

Rosa Hettmannsperger

Lernen mit multiplen Repräsentationen aus Experimenten

Ein Beitrag zum Verstehen
physikalischer Konzepte



Springer VS

Lernen mit multiplen Repräsentationen aus Experimenten

Rosa Hettmannsperger

Lernen mit multiplen Repräsentationen aus Experimenten

Ein Beitrag zum Verstehen
physikalischer Konzepte

Mit einem Geleitwort von Prof. Dr. Wolfgang Schnotz
und Prof. Dr. Andreas Müller

Rosa Hettmannsperger
Forschungswerkstatt
Bildungswissenschaften
PH Ludwigsburg
Ludwigsburg, Deutschland

Dissertation Universität Koblenz-Landau, Campus Landau Fachbereich 8, Psychologie, 2013, u.d.T. Lernen mit multiplen Repräsentationen aus Experimenten: Ein Beitrag zum Verstehen physikalischer Konzepte. Empirische Untersuchungen zu den Effekten kognitiv aktivierender Aufgaben in der Strahlenoptik

Erstgutachter: Professor Dr. Wolfgang Schnotz
Zweitgutachter: Professor Dr. Andreas Müller
Prüfungsvorsitzender: Professor Dr. Manfred Schmitt

Tag der Prüfung: 13.12.2013

OnlinePLUS Material to this book can be downloaded from
<http://www.springer>

ISBN 978-3-658-07435-7

ISBN 978-3-658-07436-4 (eBook)

DOI 10.1007/978-3-658-07436-4

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer VS

© Springer Fachmedien Wiesbaden 2015

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer VS ist eine Marke von Springer DE. Springer DE ist Teil der Fachverlagsgruppe Springer Science+Business Media.
www.springer-vs.de

Geleitwort

Die vorliegende Arbeit befindet sich an der Schnittstelle zweier zentraler Forschungsthemen der Instruktionspsychologie und der Didaktik der Naturwissenschaften: der Förderung von Konzeptwechsel („conceptual change“) und der Verwendung multipler Repräsentationen zur Förderung von Prozessen des Verstehens einschließlich Konzeptwechsels. Experimente gehören zu den wichtigsten Mitteln zur Initiierung von Konzeptwechsel im naturwissenschaftlichen Unterricht, da in ihnen Prä- und Fehlkonzepte („misconceptions“) von Lernenden in augenfälliger Weise in einen Widerspruch zur Erfahrung gebracht werden können. Allerdings zeigt die Forschung, dass Konzeptwechsel auf der Basis von Unterrichtsansätzen dieser Art keineswegs garantiert ist. Somit stellt sich die Frage nach den fördernden und hemmenden Einflüssen auf Prozesse des Konzeptwechsels anhand von naturwissenschaftlichen Experimenten.

Gleichzeitig zeigt die aktuelle naturwissenschaftsdidaktische Forschung, dass für ein angemessenes Verständnis von Experimenten und der ihnen zugrundeliegenden Konzepte die Nutzung unterschiedlicher kohärenter Repräsentationsformen wesentlich ist. Beim Durchführen und Verstehen von Experimenten müssen Daten von physikalischen Phänomenen mittels unterschiedlicher externer Repräsentationen (wie Text, realistische, schematische und logische Bilder oder Formeln) dargestellt werden, um sie für individuelle und kooperative Denk- und Lernprozesse verfügbar zu machen. Die verschiedenen Repräsentationsformen unterscheiden sich in ihrer Ausdrucksmächtigkeit für verschiedene Zwecke. Verständnis und damit Konzeptwechsel durch Experimente beruht also in besonderem Grad darauf, dass Lernende in der Lage sind, die für bestimmte Zusammenhänge die am besten geeigneten Darstellungsformen zu finden, adäquat zu nutzen, und Kohärenz der verschiedenen Repräsentationsformen herzustellen.

An dieser Stelle setzt die Arbeit von Frau Hettmannsperger an, indem sie am Beispiel des Lernens im Bereich der Strahlenoptik die Maßnahmen zur Stimulation von Konzeptwechsel in Verbindung mit der Nutzung von multiplen Repräsentationen untersucht. Die Autorin betritt in mehrfacher Hinsicht wissenschaftliches Neuland. Neben der Entwicklung von kognitiv aktivierenden Lernaufgaben mit hoher curricularer Validität, welche die gezielte Verwendung

von multiplen Repräsentationen anregen sollen und die sorgfältig mit psychometrischen Mitteln überprüft werden, entwickelt Frau Hettmannsperger einen Konzept-Test zur Strahlenoptik, der Präkonzepte und insbesondere Fehlkonzepte in dieser Domäne diagnostiziert und dabei ausdrücklich auf physikalische Fachterminologie und die Verwendung von Formeln verzichtet, da diese üblicherweise erst nach dem Unterricht verstanden werden können. Damit ist dieser Konzept-Test zur Prüfung von individuellen Lernvoraussetzungen in Untersuchungen mit Prä-Post-Design geeignet und aus fachdidaktischer Sicht besonders deshalb sehr bedeutsam, weil die bisherige Forschung in diesem Gebiet eine hohe Zahl von Fehlkonzepten festgestellt hat.

Frau Hettmannsperger zeigt, dass die entwickelten Aufgaben das Lernen aus Experimenten mit multiplen Repräsentationen wirksam fördern und dabei auch Prozesse des Konzeptlernens und Konzeptwechsels im gleichen Maße unterstützt wie ein Unterricht, der nur auf Konzeptwechsel orientiert ist. Dabei ist hervorzuheben, dass diese Förderung unter curricular validen Randbedingungen analysiert wird und unterrichtspraktische Erwägungen mit einschließt. Die vorliegende Arbeit kann für zukünftige Untersuchungen auf diesem Gebiet als vorbildlich gelten. Ihr durchgängig sehr hohes methodisches Niveau machen sie zu einer außerordentlich wertvollen Lektüre sowohl für Forschungsinteressierte als auch für diejenigen, denen vor allem an der Verbesserung der Unterrichtspraxis gelegen ist.

Schließlich hat die Arbeit von Frau Hettmannsperger noch ein besonderes, auch für andere Forschungsprojekte der Lehr-Lern-Forschung interessantes Charakteristikum vorzuweisen, nämlich die interdisziplinäre Kooperation in einem „Projekt-Tandem“ von Pädagogischer Psychologie (Rosa Hettmannsperger) und Fachdidaktik (Jochen Scheid) im Rahmen der DFG-Graduiertenkollegs Unterrichtsprozesse¹. Erklärtes Ziel dabei ist, durch komplementäre und synergistische Bearbeitung der Forschungsfragen einen sichtbaren Mehrwert zu erzielen. Auch hierfür ist diese Arbeit ein hervorragendes Beispiel, von einem Austausch bei der Nutzung der Datensätze (z.B. Konzepttest, curriculare Validierung), über ihre Erarbeitung von „pedagogical content knowledge“ bis hin zu der durchgängigen Verbindung von Unterrichts- und Forschungsperspektive. Mit Blick auf die intensive aktuelle Diskussion einer Verbindung von Anwendungs- und Grundlagen-orientierter Forschung insbesondere der Pädagogischen Psychologie und der Fachdidaktiken liefert die geleistete Arbeit ein konkretes, auch für andere

1 Universität Koblenz-Landau, Campus Landau, <http://www.upgrade.uni-landau.de>; DFG GK 1561

Projekte dieser Art stimulierendes Vorbild. Der hier als Buch vorgelegten Dissertationsschrift von Rosa Hettmannsperger sind deshalb viele aufmerksame, theoretisch wie unterrichtspraktisch interessierte Leser zu wünschen, die Theorie, Methodik und Ergebnisse hinein in die zukünftige Forschung, Entwicklung und Praxis in dem Gebiet tragen werden.

Andreas Müller (Genf) und Wolfgang Schnotz (Landau)

*Meinen Eltern und
meinem Mann*

*Quand les vérités sont évidentes et absolument contradictoires,
tu ne peux rien, sinon changer ton langage.*

Antoine de Saint-Exupéry, La Citadelle

Danksagung

Bei der Arbeit an meiner Dissertation wurde ich beruflich wie privat auf vielfältige Weise von anderen unterstützt.

An erster Stelle möchte ich mich bei Professor Schnotz und Professor Müller für die Möglichkeit bedanken, im DFG-Graduiertenkolleg „Unterrichtsprozesse“ zu promovieren. Herzlichen Dank für die Betreuung meiner Arbeit und Ihre fachlichen Ratschläge während dieser Zeit; insbesondere Professor Müller danke ich nochmals ganz besonders für die engagierte Betreuung bei der Entwicklung der Unterrichtsmaterialien und Testinstrumente.

Mein uneingeschränkter Dank gilt meinen Eltern Ulrike und Reinhold Hettmannsperger sowie meinem Mann Denis Lippolt, die stets für mich da waren. Ich danke Euch von ganzem Herzen für Euren Rat, Eure Geduld und den Rückhalt während der arbeitsintensiven Phase der Promotion.

Ein weiterer großer Dank gebührt allen meinen ehemaligen Kollegen des Graduiertenkollegs für die kollegiale Zusammenarbeit und die vielen Stunden im Büro, in denen wir mitunter bis (spät) abends gearbeitet und nicht selten herzlich zusammen gelacht haben. Meinem Projektpartner Jochen Scheid danke ich besonders für die stets gute kollegiale Zusammenarbeit.

Mein besonderer Dank gilt meinem Mann, Kathrin Bürger, Sibel Telli, Loredana Mihalca und Rahel Grüninger für fachliche Diskussionen.

Vielen Dank auch an die Lehrerinnen und Lehrer, welche die Studie bereitwillig im Unterricht umgesetzt haben.

Zu danken habe ich ebenfalls den studentischen Hilfskräften, die mich bei der Durchführung des Projekts unterstützt haben. Zu nennen sind hier vor allem Daniel Tischer, Jasmin Beyle und Christina Fianke. Mein ganz besonderer Dank gilt Daniel Tischer, der über zwei Jahre kompetent und zuverlässig an der Dateneingabe und Testauswertung mitgearbeitet hat.

Meiner Mutter und Lilli Lippolt danke ich für das konzentrierte Korrekturlesen der Dissertationsschrift. Nicht zuletzt möchte ich Christina Schiefer, meinem Mann und meinem Vater dafür danken, dass sie mir im Herbst 2009 die Zuversicht schenkten, fachliches Neuland zu erkunden.

Abstract

Students' naïve concepts of physics are a well-known reason for learning difficulties in science education research. Naïve concepts, either stemming from everyday experience or developing in lessons at school, often differ from scientific explanations. Within learning processes, they interact continuously with students' understanding (internal representations) of scientific schemata and concepts taught in lessons. Students' internal representations and mental models influence the external representations they create while solving tasks. Especially in physics, problem-solving regularly requires learners to find an appropriate and useful representation of a given problem or task. To solve the task, they then continue to operate on the representation they have found.

This work aims at investigating the mutual relations between widespread naïve students' concepts, external representations, and the construction of mental models in the context of learning from physics experiments in middle-school classrooms. For this reason, it refers to two recent strands of research in science education: fostering representational competence via learning with multiple representations and conceptual change. Considering the theoretical implications and related empirical results from these strands of research, instructions were developed that aimed at supporting students both to create scientifically appropriate representations and successfully operate on self-developed and given external representations.

To evaluate the effectiveness of the developed instructions, two quasi-experimental conditions were implemented that differed with regard to the extent of cognitive activation while dealing with (multiple) representations in ray optics. $N = 10$ teachers from ten different schools and $N = 21$ classes participated in the study. The analyses were based on data from $N = 443$ students attending German "Gymnasium" (secondary-school track for high achievers) or comprehensive schools.

Students in the treatment condition worked on cognitively activating tasks that required them to overcome widespread naïve students' concepts in ray optics. Students in the control condition worked on similar representations and were encouraged to overcome the same widespread naïve students' concepts. However, the control group was not explicitly activated to create own representations, to re-

think their representations, or to operate on them. Multilevel analyses investigated whether significant effects could be ascribed to the treatment.

The results indicate that the use of challenging learning strategies (cognitively activating tasks) while dealing with (multiple) representations leads to a significant but small improvement of physics achievement in a test immediately after the treatment for students at a German “Gymnasium”. However, there were not any significant effects in the learning increase of conceptual understanding or in the development of students’ motivation when comparing the treatment and control groups.

Furthermore, the results referring to the learning increase in conceptual understanding were compared with results from a related study (Scheid, 2013) from the same research project. The outlined comparison revealed that representational tasks addressing widespread naïve students’ concepts lead to a significant improvement of conceptual understanding for students from both types of schools even two months after the intervention.

Zusammenfassung

Eine zentrale Ursache für Lernschwierigkeiten im Physikunterricht besteht darin, dass Schüler in den Fachunterricht naive Vorstellungen mitbringen oder auch erst im Unterricht bilden, welche häufig von etablierten wissenschaftlichen Erklärungen abweichen. Diese naiven Schülervorstellungen interagieren mit den zu erlernenden wissenschaftlichen Konzepten im Physikunterricht. Die internen Vorstellungen und mentalen Modelle der Lernenden schlagen sich dabei auch in den externen Repräsentationen nieder, welche die Lernenden bei der Aufgabebearbeitung erstellen. Insbesondere in der Physik erfordert Problemlösen, eine adäquate und zweckmäßige physikalische Repräsentation zu finden, auf die anschließend zur Aufgabebearbeitung weitere Operationen angewandt werden sollen.

Diese Arbeit zielt darauf, die Wechselbeziehung von weitverbreiteten naiven Schülervorstellungen, externen Repräsentationen und der Konstruktion interner mentaler Modelle im Kontext des Lernens aus physikalischen Experimenten in der Mittelstufe in den Blick zu nehmen. Sie referiert somit auf empirische Ergebnisse von zwei jüngeren Forschungstraditionen im Bereich des naturwissenschaftlichen Denkens: die Förderung von Repräsentationskompetenz durch das Lernen mit multiplen Repräsentationen und das Erzielen von Wissensveränderung, auch bekannt als Konzeptwechsel.

Vor dem Hintergrund der genannten Forschungsrichtungen wurden kognitiv aktivierende Instruktionen entwickelt, die weitverbreitete naive Schülervorstellungen in der Strahlenoptik berücksichtigen und Lernende darin unterstützen sollen, wissenschaftlich zweckmäßigere Repräsentationen in der Strahlenoptik zu bilden sowie kompetent mit gegebenen und selbst erstellten externen Repräsentationen zu operieren.

Zur Untersuchung der Lernwirksamkeit der entwickelten Aufgaben wurden zwei Versuchsbedingungen entwickelt, die sich im Ausmaß der kognitiven Aktivierung im Umgang mit (multiplen) Repräsentationen unterscheiden. An der Studie nahmen $N = 10$ Lehrkräfte aus zehn unterschiedlichen Schulen von $N = 21$ Parallelklassen an Gymnasien und Gesamtschulen teil. Die Daten von $N = 443$ Schülern gingen in die Analyse ein. Die Treatmentbedingung beinhaltete kognitiv

aktivierende Aufgabenstellungen, welche die Überwindung von weitverbreiteten Schülervorstellungen in der Strahlenoptik erfordern. In der Kontrollbedingung wurden die gleichen naiven Schülervorstellungen thematisiert und auch vergleichbare Repräsentationsformen verwendet. Die Kontrollgruppe wurde jedoch nicht gezielt kognitiv aktiviert, eigene Repräsentationen zu erstellen, zu reflektieren und mit gegebenen und selbst erstellten Repräsentationen zu operieren. Mittels Mehrebenenanalysen wurde untersucht, ob Effekte auf das Treatment zurückgeführt werden können.

Die Ergebnisse weisen darauf hin, dass der Einsatz anspruchsvoller Lernstrategien beim Umgang mit (multiplen) Repräsentationen (kognitive Aktivierung) zu einer signifikanten, wenn auch geringen Verbesserung der Physikleistung bei Gymnasiasten in einem auf den Unterricht folgenden Test führt, nicht jedoch zu einem höheren Lernzuwachs bezüglich des konzeptuellen Verständnisses oder zu einer Steigerung der Motivation.

Ein Vergleich mit einer inhaltlich verwandten Studie (Scheid, 2013) aus dem gleichen Forschungsprojekt belegt darüber hinaus, dass repräsentationsbezogene Aufgaben, welche naive Schülervorstellungen thematisieren das konzeptuelle Verständnis der Schüler nachweislich auch zwei Monate signifikant verbessern.

Inhaltsverzeichnis

Einleitung und Problemaufriss	23
Zielsetzung und Vorgehen	29
1 Empirische und theoretische Grundlagen	31
1.1 Die Rolle von Repräsentationen in kognitiven Lernprozessen	31
1.1.1 Repräsentationen, Repräsentationsformen und Repräsentationskompetenz	31
1.1.2 Gedächtnispsychologische Grundlagen	36
1.1.2.1 Gedächtnissysteme	36
1.1.2.2 Die Cognitive-Load-Theorie	39
1.1.2.3 Das Arbeitsgedächtnis nach Baddeley	40
1.1.2.4 Mentale Modelle	41
1.1.3 Modelle der kognitiven Verarbeitung von Texten, Bildern und Sprache.....	43
1.1.3.1 Die duale Codierungstheorie von Paivio.....	43
1.1.3.2 Die kognitive Theorie des multimedialen Lernens	44
1.1.3.3 Das integrative Rahmenmodell des Text- und Bildverstehens	46
1.1.3.4 Pädagogische Implikationen	49
1.1.4 Lernen mit multiplen Repräsentationen	51
1.1.4.1 Funktionen von multiplen Repräsentationen.....	51
1.1.4.2 Empirische Befunde aus Experten-Novizen- Vergleichen	54
1.1.4.3 Repräsentation der Ausgangslage, Problemlöseprozesse und Wissenserwerb.....	59
1.1.4.4 Fazit und Ausblick auf die Bedeutung des Vorverständnisses	64

1.2 Die Rolle von Konzeptwechsln in Lehr-Lernprozessen	65
1.2.1 Vorwissen, Schülervorstellungen und Konzeptwechsel.....	65
1.2.2 Kognitionspsychologische Theorien zum Konzeptwechsel.....	67
1.2.2.1 Konzeptwechsel durch die Überwindung kognitiver Konflikte	67
1.2.2.2 Konzeptwechsel durch die Modifikation mentaler Modelle.....	71
1.2.2.3 Konzeptwechsel durch die Überwindung von Kategorisierungsfehlern	74
1.2.2.4 Konzeptwechsel durch die Reorganisation von Wissensfragmenten	76
1.2.2.5 Fazit zu den Konzeptwechseltheorien und der praktischen Förderung	78
1.3 Interaktionen: Lernen aus Experimenten unter der Perspektive des Umgangs mit multiplen Repräsentationen zur Förderung des konzeptuellen Verständnisses	79
1.3.1 Empirische Befunde: Multiple Repräsentationen und Konzeptwechsel	79
1.3.2 Fazit und weiterer Forschungsbedarf	86
1.3.3 Kognitive Aktivierung	87
1.3.4 Lernen aus Experimenten	95
1.4 Integrierte Betrachtung: Physiklernen mit multiplen Repräsentationen aus Experimenten zur Förderung des konzeptuellen Verständnisses.....	103
 2 Empirische Untersuchungen.....	107
 2.1 Theoretische Vorüberlegungen zum Lerngegenstand	107
2.1.1 Gründe für die Wahl des Lerngegenstands	107
2.1.2 Schülervorstellungen im Kontext der Bildentstehung bei der Sammellinse.....	112
2.2 Pilotstudie	122
2.2.1 Zielsetzung der Pilotstudie	122
2.2.2 Fragestellung und Hypothesen.....	123
2.2.3 Stichprobe und Design	123
2.2.4 Durchführung und Unterrichtsmaterial	124
2.2.5 Variablen und Erhebungsinstrumente.....	126

2.2.5.1	Leistungstest.....	126
2.2.5.2	Konzepttest.....	129
2.2.6	Auswertung und Ergebnisse	131
2.2.6.1	Itemstatistiken zum Leistungstest	131
2.2.6.2	Itemstatistiken zum Konzepttest	132
2.2.6.3	Exkurs: Berechnung einer polychorischen Korrelationsmatrix für ordinalskalierte Daten.....	133
2.2.6.4	Erste Hypothese: Veränderungen durch die Intervention	135
2.2.6.5	Zweite Hypothese: Wirkung des Treatments auf die Physikleistung	137
2.2.6.6	Dritte Hypothese: Wirkung des Treatments auf das konzeptuelle Verständnis	139
2.2.7	Diskussion der Pilotstudie und Konsequenzen für die Hauptstudie	141
2.3	Hauptstudie	144
2.3.1	Forschungsfragen und Hypothesen.....	144
2.3.2	Stichprobe und Design.....	147
2.3.3	Operationalisierung des Treatments	150
2.3.3.1	Genereller Überblick.....	150
2.3.3.2	Operationalisierung des Treatments in der ersten Unterrichtsstunde.....	154
2.3.3.3	Treatmentvariation in der ersten Unterrichtsstunde ..	155
2.3.3.4	Operationalisierung des Treatments in der zweiten Unterrichtsstunde	156
2.3.3.5	Operationalisierung des Treatments in der dritten Unterrichtsstunde.....	159
2.3.3.6	Operationalisierung des Treatments in der vierten Unterrichtsstunde	162
2.3.3.7	Treatmentvariation in der vierten Unterrichtsstunde .	164
2.3.3.8	Operationalisierung des Treatments in der fünften Unterrichtsstunde	167
2.3.3.9	Operationalisierung des Treatments in der sechsten Unterrichtsstunde.....	172
2.3.4	Variablen und Erhebungsinstrumente.....	175
2.3.4.1	Überblick.....	175
2.3.4.2	Leistungstest.....	176
2.3.4.3	Konzepttest.....	179

2.3.4.4	Motivationsfragebogen	181
2.3.4.5	Erfassung der Kovariaten	182
2.3.4.6	Einflussfaktoren bei der Anwendung des Lehrmaterials	185
2.3.5	Ergebnisse zur Messung der abhängigen Variablen	186
2.3.5.1	Itemstatistiken zum Leistungstest	186
2.3.5.2	Faktorenanalyse des Leistungstests	188
2.3.5.3	Itemstatistiken zum Konzepttest	195
2.3.5.4	Raschanalyse des Konzepttests	202
2.3.5.5	Kreuzvalidierung des Konzepttests	207
2.3.5.6	Itemstatistiken zum Motivationsfragebogen	215
2.3.5.7	Diskussion der Ergebnisse	216
2.3.5.8	Fazit zur ersten Zielsetzung: Entwicklung eines tragfähigen Konzepttests	219
2.3.6	Ergebnisse zu den Kovariaten	221
2.3.6.1	Ergebnisse zu kognitiven Fähigkeiten	221
2.3.6.2	Ergebnisse zu vorherigen Schulleistungen	226
2.3.6.3	Zusammenhang zwischen Schulleistungen und kognitiven Fähigkeiten	227
2.3.6.4	Diskussion der Analysen zu den Kovariaten	228
2.3.7	Einflussfaktoren bei der Anwendung des Lehrmaterials	229
2.3.7.1	Wahrgenommenes Lehrerengagement aus Schülersicht	229
2.3.7.2	Auswertung des Lehrernotizhefts	233
2.3.7.3	Diskussion zu den Einflussfaktoren bei der Anwendung des Lehrmaterials	237
2.3.8	Einblick in die Lernprozesse beim Umgang mit Repräsentationen	238
2.3.8.1	Auswertung der Arbeitsblätter 4 und 7 zur Bildkonstruktion	238
2.3.8.2	Diskussion der Ergebnisse zur Analyse der Lernprozesse	244
2.3.9	Ergebnisse zu den untersuchten Hypothesen	245
2.3.9.1	Methodisches Vorgehen: Mehrebenenanalyse	245
2.3.9.2	Methodisches Vorgehen: Modellentwicklung	250
2.3.9.3	Erste Hypothese: Wirkung des Treatments auf die Physikleistung	274

2.3.9.4	Zweite Hypothese: Wirkung des Treatments auf das konzeptuelle Verständnis	285
2.3.9.5	Dritte Hypothese: Wirkung der Treatmentvariation auf das konzeptuelle Verständnis	292
2.3.9.6	Vierte Hypothese: Wirkung des Treatments auf die Schülermotivation.....	299
2.3.9.7	Fünfte Hypothese: Vergleich inhaltlich aufeinander bezogener Studien	306
2.3.9.8	Vertiefende Analysen zu Aptitude-Treatment-Interaktionen	315
2.3.9.9	Zusammenfassung der Ergebnisse zur Untersuchung der Hypothesen 1-5	318
2.3.9.10	Diskussion der Ergebnisse zur Untersuchung der Hypothesen 1-5.....	320
3	Gesamtdiskussion	325
3.1	Integration der Befunde und Kritik am Vorgehen.....	325
3.2	Ausblick	327
3.3	Handlungsempfehlungen für die Gestaltung von Lernumgebungen.....	330
	Abkürzungen	335
	Tabellenverzeichnis.....	339
	Abbildungsverzeichnis.....	343
	Literaturverzeichnis	349
	Verzeichnisse für den Anhang.....	361

Einleitung und Problemaufriss

Jede Form des Lernens basiert auf dem Umgang mit Sprache, Bildern oder auf der Auseinandersetzung mit konkreten Gegenständen aus der Umwelt. Auch bevor sich die Kognitive und Pädagogische Psychologie, Bildungswissenschaft und Didaktik wissenschaftlich mit dem Thema Lernen auseinandergesetzt haben, verständigten sich Lehrende und Lernende mittels Sprache, Symbolen oder Bildern und stellten sich Sachverhalte mental vor. Worte, Bilder und Symbole teilen die Gemeinsamkeit, für einen anderen Gegenstand oder ein Ereignis zu stehen, den sie „re-präsentieren“. Komplexe Sachverhalte zu begreifen, erfordert in der Regel unterschiedliche Repräsentationsformen (z.B. Texte und Diagramme) mit äquivalenter oder auch mit komplementärer Bedeutung zu koordinieren und ineinander zu übersetzen. Diese Fähigkeit, Phänomene auf unterschiedliche Weise zu repräsentieren und zwischen verschiedenen Repräsentationsformen hin- und her zu wechseln (vgl. Dolin, 2007, S. 77) wird auch als Repräsentationskompetenz bezeichnet.

