. (2014). *Kodierung von Unterrichtsvideos nach dem forschend-entwickelnden Unterrichtsverfahren* (Schriftliche Hausarbeit). Universität Regensburg, Fachdidaktik Physik.
- Hartig, J. & Rakoczy, K. (2010). Handbuch Statistik, Methoden und Evaluation. In H. Holling & B. Schmitz (Hrsg.), (S. 538-547). Hogrefe Verlag GmbH & Co. KG, Göttingen.
- Hattie, J. (2009). *Visible Learning - A Synthesis of Meta-Analyses Relating to Achievement*. Routledge.
- Helaakoski, J. & Viiri, J. (2014). Quality of Instruction in Physics- Comparing Finland, Germany and Switzerland. In H. E. Fischer, P. Labudde, K. Neumann & J. Viiri (Hrsg.), (S. 93-110). Waxmann, Münster, New York.
- Heller, K. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. Beltz Test GmbH.
- Helmke, A. (2007). Lernprozesse anregen und steuern. Was wissen wir über Klarheit und Strukturiertheit? *Pädagogik (Weinheim)*, 59 (6), 44-47.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität - Diagnose, Evaluation und Verbesserung des Unterrichts*. Klett/Kallmeyer.
- Helmke, A. (2015, April). *Persönliche Mitteilung*.
- Herbers, R. (1990). *Konzeption eines Spiralmodells zur Behandlung der chemischen Schadstoffe im Chemieunterricht verschiedener Jahrgangsstufen basierend auf den Ergebnissen einer empirischen Untersuchung*. Dissertation. Universität Münster.
- Herweg, C. (2008). *Zielorientierung im deutschen und schweizerischen Physikunterricht – eine Videostudie –* (Dissertation, Philosophische Fakultät der Christian-Albrechts-Universität zu Kiel). Zugriff auf http://eldiss.uni-kiel.de/macau/receive/dissertation_diss_00002662
- Hmelo-Silver, C. E. (2004). Problem-Based Learning: What and How Do Students Learn? *Educational Psychology Review*, 16, No. 3, 235-266.
- Hmelo-Silver, C. E., Duncan, R. G. & Chinn, C. A. (2007). Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42(2), 99-107.
- Huber, O. (2009). *Das psychologische Experiment: Eine Einführung* (5. Aufl.). Hogrefe Verlag GmbH & Co. KG, Göttingen.

- Hyde, J. S. (2005, September). The Gender Similarities Hypothesis. *American Psychologist*, 60 (6), 581-592.
- Jong, T. D. & Joolingen, W. R. V. (1998). Scientific Discovery Learning with Computer Simulations of Conceptual Domains. *Review of Educational Research*, 68, 179-201. doi: DOI:10.3102/00346543068002179
- Kessels, U. (2015). Zur Kompatibilität von Geschlechteridentität, MINT-Fächern und schulischem Engagement: Warum wählen Mädchen seltener Physik und machen häufiger Abitur als Jungen? In S. Bernholt (Hrsg.), *Heterogenität und Diversität - Vielfalt der Voraussetzungen im naturwissenschaftlichen Unterricht* (Bd. 35, S. 19-30). Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Bremen 2014. Kiel: IPN.
- Kirschner, P. A., Sweller, J. & Clark, R. E. (2006). Why minimal Guidance during Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, 41(2), 75-86.
- Klos, S. (2008). *Kompetenz im naturwissenschaftlichen Anfangsunterricht* (H. Niederer, H. Fischler & E. Sumfleth, Hrsg.). Logos Verlag Berlin GmbH.
- Kultusministerkonferenz. (2005). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss - Beschluss vom 16.12.2004* (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland., Hrsg.). Luchterhand, Wolters Kluwer Deutschland GmbH, München, Neuwied.
- Kunter, M. & Voss, T. (2011). Professionelle Kompetenz von Lehrkräften. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.), (S. S.85-113). Münster u.a.: Waxmann.
- Lienert, G. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Psychologie Verlags Union, Weinheim.
- Lindemann, H. (1989). Strukturen von Lern- und Unterprozessen im experimentellen Chemieunterricht. *Naturwissenschaften im Unterricht.*, 41, 20-27.
- Lipowsky, F. (2009). Pädagogische Psychologie. In E. Wild & J. Möller (Hrsg.), (S. S.73-101). Springer Verlag.
- Magliaro, S. G., Lockee, B. B. & Burton, J. K. (2005). Direct Instruction Revisited: A Key Model for Instructional Technology. *ETR&D*, 53 No.4, 41-55.