Lernen in den Naturwissenschaften erfordert diese Kompetenz in besonderem Maße. Ergebnisse der Expertiseforschung belegen, dass Experten und Novizen im Bereich der Physik und Chemie nicht nur qualitativ unterschiedliche Problemrepräsentationen entwickeln, sondern auch in stärkerem Maß unterschiedliche Repräsentationsformen aufeinander beziehen (vgl. Larkin, 1983; Larkin & Simon, 1987; Kozma & Russel, 1997; Plötzner & Spada, 1998; Savelsbergh, Ferguson-Hessler & de Jong, 1997; Tabachneck, Leonardo & Simon, 1994).

Der kommunikative Aspekt des Austauschs zwischen Lehrendem und Lernendem mittels kulturell¹ oder auch fachspezifisch gebundenen² Darstellungsformen (Sprache, Symbole, bildliche Darstellungen) zeigt auf, dass Lernen in sozialen Situationen stattfindet. Der Lernende wird in eine Expertenkultur eingeführt und erwirbt durch kompetente Interaktionspartner fachlich relevante und akzeptierte Sicht-, Denk- und Handlungsweisen gegenüber dem jeweiligen Lerngegenstand (vgl. Schnotz, 2006a, S. 123). Die Kooperation zwischen Individuen

1 z.B. verbale Darstellungen

2 z.B. mathematische Darstellungen

zeichnet sich dabei durch den Gebrauch materieller und kognitiver Werkzeuge aus (vgl. Wygotski 1986/1964, S. 44 f.). Schnotz (2006a) schreibt hierzu:

„Wenn ein Lehrender (Experte) das Handeln eines Lernenden (Novizen) durch Hilfestellungen unterstützt, so erfolgt die Handlungssteuerung auf zweierlei Weise:

- (1) durch Hilfen des Lehrenden, die das Bewusstsein des Lernenden durch Kommunikation, also interpsychisch – beeinflussen, und
- (2) durch Selbststeuerung des Lernenden, der die Handlung teilweise schon selbstständig – also intrapsychisch gesteuert – vollzieht“ (S. 44).

Beide Aspekte lassen sich auf den Umgang mit Repräsentationen beziehen:

Einerseits stellen die Lehrenden externe Repräsentationen in Form von Erklärungen, realistischen und schematischen Bildern zur Verfügung, andererseits bilden sich die Lernenden selbst interne und externe Repräsentationen zum Umgang mit konkreten Gegenständen (z.B. beim Experimentieren in Kleingruppen), verarbeiten die Repräsentationen, die ihnen im Unterricht zur Verfügung gestellt werden und fertigen eigene Erklärungen, Skizzen, Tabellen oder schematische Darstellungen an, wenn sie naturwissenschaftliche Probleme bearbeiten.

Kern der Selbsttätigkeit der Lernenden ist die kognitive Verarbeitung der Lehrinhalte in der Auseinandersetzung mit verschiedenen Medien sowie im Austausch mit Lehrenden und Mitschülern.

Die Förderung von Repräsentationskompetenz ist dabei an der Schnittstelle zwischen einer soziokonstruktivistischen Perspektive auf Lernprozesse, bei der Lernen als kulturell vermittelten Werkzeuggebrauch und als Enkulturation in eine Fachkultur (vgl. Wygotski, 1986/1964, S. 44 f.) verstanden wird, sowie einer kognitionspsychologischen Sicht auf die mentale Verarbeitung des Gelernten, zu verorten.

Da der kompetente Umgang mit Repräsentationen essentiell für das Verstehen und Betreiben von Naturwissenschaften ist, kann Repräsentationskompetenz als ein Kernelement naturwissenschaftlicher Kompetenz verstanden werden. „Naturwissenschaftliche Kompetenz wiederum ist eine zentrale Voraussetzung, um an einer von Naturwissenschaften und Technik geprägten Welt teilhaben zu können“ (Rönnebeck, Schöps, Prenzel, Mildner & Hochweber, 2010, S. 177). Internationale Vergleichsstudien wie TIMSS oder PISA (vgl. Baumert, et al., 2000; Rönnebeck et al., 2010) zeigen, dass es deutschen Schülern³ schwerfällt, das in der Schule

3 Wenn in dem folgenden Text von „Schülern“ und „Lehrern“ gesprochen wird, sind um der sprachlichen Einfachheit willen, Schülerinnen und Lehrerinnen selbstverständlich eingeschlossen.

erworbene Wissen auf Probleme aus alltäglichen, schulischen oder beruflichen Situationen anzuwenden. Die Förderung von Repräsentationskompetenz kann hier ein zentraler Baustein sein, welcher Schüler darin unterstützt, eine fundierte naturwissenschaftliche Grundbildung zu erwerben. Eine solche Grundbildung zeichnet sich dadurch aus, dass Schüler

- die „charakteristischen Eigenschaften sowie die Bedeutung der Naturwissenschaften in unserer heutigen Welt“ verstehen (zit. n. Prenzel, Carstensen, Frey, Drechsel & Rönnebeck, 2007, S. 39),
- naturwissenschaftliches Wissen anwenden können, um Fragestellungen zu erkennen, die mit Hilfe naturwissenschaftlicher Methoden untersucht und bearbeitet werden können,
- naturwissenschaftliche Phänomene beschreiben und erklären können,
- naturwissenschaftliche Evidenzen interpretieren können und aus Belegen Schlussfolgerungen ziehen zu können (vgl. ebd., S. 38).

Die Förderung von Repräsentationskompetenz greift dabei an den drei naturwissenschaftlichen Teilkompetenzen naturwissenschaftlicher Grundbildung an, die nach PISA (vgl. Rönnebeck et al., 2006, S. 69) wie folgt definiert sind:

1. naturwissenschaftliche Fragestellungen erkennen
2. naturwissenschaftliche Phänomene erklären
3. naturwissenschaftliche Evidenz nutzen

Für jede dieser drei Teilkompetenzen ist der Umgang mit Repräsentationen in unterschiedlichen Formaten (verbale Beschreibungen, Skizzen, Tabellen, schematische Abbildungen, Graphen und Formeln) zentral. Um Fragestellungen zu erkennen, die mit naturwissenschaftlichen Methoden untersucht werden können, müssen Schüler Schlüsselwörter identifizieren können, die sie mit naturwissenschaftlichen Informationen verbinden. Um Phänomene naturwissenschaftlich zu beschreiben oder Änderungen vorherzusagen, ist der Umgang mit Sprache, Symbolen aber auch mit Diagrammen und Tabellen erforderlich. Um naturwissenschaftliche Evidenz zu nutzen, müssen Schüler in der Lage sein, Schlussfolgerungen zu ziehen, was häufig die Konstruktion eines mentalen Modells, also einer geistigen Repräsentation des Sachverhaltes, erfordert. Unter einem mentalen Modell wird hierbei ein hypothetisches Konstrukt verstanden, das Gegebenheiten aus der Umwelt abbildet, wobei in dem Modell Informationen sowohl reduziert als auch elaboriert werden (vgl. Dutke, 1994, S. 76). Zudem müssen sie in der Lage

sein, ihre Schlussfolgerungen auch zu kommunizieren. Je nach Anforderung der Aufgabe ist hier der Umgang mit unterschiedlichen Repräsentationen (Grafiken, Tabellen, verbalen Beschreibungen oder Formeln) nötig.

Repräsentationskompetenz ist im dänischen Lehrplan neben empirischer Kompetenz⁴, Modellkompetenz⁵ und Perspektiven-Kompetenz⁶ als Zielkompetenz formuliert. Ein kritischer Einwand könnte nun sein, dass jedweder Wissensaustausch und jedweder Lerninhalt an irgendeine Darstellungsform, sei sie verbal oder bildlich, gebunden ist. Keineswegs trivial ist es jedoch, kompetent und überdies oft auch simultan mit verschiedenen Repräsentationen umzugehen. Dies wird insbesondere ersichtlich, wenn man sich bewusst macht, welche Anforderungen hiermit einhergehen, z.B. die Kompetenzen unterschiedliche Repräsentationsformen mit äquivalenter Bedeutung zu koordinieren und ineinander zu übersetzen.

Dass der kompetente Umgang mit verschiedenen Repräsentationsformen Lernanstrengung und Übung erfordert, zeigt sich auch an den Lernschwierigkeiten, die auftreten, wenn Schüler sich mit wissenschaftlichen Konzepten auseinandersetzen. Diese schlagen sich auch in den externen Repräsentationen nieder (vgl. Cox, 1999), welche die Lernenden erstellen. Solche Lernschwierigkeiten können das Lösen von Aufgaben behindern, da sie die Lernenden daran hindern relevante Lösungsstrategien abzurufen (vgl. Larkin, 1983; Ohlsson, 1992; Zhang, 1997). An den Lernschwierigkeiten der Schüler wird deutlich, dass die Lernenden wissenschaftliche Konzepte nicht in aller Konsequenz zu Ende denken und ihnen Inkonsistenzen zwischen ihren Vorstellungen und wissenschaftlichen Konzepten nicht auffallen (vgl. diSessa, 2008).

Eine zentrale Ursache für Lernschwierigkeiten im Physikunterricht besteht darin, dass Schüler, obgleich sie Novizen sind, bereits vor Beginn des Fachunterrichts über Vorstellungen verfügen, die oft auf ihren Alltagserfahrungen basieren (vgl. Wiesner, 1992a, 1992b; Duit, 1993; Langley, Ronen & Eylon, 1997). Diese Schülervorstellungen weichen häufig von den etablierten wissenschaftlichen Erklärungen ab (vgl. Vosniadou & Brewer, 1992, S. 536) und interagieren mit dem neuen Wissen, welches im Physikunterricht vermittelt wird. Auf diese Weise beeinflussen sie das Erlernen wissenschaftlicher Konzepte und können den Wissenserwerb nachhaltig behindern. Dabei zeigen sich Schülervorstellungen gegenüber

4 Fähigkeit zu beobachten, zu beschreiben, zu experimentieren und zu messen.

5 Fähigkeit Komplexität zu reduzieren, Kausalitäten zu bestimmen und verschiedene Modelle zu entwickeln sowie zu verwenden.

6 Fähigkeit die Rolle von Naturwissenschaften in der Gesellschaft zu reflektieren sowie wissenschaftliches Wissen in Relation zu anderem Wissen zu setzen.

Änderungen oft als widerständig (vgl. ebd., S. 536). Klassische Lehrstrategien erweisen sich oft als wenig effektiv darin, die Vorstellungen der Lernenden zu verändern, weil diese tief in den Alltagserfahrungen verankert sind. Lernprozesse, die eine Wissensveränderung beinhalten, sind daher in der Regel recht zeitaufwändig. Sie wurden im Rahmen von Theorien über Konzeptwechsel aus Perspektive der Entwicklungspsychologie (vgl. Piaget 1953/1979, 1973/1996; Carey, 1985), der Kognitionspsychologie (vgl. diSessa, 1988, 1993) und der Science Education Research seit über 30 Jahren erforscht (vgl. Wiesner, 1986; Wiesner, 1992a, 1992b, 1994; Reiner, Slotta, Chi & Resnick, 2000). Lernenden wissenschaftliche Theorien und Konzepte zu vermitteln, erfordert es, das Denken der Lernenden zu kennen und Instruktionen entsprechend anzupassen (vgl. Tyson, Venville, Harrison & Treagust, 1997; Özdemir & Clark, 2007). Instruktionen, die das Vorwissen der Schüler berücksichtigen, können Schüler darin unterstützen, wissenschaftlich zweckmäßigere Konzepte zu entwickeln. So konnten Guzzetti, Snyder, Glass und Gamas (1993) in einer Metaanalyse zeigen, dass Vermittlungsstrategien, in denen Konflikte zwischen den vorunterrichtlichen Vorstellungen und den zu lernenden wissenschaftlichen Vorstellungen herausgearbeitet worden sind, solchen Strategien statistisch signifikant überlegen waren, in denen das nicht der Fall war.

Zielsetzung und Vorgehen

Diese Arbeit zielt darauf, die Wechselbeziehung von (naiven) Schülervorstellungen, externen Repräsentationen und der Konstruktion interner mentaler Modelle im Kontext des Lernens aus physikalischen Experimenten in den Blick zu nehmen.

Vor dem Hintergrund der Forschungsergebnisse zur Wissensveränderung und zur Förderung von Repräsentationskompetenz wurden kognitiv aktivierende Instruktionen entwickelt, die weitverbreitete naive bzw. wissenschaftlich inadäquate Schülervorstellungen in der Strahlenoptik berücksichtigen und Lernende darin unterstützen, wissenschaftlich zweckmäßigere Repräsentationen zu bilden sowie kompetent mit gegebenen und selbst erstellten Repräsentationen zu operieren.

Wenn es gelänge, Lernenden durch kognitive Aktivierung eine tiefere Verarbeitung zu ermöglichen, bei der Lernende geeignete interne Repräsentation in Form von Konzepten und Begriffen (mathematische Formeln, statische Bilder und Diagramme etc.) aufbauen und mit diesen operieren, sollten sie in der Lage sein, kohärente mentale Modelle zu bilden. Diese ermöglichen es dann, den Lerninhalt so zu verstehen, dass die Lernenden wissenschaftliche Konzepte erwerben und auf dieser Basis Problemlöseaufgaben erfolgreich bearbeiten können.

Zum Zweck einer solchen kognitiven Aktivierung wurden in dieser Studie experimentbezogene Aufgaben entwickelt, welche darauf zielen, die Lernenden prüfen zu lassen, ob ihr Wissen auf fachlich angemessenen und relevanten Konzepten (anstelle auf Schülervorstellungen) basieren, wobei die Lösung der Aufgaben regelmäßig erfordert, verschiedene Repräsentationen aufeinander zu beziehen.

Anschließend wurde empirisch untersucht, ob die entwickelten Aufgaben erstens Wissen und Problemlösen beim Umgang mit fachbezogenen Repräsentationen in der Strahlenoptik und zweitens das konzeptuelle Grundverständnis in der Strahlenoptik fördern. Die Analyse möglicher Auswirkungen auf die Entwicklung der Motivation im Fach Physik stellt einen Nebenaspekt dieser Arbeit dar.

Hierzu wurde im Schuljahr 2010/2011 eine quasi-experimentelle Studie mit dem Lerninhalt Strahlenoptik in den Klassenstufen 7 und 8 an acht Gymnasien und zwei Gesamtschulen in Rheinland-Pfalz durchgeführt. Wissen und Problem-

lösen bei repräsentationsbezogenen Aufgaben sowie das konzeptuelle Verständnis und die Motivation der Schüler im Fach Physik wurde jeweils vor der Studie, unmittelbar danach sowie zwei Monate später erhoben.

Im ersten Teil dieser Arbeit werden die in der Einleitung umrissenen theoretischen Rahmenbezüge erklärt, vor deren Hintergrund die empirische Studie geplant wurde. Darüber hinaus werden empirische Befunde zum Lernen mit multiplen Repräsentationen, zur Förderung des konzeptuellen Verständnisses und zum Thema kognitive Aktivierung vorgestellt und der Zusammenhang zwischen dem Umgang mit multiplen Repräsentationen und der Förderung des konzeptuellen Verständnisses erläutert.

Im zweiten Teil der Arbeit wird zunächst auf die spezifischen Charakteristika des Lerninhalts Strahlenoptik eingegangen und die Wahl des Lerninhalts begründet. Anschließend werden die Forschungsfragen und Hypothesen vorgestellt, das eingesetzte Unterrichtsmaterial, die Stichprobe, die Messinstrumente und das Vorgehen bei der Auswertung der verwendeten Tests und Fragebögen beschrieben. In weiteren Kapiteln werden die statistischen Analysemethoden dargestellt und das mathematische Modell zur Analyse der Daten erläutert sowie die Ergebnisse präsentiert.

Im dritten Teil folgt die kritische Diskussion der Ergebnisse. Abschließend werden Möglichkeiten und Grenzen der Studie, offene Fragen für künftige Forschungsprojekte und Konsequenzen für die Unterrichtspraxis aufgezeigt.

1 Empirische und theoretische Grundlagen

1.1 Die Rolle von Repräsentationen in kognitiven Lernprozessen

1.1.1 Repräsentationen, Repräsentationsformen und Repräsentationskompetenz

Eine Repräsentation eines Gegenstands oder eines Ereignisses ist zunächst eine Darstellung. In ihrer Funktion als Darstellung kann eine Repräsentation als ein Objekt definiert werden, das in einem jeweiligen Kommunikationszusammenhang für diesen Gegenstand bzw. dieses Ereignis steht. Der grundsätzliche Charakter einer Repräsentation besteht also darin, dass eine Repräsentation auf etwas anderes *verweist* (vgl. Teichert, 2006, S. 109; Schnotz, Baadte, Müller & Rasch, 2011, S. 216); sie hat eine Bedeutung. Das Foto eines Versuchsaufbaus für ein physikalisches Experiment verweist beispielsweise auf die Anordnung der technischen Bestandteile des Experiments und die Abstimmung der Bestandteile aufeinander zum Zweck der Versuchsdurchführung. Ein und dasselbe Objekt, hier der Versuchsaufbau, kann durch verschiedene Repräsentationen dargestellt werden; so kann der Versuchsaufbau statt durch ein Foto durch eine Anleitung in Worten repräsentiert werden.

Darstellungen besitzen die Funktion von Zeichen und weisen als solche eine dreistellige Relation auf (vgl. Metzinger, 2010, S. 24).

- g (Repräsentanda); hier der Versuchsaufbau
- R (Repräsentation); hier Fotografie des Versuchsaufbaus oder Beschreibung des Versuchsaufbaus in Worten. Zwischen $R(g)$ und g besteht eine Abbildungsrelation, die festlegt, welche Eigenschaften von g welchen Eigenschaften $R(g)$ entsprechen (vgl. Schnotz, 1994; Herrmann, 1993).
- Die dritte Komponente der Relation betrifft „b“ (den Repräsentationsvorgang durch den Träger einer Repräsentation); hier Repräsentationsvorgang durch einen Physiklehrer; „ gbR “ bedeutet in diesem Fall: der Versuchsaufbau wird durch eine Beschreibung des Lehrers repräsentiert. Der Versuchsaufbau ist hier das intentionale Objekt, auf das der Physiklehrer mental gerichtet ist.

Die Repräsentation ist der Vorgang durch den „g“, der Versuchsaufbau, repräsentiert wird; „g“ und „b“ können niemals identisch sein, der Versuchsaufbau als dreidimensionales Objekt im Klassenraum oder der Physiksammlung ist weder die verbale Beschreibung noch das Foto des Versuchsaufbaus. Repräsentationen erfolgen daher immer unter einer bestimmten Perspektive, d.h. es werden nie alle Eigenschaften des repräsentierten Gegenstandes abgebildet. Derselbe Gegenstand kann grundsätzlich unterschiedlich repräsentiert werden, wobei Repräsentationen, die untereinander nicht informationsäquivalent sind, sich für unterschiedliche Zwecke unterschiedlich gut eignen können (vgl. Schnotz, 1994, S. 145).

Zeichen werden zur Kommunikation genutzt. Der Produzent eines Zeichens, z.B. die lehrende Person, „meint“ etwas, das er oder sie mitteilen möchte. Er oder sie hat einen bestimmten Inhalt im Kopf, den er oder sie einem Empfänger, hier dem Schüler, übermitteln will. Der Empfänger nutzt das Zeichen, um das Wissen über den repräsentierten Inhalt zu rekonstruieren. Ein Schüler kann auf Basis der Beschreibung in Textform den Versuchsaufbau so rekonstruieren, dass dieser dem Versuchsaufbau, den sich die lehrende Person vorgestellt hatte, entspricht. Er oder sie kann sich gegebenenfalls auch ein Foto der Experimentieranordnung, das der Produzent, z.B. der Autor eines Schulbuchs, erstellt hat, ansehen und den Versuch nachbauen.

Anhand dieses Beispiels wird deutlich, dass Repräsentationen hinsichtlich zweier Gesichtspunkte unterschieden werden können. Repräsentationen können danach differenziert werden, ob es sich um beschreibende (deskriptive) oder um bildliche (depiktionale) Repräsentationen handelt und danach, ob eine Repräsentation intern (mental) oder extern als Zeichen auf Papier, als Verkehrsschild, usw. (semiotisch) existiert.

Beschreibende (deskriptive) Repräsentationen bestehen aus Symbolen, also Zeichen, die keine Ähnlichkeit mit ihrem Referenten haben, bildliche (depiktionale) Repräsentationen hingegen repräsentieren ihren Referenten mittels einer räumlichen Konfiguration. Sie bestehen aus Ikonen, also aus Zeichen, die in irgendeiner Form der Ähnlichkeit zu ihrem Referenten stehen, strukturell mit dem Referenzobjekt übereinstimmen und daher mit dem Referenzobjekt assoziiert werden (Schnotz, 2005, S. 52). Deskriptive Repräsentationen eignen sich insbesondere für die Darstellung abstrakter Wissensinhalte. Mit ihnen können abstrakte Konzepte logisch verknüpft werden (vgl. ebd., S. 53).

Beispiel: *„Wenn sich ein Gegenstand im Abstand der doppelten Brennweite vor einer Sammellinse befindet, dann wird er gleich groß abgebildet.“*

In diesem Beispiel wird der Oberbegriff „Gegenstand“, der Begriff der Brennweite und der Begriff Sammellinse mit einer Größenrelation durch eine Wenn-

dann-Beziehung verknüpft, um eine allgemeine Aussage zu treffen, unter der alle Fälle fallen, für welche die genannten Bedingungen gelten. Wollte man den gleichen Inhalt bildlich darstellen, benötigte man mehr als ein Bild: in einer einzigen bildlichen Repräsentation ist es nur möglich, einen bestimmten Gegenstand darzustellen oder eine bestimmte Brennweite darzustellen und auch die Wenn-dann-Verknüpfung ließe sich nicht in einer einzigen depiktionalen Repräsentation alleine darstellen.

Depiktionale Repräsentationen hingegen enthalten sehr viele Informationen auf einmal, sie „beanspruchen bezogen auf eine spezifische Klasse von Informationen immer Vollständigkeit“ (vgl. Schnotz et al., 2011, S. 219). Dies drückt sich auch in der Redewendung aus: *Ein Bild sagt mehr als tausend Worte*. Das Foto der Versuchsanordnung enthält z.B. auch Informationen darüber, um welchen leuchtenden Gegenstand es sich handelt, z.B. um eine Kerze, wie diese Kerze geformt ist und wie sie befestigt wurde, in welcher Größenrelation sie zu der Linse steht, usw. Ist die Kerze angezündet und steht der Schirm an der richtigen Stelle, wird direkt ersichtlich, ob das Bild vergrößert, verkleinert oder gleich groß abgebildet wurde. Will man den Versuch nachbauen, kann man sich sehr leicht an die Vorlage halten und braucht nicht lange zu überlegen, wie die Teile montiert werden müssen. Aus depiktionalen Repräsentationen können daher Informationen leicht „abgelesen werden“, sie werden direkt ersichtlich. Sie weisen eine hohe „inferenzielle“ Nutzungseffizienz auf (vgl. ebd., S. 219), d.h. sie eignen sich gut für konkrete und (fall)spezifische Schlussfolgerungen. Externe Repräsentationen verweisen insofern auf interne Repräsentationen, als dass die externen semiotischen Repräsentationen ein Subjekt benötigen, welches die Bedeutung des Zeichens erkennen kann.

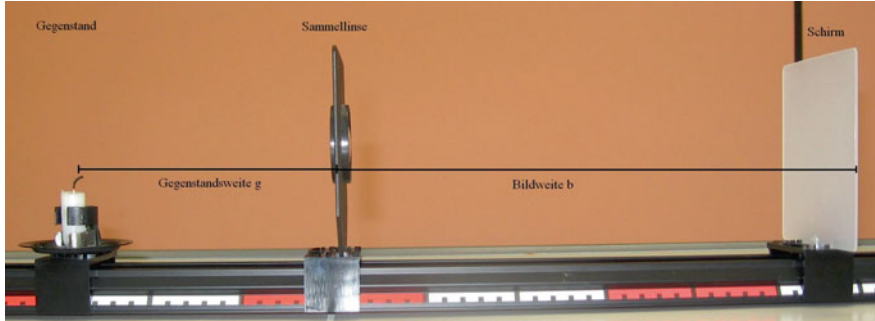
Des Weiteren werden Repräsentationsformen, z.B. in der Fachdidaktik, hinsichtlich ihres Abstraktionsgrades geordnet. So unterscheidet Leisen (1998, S. 9) zwischen der Phänomenebene, der Modellebene und der Theorieebene.

Wie an dem Beispiel der Repräsentation eines optischen Versuchsaufbaus deutlich wird, lässt sich ein korrektes und möglichst umfassendes Verständnis am besten erreichen, indem die depiktionalen und die deskriptiven Repräsentationen aufeinander bezogen werden. So kann mental im Abgleich der beiden Repräsentationen anhand der Beschreibung des Versuchsaufbaus als Text und des Foto des Versuchsaufbaus überprüft werden, ob der Versuchsaufbau der Beschreibung des Textes entspricht und umgekehrt. Soll der Zweck des Versuchsaufbaus, die Untersuchung und Erklärung eines physikalischen Phänomens durch das entsprechende Experiment, erfasst werden, kommen die verschiedenen Abstraktionsebenen ins Spiel. So kann ein umfassendes Verständ-

nis eines physikalischen Sachverhaltes dadurch erzielt werden, dass eine Person den Sachverhalt auf der Phänomenebene, der Modellebene und der Theorieebene erklären kann und hierbei zwischen verschiedenen Repräsentationen mit unterschiedlichem Abstraktionsgrad wechselt sowie verschiedene Repräsentationen mit unterschiedlichem Abstraktionsgrad aufeinander bezieht. Eine Person, die hierzu in der Lage ist, kann z.B. aus dem Verhältnis von Gegenstandsweite und Bildweite im Versuchsaufbau den Abbildungsmaßstab eines reellen Bildes mittels einer Sammellinse bestimmen oder umgekehrt den Versuch so aufbauen, dass eine bestimmte Bildgröße erzielt wird.

Ersichtlich wird auch, dass sich verschiedene Repräsentationsformen für unterschiedliche Aufgaben unterschiedlich gut eignen oder sich in ihrem Informationsgehalt ergänzen. Die Verhältnisse von Gegenstandsgröße (G) zu Bildgröße (B) und Gegenstandsweite (g) zu Bildweite (b) lassen sich kompakt und übersichtlich in der folgenden Formel zum Abbildungsmaßstab (A) darstellen:

$$A = \frac{B}{G} = \frac{b}{g}$$



Montieren Sie eine Kerze inklusive Kerzenteller, eine Sammellinse und einen Schirm hintereinander auf einer optischen Bank. Befestigen Sie je eine Klammer am Kerzenteller, an der Linse und an dem Schirm. Kerze, Linse und Schirm werden in der genannten Reihenfolge durch Aufstecken auf der optischen Bank montiert. Durch das Verschieben der Klammern auf der Bank können die Abstände zwischen Kerze und Linse sowie zwischen Linse und Schirm eingestellt werden. Den Abstand zwischen Kerze und Linse bezeichnet man als Gegenstandsweite (g), den Abstand zwischen Linse und Schirm als Bildweite (b).

Abbildung 1: Beschrifteter Versuchsaufbau zur Entstehung reeller Bilder bei der Sammellinse und schriftliche Erläuterung des Versuchsaufbaus

Eine Zeichnung hingegen, die alle verschiedenen Abbildungsvarianten erfasst, wäre sehr komplex und könnte den Sachverhalt doch nicht in der Allgemeinheit der oben genannten Gleichung ausdrücken.

Für das Verständnis eines physikalischen Versuchsaufbaus ist eine Abbildung sicherlich hilfreicher als eine Beschreibung in Textform. Bei einem komplizierten Versuchsaufbau oder für eine Zielgruppe mit geringem Vorwissen ist eine Kombination aus einer beschrifteten Fotografie und Text einer Repräsentationsform alleine (nur Foto bzw. nur Text) überlegen und in der Praxis naturwissenschaftlich-technischen Unterrichts unentbehrlich (siehe Abbildung 1).

In aller Regel befassen sich Lernende bei der Auseinandersetzung mit einem Fachgebiet, also nicht nur mit einer Repräsentationsform. Vielmehr müssen sie verschiedene Repräsentationsformen (z.B. Text und Bild, Diagramm und Formeln) aufeinander beziehen. Ganz besonders in den Naturwissenschaften bezieht sich Repräsentationskompetenz auf ein Set an Fähigkeiten und Praktiken, die mit der Verwendung einer Vielfalt visueller Repräsentationen verbunden sind, welche als Denkwerkzeuge, zur Kommunikation und zur Konzeptualisierung wissenschaftlicher Konzepte eingesetzt werden (vgl. Kozma & Russell, 2005). Repräsentationskompetenz beinhaltet die Fähigkeiten:

- unterschiedliche Repräsentationsformen mit äquivalenter Bedeutung zu koordinieren und ineinander zu übersetzen (vgl. Ainsworth, 1999; Dolin, 2007);
- diejenige Repräsentationsform zu wählen, die den Sachverhalt am „einfachsten“ darstellt (Peterson, 1996; Tabachneck-Schijf & Simon, 1996). Praktisch bedeutet dies: eine Person, die erkennt, welche Repräsentation sich in welcher Situation zum Erreichen eines Ziels eignet, ist nicht nur schneller bei der Bewältigung einer Aufgabe, sondern macht sich das Leben auch leichter.
- Repräsentationen im Hinblick auf ihre Angemessenheit und Zweckmäßigkeit hin zu beurteilen,
- neue Repräsentationen zu kreieren und
- Informationen aus unterschiedlichen Repräsentationen zu vergleichen sowie
- Repräsentationen bei der Lösung von Problemen zu verwenden, aus Repräsentationen logische Schlüsse zu ziehen und Repräsentationen zu nutzen, um Vorhersagen zu treffen (vgl. Gilbert, 2008; Kozma & Russell, 2005).

Um zu erkennen, wie ein kompetenter Umgang mit Repräsentationen gefördert werden kann, ist es notwendig, wenn auch nicht hinreichend, die gedächtnis-

psychologischen Grundlagen des Lernens der Informationsverarbeitung beim Lernen mit multiplen Repräsentationen zu kennen.

1.1.2 Gedächtnispsychologische Grundlagen

1.1.2.1 Gedächtnissysteme

Wenn Lernende mit Informationen aus der Umwelt konfrontiert werden, so durchlaufen die Informationen in den unterschiedlichen menschlichen Gedächtnissystemen verschiedene Instanzen der kognitiven Verarbeitung. Diese Instanzen bilden den wissenschaftlichen Hintergrund, vor dem die Bedeutung von Repräsentationen für kognitive Lernprozesse in den Blick genommen werden kann, welche in Modellen der kognitiven Verarbeitung von Texten und Bildern thematisiert wird.

Die folgenden grundlegenden Gedächtnissysteme in der menschlichen Informationsverarbeitung gilt es zu unterscheiden:

1. das sensorische Register (ältere Bezeichnung: Ultrakurzzeitgedächtnis),
2. das Arbeitsgedächtnis (ältere Bezeichnung: Kurzzeitgedächtnis) und
3. das Langzeitgedächtnis (vgl. Atkinson & Shiffrin, 1968; Mietzel 2007; zit. n. Zumbach, 2010, S. 72).

Das sensorische Register: Im sensorischen Register werden die Informationen aus unserer Umwelt in einer ersten Instanz aufgenommen und (sehr) kurzzeitig (im Bereich weniger hundert Millisekunden) gespeichert (vgl. Kluwe, 1992, zit. n. Zumbach 2010, S. 72). Gelenkt von Aufmerksamkeitsprozessen werden bestimmte Informationen ausgewählt und ins Arbeitsgedächtnis überführt.

Das Arbeitsgedächtnis: Im Arbeitsgedächtnis findet die eigentliche Informationsverarbeitung statt. Informationen, welche im Arbeitsgedächtnis weiterverarbeitet wurden, können anschließend dauerhaft verbal, bildhaft oder auditiv im Langzeitgedächtnis gespeichert werden.

Das Langzeitgedächtnis: Im Langzeitgedächtnis ist das Wissen in Form von Propositionen, Schemata oder Skripts organisiert (vgl. Atkinson & Shiffrin, 1968; Shiffrin, 1977; zit. n. Zumbach 2007, S. 72) und kann zu einem späteren Zeitpunkt wieder abgerufen werden (vgl. Niegemann et al., 2008, S. 43).

Propositionen: Unter einer Proposition wird die kleinste, abstrakte Wissenseinheit verstanden, die einen Sachverhalt beschreibt und die es ermöglicht, zwischen

unterschiedlichen Repräsentationen von Wissen hin und her zu wechseln, beispielsweise von Begriffen zu Bildern und umgekehrt (vgl. Homberger, Sachwörterbuch zur Sprachwissenschaft, 2003/Proposition).

Bei Propositionen handelt es sich um komplexe Symbole, die aus einfacheren Symbolen zusammengesetzt sind. Eine Proposition beinhaltet eine semantische Prädikat-Argument-Relation: ein Symbol, das Prädikat, steht für den Bezug, ein weiteres Symbol bzw. mehrere weitere Symbole, das Argument bzw. die Argumente, stehen für die Funktionseinheiten.

Die Argumente werden durch das Prädikat verbunden. Beispielsweise wird mit dem Begriff „nähen“ gleichzeitig der Agens (jemand, oder etwas näht) und das Objekt (alles, was genäht werden kann) sowie das Instrument (alles, mit dem genäht werden kann) verbunden. Im Kontext der mentalen Repräsentation von Propositionen, werden Propositionen als „Beschreibungen in einer hypothetischen mentalen Sprache“ verstanden (zit. n. Schnotz et al., 2011, S. 222), welche Inhalte auf einer konzeptuellen Ebene, d.h. unabhängig von einem spezifischen Wortlaut oder der Syntax repräsentieren (vgl. ebd.).

Schemata: Bei Schemata handelt es sich um kognitive Konstrukte, die es ermöglichen mehrere kleinere Informationseinheiten in einer großen Informationseinheit zu kategorisieren. Eine solche Informationseinheit kann dann im Ganzen im Langzeitgedächtnis gespeichert werden (vgl. Sweller, 2005, S. 20). Schemata bilden einerseits Wissen ab und steuern andererseits in ihrer Funktion, Wissen zu ordnen und zu organisieren, den Erwerb neuer Informationen (vgl. Niegemann et al., 2008, S. 44).

Konstruktion von Schemata. Für die Konstruktion von Schemata sind zwei kognitive Prozesse besonders bedeutsam (vgl. Rey, 2009, S. 40):

- die Elaboration (neue Informationen erhalten eine Bedeutung, indem sie mit dem Vorwissen der Lernenden verknüpft werden, wodurch neue Informationen strukturiert und besser verstanden werden) und
- die Induktion (aus konkreten Lernerfahrungen werden übergeordnete Schemata abstrahiert, was Lernende darin unterstützt, Probleme zu lösen, die sich von Aufgaben unterscheiden, die in der Lernphase bearbeitet wurden).

Automatisierung von Schemata: Neben der Konstruktion von Schemata spielt auch die Automatisierung von Schemata eine wichtige Rolle. Sobald ein Schema erworben wurde, kann es durch Übung über einen langen Zeitraum hinweg automatisiert werden. Wenn eine Person beispielsweise die Buchstaben des Alphabets und ihre Kombination zu Wörtern und Sätzen erlernt hat, also Lesen gelernt

hat, kann sie durch Lesen neues Wissen erwerben ohne sich bewusst auf die Kombination von Buchstaben zu einzelnen Wörtern konzentrieren zu müssen (vgl. Sweller, 2005, S. 21).

Erlernete Fähigkeiten in komplexen Fachgebieten setzen unzählige Schemata voraus, die im Langzeitgedächtnis bereit gehalten werden (vgl. Sweller, 2005, S. 21).

Skripte: Skripte sind eine spezielle Form von Schemata (Schank & Abelson, 1977, zit. n. Niegemann et al., 2008, S. 44). „Sie erfassen prototypische Ereignisse, in denen meist stereotype Handlungssequenzen auftreten.“ (zit. n. Niegemann et al., 2008, S. 44). Ein typisches Skript ist z.B. ein Restaurantbesuch, dessen Ablauf angefangen von der Wahl des Tisches bis zur Bezahlung sich immer gleicht.

Die zentrale Instanz für die Verarbeitung von neuen Informationen ist das Arbeitsgedächtnis. Beim Wissenserwerb und Verstehen komplexer Zusammenhänge, wie sie das Physiklernen erfordert, sind einige „Besonderheiten“ (zit. n. Zumbach, 2010, S. 72) des Arbeitsgedächtnisses zu beachten.

Die Kapazität des Arbeitsgedächtnisses ist hinsichtlich des Umfangs begrenzt. Bei einfachen Behaltensaufgaben können 7 ± 2 Informationseinheiten gespeichert werden. Beim Arbeitsgedächtnis handelt es sich jedoch nicht um ein isoliertes System. So steht es beim Informationserwerb mit dem Langzeitgedächtnis in Verbindung. Der Umfang an Informationseinheiten, die dem Arbeitsgedächtnis zur Verfügung stehen, kann nämlich durch Chunking erweitert werden. Beim Chunking werden die einzelnen „neuen“ Informationen zu übergeordneten Informationseinheiten zusammengefasst. Chunking ist ohne Informationen, die im Langzeitgedächtnis gespeichert sind, nicht möglich. Chunks sind entweder direkt im Langzeitgedächtnis abgespeichert oder es stehen dort Strukturen zur Verfügung, aus denen Chunks gebildet werden können (vgl. Sweller, 2005, S. 24). In der Konsequenz können zwar nicht mehr als $7+2$, maximal also 9 Informationseinheiten, im Arbeitsgedächtnis präsent sein. Der Umfang an Informationen, die durch Chunking erweitert wurden, ist jedoch offen (vgl. Kluwe, 1992, zit. n. Zumbach 2010, S. 72).

Beim Lernen von neuen Informationen kann es jedoch vorkommen, dass zu viele Informationen gleichzeitig verarbeitet werden müssen. Die Kapazität des Arbeitsgedächtnisses ist sowohl hinsichtlich des Umfangs als auch der Zeit begrenzt.

1.1.2.2 Die Cognitive-Load-Theorie

Ein Erklärungsmodell solcher kognitiven Belastungen bietet die Cognitive-Load-Theorie (vgl. Sweller, 1994). Gemäß der Theorie werden drei unterschiedliche Formen der kognitiven Belastung unterschieden: a) „Intrinsic Cognitive Load“, b) „Extraneous Cognitive Load“ und c) „Germane Cognitive Load“.

Quelle des Intrinsic Cognitive Load ist das Lernmaterial selbst. Je komplexer und schwieriger die Inhalte sind, desto höher der Cognitive Load. Der Schwierigkeitsgrad wird von der Elementinteraktivität bestimmt. Lernmaterial mit einer hohen Elementinteraktivität zeichnet sich dadurch aus, dass Lernende sehr viele Zusammenhänge zwischen den unterschiedlichen Inhalten herstellen müssen. Die Elementinteraktivität kann gesteuert werden, indem die Lerninhalte in Untereinheiten eingeteilt werden (Segmentierung) und in eine gut verständliche Reihenfolge gebracht werden (Sequenzierung). Dabei sollte das Vorwissen der Lernenden beachtet werden.

Der Extraneous Cognitive Load wird durch die Gestaltung und die Darstellung der Lerninhalte bestimmt. Irrelevante und redundante Informationen und zahlreiche Verweise erhöhen den Extraneous Cognitive Load. Instruktionen sollten so gestaltet werden, dass der Extraneous Cognitive Load minimiert wird. So sollten Lerninhalte, die zeitgleich bearbeitet werden müssen, nahe beieinander platziert werden oder visuelle Lerninhalte durch auditiv dargebotene Informationen ergänzt werden, um das visuelle Arbeitsgedächtnis zu entlasten. Weitere Gestaltungsempfehlungen, die der Verminderung des Extraneous Cognitive Load dienen, finden sich z.B. in Niegemann et al. (2008, S. 47, 48).

Die kognitive Belastung schließlich, die für den reinen Wissenserwerb benötigt wird, bezeichnet man als Germane Cognitive Load. Dieser kognitive Load ist gemäß der Theorie nicht veränderbar. Für den Germane Cognitive Load sollten möglichst viele Ressourcen zur Verfügung stehen, damit die Lernenden kognitive Schemata und Automatismen aufbauen können. Es empfiehlt sich, Instruktionen so zu gestalten, dass der Extraneous Cognitive Load minimiert und der Intrinsic Cognitive Load abgestimmt auf das Vorwissen der Zielgruppe durch Segmentierung und Sequenzierung so gesteuert wird, dass die Lernenden die verfügbaren Ressourcen im Arbeitsgedächtnis optimal nutzen können. Kritisch zu bewerten ist, dass es zurzeit nur möglich ist, die Gesamtbelastung zu erfassen, nicht aber die verschiedenen Formen von Cognitive Load zu unterscheiden. Des Weiteren weisen Forschungsergebnisse von Paas, van Merriënboer und Adam (1994), Paas, Tuovinen, van Merriënboer und Darrabi (2005) sowie von Schnotz und Rasch (2005, 2007) darauf hin, dass das Vorwissen der Lernenden mehr be-

achtet werden sollte. Lernenden mit höherem Vorwissen stehen relevante Schemata zur Verfügung mit denen sie komplexes Lernmaterial strukturieren können. Für Lernende mit höherer Expertise reduziert sich daher der Intrinsic Cognitive Load. Durch eine optimale Passung zwischen Expertise und Intrinsic Load können die Lernenden ihre Ressourcen für den Aufbau von Schemata und die Elaboration der Lerninhalte nutzen.

Neben der Begrenzung der Kapazität des Arbeitsgedächtnisses spielen auch die „modalitätsspezifischen“ Eigenschaften (zit. n. Zumbach, 2010, S. 73) eine zentrale Rolle beim Lernen mit verschiedenen Repräsentationen. Diese modalitätsspezifischen Eigenschaften werden im erweiterten Arbeitsgedächtnismodell nach Baddeley beschrieben.

1.1.2.3 Das Arbeitsgedächtnis nach Baddeley

Gemäß dem erweiterten Arbeitsgedächtnismodell sind vier Instanzen im Arbeitsgedächtnis zu unterscheiden (vgl. Abbildung 2):

1. die zentrale Exekutive,
2. die phonologische Schleife (phonological loop),
3. die visuelle-räumliche Notiztafel (visual-spatio sketchpad) und
4. der episodische Puffer.

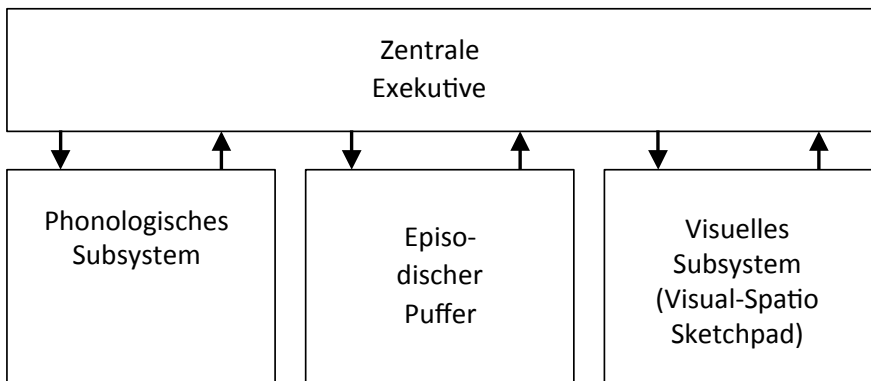


Abbildung 2: Das Arbeitsgedächtnis nach Baddeley (1988), entnommen aus Zumbach (2010, S. 74)

- Die zentrale Exekutive: Sie kontrolliert und koordiniert die anderen Subsysteme des Arbeitsgedächtnisses (phonologische Schleife und visuell-räumliche Notiztafel und episodischer Puffer). Hierbei besitzt sie keine eigene Speicherkapazität. Ihre Funktion besteht darin, die Aufmerksamkeit zu koordinieren.
- Die phonologische Schleife: In diesem Subsystem werden akustische, sprachbasierte Informationen verarbeitet. Sie ist in ihrer Kapazität auf etwa 2 Sekunden begrenzt.
- Die visuelle-räumliche Notiztafel: Sie speichert visuell und räumlich aufgenommene Informationen (z.B. Bilder oder Filme) vorübergehend ab.
- Phonologische Schleife und visuell-räumliche Notiztafel werden als weitgehend unabhängig voneinander angenommen. Verbale Informationen aus Texten werden zunächst im visuellen Subsystem verarbeitet und anschließend als auditive Informationen in der phonologischen Schleife weiterverarbeitet.
- Der episodische Puffer: Der episodische Puffer stellt eine Erweiterung des ersten Arbeitsgedächtnismodells dar, da das anfängliche Modell nicht erklären konnte, wie Informationen, die in unterschiedlichen Repräsentationsformaten vorliegen, kombiniert werden. Es handelt sich um einen kapazitätsbegrenzten Speicher im Arbeitsgedächtnis, der mit dem Langzeitgedächtnis in unmittelbarem Kontakt steht. Er fungiert als verbindende Instanz zwischen Langzeitgedächtnis und phonologischer Schleife und ermöglicht es, Inhalte, die in verschiedenen Repräsentationsformaten vorliegen, zu integrieren. Phonologische Schleife und visuell-räumliche Notiztafel sind zusätzlich aber auch direkt mit dem Langzeitgedächtnis verbunden.

1.1.2.4 Mentale Modelle

An der Schnittstelle zwischen Arbeits- und Langzeitgedächtnis stehen mentale Modelle. Bei einem mentalen Modell handelt es sich um ein internes hypothetisches Konstrukt, welches eine strukturelle oder funktionelle Analogie zu den repräsentierten Objekten bzw. zu den repräsentierten Abläufen aufweist (vgl. Dutke, 1994, S. 76). Mentale Modelle enthalten eine aufgabenorientierte thematische Auswahl an Informationen aus dem Langzeitgedächtnis, die der Aufgabenbearbeitung im Arbeitsgedächtnis dienen. In die Konstruktion mentaler Modelle fließen Informationen aus dem Langzeitgedächtnis (Vorwissen) mit ein. So kann ein mentales Modell von einem physikalischen Versuch Eigenschaften und Beziehungen enthalten, die in der Beschreibung des physikalischen Prozesses oder in der Abbildung des Versuchsaufbaus nicht enthalten sind.

Modelle, die sich in einem bestimmten Kontext als nützlich erwiesen haben, sind nur schwer zu verändern. Ihre Korrektur ist mit erheblichem kognitivem Aufwand verbunden.