- Marzano, R. J., Gaddy, B. B. & Dean, C. (2000). *What works in classroom instruction* (C. Aurora, Hrsg.). Mid-continent Research for Education and Learning (McREL).
- Maurer, C. & Rincke, K. (2013). Zielgerichtetes Experimentieren. In Sascha (Hrsg.), *Inquiry-based learning - forschendes lernen* (S. 119-121). Kiel: IPN-Verlag.
- Mayer, R. E. (2004). Should There Be a Three-Strikes Rule Against Pure Discovery Learning? *American Psychologist*, 59 No.1, 14-19.
- Meschede, N. (2014). *Professionelle Wahrnehmung der inhaltlichen Strukturierung im naturwissenschaftlichen Grundschulunterricht. Theoretische Beschreibung und empirische Erfassung*. Logos Verlag Berlin GmbH.

- Meyer, H. (2004). *Was ist guter Unterricht*. Cornelsen Verlag.
- Moosbrugger, H. & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion*. Springer Medizin Verlag, Heidelberg.
- Moosbrugger, H. & Rieß, S. (2010). Handbuch Statistik, Methoden und Evaluation. In H. Holling & B. Schmitz (Hrsg.), (S. 439-457). Hogrefe Verlag GmbH & Co. KG, Göttingen.
- Ohle, A., Fischer, H. E. & Kauertz, A. (2011). Der Einfluss des physikalischen Fachwissens von Primarstufenlehrkräften auf Unterrichtsgestaltung und Schülerleistung. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 357-389.
- Oser, F. & Baeriswyl, F. (2001). *Choreographies of teaching: Bridging instruction to learning*. Richardson, Virginia.
- Oser, F., Patry, J.-L., Elsässer, T., Sarasin, S. & Wagner, B. (1997). *Choreographien unterrichtlichen Lernens* (F. Oser, Hrsg.). Pädagogisches Institut der Universität Freiburg.
- Petzoldt, T. (2014). *Data Analysis with R - Selected Topics and Examples*. Technische Universität Dresden. Zugriff auf http://www.simecol.de/courses/elements_de.pdf
- Prenzel, M., Schöps, K., Rönnebeck, S., Senkbeil, M., Walter, O., Carstensen, C. H. & Hammann, M. (2007). Pisa 06 Die Ergebnisse der dritten internationalen Vergleichsstudie. In M. Prenzel et al. (Hrsg.), (S. 63-106). Waxmann.
- Rakoczy, K., Klieme, E., Drollinger-Vetter, B., Lipowsky, F., Pauli, C. & Reusser, K. (2007). *Studies on the educational quality of schools* (M. Prenzel, Hrsg.). Waxmann Verlag GmbH, Münster.
- Rakoczy, K., Klieme, E., Lipowsky, F. & Drollinger-Vetter, B. (2010). Strukturierung, kognitive Aktivität und Leistungsentwicklung im Mathematikunterricht. *Unterrichtswissenschaft*, 38, 229-246.
- Rasch, B., Frieze, M., Hofmann, W. & Naumann, E. (2006a). *Quantitative Methoden Band 1 Einführung in die Statistik*. Springer Medizin Verlag Heidelberg.
- Rasch, B., Frieze, M., Hofmann, W. & Naumann, E. (2006b). *Quantitative Methoden Band 2 Einführung in die Statistik*. Springer Medizin Verlag Heidelberg.
- Reyer, T. (2004). *Oberflächenmerkmale und Tiefenstrukturen im Unterricht* (H. N. und H. Fischler, Hrsg.). Logos Verlag Berlin GmbH.
- Rincke, K. (2015). Physikunterricht: Die Qualität wird in der Tiefe entschieden - Oberflächlich sichtbare Unterrichtsmerkmale geben wenig Auskunft über Unterrichtsqualität. *Blick in die Wissenschaft*, 32, 16-22.
- Rost, D. (2013). *Interpretation und Bewertung pädagogisch-psychologischer Studien*. Verlag Julius Klinkhardt, Bad Heilbrunn.
- Scheerens, J. (1992). *Effective Schooling. Research, Theory and Practice*. London: Cassell.
- Schmidkunz, H. & Lindemann, H. (1992). *Das Forschend-entwickelnde Unterrichtsverfahren - Problemlösen im naturwissenschaftlichen Unterricht* (Bd. 6). Westarp Wissenschaften.

- Schnotz, W. (2006). *Pädagogische Psychologie*. Weinheim: Beltz PVU.