Zentrale Charakteristika mentaler Modelle sind nach Dutke (1994, S. 76 f.) sowie Schnotz et al. (2011, S. 222) unter anderem:

- Mentale Modelle bilden Gegebenheiten in der Umwelt sowohl in „reduzierender“ als auch in „elaborierender“ Weise ab (zit. n. Dutke, 1994, S. 76 f.).
- Mittels mentaler Modelle können Sachverhalte in Gedanken anschaulich dynamisch simuliert werden. Sie enthalten eine „aufgabenorientierte thematische Auswahl“ (ebd.), d.h. irrelevante Details werden ausgeblendet.
- Mentale Modelle sind anschaulich, in dem Sinn, dass sie es ermöglichen, sich Sachverhalte zu vergegenwärtigen. Im Gegensatz zu einem Bild sind sie jedoch abstrakter.
- Mentale Modelle sollten daher nicht mit Wahrnehmungen oder mit Bildern von wahrgenommenen Objekten gleichgesetzt werden (vgl. Schnotz et al., 2011, S. 222). Beispielsweise kann ein mentales Modell auch einen Sachverhalt repräsentieren, der überhaupt nicht wahrgenommen werden kann (vgl. Johnson-Laird, 1983).
- Ein mentales Modell ist nicht an eine bestimmte sensorische Repräsentationsform gebunden. An einem mentalen Modell können Repräsentationen aus verschiedenen Sinnesmodalitäten beteiligt sein: „Ein mentales Modell von einer räumlichen Konfiguration kann beispielsweise über visuelle, auditive oder auch haptische Informationen konstruiert werden“ (zit. n. Schnotz et al., 2011, S. 223).
- Die Funktion mentaler Modelle besteht erstens im Verstehen von Sachverhalten. Zweitens können sie die Basis für die Planung und Steuerung von Handlungen bilden.
- Mentale Modelle basieren häufig (jedoch nicht immer) auf Analogien. Analogien stellen einen Spezialfall mentaler Modelle dar, bei dem Relationen zwischen Informationen, die bereits im Langzeitgedächtnis gespeichert sind, auf Elemente eines neuen Wissensbereichs übertragen werden.

Die beschriebenen Gedächtnissysteme und die zentrale Funktion mentaler Modelle sind die Basis der menschlichen Informationsverarbeitung. Im folgenden Abschnitt werden nun drei aufeinander aufbauende Modelle der kognitiven Verarbeitung von Texten und Bildern vorgestellt, welche diese kognitive Architektur zum Ausgangspunkt haben:

1. die duale Codierungstheorie von Paivio (1986),
2. die kognitive Theorie des Multimedialen Lernens von Mayer (2005) und
3. das integrative theoretische Rahmenmodell des Text- und Bildverstehens (Schnotz & Bannert, 2003; Schnotz, 2005).

1.1.3 Modelle der kognitiven Verarbeitung von Texten, Bildern und Sprache

1.1.3.1 Die duale Codierungstheorie von Paivio

Nach der dualen Codierungstheorie von Paivio (1986) sowie Clark & Paivio (1991) werden sprachliche Informationen und bildliche Informationen in unterschiedlichen kognitiven Systemen verarbeitet. Verbale Codes („logogens“) wie Wörter oder Sätze, werden normalerweise nur im verbalen System enkodiert, Bilder („imagens“) hingegen werden dual enkodiert, also verbal und imaginal. Obgleich die beiden Kodierungssysteme grundlegend unterschiedlich aufgebaut sind und sich auch in ihrer Funktionsweise voneinander unterscheiden, stehen sie in informationellem Austausch miteinander. Wird ein konkreter Sachverhalt beschrieben, kommt es zu einer doppelten Kodierung (vgl. Abbildung 3).

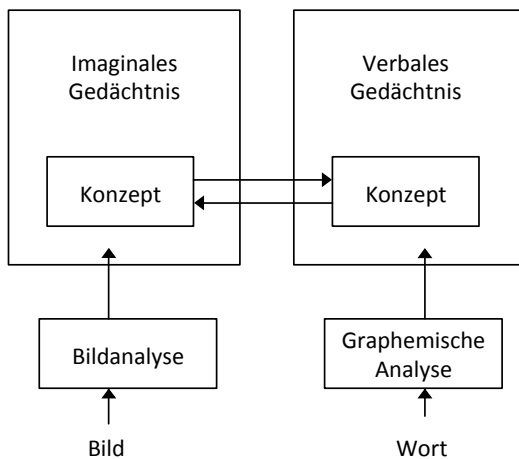


Abbildung 3: Gedächtnissysteme nach Paivio, entnommen aus Zumbach (2010, S. 75)

Die Theorie wurde oft herangezogen, um den Bildüberlegenheitseffekt zu erklären. Konkrete Bilder werden immer auch verbal enkodiert. Konkrete Begriffe, die man sich auch bildlich vorstellen kann, können bildlich enkodiert werden, sofern der Lernende hierzu eine imaginale Repräsentation aufbaut. Bei abstrakten Begriffen findet keine doppelte Kodierung statt, da man sich diese Begriffe in der Regel nicht bildlich vorstellen kann. Daher fällt es leichter, sich an Bilder von konkreten Gegenständen als an konkrete Begriffe zu erinnern. Zudem ist es generell einfacher konkrete Inhalte zu erinnern, als Inhalte in abstrakter Form aus dem Gedächtnis abzurufen. Der Bildüberlegenheitseffekt kann jedoch auch anders erklärt werden. So werden Bilder ganzheitlich enkodiert, Wörter hingegen nur sequentiell. Da die Kapazität des Arbeitsgedächtnisses begrenzt ist, vermindert ein sequentielles Speichern und Abrufen die Kapazität des Arbeitsgedächtnisses (vgl. Weidenmann, 1997).

1.1.3.2 Die kognitive Theorie des multimedialen Lernens

Die kognitive Theorie des multimedialen Lernens von Mayer (2005) greift die zentrale Annahme der dualen Codierungstheorie von Paivio (1986) auf, dass sprachliche und bildliche Informationen in unterschiedlichen Systemen verarbeitet werden. Die folgenden zentralen Kernannahmen über die menschliche Informationsverarbeitung liegen der Theorie zugrunde (vgl. Mayer, 2005, S. 31 ff.):

- Visuell/bildhafte und auditiv/verbale Informationen werden in unterschiedlichen Kanälen verarbeitet (dual-channels assumption).
- Die Verarbeitungskapazität jeder der beiden Verarbeitungskanäle ist begrenzt (limited capacity assumption), d.h. in jedem Verarbeitungssystem kann nur eine begrenzte Informationsmenge gleichzeitig verarbeitet werden. Diese Annahme ist mit der Cognitive Load Theory (Sweller, 1994) und dem Arbeitsgedächtnismodell von Baddeley (1999) konsistent.
- Lernen wird als aktiver Prozess gesehen, bei dem Menschen aktiv verschiedene kognitive Prozesse koordinieren, um eine kohärente mentale Repräsentation ihrer Erfahrungen zu generieren (active processing assumption).

Das Modell beschreibt die Verarbeitung von Bildern, gesprochenen und gedruckten Wörtern wie folgt (vgl. Abbildung 4):

- Die Informationen aus der Umwelt treten über die Augen und die Ohren ins sensorische Gedächtnis ein. Bilder können dabei im Gegensatz zu Wörtern nur über die Augen ins sensorische Gedächtnis gelangen.
- Die erfassten verbalen und bildhaften Informationen können als exaktes Abbild für sehr kurze Zeit im sensorische Gedächtnis präsent gehalten werden. Während diese beiden Schritte automatisch ohne große Anstrengung geschehen, folgt nun der erste aktive Verarbeitungsschritt, der einer bewussten Kontrolle unterliegt.
- Gelenkt durch Prozesse der Aufmerksamkeit wird eine Auswahl an relevanten Wörtern und Bildern ins Arbeitsgedächtnis überführt. Die ausgewählten Informationen bilden zunächst eine bruchstückhafte und unorganisierte Informationsbasis.
- In einem weiteren Schritt werden die Informationen nun in einer kohärenten Struktur organisiert. Hieraus resultiert ein kohärentes verbales oder bildhaftes mentales Modell. Bei diesem Prozess kann es zum Übergang zwischen visuellen und phonologischen Repräsentationen kommen; d.h. gehörte Textinformationen, die phonologisch repräsentiert wurden, werden intern auf Basis ihrer semantischen Bedeutung dargestellt. Umgekehrt können visuelle Textinformationen auch bildhaft oder phonologisch repräsentiert werden.
- Abschließend bringt der Lernende im Prozess der Integration relevantes Vorwissen aus dem Langzeitgedächtnis ins Arbeitsgedächtnis. Dort wird das in dem bildhaften und verbalen Modell gespeicherte Wissen mit dem Vorwissen des Lernenden verknüpft.
- Treten verbale Informationen durch den visuellen Kanal ein, nehmen die Wörter bei der Verarbeitung eine komplexe Route durch das System und konkurrieren um die Aufmerksamkeit mit bildlichen Informationen, die ebenfalls im visuellen Kanal verarbeitet werden (vgl. Mayer, 2005, S. 44).

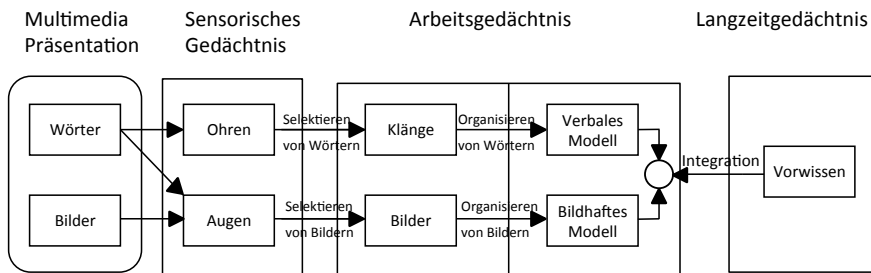


Abbildung 4: Theorie des multimedialen Lernens nach Mayer (2005) S. 43, übersetzt von der Autorin

Im Rahmen des Modells werden fünf verschiedene kognitive Prozesse voneinander unterschieden. Diese fünf Prozesse müssen dabei nicht in linearer Reihenfolge auftreten, sondern können iterativ wiederholt durchlaufen werden.

1. *Auswahl von Wörtern (selection)*: Der Lernende richtet seine Aufmerksamkeit auf die relevanten Wörter des Lerninhalts und erzeugt eine erste auditive Repräsentation innerhalb des Arbeitsgedächtnisses.
2. *Auswahl von Bildern (selection)*: Der Lernende richtet seine Aufmerksamkeit auf die relevanten Bildinformationen innerhalb des Lernmaterials und wählt relevante Bildinformationen aktiv aus, um eine visuelle Repräsentation im Arbeitsgedächtnis zu bilden.
3. *Organisation von Wörtern (organizing)*: der Lernende verbindet die ausgewählten Wörter, um ein kohärentes verbales Modell im Arbeitsgedächtnis zu generieren.
4. *Organisation von Bildern (organizing)*: Bei der Organisation von Bildern findet ein Verknüpfungsprozess zwischen ausgewählten Bildern statt, um ein kohärentes bildhaftes Modell ausbilden zu können.
5. *Integration (integrating)*: die Integration ist maßgeblich für den Lernprozess; sie kann sowohl im visuellen als auch im verbalen Arbeitsgedächtnis stattfinden und erfordert es, beide Subsysteme zu koordinieren. Bei der Koordination des bildhaften und verbalen Modells im Arbeitsgedächtnis kann der Lernende sein Vorwissen nutzen.

1.1.3.3 Das integrative Rahmenmodell des Text- und Bildverstehens

Das integrative theoretische Rahmenmodell des Text- und Bildverstehens (vgl. Schnotz & Bannert, 2003; Schnotz, 2005) basiert auf der dualen Codierungstheorie nach Paivio (1986). Ebenso wie die Theorie von Paivio (1986) und die kognitive Theorie des multimedialen Lernens von Mayer (2005) geht das integrative theoretische Rahmenmodell des Text- und Bildverstehens nach Schnotz (2005, S. 56 f.) davon aus, dass ein verbaler Kanal und ein bildlicher Kanal der Informationsspeicherung und -weiterverarbeitung existiert. Im Gegensatz zu der Theorie vom Paivio (1986) wird in diesem Modell jedoch angenommen, dass sowohl beim Text- als auch beim Bildverstehen multiple Repräsentationen geformt werden (vgl. Schnotz, 2005, S. 54). Des Weiteren bezieht das Modell ebenso wie die Theorie von Mayer (2005) die verschiedenen Subsysteme des Gedächtnisses ein: sensorisches Gedächtnis, Arbeitsgedächtnis und Langzeitgedächtnis.

Die Aufnahme und Verarbeitungskapazität sowohl des deskriptiven als auch des depiktionalen Kanals wird dabei durch die Kapazität des Arbeitsgedächtnisses begrenzt. Das Modell beschreibt die Verarbeitung von Bildern, gesprochenen und gedruckten Wörtern wie folgt (vgl. auch Abbildung 5).

- Wenn Lernende einen Text lesen oder Sprache hören, tritt die visuelle verbale bzw. auditive verbale Information durch das jeweilige Sinnesorgan (Auge oder Ohr) in das jeweilige Register (visuelles oder auditives Register) ein.
- Die Informationen werden dann durch den visuellen bzw. auditiven Kanal in das visuelle bzw. auditive Arbeitsgedächtnis weitergeleitet.
- Die Lernenden bilden nun eine mentale Repräsentation der Oberflächenstruktur des Textes. Diese Repräsentation der Oberflächenstruktur ermöglicht es den Lernenden, das Gelesene, bzw. Gehörte zu wiederholen.
- Ein verbaler bzw. auditiver Filter selektiert die verbale Information aus dem visuellen bzw. auditiven Arbeitsgedächtnis und leitet sie durch den verbalen Kanal in das propositionale Arbeitsgedächtnis weiter.
- Der Lernende konstruiert nun auf Basis der Repräsentation der Oberflächenstruktur eine propositionale Repräsentation. Diese propositionale Repräsentation enthält die Informationen des Gelesenen bzw. Gehörten auf einer konzeptuellen Ebene, d.h. die Inhalte werden unabhängig von dem spezifischen Wortlaut oder der Syntax ausgedrückt.
- Die Propositionen wiederum führen zur Bildung und Elaboration eines mentalen Modells. Hier kommt nun das Langzeitgedächtnis ins Spiel. Das allgemeine theoretische Wissen eines Individuums ist in Form kognitiver Schemata im Langzeitgedächtnis abgespeichert, die typische Zusammenhänge eines Realitätsbereichs repräsentieren. Beim Wissenserwerb fungieren diese Schemata als kognitive Instrumente zur Konstruktion propositionaler Repräsentationen und mentaler Modelle.

Die Konstruktion mentaler Modelle geht zum einen von propositionalen Repräsentationen aus, wobei die Konstruktionsprozesse von kognitiven Schemata angeleitet werden. Zum anderen werden an mentalen Modellen auch neue Informationen abgelesen und zu den bereits bestehenden propositionalen Repräsentationen im Langzeitgedächtnis hinzugefügt (vgl. Schnotz, 2006b).

Wenn ein Bild verstanden wird, tritt die visuelle bildhafte Information durch das Auge in das visuelle Register ein und wird dann durch den visuellen Kanal in das visuelle Arbeitsgedächtnis weitergeleitet. Dort kommt es zur visuellen perzeptuellen Repräsentation des Bildes. Ein visueller bildhafter Filter selektiert

die bildhaften Informationen aus dem visuellen Arbeitsgedächtnis und leitet sie durch den bildhaften Kanal weiter, wo sie zur Konstruktion und Elaboration eines mentalen Modells führen. Während die ausgewählten verbalen Informationen zunächst ins propositionale Arbeitsgedächtnis gelangen und anschließend zur Konstruktion eines mentalen Modells im Arbeitsgedächtnis führen, wird aus den ausgewählten bildlichen Informationen also direkt ein mentales Modell im Arbeitsgedächtnis gebildet. Letztere folgen somit zunächst nicht dem Weg über das propositionale Arbeitsgedächtnis. Während das visuelle Abbild sinnesspezifisch ist (visuelle Sinnesmodalität), sind mentale Modelle nicht-sinnesspezifisch, da in einem mentalen Modell Informationen aus verschiedenen Sinnesmodalitäten integriert werden können (vgl. Abbildung 5).

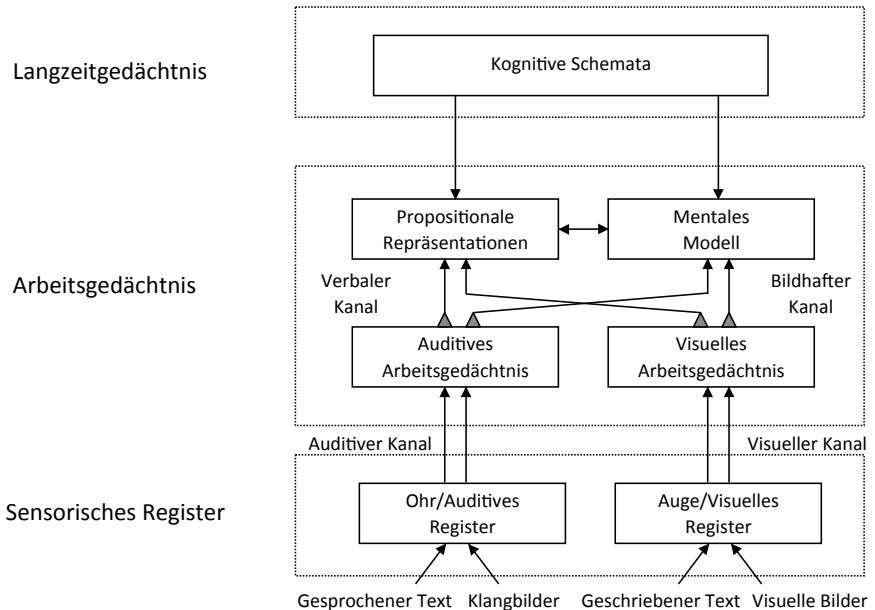


Abbildung 5: Integratives theoretisches Rahmenmodell des Text- und Bildverstehens nach Schnotz (2005), S. 57, übersetzt von der Autorin

Das integrative theoretischen Rahmenmodell des Text- und Bildverstehens (Schnotz & Bannert, 2003; Schnotz, 2005) unterscheidet sich von den anderen Modellen wie folgt:

- Im Vergleich zur dualen Codierungstheorie nach Paivio (1986) und der kognitiven Theorie des multimedialen Lernens von Mayer (2005) wird im integrativen theoretischen Rahmenmodell des Text- und Bildverstehens die Verarbeitung von Texten und Bildern detaillierter beschrieben.
- Im Gegensatz zu der traditionellen Theorie des dualen Kodierens nimmt das Rahmenwerk an, dass sowohl beim Erfassen von Texten, als auch beim Erfassen von Bildern multiple Repräsentationen geformt werden (vgl. Schnotz et al., 2011, S. 223).
- Anders als die kognitive Theorie des multimedialen Lernens von Mayer (2005) geht das integrierte Rahmenmodell davon aus, dass verbale Informationen nicht notwendigerweise mit der auditiven Modalität assoziiert werden, sondern auch in andere Sinnesmodalitäten übertragen werden können. Analoges gilt für bildhafte Informationen, die ebenfalls in andere Modalitäten, z.B. Klangbilder, überführt werden können. Ein weiterer Unterschied besteht darin, dass nach dem integrierten Rahmenmodell ein mentales Modell gebildet wird, während in dem Modell von Mayer erst ein verbales und dann noch ein bildhaftes Modell konstruiert werden, die anschließend unter Einbeziehung von Informationen aus dem Langzeitgedächtnis integriert werden (vgl. Schnotz, 2005, S. 59, 60).

1.1.3.4 Pädagogische Implikationen

Pädagogische Implikationen ergeben sich sowohl aus der begrenzten Kapazität des Arbeitsgedächtnisses, die im Kern der Cognitive Load Theorie steht als auch aus den Modellen von Mayer (2005) und Schnotz (2005). Sowohl Sweller und Chandler (1994) als auch Mayer (2005) und Schnotz (2005) leiten wesentliche Konsequenzen für die Gestaltung von Lernmaterial mit Texten und Bildern ab. Ausführliche Informationen finden sich bei den Autoren selbst, eine Überblicksdarstellung bieten Niegemann, et al. (2008, S. 45f, S. 53 f. und S. 60 f.).

Aus der Cognitive Load Theorie und der kognitiven Theorie des multimedialen Lernens, welche beide die Kapazitätsbegrenzung einbeziehen, ergeben sich unter anderen die folgenden Konsequenzen⁷:

⁷ An dieser Stelle wird auf eine Auswahl an Empfehlungen hingewiesen, die für den Kontext Lernen naturwissenschaftlicher Inhalte in der Schule relevant sind.

- Bei der Gestaltung der Lernumgebung ist darauf zu achten, dass den Lernenden nicht zu viel Material auf einmal präsentiert wird, um eine Überlastung des Arbeitsgedächtnisses zu vermeiden. Dies fällt besonders bei der gleichzeitigen Darbietung von Texten und Bildern ins Gewicht, da beide Repräsentationsformen über den visuellen Kanal verarbeitet werden. Eine Entlastung des visuellen Kanals kann erreicht werden, indem Erläuterungen zu Grafiken oder Bildern als gesprochener Text dargeboten werden (z.B. durch Erklärungen des Lehrers).
- Zu einer Überlastung des Arbeitsgedächtnisses kann es aber auch kommen, wenn zu viele Informationen gleichzeitig auf dem auditiven und dem visuellen Kanal zu verarbeiten sind. Für das Kreieren eines mentalen Modells müssen Textinformation, Bildinformation und gesprochene Informationen permanent verglichen werden, um sie zu einem kohärenten mentalen Modell zu integrieren. Werden zu viele Informationen gleichzeitig dargeboten, müssen sehr viele Informationen gleichzeitig aktiv gehalten werden. Möglichkeiten der Entlastung bestehen darin:
 - Wissensinhalte, die für das Verständnis notwendig sind, sind nahe beieinander zu platzieren (Kontiguitätsprinzip).
 - Redundante Informationen sind nicht gleichzeitig zu präsentieren (Redundanzprinzip)
 - Auf interessante Materialien, die nicht zum Verständnis des Lerninhalts beitragen ist zu verzichten (Kohärenzprinzip), so dass mehr Kapazität für die Verarbeitung des eigentlichen Verstehensprozesses zur Verfügung steht.
 - Lernumgebungen sind auf das Vorwissen der Lernenden zuzuschneiden. Während Novizen strukturierte Lernmaterialien benötigen, die durch Lernhilfen ergänzt werden, genügt fortgeschrittenen Lernern eine Informationsquelle für die Bearbeitung. Werden zusätzliche Lernhilfen angeboten, kann es zum Expertise-Revearsel-Effekt kommen: die zusätzlichen Informationen sind für die Lernenden redundant und behindern den Wissenserwerb, da aus der Fülle der Informationen die relevanten Informationen ausgewählt werden müssen.

Zusätzlich zu diesen Empfehlungen führt Schnotz (2005) drei weitere Prinzipien an:

- Präsentationsequenz-Prinzip: Wenn es nicht möglich ist, Text und Bild gleichzeitig zu präsentieren, sollte zuerst das Bild und dann der Text dargeboten werden. Im Gegensatz zu schriftlichen Texten, für welche die Information erst

vom visuellen Verarbeitungskanal in den verbalen Kanal wechseln muss bevor ein mentales Modell gebildet wird, bildet sich aus den ausgewählten bildlichen Informationen direkt ein mentales Modell im Arbeitsgedächtnis. Ein aus bildhafter Information aufgebautes mentales Modell steuert dann die Verarbeitung des Textes, so dass nicht der gesamte Text in propositionale Repräsentationen umgesetzt werden muss, was die kognitiven Ressourcen schont und ebenfalls dafür sorgt, dass mehr Kapazität für die Verarbeitung des eigentlichen Verstehensprozesses zur Verfügung steht.

- Structure-Mapping Prinzip: Stehen mehrere Bilder mit äquivalenten Inhalten zur Verfügung, sollte dasjenige ausgewählt werden, welches den Wissensseinheit am angemessensten repräsentiert.
- Prinzip der Prozesskontrolle: Beim Wissenserwerb mit besonders komplexen Texten und Grafiken, der keiner strikten Zeitbegrenzung unterliegt, sollte Text in schriftlicher Form die Grafik ergänzen. Wie bereits skizziert, kann die Grafik in ein mentales Modell überführt werden, die dann durch Informationen aus dem Text ergänzt wird. Auf diese Weise können aus dem Text relevante Informationen selektiert werden. Läge der Text auditiv vor, müsste er permanent wiederholt werden, um die Informationen aus Grafik und Text in ein kohärentes mentales Modell umzusetzen.

Die hier skizzierten pädagogischen Prinzipien weisen auf die Funktionen hin, die Repräsentationen in unterschiedlichen Formaten (geschriebene Texte, schematische und realistische Bilder, Formeln usw.) im Lernprozess zukommen. So können Texte Grafiken ergänzen, gesprochene Sprache kann Grafiken erläutern und die Darbietung einer Repräsentationsform (z.B. einer Grafik) kann die Aufmerksamkeit bei der Verarbeitung einer anderen Repräsentationsform (Text) lenken.

1.1.4 Lernen mit multiplen Repräsentationen

1.1.4.1 Funktionen von multiplen Repräsentationen

Einen systematischen Überblick über die Funktionen der Verwendung multipler Repräsentationen gibt Ainsworth (1999, 2006). Ainsworth entwickelt eine konzeptuelle Rahmentheorie zum Lernen mit multiplen Repräsentationen, in der sie drei Hauptfunktionen multipler Repräsentationen expliziert (vgl. Abbildung 6):

1. die Unterstützung komplementärer Prozesse,
2. die Begrenzung von Interpretationen und
3. die Konstruktion eines tieferen Verständnisses.

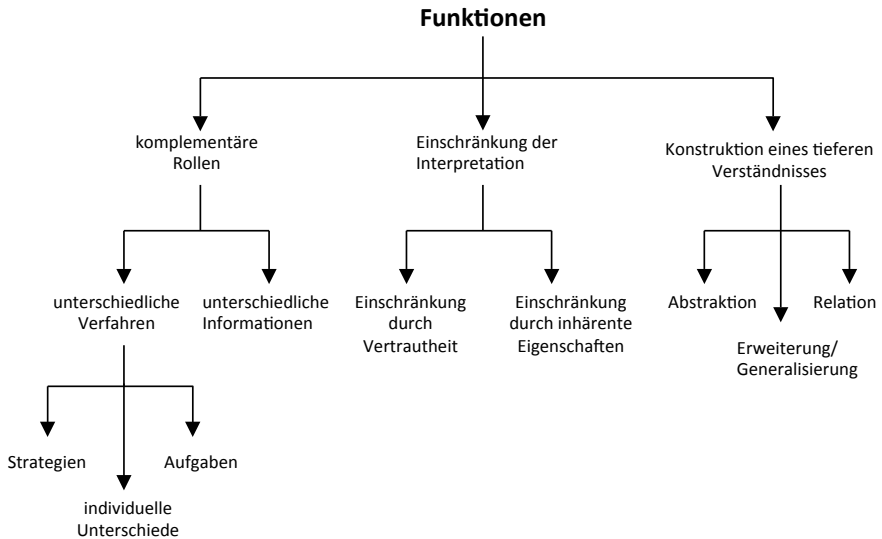


Abbildung 6: Funktionale Taxonomie multipler Repräsentationen nach Ainsworth (2006, S.6), übersetzt von der Autorin

Zu 1) *Verwendung multipler Repräsentationen zur Unterstützung komplementärer Prozesse*: Nach Ainsworth (1999) können Repräsentationen, die einen äquivalenten Inhalt aufweisen, sich dann komplementär ergänzen, sofern sie sich in ihrem Format (z.B. deskriptiv versus depiktiv) unterscheiden. Diese Ergänzung betrifft dabei sowohl die *Prozesse* als auch die *Informationen*, welche die Repräsentationen zur Verfügung stellen. So befördern Repräsentationen, welche die gleichen Informationen beinhalten, ggf. unterschiedliche Schlussfolgerungen, oder ermöglichen es, Informationen unterschiedlich leicht bzw. schwer abzulesen. Dabei können multiple Repräsentationen Lernende darin unterstützen, verschiedene Strategien zur Bearbeitung einer Aufgabe zu erproben, unter mehreren Repräsentationsformaten ein Format auszuwählen, welches die Bearbeitung der Aufgabe erleichtert oder dasjenige Repräsentationsformat zu verwenden, das ihnen individuell entgegenkommt.

Des Weiteren können unterschiedliche Repräsentationsformate eingesetzt werden, um sich inhaltlich wechselseitig zu ergänzen. Dies ist besonders dann sinnvoll, wenn es entweder zu komplex wäre, alle Informationen in einer Repräsentation darzustellen, oder wenn eine einzelne Repräsentation in einem bestimmten Format sich nicht dazu eignet, alle relevanten Informationen so darzustellen, dass diese leicht aus ihr abgelesen werden können.

Zu 2) *Die Verwendung multipler Repräsentationen zur Begrenzung von Interpretationen*: Ein weiterer Vorteil der Verwendung multipler Repräsentationen besteht darin, dass eine Repräsentation dafür genutzt werden kann, eine andere Information, die für das Verständnis des Lerninhalts zentral ist, zu unterstützen. Gemäß Ainsworth (1999, 2006) kann diese Unterstützung auf zwei Weisen erfolgen:

- Wenn Lernende mit einer Repräsentation sehr vertraut sind, kann diese bei der Interpretation einer weniger vertrauten Repräsentation hilfreich sein. So können konkrete Animationen dazu verwendet werden, das Verständnis unvertrauter Repräsentationen wie Graphen zu unterstützen.
- Des Weiteren können die Eigenschaften, die einer Repräsentationsform inhärent sind, dazu genutzt werden, die Informationen einer weiteren Repräsentation in einem anderen Format zu unterstützen. Ainsworth (2006, S. 8) nennt hier folgendes Beispiel: Die verbale Beschreibung „das Messer liegt neben der Gabel“ lässt offen, auf welcher Seite das Messer liegt. Eine bildliche Darstellung hingegen ist hier festgelegt. Werden nun beide Repräsentationen gleichzeitig dargeboten, wird die Interpretation der ersten Repräsentation durch die Darbietung der zweiten Repräsentation beschränkt. Im Fall des Versuchsaufbaus in der Optik (siehe Kapitel 1.1.1, Abbildung 1) kann z.B. eine Fotografie dabei helfen, die Reihenfolge, in der die Gegenstände auf einer optischen Bank montiert sind, eindeutig festzulegen.

Zu 3) *Die Verwendung multipler Repräsentationen zur Konstruktion eines tieferen Verständnisses*: Die Verwendung verschiedener Repräsentationen fördert die Konstruktion eines tieferen Verständnisses, wenn Lernende Informationen aus unterschiedlichen Repräsentationen integrieren, die sie nicht aus der Verwendung einer Repräsentation alleine beziehen könnten.

So kann die Verwendung multipler Repräsentationen die Abstraktion von Informationen unterstützen. Prozesse der Abstraktion beinhalten eine Reorganisation auf höherer Ebene (vgl. Ainsworth, 1999, S. 142). Dies ist z.B. der Fall, wenn Lernende begreifen, dass unterschiedlichen Repräsentationen in ver-

schiedenen Formaten (z.B. Darstellung einer Funktion als Gleichung und als Graph) die gleiche Struktur zugrunde liegt oder wenn Lernende durch die Operation mit verschiedenen Repräsentationen grundlegende Strukturen des Wissensgebietes erkennen. Im Gegensatz zur Abstraktion werden bei der Verwendung multipler Repräsentationen zur Generalisierung des bestehenden Wissens die Kenntnisse der Lernenden erweitert, ohne dass es zu einer grundlegenden Veränderung durch eine Konzeptualisierung auf einer höheren Ebene kommt. Beispielsweise können Tabellen und Graphen, die zunächst im Mathematikunterricht verwendet werden, auch im Geografie- oder Physikunterricht verwendet werden oder bestehendes Wissen, welches in einem Repräsentationsformat vorliegt, kann in ein anderes Repräsentationsformat übertragen werden (ein Geschwindigkeits-Zeit Diagramm wird in eine Tabelle übertragen oder in ein Beschleunigungs-Zeit-Diagramm). Gerade der letzte Punkt verweist auf eine weitere zentrale Funktion der Verwendung multipler Repräsentationen hin: der Einsatz multipler Repräsentationen, um Lernende gezielt zu fördern, Repräsentationen ineinander zu übersetzen.

Nachdem nun die empirisch fundierten Theorien und ihre Implikationen für die Gestaltung von Instruktionsmaterialien dargelegt wurden, geben die folgenden Kapitel einen Forschungsüberblick zum Thema Lernen mit multiplen Repräsentationen.

1.1.4.2 Empirische Befunde aus Experten-Novizen-Vergleichen

Erste Hinweise auf die Schlüsselposition von Repräsentationen für das Verständnis von Mathematik und Naturwissenschaften stammen aus der Expertiseforschung. Eine klassische Studie von deGroot (deGroot 1965, zit. n. Bransford, Brown & Cocking, 2000, S. 32) untersuchte, wie es Weltklasse-Schachmeistern gelang, ihre Kontrahenten zu besiegen. Schachmeistern und weniger erfahrenen, jedoch immer noch relativ guten Spielern wurden Beispiele von Schachspielen gezeigt und sie wurden gebeten, laut darüber nachzudenken, welche Entscheidung sie trafen, wenn sie in der Rolle einer der Spieler wären. DeGroot (1965) beobachtete, dass die Weltklassenspieler in der Lage waren, bedeutungsvolle Konfigurationen zu erkennen und die Implikationen der Situation einzuschätzen, was sie zu geschickteren Zügen befähigte.

Diese Überlegenheit der Experten, bedeutungsvolle Konfigurationen abzurufen, wurde durch Chunking erklärt. Die Schachmeister erkannten Chunks von bedeutungsvollen Informationen, die sie zusammenfassen konnten, um die strategischen Implikationen zu beurteilen. Da Novizen nicht über die hierarchisch

organisierten Wissensstrukturen im Langzeitgedächtnis verfügen, können sie nicht auf die Chunking Strategie zurückgreifen. Dieser Befund wurde in Experten-Novizen Vergleichen zu anderen Domänen wie der Radiologie (vgl. Lesgold, 1988, zit. n. Brandsford et al., 2000, S. 33) der Informatik (vgl. Ehrlich und Soloway 1984, zit. n. ebd., S. 33) oder der Mathematik (vgl. Robinson und Hayes, 1978, zit. n. ebd., S. 33) bestätigt. Beispielsweise waren Experten der Elektrotechnik in der Lage, lange Einheiten von komplexen Schaltplänen zu reproduzieren, die sie nur wenige Sekunden gesehen hatten (Egan & Schwartz, 1979, zit. n. ebd., S. 33). Die Experten erfassten verschiedene Einzelelemente von Schaltplänen (z.B. von Widerständen oder Kondensatoren), welche als Verstärker fungierten, als eine Informationseinheit. Indem sie sich an den Schaltplan eines Verstärkers erinnerten, konnten sie sich die Anordnung der einzelnen Schaltpläne merken. Die Relevanz für die Erklärung von Lernprozessen unter dem Fokus des Lernens mit multiplen Repräsentationen wird deutlich, wenn man sich vor Augen führt, dass sich die internen mentalen Repräsentationen von Experten und Novizen hinsichtlich der internen Repräsentation der Ausgangssituation unterscheiden. Die Chunks, auf welche die Experten zurückgreifen, stellen im Bereich des Schachspiels oder der Elektronik eine interne Repräsentation von depiktionalen Konfigurationen dar, auf die logische Operationen (wie Spielzüge oder das Erkennen von Funktionszusammenhängen im Bereich elektrischer Schaltungen) angewandt werden.

Eine Studie zum Experten-Novizen-Vergleich im Bereich der Physik ergab, dass Experten der physikalischen Beschreibung zur Strömung von Flüssen sowie des Gegenwinds und Rückenwinds in der Luftfahrt die gleichen mathematischen Prinzipien zugrunde legen, wie sie auch im Bereich der relativen Geschwindigkeiten angewandt werden. Die Autoren folgern, dass die Expertise auf der Fähigkeit basiert, bei der Wahrnehmung von Problemen auf Schemata zurückgreifen zu können, welche die Art und Weise anleiten, wie das Problem repräsentiert und erfasst wird (vgl. Glaser & Chi, 1988, zit. n. Brandsford et al., 2000, S. 33).

Eine weitere Studie, die sich explizit mit der Bedeutung der Problemrepräsentation beim Physiklernen unter dem Gesichtspunkt der Expertiseforschung auseinandersetzt, stammt von Larkin (1983). Larkin (1983) nahm an, dass Novizen von einer naiven Problemrepräsentation ausgehen, die sich auf konkrete Objekte der realen Welt und Prozesse bezieht, die in Echtzeit ablaufen. Experten hingegen seien dazu zusätzlich in der Lage, eine physikalische Repräsentation zu entwickeln, welche gedachte Entitäten wie Kraft oder Impuls enthielten. Sie verfügten somit über ein zusätzliches mentales Modell der Problemsituation, das aussagekräftige und bedeutungsvolle Eigenschaften zur Problemlösung beinhaltet (vgl. Larkin, 1983, S. 75). Larkin (1983) legte 11 Experten (Professoren und Doktoranden)

und 11 Novizen (Studienanfängern der Physik und der Ingenieurwissenschaften) vier physikalische Probleme vor, die sowohl mit dem physikalischen Konzept der Kraft als auch mit dem physikalischen Konzept der Energie bearbeitet werden können. Die Teilnehmer wurden gebeten, die Probleme zu lösen und dabei laut zu denken. Während die Experten systematisch das „Kraft- oder Energie-“ Schema vervollständigten, zeichnete sich das Vorgehen der Novizen durch ein simples „Ziel-Mittel“-Vorgehen aus, bei dem die Novizen Werte solange in physikalische Formeln einsetzten, bis sie die gewünschte Größe berechnet hatten. Larkin (1983) führt diese unterschiedliche Vorgehensweise auf Unterschiede der Problemrepräsentation zurück. Während die Experten über ein physikalisches Schema verfügten und systematisch die Lücken füllten, die sich auf bekannte Größen bezogen, verfügten die Novizen nicht über einer adäquate Problemrepräsentation, was sie daran hinderte, effektiv zu operieren (vgl. ebd., S. 86).

In einer Anschlussstudie (vgl. ebd., S. 87) präsentierte Larkin den Teilnehmern deutlich schwierigere Probleme. Einen zentralen Grund für die höheren Schwierigkeitsgrad der Probleme sieht Larkin (1983), darin, dass die Lösung dieser Probleme die Koordination mehrerer Schemata erforderte. Während die Experten eine wissenschaftlich angemessene physikalische Repräsentation der Ausgangslage entwickelten, suchten die Novizen auch hier nach einer Gleichung, welche die physikalische Größe beinhaltete. Ihr Vorgehen endete immer, wenn sie nicht in der Lage waren, eine Beziehung in einer physikalischen Gleichung zu ersetzen (vgl. ebd., S. 90). Eine weitere Aufgabe, die eine besondere Herausforderung beinhaltete, wurde schließlich nur noch von sechs Experten erfolgreich gelöst. Die Auswertung der Protokolle bestätigte die Ergebnisse der ersten Studie. Die Experten unternahmen zunächst die Anstrengung, eine adäquate physikalische Problemrepräsentation zu entwickeln und konstruierten auf dieser Basis die passende mathematische Repräsentation, welche dann zur Lösung führte. Sie gingen also erst dann zur Verwendung der mathematischen Repräsentation über, wenn sie eine konsistente physikalische Repräsentation gefunden hatten.

In einer dritten Studie untersuchte Larkin (1983, S. 94 f.), wie sich das explizite Vermitteln der Verwendung repräsentationaler Schemata auswirkte. Hierzu wurden $N = 10$ Studienanfängern der Biologie und Medizin beigebracht, individuell sieben Prinzipien bei der Beschreibung eines Stromkreises, der mit Gleichstrom betrieben wird, anzuwenden. Nach dieser anfänglichen Instruktion enthielt die eine Hälfte der Studenten ein Training zur Verwendung physikalischer Repräsentationen, während die andere Hälfte lernte, zur Erzielung des gewünschten Ergebnisses systematisch diese Prinzipien anzuwenden, um Gleichungen zu generieren und

diese Gleichungen algebraisch zu kombinieren. Die Experimentalgruppe erzielte hierbei signifikant bessere Ergebnisse als die Kontrollgruppe ($p < .05$). Larkin (1983) fasst die wesentlichen Charakteristika physikalischer Repräsentationen auf Basis dieser Studien wie folgt zusammen:

„The entities are technical, with meaning only in physics. The inferencing rules are qualitative. The inferencing rules are time-independent and redundant. The representation is closely associated with fundamental principles of physics. Properties of the entities are localized to those entities” (S. 97).

Ein weiterer entscheidender Unterschied in den Kompetenzen zwischen Experten und Novizen besteht in ihrer andersartigen Organisation des Wissens. Die Wissensorganisation der Experten beeinflusst in zentraler Weise, wie diese Probleme verstehen und repräsentieren. In einer Studie von Chi, Feltovich und Glaser (1981) wurden Doktoranden (Experten) und Studienanfänger (Novizen) gebeten, physikalische Probleme hinsichtlich der Ähnlichkeit der Lösungswege zu kategorisieren. Dabei zeigt sich, dass die Novizen die physikalischen Probleme auf Basis oberflächlicher Kategorien wie Schlüsselwörter oder ähnliche visuelle Darstellungen klassifizierten, während sich die Experten auf die physikalischen Gesetze bezogen, welche den Problemen zu Grunde lagen. In der Folge bestand auch nur eine geringe Überschneidung zwischen dem Kategoriensystem der Experten und dem Kategoriensystem der Novizen.

In einer Anschlussstudie wurde diese Beobachtung überprüft (vgl. ebd.). Hierzu wurden Probleme gewählt, die sich hinsichtlich dreier physikalischer Basisprinzipien ordnen ließen. Gleichzeitig unterschieden sich diejenigen Probleme, die sich in ihrer visuellen Darstellung ähnelten (oberflächliche Ähnlichkeit) gerade in Bezug auf die zugrundeliegenden physikalischen Prinzipien. Die Analyse bestätigte die Ergebnisse der ersten Studie. Ebenfalls konform mit diesem Befund war die Beobachtung, dass fortgeschrittene Lernende (mit einem mittleren Wissensniveau) sowohl teilweise oberflächliche als auch teilweise prinzipielle Kriterien anwandten.

In einer Studie von Savelsbergh, de Jong und Ferguson-Hessler (1997) wurde die Qualität mentaler Repräsentationen von Experten und Novizen in den Blick genommen. Die Stichprobe bestand aus Doktoranden und Dozenten auf Seite der Experten sowie aus Studienanfängern zweier unterschiedlicher Universitäten, die einen Einführungskurs in Elektrodynamik besucht hatten auf Seite der Novizen. Allen Teilnehmern legte man eine zentrale Formel eines physikalischen Fachgebietes und ein physikalisches Schlüsselwort vor. Die Teilnehmer wurden nun gebeten, zwei physikalische Aufgabenstellungen zu beschreiben, die mit den

dargebotenen Informationen (Formel und Schlüsselbegriff) bearbeitet werden können. Dabei sollten die Teilnehmer ihre Gedanken direkt dazu äußern. Das Design der Studie basiert auf der Annahme, dass die verbale Beschreibung der Teilnehmer wesentliche Merkmale des zugrundeliegenden mentalen Modells offenlegt. Leistungsschwache Novizen entwickelten meist nur eine einzige relativ inkohärente Problemrepräsentation. Erfolgreichere und fortgeschrittene Problemlöser elaborierten die Information und kreierten kohärentere und informationsreichere Repräsentationen (vgl. ebd., S. 282). Die Analyse der Antworten zeigte, dass sich die Antworten der Experten von denen der Novizen in zwei Faktoren unterschieden, welche die Qualität der Problemrepräsentation mitbestimmen (vgl. ebd., S. 281):

- Der erste Faktor betrifft die interne Struktur einer einzelnen Repräsentation und erhielt von den Autoren das Label „Kohärenz“.
- Der zweite Faktor betrifft die Flexibilität und umfasst die Verknüpfung zwischen multiplen Repräsentationen, z.B. die Verknüpfung von abstrakten und konkreten Repräsentationen, zur Lösung eines physikalischen Problems.

Beide Faktoren hängen mit der Konsistenz des Wissens zusammen: Eine wenig kohärente Repräsentation ist anfälliger für Inkonsistenzen als eine kohärente Repräsentation. Wenn eine Inkonsistenz auftritt, oder der Denkprozess ins Stocken kommt, erleichtert die Verwendung multipler Repräsentationen das Finden des Lösungswegs im Vergleich zur Bildung eines einzigen mentalen Modells. Hinzu kommt, dass die inkohärenten Problemrepräsentationen von leistungsschwachen Novizen die Ausgangssituation des physikalischen Sachverhalts oft nicht vollständig abbildeten.

Ergebnisse der Expertiseforschung belegen darüber hinaus, dass Experten und Novizen nicht nur qualitativ unterschiedliche Problemrepräsentationen entwickeln, sondern auch in stärkerem Maß unterschiedliche Repräsentationsformen aufeinander beziehen.

Tabachneck, Leonardo und Simon (1994) berichten, dass es Novizen, welche mit multiplen Repräsentationen im Bereich der Ökonomie lernten, nicht gelang, Informationen aus linearen Graphen und geschriebener Information ineinander zu übersetzen, wohingegen Experten beide Informationen eng aneinander koppelten. Offensichtlich befähigte das tiefere Verständnis die Experten dazu, die unterschiedlichen repräsentationalen Formate zu integrieren.

Kozma und Russell (1997) führten zwei Studien durch, in denen untersucht wurde, wie Novizen und Experten der Chemie mit fachspezifischen

Repräsentationen umgehen. In beiden Studien wurde das Vorgehen von 10 Studienanfängern der Chemie dem Vorgehen 11 ausgebildeter Chemiker gegenübergestellt.

In einer ersten Studie wurden die Teilnehmer gebeten, 14 chemische Sachverhalte, die in unterschiedlichen Repräsentationsformaten vorlagen (Graphen, Animationen, Gleichungen oder Videosegmente) am Computer zu gruppieren und ihre Entscheidung zu begründen. Das Vorgehen der Experten unterschied sich von den Novizen in folgenden Punkten: Die Experten gruppieren größere Einheiten, die Einheiten der Experten enthielten unterschiedliche Repräsentationsformate und die Experten begründeten ihre Wahl durch tiefergehende konzeptuelle Erklärungen.

In einer zweiten Studie wurden die Teilnehmer gebeten, eine Reihe von Repräsentationen, die in unterschiedlichen Formaten vorlagen, ineinander zu transformieren (beispielsweise sollte eine Gleichung durch einen Graphen dargestellt werden). Den Experten gelang diese Aufgabe deutlich besser als den Novizen. Insbesondere waren sie den Novizen bei Aufgaben überlegen, die es erforderten, einen bildlich dargebotenen Sachverhalt verbal zu beschreiben. Hier gelang es den Experten im Gegensatz zu den Novizen, jede gegebene symbolische Repräsentation in eine wissenschaftlich korrekte verbale Repräsentation umzusetzen.

Zusammenfassung: Der Überblick zu Befunden der Expertiseforschung zeigt auf, dass für das erfolgreiche Problemlösen zwei Aspekte essentiell sind: das Erfassen des Problems in einer internen mentalen Repräsentation der Ausgangslage und das Finden geeigneter Operationen zur Problemlösung. Entsprechend der Definition von Repräsentationskompetenz von Dolin (2007) lässt sich die Überlegenheit eines flexiblen Umgangs mit multiplen Repräsentationen belegen (z.B. die Wahl geeigneter Repräsentationsformate oder die Fähigkeit verschiedene Repräsentationen aufeinander zu beziehen bzw. ineinander zu übersetzen).

1.1.4.3 Repräsentation der Ausgangslage, Problemlöseprozesse und Wissenserwerb

Ohlsson (1992) analysiert in seiner „representational change theory“, wie diese beiden Aspekte, nämlich das Erfassen des Problems in einer internen mentalen Repräsentation der Ausgangslage und das Finden geeigneter Operationen beim Problemlösen zusammenhängen. Gemäß der Theorie wirkt die Repräsentation des Ausgangszustands wie ein Gedächtnistrigger, der relevantes Wissen aus

dem Langzeitgedächtnis aktiviert. Eine geeignete Repräsentation des Ausgangszustandes zeichnet sich dadurch aus, dass das aktivierte Wissen geeignete Operationen und Aktionen zur Problembewältigung beinhaltet, während eine inadäquate Repräsentation des Ausgangsproblems Wissen aktiviert, das zur Lösung der Aufgabe nicht dienlich ist. Um ein Problem zu bewältigen, muss der Lernende daher eine geeignete Repräsentation des Ausgangszustandes anstreben und ungünstige Repräsentationen erkennen und umwandeln. Hierfür bestehen gemäß Ohlsson (1992) unterschiedliche Wege der Veränderung von Repräsentationen. Der Lernende kann die Information über das Problem durchdenken, neue Informationen hinzufügen, Bedingungen im Denken mental lockern, um eine andere Sichtweise zu erlangen oder die Informationen re-inkodieren, um zentrale Aspekte der Repräsentation des Ausgangszustands neu zu interpretieren. Die Veränderung der Repräsentation des Ausgangszustands führt zu einer Aktivierung von anderen Wissensbeständen aus dem Langzeitgedächtnis und damit zur Aktivierung anderer Operationen, welche die Lösungsstrategie optimieren.