- Seidel, T., Prenzel, M., Rimmele, R., Dalehefte, I. M., Herweg, C., Kobarg, M. & Schwindt, K. (2006). Blicke auf den Physikunterricht. Ergebnisse der IPN Videostudie. *Zeitschrift für Pädagogik*, 52 No.6, 799-821.
- Seidel, T., Rimmele, R. & Dalehefte, M. (2003). Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“. In T. Seidel, M. Prenzel, R. Duit & M. Lehrke (Hrsg.), (S. 317-388). IPN Leibniz-Institut für die Pädagogik der Naturwissenschaften.
- Seidel, T. & Shavelson, R. J. (2007). Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results. *Review of Educational Research*, 77 (4), 454-499.
- Seiwert, L. J. (1984). *Das 1 x 1 des Zeitmanagement*. GABAL Verlag GmbH, Speyer.
- Snijders, T. & Bosker, R. (2012). *Multilevel Analysis - An Introduction to Basis and Advanced Multilevel Modeling* (2. Aufl.). SAGE Publications. London.
- Sommer, K. (2014, Juni). *Persönliche Mitteilung*.
- Thalheimer, W. & Cook, S. (2002, August). *How to calculate effect sizes from published research: A simplified methodology*. Zugriff auf http://work-learning.com/effect_sizes.htm (Erhalten im Juni 2015)
- Trendel, G., Wackermann, R. & Fischer, H. E. (2007). Lernprozessorientierte Lehrerfortbildung in Physik. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 9-31.
- Trepke, C., Seidel, T. & Dalehefte, I. M. (2003). Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“. In T. Seidel, M. Prenzel, R. Duit & M. Lehrke (Hrsg.), (S. 201-228). IPN Leibniz-Institut für die Pädagogik der Naturwissenschaften.
- Tymms, P. (2004). But what does it mean? The use of effect sizes in educational research. In I. Schagan & K. Elliot (Hrsg.), (S. 55-66). National Foundation for Educational Research.
- van Joolingen, W. (1999). Cognitive tools for discovery learning. *International Journal of Artificial Intelligence in Education*, 10, 385-397.
- Wackermann, R. (2008). *Überprüfung der Wirksamkeit eines Basismodell-Trainings für Physiklehrer* (H. Niederer, H. Fischler & E. Sumfleth, Hrsg.). Logos Verlag Berlin GmbH.
- Wackermann, R., Hater, Simic, Pieper & Priemer, B. (2012, Juni). *Anleitung zur Videoanalyse: Bochumer Tiefenstrukturmodell zum Experimentieren im Physikunterricht*. (Didaktik der Physik, Ruhr-Universität Bochum)
- Wackermann, R. & Priemer, B. (2012). Tiefenstrukturen im Physikunterricht mit Schülerexperimenten. In G. für Didaktik der Chemie und Physik (Hrsg.), *Inquiry-based Learning – Forschendes Lernen* (S. 218-220). Sascha Bernholt, Jahrestagung in Hannover 2012. Kiel: IPN.
- Wackermann, R., Trendel, G. & Fischer, H. E. (2010). Evaluation of a Theory of Instructional Sequences for Physics Instruction. *International Journal of Science Education*, No. 7, 963-985.

- Wagner, B. (1999). *Lernen aus Sicht der Lernenden* (Bd. 780). Peter Lang. Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Wien.
- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1993). Toward a Knowledge Base for School Learning. *Review of Educational Research*, 63 (3), 249-294.
- Wellenreuther, M. (2005). *Lehren und Lernen - aber wie? Empirisch-experimentelle Forschungen zum Lehren und Lernen im Unterricht*. 2. korrigierte und überarbeitete Auflage. Baltmannsweiler: Schneider-Verl. Hohengehren.
- Wiater, W. (2012). *Unterrichtsprinzipien*. Auer Verlag.
- Zander, S., Krabbe, H. & Fischer, H. E. (2013). Lernzuwächse in Mechanik im Rahmen der Lehrerfortbildung „Sequenzierung von Lernprozessen“. In S. Bernholt (Hrsg.), *Inquiry-based Learning -Forschendes Lernen* (Bd. 33, S. 503-505). Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Hannover 2012. Kiel: IPN.
- Zander, S., Krabbe, H. & Fischer, H. E. (2015). Guter Physikunterricht für schwächere Schülerinnen und Schüler. In S. Bernholt (Hrsg.), *Heterogenität und Diversität - Vielfalt der Voraussetzungen im naturwissenschaftlichen Unterricht* (Bd. 35, S. 390-392). Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Bremen 2014. Kiel: IPN.