Eine Studie von Gentner und Gentner (1983) nähert sich dem Thema der Problemrepräsentation und der damit verbundenen Aktivierung geeigneter Operationen aus Richtung der Forschung zur Rolle mentaler Modelle. Die Autoren untersuchten, wie unterschiedliche mentale Modelle von Elektrizität auf die Schlussfolgerungsleistung von Studenten wirken: Die Studenten sollten beurteilen, wie sich die Stromstärke einer Schaltung gegenüber einer Vergleichsschaltung ändert. Dabei wurde ebenfalls erhoben, welche Vorstellung die Studenten von Elektrizität haben („flowing-fluid“ versus „moving-crowd“ Modell). Die Autoren prüften folgende Hypothesen:

1. Die Auswirkung paralleler Batterien kann leichter aus dem „flowing-fluid“ Modell abgeleitet werden.
2. Die Wirkung von parallelen Widerständen kann leichter aus dem „moving-crowd“ Modell abgeleitet werden.

Im Ergebnis wurde die Hypothese der Autoren bestätigt: Die Verwendung unterschiedlicher Analogien führt zu systematischen Unterschieden im Schlussfolgern (Gentner & Gentner, 1983, S. 118): Die Gruppe mit dem „flowing-fluid“ Modell löste die Batterieaufgabe besser als die Widerstandsaufgaben und die „moving-crowd“ Gruppe löste die Widerstandsaufgaben besser als die „flowing-fluid“ Gruppe.

Für das Lernen in den Naturwissenschaften stellen mentale Modelle eine zentrale Form der Problemrepräsentation dar. Das Verständnis der Ausgangssituation, welches sich in einer komplexen internen Repräsentation niederschlägt,

wirkt sich entscheidend auf die weiteren Operationen aus. Die Studie belegt, dass das mentale Modell der Ausgangssituation maßgeblich die logischen Schlüsse beeinflusste, die aus dem Modell gezogen wurden.

Die Repräsentation des Ausgangszustandes wird in jedem Fall intern repräsentiert, kann jedoch zusätzlich auch in externalen Repräsentationen, wie schriftlichen Symbolen, Zeichnungen, Diagrammen oder Objekten mit repräsentationaler Funktion (wie z.B. einem Abakus) festgehalten werden. In Lehr-Lernprozessen kann es hilfreich sein Lernende durch explizite Arbeitsaufträge hierzu aufzufordern. Durch die externe Repräsentation der Ausgangslage kann das Arbeitsgedächtnis entlastet werden (vgl. Zhang, 1997; Schnotz et al., 2011), so dass mehr Kapazität für die Anwendung notwendiger Operationen zur Verfügung steht.

Eine Studie aus dem Grundschulbereich, die sich mit unterschiedlichen Formen der externen (und damit verbundenen) internen Repräsentationen eines Problems befasst, stammt von Hardy, Schneider, Jonen, Stern und Möller (2005). Sie untersuchten in einem Experiment mit 98 Drittklässlern wie naturwissenschaftliches Denken und Schlussfolgern im Umgang mit Diagrammen gefördert werden kann. Einer Gruppe von Kindern wurde in einer Unterrichtseinheit zum Thema Schwimmen und Sinken das Konzept der Dichte anhand einer Balkenwaage erklärt, wobei Volumen und Masse auf den beiden Waagschalen lagen. Eine Vergleichsgruppe bildete sich eigene Repräsentationen zu Volumen und Masse, wobei ihnen Materialien wie Buntkarton in unterschiedlichen Farben und Formen sowie Knöpfe und Buntstifte zur Verfügung gestellt wurden. Die Kinder konnten also ihr individuelles Verständnis der Situation direkt zum Ausdruck bringen. Da die selbst konstruierten Repräsentationen der Kinder unvollständig oder fehlerhaft sein können, wurden die Schüler bei der Konstruktion der Repräsentationen vom Lehrer unterstützt.⁸ Nach der Unterrichtseinheit bearbeiteten die Kinder einen Test zum Thema Schwimmen und Sinken und drei Tests zu proportionalem Denken. 56 Kinder wurden aus den 98 teilnehmenden Kindern ausgewählt und 5 Monate nach der Untersuchung interviewt. Zwischen den für die Interviews *ausgewählten* Kindern, die mit der Balkenwaage gearbeitet hatten und den Kindern, die eigene Repräsentationen erstellt hatte, bestand dabei keinerlei Leistungsunterschied in den Posttests. (Daher kann das Ergebnis der Studie nicht auf Gruppenvorteile zurückgeführt werden). In Einzelinterviews erhielten die Kinder 10 Fragen über die Integration von Masse und Volumen in das Konzept der Dichte und 6 Fragen zu

⁸ Die Kinder wurden den beiden Versuchsbedingungen zugeteilt, so dass beide Gruppen mit den gleichen Voraussetzungen starteten.

einem Transferthema, hier die Integration des Verhältnisses von zurückgelegtem Weg und Zeit in das Konzept der Geschwindigkeit. Die Aufgabe bestand darin, Graphen in einem Koordinatensystem zu interpretieren, in denen die integrierten Größen dargestellt waren. Die Kinder der ersten Gruppe konnten lineare Graphen, die Dichte betreffend, besser interpretieren als Kinder der zweiten Gruppe. Bei der Interpretation von linearen Graphen in dem Transferkontext (Geschwindigkeit) verschwand dieser Vorteil fast gänzlich.

Wenn Schüler mit einem geeigneten Modell lernen, fördert dies offenbar ihre domänenspezifische Problemlösekompetenz. Externe Repräsentationen, welche die Kinder unterstützen, die Ausgangslage adäquat zu repräsentieren, fördern dabei offenbar das Ziehen richtiger Schlussfolgerungen aus abstrakten auf den Lerninhalten aufbauenden Repräsentationen, wie die Interpretation linearer Graphen. Diese Ergebnisse belegen einmal mehr die Domänenspezifität von Lerneffekten, wie sie in der neueren Instruktionspsychologie berichtet werden (vgl. Klauer & Leutner, 2007, im Überblick).

Eine Studie von Zhang (1997) mit erwachsenen Versuchsteilnehmern bestätigt die Bedeutung der externen Repräsentation der Ausgangslage für schlussfolgerndes Denken und Problemlösen. In einem Experiment mit Universitätsstudenten wurde der Einfluss unterschiedlicher Repräsentationsformen auf das Problemlöseverhalten am Beispiel eines einfachen Strategiespiels untersucht. 80 Studenten spielten Drei-gewinnt (Tic-Tac-Toe) gegen einen perfekten Computergegner. Das Spiel basiert auf den Operationen mit einer schematischen Repräsentation, welche unterschiedlich dargestellt wurde. Je Bedingung wurden 20 Studenten per Zufall den vier Experimentalbedingungen zugeteilt:

- In Bedingung 1 gewann der Spieler, der zuerst drei Kreise in einer Linie belegte.
- In Bedingung 2 gewann, wer in einer Reihe drei Ziffern, die in der Summe 15 ergeben, ankreuzte.
- In Bedingung 3 konnte ein Sieg erzielt werden, indem drei Kreise gewählt wurden, welche die gleiche geometrische Form enthielten.
- In Bedingung 4 gewann der Spieler, der drei Figuren wählte, welche die gleiche Farbe enthielten.

Die vier Repräsentationsformen des Tic-Tac-Toe-Spiels sind logisch äquivalent, können also ineinander überführt werden. Jede der Repräsentation besteht aus neun Elementen und acht Gewinnmöglichkeiten.

Die Teilnehmer spielten zwei Runden, in der ersten Runde wurden die Teilnehmer angewiesen, drei Elemente in einer Reihe zu belegen. Vor Beginn der zweiten Runde wurden die Teilnehmer darauf hingewiesen, dass das Spiel entweder unentschieden ausgeht oder der Computer gewinnt. Sie erhielten die Aufgabe, den Computer am Sieg zu hindern. Sobald ihnen dies zum zehnten Mal gelang, endete das Experiment, andernfalls endete das Experiment nach 50 Runden.

Während alle Teilnehmern (100%) in der ersten Experimentalbedingung „Linie“ es schafften, den Computer zehn Mal am Sieg zu hindern, gelang dies nur 80 % der Teilnehmer in der Bedingungen 2 „Farbe“ und nur 45 % der Teilnehmer in den Bedingungen 3 „Form“ und Bedingung 4 „Zahl“.

Des Weiteren wurde analysiert, mit welchem Spielzug die Teilnehmer in der ersten Runde das Spiel eröffneten. Der geschickteste Eröffnungszug ist die Wahl des mittleren Feldes in der Linienrepräsentation, bzw. der „5“ in der Zahlrepräsentation, da sich hier jeweils vier Gewinnmöglichkeiten ergeben. Die zweitbeste Wahl besteht im Ankreuzen eines Eckfeldes. In der Linienrepräsentation des Spieles wurde nun die Wahl, das Kreuz in die Mitte zu setzen, sehr viel häufiger getroffen als dies bei einer zufälligen Wahl eines Feldes zu erwarten gewesen wäre. In der Zahlrepräsentation hingegen wurde die äquivalente Wahl (5) kaum häufiger getroffen als es bei einer zufälligen Wahl eines Feldes zu erwarten gewesen wäre. Offensichtlich fiel es den Probanden in der Zahlrepräsentation sehr viel schwerer, den strategisch besten Eröffnungszug zu erkennen als in der Linienrepräsentation. Im Hinblick auf die beiden anderen Repräsentationsformen bestätigte sich die Reihenfolge der Schwierigkeit:

- die Linienrepräsentation war einfacher als Farbrepräsentation und
- die Farbrepräsentation war einfacher als die Formrepräsentation.

Die Analyse der Lösungszeit der ersten zehn Spielzüge ergab einen signifikanten Haupteffekt in den Unterschieden zwischen den Bedingungen. Die Zeit, die für die Entscheidung benötigt wurde, war kürzer in der Linien-Bedingung als in der Zahl-Bedingung. Offenbar wurden die externen Repräsentationen Linie und Zahl unterschiedlich verarbeitet.

Auch dieses Experiment bestätigt, dass das Problemlöseverhalten der Teilnehmer in entscheidender Weise durch die direkt verfügbaren Informationen bestimmt wird, welche in der Repräsentation der Problemstellung enthalten sind. Eine geeignete externe (und damit auch intern verarbeitete) Repräsentation erleichtert offenbar das Operieren mit der Repräsentation und damit das Finden geeigneter Lösungsstrategien.

1.1.4.4 Fazit und Ausblick auf die Bedeutung des Vorverständnisses

Für erfolgreiches Schlussfolgern und Problemlösen sind zwei Faktoren entscheidend: Das Finden einer zutreffenden und zweckmäßigen Repräsentation der Ausgangslage und die Anwendung von Lösungsstrategien und Regeln auf das Lernmaterial, so dass schrittweise über Zwischenstufen der Zielzustand erreicht werden kann.

Eine gute Repräsentation der Ausgangslage zeichnet sich dadurch aus, dass mit ihr leicht diejenigen Operationen durchgeführt werden können, die zum Erreichen des Ziels führen. In die Repräsentation der Ausgangssituation fließt das Vorwissen und Vorverständnis (vgl. Gentner & Gentner, 1983) der Lernenden mit ein. Externe (bildliche) Repräsentationen können die Lernenden, sofern sie wissenschaftlich adäquat sind, dadurch entlasten, dass die Lernenden nicht alle relevanten Informationen innerhalb des Arbeitsgedächtnisses präsent halten müssen. Voraussetzung für die entlastende und unterstützende Funktion externer Repräsentationen ist jedoch, dass die externe Repräsentation wissenschaftlich angemessen und für die Bearbeitung der Aufgabe relevant ist. Insbesondere in der Physik erfordert Problemlösen, dass eine adäquate und zweckmäßige physikalische Repräsentation gefunden werden muss (diese ist oft bildlich). Auf diese Repräsentation können dann mathematische Operationen angewandt werden. Oft stellt sich aber bereits das Problem, dass die Lernenden erst gar nicht von einem adäquaten Vorverständnis ausgehen. In der Regel bringen Schüler, obgleich sie Novizen sind, in der Regel bereits vor Beginn des Fachunterrichts erworbenes Wissen über den Sachverhalt mit, das sich aus fachlicher Sicht später oft als unvollständig und fehlerhaft erweist (vgl. Wiesner, 1992a; 1992b; Langley et al., 1997; Duit, 1993, 2001; Duit, Roth, Komorek & Wilbers, 2001). In diesem Fall ist es nötig, das vorhandene Wissen der Schüler durch Umlernen zu verändern (Schnotz 2006b, S. 75). Insbesondere gewinnen Lernende im Bereich der Physik oft Konzepte aus ihrer Alltagserfahrung. Dieses physikalische Vorverständnis interagiert mit dem neuen Wissen, welches im Physikunterricht vermittelt wird und wirkt sich auf das Erlernen wissenschaftlicher Konzepte aus. Dabei erweisen sich diese naiven Konzepte gegenüber Änderungen oft als widerständig. In der Lern- und Entwicklungspsychologie sowie Science Education Research wird diese Thematik in den letzten Jahrzehnten unter dem Begriff Konzeptwechsel (Conceptual Change) diskutiert.

Jüngere Forschungsarbeiten nehmen die repräsentationalen Anforderungen beim Erwerb des konzeptuellen Verständnisses in den Blick. Im folgenden Kapitel wird zunächst ein Überblick zu den die theoretischen Grundlagen von Konzeptwechseltheorien gegeben. An späterer Stelle wird der Bezug zum Lernen

mit Repräsentationen hergestellt und anschließend empirische Ergebnisse zum Thema Aufbau von Konzeptwissen und Lernen mit multiplen Repräsentationen vorgestellt.

1.2 Die Rolle von Konzeptwechseln in Lehr-Lernprozessen

1.2.1 Vorwissen, Schülervorstellungen und Konzeptwechsel

Verschiedene Theorierichtungen zum Konzeptwechsel gehen von unterschiedlichen Annahmen aus, durch welche Entwicklungsmechanismen und unter welchen Bedingungen ein Konzeptwechsel stattfindet und wie dieser gefördert werden kann (vgl. Schnotz, 2006b). Angesichts der Vielzahl von unterschiedlichsten Theorien darüber, woher naive Konzepte stammen und wie Konzepte verändert werden können, ist es kaum möglich, dieser „beeindruckenden“ und „verwirrenden“ „Vielfalt von Positionen gerecht zu werden“ (zit. n. Stark 2002, S. 4). Daher werden im Folgenden die bekanntesten Theorien im Überblick dargestellt und zentrale Kritikpunkte an den jeweiligen Ansätzen benannt. Um die vorgestellten Theorien zu ordnen, werden zunächst die klassischen Ansätze beschrieben, anschließend einige Weiterentwicklungen aus kognitionspsychologischer Perspektive vorgestellt und abschließend pädagogische Implikationen aufgezeigt.

Eine umfassende Theorie zur Initiierung von Wissensveränderungen liegt bislang nicht vor (vgl. Tyson, et al., 1997). Konsens besteht jedoch über folgende Punkte (vgl. Vosniadou & Brewer, 1992; Tyson et al., 1997; Özdemir & Clark, 2007):

- Lernende gewinnen Konzepte aus Ihrer Alltagserfahrung. Sie kommen in den Unterricht mit einem Set an Konzepten über Naturphänomene und -ereignisse.
- Die anfänglichen Vorstellungen der Lernenden weichen häufig von den etablierten wissenschaftlichen Erklärungen ab (vgl. Vosniadou & Brewer, 1992, S. 536).
- Die naiven Vorstellungen der Lernenden beeinflussen das formale Erlernen wissenschaftlicher Konzepte. Dabei erweisen sich die naiven Konzepte gegenüber Änderungen oft als widerständig (vgl. Vosniadou & Brewer, 1992, S. 536).
- Mit klassischen Lehrstrategien ist es oft nicht möglich, die naiven Vorstellungen der Lernenden zu verändern, weil diese tief in den Alltagserfahrungen verankert sind. Lernprozesse, die einen Konzeptwechsel beinhalten, sind daher in der Regel recht zeitaufwändig.

- Um Lernenden wissenschaftliche Theorien und Konzepte zu vermitteln, ist es nötig, das Denken der Lernenden zu kennen und Instruktionen entsprechend anzupassen (vgl. Tyson et al., 1997; Özdemir & Clark, 2007).

Je nach theoretischer Ausrichtung werden die Vorstellungen, welche die Schüler in den Unterricht mitbringen, mit unterschiedlichen Begriffen bezeichnet: Schülervorstellungen (Wiesner, 1992a, 1992b), Alltagsvorstellungen (Duit, 1993), Fehlvorstellungen (Plötzner & Spada, 1998) bzw. „misconceptions“ (Chi, 1992; Chi, Slotta & DeLeeuw, 1994), „alternative frameworks“ (Driver & Easley, 1978), Präkonzepte (Ausubel, 1968), intuitives Wissen (Vosniadou & Brewer, 1992, S. 580) oder phänomenologische Primitive, p-Prims genannt (vgl. diSessa 1983, 1988, 1993). Den verschiedenen Begrifflichkeiten liegen dabei teilweise unterschiedliche theoretische Annahmen zugrunde (vgl. Wodzinski, 1996/2004, S. 23).

Die verschiedenen Termini zur Bezeichnung der Lernschwierigkeiten der Schüler werden wie folgt verwendet. Bei den Darstellungen der unterschiedlichen Theorien zur Förderung eines Konzeptwechsels (im nächsten Kapitel) wird jeweils der Begriff verwendet, den die Autoren selbst benutzen, um eine sachgerechte Darstellung der jeweiligen Theorie zu ermöglichen.

In den Kapiteln, in denen es um die empirische Studie geht, die im Rahmen der Dissertation durchgeführt wurde, fiel die Entscheidung für den Begriff „Schülervorstellung“. Die Nachteile des Begriffs liegen darin, dass er sich auf den Kontext der Schule und des entsprechenden Altersbereichs bezieht, obwohl ähnliche Vorstellungen auch bei Erwachsenen außerhalb des Schulunterrichts gefunden werden können (vgl. Zeyer, 1981).

Für den Begriff Schülervorstellungen spricht, dass er im Gegensatz zu den Begriffen Fehlkonzepthen oder Fehlvorstellung keine Wertung vornimmt.

Die Einordnung in den Kontext der Schule ist im Rahmen dieser Arbeit akzeptabel, weil sich die hier durchgeführte Studie auf den Umgang mit weitverbreiteten Vorstellungen von Mittelstufenschülern bezieht, die aus der fachdidaktischen Forschung bekannt sind.

Des Weiteren lässt der Begriff im Gegensatz etwa zu den Begriffen „Alltagsvorstellungen“, oder“ Präkonzepte“ offen, woher die Vorstellungen der Schüler stammen und ob diese zum Vorwissen der Schüler zählen, also vor Beginn des Unterrichts vorhanden waren oder erst im Physikunterricht selbst entstehen.

Der im Rahmen der empirischen Studie verwendete Begriff der Schülervorstellungen umfasst ausdrücklich neben Vorstellungen aus dem Vorwissen der Schüler auch im Unterricht gebildete Vorstellungen, die den fachlichen Anforderungen nicht genügen.

Nicht alle Vorstellungen, welche Schüler in den Unterricht mitbringen, müssen fehlerhaft oder mit wissenschaftlichen Konzepten unvereinbar sein. Daher wurde in der Zusammenfassung und in der Einleitung von naiven Schülervorstellungen gesprochen. Auf diesen Zusatz wird im Folgenden verzichtet: Wenn im zweiten Teil dieser Arbeit davon die Rede ist durch entsprechende Interventionen „weitverbreitete Schülervorstellungen“ verändern zu wollen, bezieht sich dies auf Vorstellungen, die einem wissenschaftlichen Verständnis entgegen stehen oder so unklar und vage sind, dass sie einer fachlichen Präzisierung bedürfen.

1.2.2 Kognitionspsychologische Theorien zum Konzeptwechsel

1.2.2.1 Konzeptwechsel durch die Überwindung kognitiver Konflikte

Einer der frühesten und bekanntesten Ansätze, die sich mit der Reorganisation von Konzepten befasst, ist Piagets Theorie der kognitiven Entwicklung (vgl. Piaget 1953/1979, 1973/1996). Nach Piagets Theorie (1953/1979) lässt sich die kognitive Entwicklung von Menschen durch universale Denkprozesse charakterisieren. Gemäß Piagets Ansatz (1953/1979) vollzieht sich die kognitive Entwicklung durch Veränderung übergreifender kognitiver Strukturen. Entsprechend erfolgt die Reorganisation von Wissen domänenübergreifend. Piaget (1973/1996) unterscheidet zwischen:

- funktionalen Invarianten (Strukturierung, Adaptation und Äquilibration), die über die gesamte Entwicklung konstant bleiben und allgemein die Lernprozesse lernfähiger Spezies beschreiben,
- „entwicklungsgesetzmäßige“ Varianten (Schemata und Strukturen), die sich entsprechend der verschiedenen Denkstadien verändern und also bestimmten „Gesetzmäßigkeiten“ der Entwicklung folgen,
- konkreten Inhalten, die von Individuum zu Individuum und von Augenblick zu Augenblick variieren. Für Piaget sind die Inhalte nur sekundär von Interesse, da sie zu wenig Regelmäßigkeit zeigen.

Unter der *funktionalen Invariante* der Strukturierung versteht Piaget die Tendenz, einzelne Beobachtungen in kohärente Wissenssysteme zu integrieren. Die funktionale Invariante der Adaptation hingegen beschreibt die Tendenz, durch die gegenseitige Anpassung von Organismus und Weltwahrnehmung (vgl. Flammer, 2003, S. 117) das eigene Denken und Verhalten mit der Umwelt in Einklang zu bringen.

Die Adaptation kann durch die beiden komplementären Mechanismen der *Assimilation* und der *Akkomodation* beschrieben werden. Bei der Assimilation werden Erfahrungswerte an die eigene geistige Struktur angepasst, die Wahrnehmung der Umwelt wird an die kognitiven Möglichkeiten des Organismus angeglichen (vgl. ebd., S. 117). Umgekehrt passt sich bei der Akkomodation der Organismus an die Realität an. Die kognitive Organisation wird also so verändert, dass sie mit den neuen Erfahrungen aus der Umwelt in Einklang steht (vgl. ebd., S. 118).

Aus den beiden elementaren funktionalen Invarianten der Strukturierung und der Adaptation ergibt sich als dritte funktionale Invariante die Äquilibration. Äquilibration ist das Streben des Menschen nach einem kognitiven Gleichgewicht zwischen der kognitiven Organisation und der Umwelt. In der alltäglichen Interaktion des Individuums mit der Umwelt wechseln sich die beiden Adaptationsprozesse der Assimilation und Akkomodation ständig ab. Im Wechselspiel aus beiden Prozessen wird ein jeweils neuer, höherer Gleichgewichtszustand erreicht. Die kognitive Äquilibration ist daher der Motor der Entwicklung.

Posner, Strike, Hewson und Gertzog (1982, S. 212) greifen die Idee, von kognitiven Konflikten als Motor der Entwicklung in ihrer Theorie zum Konzeptwechsel auf. Sie betrachten Lernen als eine Form von vernunftgeleitetem Forschungsprozess, bei dem Lernende Annahmen auf Basis verfügbarer, evidenter und einleuchtender Aussagen als erwiesen akzeptieren. Dabei stellen sie sich in erster Linie die Frage, wie sich die Konzepte der Lernenden unter dem Einfluss von neuen evidenten Informationen ändern. In Anlehnung an Piaget unterscheiden die Autoren zwei Formen des Konzeptwechsels: Bei der Assimilation, die auch den ersten Schritt eines länger währenden Lernprozesses bilden kann, nutzen die Lernenden ihre bestehenden Konzepte, um neue Phänomene zu verstehen. In denjenigen Fällen jedoch, in denen die bisherigen Konzepte sich nicht dafür eignen, neue Phänomene zu begreifen, müssen die Lernenden ihre bisherigen Konzepte reorganisieren oder durch neue Konzepte ersetzen. Diese tiefgreifende Form des Konzeptwechsels bezeichnen die Autoren Piaget folgend als Akkomodation (vgl. Posner et al., S. 212). Eine wichtige Voraussetzung für einen Konzeptwechsel im Sinn einer Akkomodation besteht darin, dass die Lernenden mit den bisherigen Konzepten unzufrieden sind. Stehen ihnen darüber hinaus alternative und verständlichere Konzepte zur Verfügung, die einleuchtender und ergiebiger zur Bearbeitung von Problemstellungen sind oder neue Verständnis- bzw. Lösungsansätze bereitstellen, findet ein Konzeptwechsel statt. Der Motor des Konzeptwechsels besteht also in kognitiven Konflikten. Gemäß dieser Theorie ist ein Konzeptwechsel an die folgenden vier Bedingungen geknüpft (vgl. Posner et al., S. 214; Strike & Posner 1982, S. 235):

1. Unzufriedenheit mit bestehenden Konzepten: Erst wenn die Lernenden vor einer Reihe ungelöster Probleme stehen oder daran zweifeln, dass ihre bisherigen Konzepte für die Lösung dieser Probleme ausreichen und daher das Vertrauen in die bisherigen Konzepte erschüttert wurde, entsteht die Bereitschaft, neue Konzepte zu akzeptieren.
2. Verständlichkeit der neuen alternativen Konzepte. Die Lernenden müssen in der Lage sein, zu begreifen wie neue Erfahrungen durch alternative Konzepte strukturiert werden können, um die Möglichkeiten, die sich aus diesen alternativen Konzepten ergeben, ausschöpfen zu können. Neue Konzepte müssen das Potential aufweisen, neue Erfahrungen so zu strukturieren, so dass das Potential der neuen Konzepte ausgeschöpft werden kann. Voraussetzung für Verständlichkeit ist eine ausreichende Wissensbasis, die es ermöglicht, neue Inhalte in bisherige Wissensbestände zu integrieren.
3. Plausibilität der neuen alternativen Konzepte. Die neuen Konzepte müssen glaubwürdig wirken. Glaubwürdigkeit resultiert auch aus der Konsistenz des Konzepts mit anderem Wissen, d.h. die neuen Konzepte dürfen nicht in einem Widerspruch zu ontologischen und epistemologischen Grundüberzeugungen der Lernenden stehen. Verständlichkeit ist eine notwendige, jedoch keine hinreichende Voraussetzung für Plausibilität (vgl. Posner et al., 1982, S. 216).
4. Fruchtbarkeit der neuen alternativen Konzepte. Die neuen Konzepte sollten die Möglichkeit eines ergiebigen Forschungsprogramms eröffnen. Sie sollten über die Lösung von Problemen, welche mit den bisherigen Konzepten nicht bewältigt werden konnten, hinaus, weitere Ansätze für das Verständnis von Problemen versprechen, neue Forschungsgebiete eröffnen, es ermöglichen, neue Erfahrungen zu erklären, zu neuen Einsichten zu gelangen. In dieser Weise sollten sie vermitteln hilfreicher bei der Erklärung von Sachverhalten und bei der Lösung von Problemen zu sein. Fruchtbarkeit hat Plausibilität und Verständlichkeit zur Voraussetzung.

Instruktionen sollten darauf zielen, Lernenden eine Brücke zu bauen, die es ihnen ermöglicht, von ihren bisherigen Konzepten zu neuen alternativen Konzepten zu gelangen (vgl. Strike & Posner, 1982, S. 238). Demgemäß geben die Autoren folgende Empfehlungen zu Lehrstrategien (vgl. Posner et al., 1982, S. 225, 226):

- Unterricht sollte darauf zielen, kognitive Konflikte bei den Lernenden hervorzurufen, indem Lernende mit unerwarteten Phänomenen konfrontiert werden.
- Lehrende sollten Zeit für die Diagnose von Denkfehlern und zur Identifikation von Widerständen gegen Akkomodationsprozesse einplanen.

- Im Unterricht sollten Strategien entwickelt werden, wie mit Denkfehlern der Lernenden und Widerständen gegen Akkomodationsprozesse umgegangen werden kann.
- Lernende sollten darin unterstützt werden, Lerninhalte auf verschiedene Weise zu repräsentieren „verbal, mathematisch, konkret-praktisch oder bildlich“ (zit. n. Posner et al., 1982, S. 226, Übersetzung durch die Autorin) sowie Lerninhalte von einem Modus der Repräsentation in einen anderen zu übersetzen.
- Es sollten Techniken entwickelt werden, welche die Lehrenden darin unterstützen, die Präkonzepte der Schüler zu bestimmen und ihnen dabei helfen, die Lernenden auf den Weg des Konzeptwechsels zu bringen.

Die klassischen Ansätze von Piaget (1953/1979, 1973/1996) sowie von Posner et al., 1982 bzw. Strike und Posner (1982) wurden nachhaltig kritisiert.

Der zentrale Kritikpunkt an Piagets Theorie (1953/1979, 1973/1996) besteht darin, dass sich die kognitive Entwicklung nicht domänenübergreifend vollzieht, sondern als domänenspezifischer Prozess der Umstrukturierung betrachtet werden sollte (vgl. Carey, 1985). Nach Carey (1985, S. 273) starten Kinder mit einigen wenigen theorieähnlichen Strukturen, die in intuitive Theorien eingebettet sind. So sind z.B. nach Carey (1985) das Konzept „Gewicht“ und das Konzept „Dichte“ in eine intuitive Theorie über Masse eingebettet. In Lernprozessen kann sowohl eine schwache als auch eine starke Form der Umstrukturierung stattfinden. Zu einer völlig neuen Konstruktion von Theorien oder dem Entstehen neuer Paradigmen aus bestehenden Strukturen kommt es jedoch nur durch die starke Umstrukturierung. Diese Form der Umstrukturierung wird nicht als Ergebnis der logischen Denkfähigkeit des Kindes per se betrachtet, sondern als Produkt des wachsenden Wissens über eine Domäne gesehen.

Was die Strategie der Erzeugung kognitiver Konflikte zur Förderung eines Konzeptwechsels nach Strike und Posner (1982) anbelangt, belegen empirische Studien, dass das Auslösen kognitiver Konflikte häufig nicht zum gewünschten Erfolg führt.

- Forschungsergebnisse von Chan, Burtis und Bereiter (1997) weisen darauf hin, dass zwischen externalen und internalen Konflikten unterschieden werden muss. Nicht jede widersprüchliche Information führt auch zur Umstrukturierung von Konzepten. In etlichen Fällen werden die neuen Informationen in bestehende Wissensbestände integriert, ohne dass ein kognitiver Konflikt entsteht.

- Ein weiteres Problem besteht darin, dass nicht jeder Konflikt, der als bedeutungsvoller Konflikt intendiert war, auch als solcher wahrgenommen wird (vgl. Dreyfus, Jungwirth & Eliovitch, 1990). Ohne die Relevanz der neuen Informationen zu erkennen, verspüren die Schüler nicht die Neugier, sich tiefgehend mit den Lerninhalten auseinanderzusetzen und sind in der Konsequenz hierzu auch nicht ausreichend motiviert.
- Die Wahrscheinlichkeit, einen kognitiven Konflikt zu erleben, sinkt ebenfalls, wenn die Lernenden nicht in der Lage sind die neuen Fakten zu bewerten oder zu erkennen, dass ihre bisherigen Annahmen den neuen Fakten widersprechen (vgl. Kuhn, Amsel & O'Loughlin, 1988; Kuhn, 1991).
- Schließlich ist es auch möglich, dass zwar eine verständliche Alternative zur Verfügung steht, diese jedoch nicht besonders plausibel erscheint. In diesem Fall verstehen die Lernenden die neuen Konzepte, ohne ihnen jedoch zu trauen, was ebenfalls nicht zum Konzeptwechsel führt (vgl. Schnotz, Vosniadou & Carretero, 1999, S. 14).

1.2.2.2 Konzeptwechsel durch die Modifikation mentaler Modelle

Aus Perspektive der neueren kognitiv entwicklungspsychologischen Forschung ist zu beachten, dass Lernende allgemeine epistemologische Grundannahmen zu verschiedenen Inhaltsbereichen ausbilden, welche die Veränderung einzelner inhaltlicher Aspekte mit beeinflussen (vgl. Vosniadou & Brewer, 1992; 1994). Die Konzeptwechseltheorie von Vosniadou und Brewer (1992, 1994) berücksichtigt, dass Wissensveränderung von domänenspezifischem Wissen vor dem Hintergrund dieser allgemeinen Vorstellungen z.B. über das Wesen von Naturwissenschaften zu betrachten ist. Gemäß der Theorie von Vosniadou und Brewer (1992, 1994) konstruieren Kinder mentale Modelle auf Basis ihrer Alltagserfahrung. Da Konzepte in umfassendere theoretische Strukturen eingebettet sind, unterscheiden die Autoren zwischen domänenspezifischen Theorien und allgemeinen Rahmentheorien.

Allgemeine Rahmentheorien bestehen aus ontologischen und epistemologischen Überzeugungen (vgl. Krüger, 2007, S. 87). Rahmentheorien stellen Lernenden relativ kohärente Erklärungssysteme zur Verfügung. Sie entwickeln sich sehr früh in der Lerngeschichte und werden über Jahre hinweg oft im Alltag bestätigt. Eigene Erfahrungen und kulturell vermittelte Informationen werden nun vor dem Hintergrund dieser Rahmentheorien interpretiert. Aus dieser Interpretation resultiert ein Muster von Annahmen, aus dem sich nun eine domänenspezi-

fische Theorie bildet (Stark, 2002, S. 5). Spezifische Theorien bestehen aus sich wechselseitig bedingenden Propositionen oder Überzeugungen über Phänomene des jeweiligen Inhaltsbereichs. Domänenspezifische Theorien sind in allgemeine Rahmentheorien eingebettet. Die allgemeinen Rahmentheorien bilden den ontologischen und epistemologischen Rahmen, der die domänenspezifischen Konzepte mitbestimmt.

Im Fall spezifischer Theorien über Physik beziehen sich die Propositionen der spezifischen Theorie auf die Eigenschaften und das Verhalten physikalischer Objekte. Naive Rahmentheorien über Physik bestehen aus ontologischen und epistemologischen Grundprinzipien, die den Prozess des Erwerbs von Wissen über die physikalische Welt anleiten. Vosniadou (1994) stützt sich hier beispielsweise auf Forschungsergebnisse von Spelke (1991) (zit. n. Vosniadou, 1994, S. 47). Spelke (1991) beschreibt fünf Grundbedingungen des Verhaltens physikalischer Objekte, mit denen Kinder bereits im Alter von wenigen Monaten rechnen: Kontinuität, Solidität, keine Interaktion über Distanz hinweg, Schwerkraft und Trägheit. Solche Grundannahmen, die eine naive Rahmentheorie konstituieren, sind unbewusst und werden daher auch nicht in Form von Hypothesentests überprüft (vgl. Vosniadou, 1994, S. 47). Die frühkindlichen Rahmentheorien ermöglichen es den Kindern dadurch, dass sie ihnen intuitives Wissen über die physikalische Welt bereitstellen, sich in der Welt zurechtzufinden und bilden somit die Basis für weitere Lernprozesse. In späteren Lernprozessen können sie jedoch den Erwerb von wissenschaftlichen Konzepten behindern, weil wissenschaftliche Erklärungen oft die fundamentalen Prinzipien der intuitiven Physik verletzen, die sich im Alltag bewährt haben. Wenn Lernende wissenschaftliche Informationen interpretieren, entstehen oft kognitive Konflikte, welche die Lernenden auflösen müssen. Bei einer reinen Assimilation neuer Informationen kommt es leicht zu inkonsistenten Erklärungen (vgl. Stark, 2002, S. 6). Die Lernenden versuchen, diese Inkonsistenzen aufzulösen, indem sie ihre bisherigen Konzepte und die neuen Informationen in einem synthetischen Modell zusammenfügen. Diese synthetischen Modelle sind das Ergebnis der kreativen Versuche der Lernenden, kohärente Konzepte zu entwickeln, in denen das bisherige alltagsbasierte Wissen und die neuen Informationen integriert werden. Sie erscheinen dabei manchmal konsistenter als sie es tatsächlich sind.

Lernen in den Naturwissenschaften erfordert daher die radikale Reorganisation bestehender konzeptueller Strukturen, die Bildung qualitativ anderer Repräsentationen und nicht nur eine Anreicherung bestehender Konzepte mit neuem Wissen (vgl. Vosniadou, Ioannides, Dimitrakopoulou & Papademetriou, 2001, S. 384).

Bevor Kinder Wissen von Erwachsenen vermittelt bekommen, entwickeln sie zunächst Initialmodelle, die auf ontologischen und epistemologischen Grundannahmen aus der frühen Kindheit basieren. Ein solches Initialmodell ist z.B. die Vorstellung, die Erde sei eine Scheibe. Das Scheibenmodell entspricht der Alltagserfahrung und ist vom wissenschaftlichen Modell der Erde als Kugel noch unbehelligt.

In weiteren Lernprozessen nehmen die Kinder kulturell vermittelte Wissensinhalte, die von Erwachsenen stammen, in die Initialmodelle mit auf und formen auf diese Weise die Initialmodelle zu synthetischen Modellen um. In diesen synthetischen Modellen versuchen die Kinder aus den Informationen, die sie von Erwachsenen und aus ihren Alltagserfahrungen erhalten, kohärente mentale Modelle aufzubauen, welche sie in konsistenter Weise verwenden können (vgl. Vosniadou & Brewer, 1992, S. 580).

Zunächst integrieren die Kinder bei diesen synthetischen Modellen (vgl. ebd., S. 579) kulturell vermittelte Informationen in einer Weise, die es ihnen ermöglicht, so viele ihrer Vorannahmen beizubehalten wie möglich. Ein solches Modell ist z.B. die Vorstellung, es gäbe zwei Erden – eine Scheibe, auf der die Menschen leben und eine zweite Erde als Körper im Raum. Die wissenschaftliche Information wird hier einfach dem bisherigen Konzept hinzugefügt.

Fortgeschrittene synthetische Modelle zeichnen sich dadurch aus, dass einige der Vorannahmen revidiert wurden. Die Revision einer Vorannahme befreit die Kinder von den Begrenzungen, welche die Vorannahme über die mentalen Modelle der Lernenden verhängt und ermöglicht es, eine andere Klasse mentaler Modelle zu bilden.

Solche Modelle sind z.B. die Vorstellung, die Erde sei eine Hohlkugel, bei der die Menschen auf einer Ebene in der Kugel leben oder das Modell der Erde als abgeflachte Kugel. Zwar haben die Kinder das Konzept der Schwerkraft noch nicht komplett verstanden, jedoch die Annahme revidiert, dass der Untergrund der Erde flach sein muss. Gegen Ende der Grundschulzeit haben die meisten Kinder schließlich ein astronomisch korrektes Erdmodell entwickelt.

Bei Konzeptwechseln handelt es sich nach dieser Theorie also um einen langsamen und allmählichen Prozess, der von Initialmodellen über synthetische Modelle zu wissenschaftlich korrekten Modellen führt. Vosniadou unterscheidet hier die folgenden verschiedenen Formen des Konzeptwechsels:

- Bei der Anreicherung der bisherigen konzeptuellen Strukturen („enrichment“) werden neue Informationen zu bestehenden theoretischen Rahmentheorien hinzugefügt. Sind diese Informationen konsistent mit dem Vorwissen, fällt die Anreicherung den Lernenden leicht.

- Zur Revision kommt es, wenn die neuen Lerninhalte inkonsistent zu bestehenden Überzeugungen, Vorannahmen oder den relationalen Strukturen einer Theorie sind. Die Revision einer bestehenden Theorie ist dabei leichter zu vollziehen als die Revision einer Rahmentheorie. Schwierig wird es, wenn die Überzeugungen, die eine spezifische Theorie konstituieren, in Abhängigkeit einer Rahmentheorie konstruiert wurden. In diesem Fall kann ein Konzeptwechsel nur vollzogen werden, wenn auch die Rahmentheorie revidiert wird. Die Veränderung einer Rahmentheorie ist aus zwei wesentlichen Gründen schwer zu erreichen, erstens sind sich die Lernenden der impliziten ontologischen und epistemologischen Grundannahmen nicht bewusst und zweitens bilden die Grundannahmen der Rahmentheorie ein relativ kohärentes System von Erklärungen, das sich auf Alltagserfahrungen gründet, die über Jahre bestätigt wurden. Diese zu verändern bedeutet, „tief verwurzelte Erfahrungen zu erschüttern“ (zit. n. Krüger, 2007, S. 87).

1.2.2.3 Konzeptwechsel durch die Überwindung von Kategorisierungsfehlern

Ein weiterer kognitionspsychologischer Ansatz über die Natur naiver Konzepte oder Fehlkonzepte („naive or (mis-)conceptions“, zit. n. Chi, 1992, S. 139) wurde von Chi (1992) entwickelt. Die Theorie geht von der Annahme aus, dass Individuen während ihrer kognitiven Entwicklung drei wesentliche ontologische Kategorien erwerben: Dinge, Prozesse und mentale Zustände. Dinge, z.B. Felsen und Gebäude, weisen Eigenschaften auf, wie „ist rot“ oder „nimmt Raum ein“ auf. Prozessen, z.B. Pausen oder Unwetter, werden Eigenschaften wie „dauert eine Stunde“ oder „geschah gestern“ zugeschrieben. Mentalen Zuständen, wie Träumen oder Bildern, werden Eigenschaften wie „ist wahr“ oder „handelt von“ zugeordnet. Diese übergeordneten Kategorien (Dinge, Prozesse und mentale Zustände) werden in weitere Unterkategorien eingeteilt. Solche Subkategorien, die verschiedenen ontologischen Kategorien angehören, haben keine gemeinsamen Eigenschaften; auch Subkategorien, die derselben ontologischen Kategorie angehören, sind ontologisch verschieden. Dinge, die zur gleichen Hauptkategorie, aber zu verschiedenen Subkategorien zählen, können jedoch ontologische Eigenschaften höherer Ordnung teilen, so können z.B. ein Hund (Hauptkategorie Ding, Subkategorie Naturdinge und Tier) und ein Messer (Hauptkategorie Ding, Subkategorie: Artefakte) eine gleiche übergeordnete Eigenschaft aufweisen, etwa die Eigenschaft „ist braun“. Entitäten unterschiedlicher ontologischer Kategorien hingegen können keine gemeinsamen übergeordneten Eigenschaften besitzen.

„Fehlkonzepte“ (misconceptions) entstehen nach Chi, Slotta und deLeeuw (1994), wenn ein Individuum ein Konzept, das eigentlich einer bestimmten ontologischen Kategorie angehört, einer anderen Kategorie zuordnet. Beispielsweise subsumieren Individuen das Konzept „Hitze“ oft unter der Kategorie „Ding“ statt unter der Kategorie „Prozess“. Folglich betrachten die Lernenden Hitze als Eigenschaft, die durch Kontakt von einem Objekt auf ein anderes übertragen werden kann. Analoges gilt für den Begriff der Kraft, der Kraftbegriff in der Physik fällt in die Kategorie der Prozesse, bzw. unter die Subkategorie der Interaktion zwischen Objekten. Lernende ordnen den Begriff der Kraft jedoch oft der Kategorie der Dinge bzw. der Eigenschaften von Dingen zu (vgl. Chi & Slotta, 1993, S. 255). Kraft wird als eine Form von Impetus verstanden, den ein Körper vermittelt oder als intensive Größe, die ein Körper besitzt (ähnlich der Geschwindigkeit). Die Entstehung von Fehlkonzepten resultiert gemäß dieser Theorie also aus ontologischen Kategorisierungsfehlern.

Eine wesentliche Schwierigkeit dieses Ansatzes besteht in der mehr oder weniger „willkürlichen“ Festlegung ontologischer Kategorien (vgl. Duit, 1999, S. 268; Schnotz et al., 1999, S. 16). So kritisiert Duit (1999), dass die Änderung des konzeptuellen Verständnisses des Kraftbegriffs elaborierterer Kategorien bedarf als die von Chi und Slotta (1993) verwendeten. Des Weiteren wird der Erwerb von Schlüsselkonzepten durch den Wechsel von einer Kategorie zu einer anderen oft nicht adäquat beschrieben. In Bezug auf das Alltagskonzept von Hitze müsse, so Duit (1999), das Hitzekonzept differenziert werden und die physikalische Erklärung der Hitze in die Konzepte der Temperatur, Wärmeenergie und Entropie eingeordnet werden. Eine weitere Schwierigkeit des Ansatzes besteht darin, dass die Theorie nur schwer erklären kann, warum ein Konzeptwechsel zwischen Kategoriengrenzen schwieriger zu erreichen ist, ist als ein Konzeptwechsel innerhalb von Kategorien (vgl. Schnotz et al., 1999, S. 16). Vosniadou (1994) argumentiert, dass die Theorie nur eine syntaktische und nicht eine semantische Erklärung eines Konzeptwechsels zur Verfügung stelle. Den Vorteil ihres eigenen Ansatzes sieht sie darin, dass die Sichtweise von Konzepten, die in allgemeine Rahmentheorien eingebettet seien, einen Konzeptwechsel umfassender erklären könne. So werde in Vosniadous Rahmentheorieansatz nicht nur der Wechsel ontologischer Kategorien berücksichtigt, sondern auch die Veränderung epistemologischer Grundannahmen.

1.2.2.4 Konzeptwechsel durch die Reorganisation von Wissensfragmenten

Eine andere kognitionspsychologische Richtung charakterisiert das anfängliche Wissen der Lernenden als Sammlung unstrukturierter Fragmente, die mehr oder weniger beziehungslos nebeneinander existieren (vgl. diSessa, 1988, 1993; Hunt & Minstrell, 1994; Linn, Eylon, & Davis, 2004; Clark, 2006)

Einer der bekanntesten Ansätze, der sich explizit mit Physiklernen befasst, stammt von diSessa (1983, 1988, 1993). Dieser Ansatz nimmt an, dass das naive „Alltagswissen“ der Lernenden anfänglich aus unstrukturierten isolierten Bruchstücken besteht. Diese Wissensfragmente bezeichnet er als phänomenologische Primitive (p-Prims). Dabei handelt es sich um kognitive Schemata, die eine oberflächliche Interpretation von Beobachtungen zusammenfassen. P-Prims sind relativ begrenzte Abstraktionen von einfachen und gewöhnlichen Phänomenen (vgl. diSessa 1983, S. 32). Es handelt sich um Zusammenstellungen von (wieder-)erkennbaren Erscheinungen, welche die Sicht auf die Welt beschreiben und teilweise auch Erklärungen zur Verfügung stellen. Dementsprechend orientieren sie sich eng an der Beobachtung und weisen nur einen geringen Grad an Abstraktion und Theoriebildung auf (vgl. diSessa 1983, S. 16). Wenn Lernende beginnen, sich mit dem Fach Physik auseinanderzusetzen, beginnen sie mit einem naiven physikalischen Verständnis, das aus einer Ansammlung von Phänomenen besteht, über das, was sie in der Welt beobachten. Diese p-Prims sind jedoch nicht hierarchisch geordnet und nur lose miteinander verbunden. Sie werden in dem Moment aktiviert, indem eine Situation eintritt, in der sie Relevanz aufweisen.

P-prims entstehen, z.B. durch unser intuitives Verständnis für Mechanik, indem wir unserer Interaktionen mit unserer physikalischen Umgebung reflektieren, z.B. wenn wir einen Gegenstand stoßen, halten oder werfen. Solche impliziten Vorannahmen beeinflussen die Schlussfolgerungen der Lernenden, wenn sie die physikalische Welt analysieren (vgl. diSessa; 1993, S. 112).

Aufgrund mangelnder metakognitiver Fähigkeiten fällt es den Lernenden zunächst nicht auf, dass ihr Wissen fragmentarisch und wenig kohärent ist. Metakognition bedeutet in diesem Kontext, das Bewusstsein, dass Wissen nur den Charakter vorläufiger Hypothesen aufweist, die zu einem späteren Zeitpunkt auch revidiert werden können. Da Novizen ein solches metakonzeptuelles Bewusstsein über den epistemologischen Status des eigenen Wissens häufig fehlt, werden Theorie und Beobachtung nicht ausreichend koordiniert. (vgl. Schnotz et al., 1999, S. 15; Schnotz, 2006b, S. 77). Häufig überprüfen Lernende zu Beginn also nicht, ob ihr themenbezogenes Wissen ein kohärentes Ganzes bildet und bemerken daher auch nicht, dass ihr Wissen bruchstückhaft oder teilweise widersprüchlich ist. In

der Folge werden unerwartete Beobachtungen entweder ignoriert oder dahingehend uminterpretiert, dass sie mit den bisherigen theoretischen Annahmen kompatibel sind.

Die anfänglichen Konzepte werden im Lernprozess allmählich abstrakter. Unwesentliche situationsspezifische Details, die nicht dazu beitragen, den physikalischen Sachverhalt zu erklären, werden weggelassen. Dieser Prozess kann ausgehend von den p-Prims starten: so können einige der p-Prims mit formalen wissenschaftlichen Konzepten in der Physik kompatibel sein als andere und somit bestätigt werden. Sie erhalten in der Folge eine höhere Priorität und werden bereitwillig bei physikalischen Analysen und Erklärungen verwendet (vgl. diSessa, 1983, S. 16).

Diese Konzepte werden nun zunehmend auf alle Fälle angewandt, die sich physikalisch unter dem Sachverhalt subsumieren lassen (vgl. diSessa, 1983; S. 23). Des Weiteren spielen sie nicht mehr die gleiche naive Rolle von sich für sich selbst sprechenden Erklärungen und müssen anderen Analysen den Vortritt lassen, wenn sich diese als geeigneter erweisen (vgl. diSessa, 1983, S. 32).

Ein Konzeptwechsel findet also statt, wenn es Lernenden gelingt, phänomenologische Prinzipien in komplexere und adäquatere konzeptuelle Strukturen einzubinden und das Wissen zu systematisieren. Aus dem anfänglich nur wenig kohärenten Alltagswissen entsteht somit zunehmend ein zusammenhängendes Ganzes.

Die anfänglich vorhandenen p-Prims haben als Wissensfragmente nichts mit der Systematik von Theorien gemein (vgl. diSessa, 1988, S. 50). Hierin unterscheidet sich diSessas Ansatz also deutlich gegenüber dem Vosniadous. Nach diSessas Theorie werden während des Prozesses des Konzeptwechsels, die Elemente und die Interaktionen zwischen den Elementen revidiert und durch Hinzufügen, Verwerfen und Reorganisation verändert. Hierdurch verdichtet und verstärkt sich das Wissensnetzwerk der Lernenden. Aus dieser Perspektive wird unter einem Konzeptwechsel also ein evolutionärer Prozess verstanden, der Stück für Stück stattfindet und nicht als Prozess, in welchem naive Theorien durch fortgeschrittenere Theorien ersetzt werden.

Der Autor betont, dass der Unterschied zwischen Novizen und Experten weniger in der Art des Wissens oder den Inhalten bestünde, sondern vielmehr in der Organisation des Wissens (vgl. diSessa 1983, S. 32). Das Wissen der Novizen reiche strukturell nicht hin, ein physikalisches Phänomen durch systematische prinzipiengeleitete Zusammenhänge zu erklären. Experten hingegen besäßen ein erheblich tieferes und komplexeres System von Wertigkeiten (priority System) (vgl. ebd., S. 33).

Neuere Forschungsergebnisse bestätigen diSessas Annahmen. Southerland, Abrams, Cummins und Anzelmo (2001) erforschten die Wissensstrukturen von Lernenden im Fach Biologie. Sie fanden heraus, dass p-Prims eine größere Erklärungskraft bezüglich der Veränderung von Konzepten zu biologischen Phänomenen aufweisen als Theorien, die annehmen, Konzepte seien in Rahmentheorien eingebettet. Eine Längsschnittstudie von Clark (2006) zeigt, dass das Verständnis von Hitze und Temperatur durch die Perspektive von miteinander verbundenen Elementen erklärt werden kann. Özdemir und Clark (2007) untersuchten die Wissensstrukturen von Lernenden vom Kindergartenalter bis zu High School und bestätigten, dass Wissensstrukturen am besten durch eine „Wissens-als-Elemente-Perspektive“ beschrieben werden kann.

1.2.2.5 Fazit zu den Konzeptwechseltheorien und der praktischen Förderung

Je nachdem, welche psychologischen Mechanismen als entscheidend für einen Konzeptwechsel angesehen werden, resultieren unterschiedliche Konsequenzen der praktischen Förderung eines Konzeptwechsels (vgl. Schnotz, 2006b, S. 79). Wenn die naiven Vorstellungen der Lernenden darauf zurückgeführt werden, dass die Lernenden mentale Modelle konstruieren, welche von unpassenden Rahmentheorien beschränkt werden, sollten Instruktionen darauf zielen, diese epistemologischen und ontologischen Grundannahmen zu verändern. Naive Konzepte, welche die Lernenden in die Lernsituation einbringen, sollten im Licht der neuen theoretischen Annahmen reinterpretiert werden.

Andere Schlussfolgerungen ergeben sich, wenn man davon ausgeht, dass die anfänglichen Vorstellungen der Lernenden inkohärent sind, da es den Lernenden an den metakognitiven Fähigkeiten mangelt, den epistemologischen Status des eigenen Wissens zu überprüfen. Dann sollten die Lernenden dazu angeregt werden, zwischen theoretischen Annahmen und Beobachtungen zu unterscheiden sowie die Kohärenz der verschiedenen Teilbereiche des Wissens kritisch zu durchdenken.

Wird die Ursache für Schwierigkeiten, einen Konzeptwechsel zu vollziehen, in einer inadäquaten Kontextualisierung gesehen, müssen die Lernenden darin unterstützt werden, zwischen verschiedenen Kontexten zu differenzieren und zu erkennen, in welchem Kontext ihr Wissen relevant ist. Zudem sollte darauf geachtet werden, dass die Lerninhalte in einem bedeutungsvollen Kontext eingebettet sind, welcher die Lernenden persönlich anspricht.

Für die praktische Gestaltung von Lehr-Lernumgebungen ergeben sich im Rückblick auf die empirischen Untersuchungen, die von den Autoren der unterschiedlichen Ansätze durchgeführt wurden folgende Bedingungen, welche für das Erreichen eines Konzeptwechsels hilfreich sind (vgl. Schnotz, 2006b, S. 80):

- Die Lerninhalte werden auf die vorhandenen Lernvoraussetzungen abgestimmt. Die Abstimmung auf das Vorwissen der Lernenden zielt darauf, den Lernenden bewusst zu machen, dass sie ihre bisherigen Annahmen überprüfen müssen, ohne sie zu stark überfordern, so dass die neuen Erklärungen eine echte verständliche Alternative bieten (diese Konsequenz kann mehr oder weniger aus allen vorgestellten Ansätzen gezogen werden).
- Die bisherigen Sichtweisen der Lernenden werden reflektiert und im Licht der neuen Annahmen reinterpretiert, so dass die Lernenden ihr Denken über einen Sachverhalt von inadäquaten Rahmenannahmen befreien können (siehe Rahmentheorieansatz).
- Die Lernenden werden aktiviert, ihre bisherigen Annahmen kritisch zu durchdenken und ihre Wissensbestände auf Kohärenz hin zu überprüfen (siehe Konzeptwechsel als Reorganisation von Wissensfragmenten).
- Das Wissen wird als Werkzeug betrachtet, das sich in bestimmten Kontexten bewähren muss.

1.3 Interaktionen:

Lernen aus Experimenten unter der Perspektive des Umgangs mit multiplen Repräsentationen zur Förderung des konzeptuellen Verständnisses

1.3.1 Empirische Befunde: Multiple Repräsentationen und Konzeptwechsel

Im letzten Kapitel wurde ein Überblick zum theoretischen und empirischen Forschungsstand von Konzeptwechseltheorien gegeben. Dabei hatte sich gezeigt, dass bislang keine abschließende Erklärung dazu besteht, woher (naive) Schüler Vorstellungen stammen und wie diese Vorstellungen verändert werden können. Dennoch lassen sich aus den bestehenden Ergebnissen einige Empfehlungen ableiten, wie mit solchen Vorstellungen umgegangen werden kann.

Ergebnisse der Expertiseforschung und Studien über die Auswirkung unterschiedlicher Repräsentationen auf Problemlöseprozesse und schlussfolgerndes

Denken belegen, dass eine wissenschaftlich adäquate und für den Kontext relevante Repräsentation der Ausgangslage Problemlöseprozesse vereinfacht, beschleunigt und zu einer höheren Erfolgsquote führt.

Einige jüngere Studien nehmen die repräsentationalen Anforderungen beim Erwerb wissenschaftlicher Konzepte in den Blick. In einer Studie von Plötzner und Spada (1998) wurde der Einfluss von Schülervorstellungen auf den Umgang mit Repräsentationen untersucht. Plötzner und Spada (1998) analysierten die Konstruktion quantitativer Problemrepräsentationen auf Basis qualitativer Schlussfolgerungen. Sie entwickelten ein psychologisch motiviertes kognitives Simulationsmodell namens Sepia im Bereich der klassischen Newton'schen Mechanik. Mit Hilfe des Simulationsmodells sollte analysiert werden, wie qualitative Problemrepräsentationen zum Vorteil quantitativer Problemrepräsentationen genutzt werden können. Die Entwicklung des Sepia-Modells verfolgte vier Ziele:

1. Die *Untersuchung von Wissensstrukturen*, die den quantitativen und qualitativen Problemrepräsentationen zu Grunde liegen.
2. Die *Identifikation von Schlussfolgerungsmechanismen*, die es ermöglichen, qualitative und quantitative Problemrepräsentationen zu konstruieren und zu koordinieren.
3. Die *Untersuchung, welche fehlerhaften qualitativen Konzepte* die Autoren sprechen in diesem Kontext von „Fehlvorstellungen“ (misconceptions) die Konstruktion qualitativer und quantitativer Problemrepräsentationen beeinflussen.
4. Die Exploration wie *die Gestaltung von Lernumgebungen* von den Einsichten, die auf Basis von Sepia gewonnen wurden, profitieren kann.

Zur Bewertung der Vorhersagekraft von Sepia wurde eine empirische Studie mit 84 Schülern eines technischen Gymnasiums durchgeführt. Im Rahmen der Studie bearbeiteten die Schüler Aufgaben zur klassischen Mechanik. Während der Aufgabebearbeitung verfügten die Teilnehmer über zwei Tabellen. Die erste Tabelle beinhaltet alle quantitativen Angaben, die für die gestellte Fragestellung relevant sind; die zweite Tabelle beinhaltete eine kurze Zusammenfassung über die Zerlegung von Vektoren in ihre Komponenten. Die Teilnehmer wurden angehalten, nur die zur Verfügung gestellten Materialien zu nutzen und die einzelnen Bearbeitungsschritte während des Lösungsprozesses möglichst detailliert zu notieren. Eine der zentralen „Fehlvorstellungen“, welche einer erfolgreichen Aufgabebearbeitung im Weg stehen, ist das Impetuskonzept. Lernende, die vom Impetus-

konzept ausgehen, nehmen an, dass ein Körper, der sich gleichbleibend bewegt, über eine ihm innewohnende Größe verfügt, die seine Bewegung aufrechterhält. Gemäß dem Impetuskonzept verlangsamt sich ein Körper dadurch, dass sich diese Größe „aufbraucht“. Viele Schüler identifizieren den Impetus mit der Kraft.

Das wissenschaftlich adäquate Konzept lässt sich in den drei Newton'schen Axiomen zusammenfassen:

1. *dem Trägheitsprinzip*: Ein Körper beharrt im Zustand der Ruhe oder der gleichförmig geradlinigen Bewegung, solange keine äußeren Einflüsse auf ihn wirken. Die Geschwindigkeit eines solchen sich „frei“ bewegendes Körpers ist nach Betrag und Richtung konstant.
2. *dem Beschleunigungsprinzip*: Durch einwirkende Kräfte erfährt ein Körper eine Beschleunigung, die der Kraft proportional ist und deren Richtung besitzt und
3. *dem Wechselwirkungsprinzip* (actio = reactio): Übt ein Körper A auf einen Körper B eine Kraft aus (actio), so übt auch B auf A eine Kraft aus, die Gegenkraft (reactio) genannt wird. Die Gegenkraft ist gleich der ersten Kraft ist und wirkt dieser entgegengesetzt.

Die Diagnose, dass ein Schüler vom Impetuskonzept ausgeht, wurde immer dann gestellt, wenn der Schüler sich mündlich oder schriftlich auf eine Kraft in Richtung der Bewegung bezog. Beispielsweise erhielten die Schüler die Skizze der Wurfbahn eines senkrechten Wurfes und sollten zeichnerisch und in Worten erklären, welche Kräfte bzw. welche Kraft bei einem senkrechten Wurf an verschiedenen Stellen der Wurfbahn einwirken bzw. einwirkt. Schüler, die vom Impetuskonzept ausgingen, zeichneten häufig eine nach oben gerichtete Kraft in Bewegungsrichtung des Körpers ein. Tatsächlich wirkt an jeder Stelle der Wurfbahn nur die nach unten gerichtete Gewichtskraft, die gemäß des zweiten Newton'schen Axiomes auch die Ursache der Geschwindigkeitsänderung des Körpers ist.

Die Analyse der Aufgabenbearbeitung der Schüler zeigte, dass die Anzahl korrekter Lösungen signifikant sank, wenn diese vom Impetuskonzept ausgingen ($\chi^2 = 13,26$, $df = 2$, $p < .01$). Lag ein Impetuskonzept vor, waren entweder Lösung oder der Lösungsweg fehlerhaft oder die Schüler waren überhaupt nicht dazu in der Lage, eine Lösung zu finden. 12 Schüler wandten unzulässige algebraische Transformationen an, die nicht zur richtigen Lösung führten. Drei Schülern gelang es trotz Impetuskonzept auf die richtige rechnerische Lösung zu kommen. Einer dieser drei Schüler ließ das Impetuskonzept außer Acht, zwei dieser drei Schüler erzielten ein rechnerisch korrekte Lösung, indem sie ausgehend vom

Impetuskonzept eine korrekte Repräsentation der Kraft entwickelten und anschließend die richtigen algebraischen Operationen anwandten. 25 Schüler (89%) mit einem Impetuskonzept waren nicht dazu fähig, die Ergebnisse ihrer qualitativen Schlussfolgerungen algebraisch auszudrücken. Die Fehlkonzepte der Schüler beeinflussten also offenbar entscheidend die qualitative und die quantitative Problemrepräsentation. Da die quantitative Problemrepräsentation der Schüler auf ihrem qualitativen Verständnis beruht, konnten die Schüler entsprechend seltener geeignete Operationen zur Problemlösung anwenden, was den Lösungserfolg minderte. Die Ergebnisse weisen darauf hin, dass ein gutes konzeptuelles Grundverständnis eine wichtige Basis für darauf aufbauende abstrakte Lerninhalte bildet. Instruktionen, die hier ansetzen, können Lernende darin unterstützen effektivere Strategien zur Problemlösung zu finden.

In eine ähnliche Richtung weisen auch die empirischen Ergebnisse von Wilhelm (2005). Der Autor entwickelte einen Kinematik/Dynamik Lehrgang mit Hilfe dynamisch bildhafter Repräsentationen und graphischer Modellbildung zur Veränderung von Schülervorstellungen und evaluierte diesen durch den Einsatz verschiedener Tests: Leistungstestaufgaben im Bereich der Kinematik, das Force Concept Inventory (FCI) von Hestenes, Wells und Swackhamer (1992), der das konzeptuelle Verständnis des Kraftbegriffs in unterschiedlichen Kontexten erfasst, Concept-Maps zum qualitativen Verständnis in Bezug auf Newtons Bewegungslehre sowie durch den Maryland Physics Expectations Survey (MPEX), der darauf zielt, Vorstellungen über Physik als wissenschaftliche Disziplin und Physiklernen zu erheben. In einer Treatmentgruppe (17 Klassen) wurden kinematische Größen intensiv behandelt, dabei wurden zweidimensionale Bewegungen, der zeitliche Verlauf relevanter Größen und die Beziehung zwischen den Größen auf verschiedene Abstraktionsniveaus ausführlich thematisiert (vgl. ebd., S. 254). Eine Kontrollgruppe (13 Klassen), erhielt herkömmlichen Unterricht, in dem sich die Schüler intensiv mit eindimensionalen Bewegungen auseinandersetzten, die in Form von Graphen dargestellt waren, die es zu interpretieren galt. Im Ergebnis zeigte sich, dass das Unterrichtskonzept der Treatmentgruppe den Schülern half, physikalische Konzepte besser zu verstehen (vgl. ebd., S. 255). So konnte zwar im MPEX-Test zur Erhebung von Vorstellungen von Physik und in Leistungstestaufgaben, in denen die Schüler zu einer beschriebenen Bewegung den passenden Graphen (Zeit-Geschwindigkeits- oder Zeit-Beschleunigungsgraph) auswählen mussten, kein Unterschied festgestellt werden, in Leistungstestaufgaben, zur Richtung der Beschleunigung (z.B. senkrechter Wurf mit Richtungsumkehr) und zur Beschleunigung bei einer zweidimensionalen Bewegung und zu Kurvenfahrten, schnitt die Treatmentgruppe signifikant besser ab und zeigte eine an-

gemessenerer Richtungsvorstellung. Das bessere Verständnis der Treatmentgruppe wurde durch die Testergebnisse im FCI bestätigt. Auch in den Konzeptmaps, die bei einer Teilstichprobe ($N = 63$) eingesetzt wurden, zeigten Schüler der Treatmentgruppe mehr Verständnis als traditionell unterrichtete Klassen.

Prain und Waldrup (2006) befragten in einer explorativen Studie Schüler der vierten bis sechsten Klassenstufe⁹ und ihre Lehrer ($N = 20$) zum Thema Verwendung multipler Repräsentationen von physikalischen Konzepten, wie elektrische Stromkreise oder Kollision und Fahrzeugsicherheit (vgl. ebd., S. 1848 f.). Die Datenerhebung erfolgte mittels quantitativer (Fragebogenauswertung) und qualitativer Methoden (Auswertung von Interviews). In einer Teilstudie wurden aus der ursprünglichen Stichprobe von 20 Klassen zudem sechs Klassen und ihre Lehrer ausgewählt und im Unterricht beobachtet. Die Ergebnisse der Autoren (vgl. ebd., S. 1850 ff.) weisen darauf hin, dass das Verständnis der Verbindung zwischen unterschiedlichen Repräsentationsformaten wie Graphen und verbalen Beschreibungen eine zentrale Funktion im Lernprozess zukommt. Während einige Schüler in der Lage waren, unterschiedliche Repräsentationen aufeinander zu beziehen und ineinander zu übersetzen, benötigten andere Schüler Unterstützung beim Erkennen der Interpretation und Übersetzung verschiedener Repräsentationsformen. Im Hinblick auf die Förderung des konzeptuellen Verständnisses ist besonders das Ergebnis hervorzuheben, dass Schüler, welche die Verbindung zwischen verschiedenen Repräsentationsformen verstanden hatten, auch über ein höheres konzeptuelles Verständnis verfügten. Die Autoren vermuten, dass die Kompetenz Repräsentationen ineinander zu übersetzen, zu einem guten Teil auch auf ein generell besseres Verständnis von Übereinkünften der Verwendung und Funktion von Repräsentationen zurückgeführt werden könnte.

Waldrup, Prain und Carolan (2010) beobachteten Schüler der 8. Klassenstufe beim Erstellen von multimodalen Repräsentationen in einer Lerneinheit über Kraft und Materie; insbesondere ging es um Teilchenmodelle zu Festkörpern, Flüssigkeiten und Gasen. Die Autoren beobachteten, dass die beteiligten Lehrkräfte in Diskussionen über die bestmögliche Form der Repräsentation wissenschaftlicher Konzepte eine Vermittlerrolle einnahmen. Des Weiteren verwendeten die Lehrer die von den Schülern generierten Repräsentationen, um das konzeptuelle Verständnis ihrer Schüler einzuschätzen. Zudem zeigten sich in einer qualitativen Analyse, dass Repräsentationen, die von Schülern selbst erstellt wurden, den Erwerb von konzeptuellem Wissen insbesondere dann fördern können, wenn die

9 Die Anzahl der teilnehmenden Schüler wird in dem Artikel nicht genannt.

Lehrenden Schüler darin unterstützen, auf Klarheit, Kohärenz und Angemessenheit der Repräsentationen zu achten.

Hubber, Tytler und Haslam (2010) belegten in einer qualitativ ausgewerteten Videostudie, dass ein repräsentationaler Fokus Lehren und Lernen des physikalischen Kraftbegriffs im Bereich der Mechanik unterstützte. Die Autoren zeichneten 12 Unterrichtsstunden von drei Lehrern auf, führten Interviews mit den teilnehmenden Lehrern und Schülern durch, sammelten Arbeiten der Schüler ein und erstellten Notizen zu den Feldbeobachtungen. Die Analyse zielte darauf, Belege für das produktive Generieren von multiplen Repräsentationen und die Verwendung von multiplen Repräsentationen zur Kommunikation über wissenschaftliche Inhalte zu finden, welche zu einem nachweislichen Lernerfolg führten (vgl. ebd., S. 10). Die Lehrer wurden explizit angehalten, im Unterricht multiple Repräsentationen zum Begriff der Kraft und der Bewegung einzuführen und diese Repräsentationen als Denkwerkzeuge zu nutzen, die Schüler darin zu unterstützen, ihre eigenen Repräsentationen zu generieren sowie die Angemessenheit von Repräsentationen zu diskutieren und zu bewerten (vgl. ebd.). Für die praktische Umsetzung dieser Empfehlungen, die darauf zielen den Schwerpunkt im Unterricht auf das Lernen mit verschiedenen Repräsentationen zu setzen, waren die Lehrer selbst verantwortlich. Die Auswertung der Interviews mit den Lehrern ergab, dass die Empfehlungen in der Wahrnehmung der Lehrer zu einer offeneren Herangehensweise im Unterricht und einem höheren Anteil an fachlichen Diskussionen führte, die Aufmerksamkeit der Schüler auf das Sprechen über Physik und Übereinkunft im Umgang mit Repräsentationen lenkte und zu einer höheren Vielfalt an Erklärungsmodi führte. Die Analyse der Videos belegte zudem, dass sich in vielen Unterrichtssequenzen die Verwendung multipler Repräsentationen automatisch aus dem Unterrichtsverlauf und den sozialen Interaktionen ergab. Der Sprachgebrauch der Lehrer und Schüler zeichnete sich durch einen hohen Anteil an Erklärungen aus. Die Analyse der Diskussionen der Schüler ergab, dass die Äußerungen der Schüler reich an Bezügen zu wissenschaftlichen Konzepten waren. Die Auswertung der Schülerinterviews belegte, dass die Schüler sich des Nutzens unterschiedlicher Repräsentationen bewusst waren und deren Verwendung schätzen lernten. Einschränkend muss jedoch festgehalten werden, dass keine Kontrollgruppe vorliegt, die einen direkten Vergleich zur Bewertung der Effektivität des Ansatzes ermöglicht.

Eine Studie von Mortimer und Buty (2009) befasste sich mit Lernschwierigkeiten beim Umgang mit der Repräsentation des „Unendlichen“ in einer Lehr-

einheit zur Optik.¹⁰ Die Repräsentation des Unendlichen meint in diesem Zusammenhang folgendes: In einem Versuchsaufbau zur Bildentstehung bei der Sammellinse kann entweder die Bildposition oder die Position des Gegenstands gegen Unendlich gehen. Eine schematische Repräsentation, welche diesen Zusammenhang verdeutlicht, befindet sich in dem Artikel der Autoren (vgl. ebd., S. 231). Im Rahmen der Studie erstellten Schüler der elften Klassenstufe in Gruppenarbeit graphische Repräsentationen zum Verlauf von Lichtstrahlen durch eine Sammellinse. Die Autoren videografierten Episoden aus 15 Lernaktivitäten einer Schulklasse. Das Material umfasst neun Unterrichtsstunden und zwei Experimentierstunden. Drei Episoden, in denen ein Schüler deutliche Verständnisfortschritte machte, wurden detailliert ausgewertet. Hierzu wurde das Transkript einer Interaktion analysiert, bei der ein Schüler, der deutliche Erkenntnisfortschritte machte, sich mit einem Schüler austauscht, der bereits verstanden hatte, dass entweder die Bildposition oder die Position des Gegenstands gegen Unendlich gehen kann. Die anfänglichen Verständnisschwierigkeiten des Schülers resultierten aus der Problematik, dass das reelle Bild so weit entfernt war, dass er es nicht in einer schematischen Repräsentation darstellen konnte. Zu Erkenntnisfortschritten kam es, als der Schüler begann, Verbindungen zwischen den physikalischen Modellen und Theorien und den konkreten Objekten und Abläufen herzustellen. Dabei wechselte er zwischen der schematischen Repräsentation und der alltagssprachlichen Beschreibung hin und her. In dem Moment, als es ihm gelang, das reelle Bild in unendlicher Entfernung zu repräsentieren, konnte er besser mit dem Sachverhalt umgehen, was zu einer Wissensveränderung führte, die sich auf das konzeptuelle Verständnis der Bildentstehung auswirkte. Wirksame Lernmechanismen, welche den Verständnisfortschritt erklären, sehen die Autoren in:

- dem Wechsel zwischen der Welt der Theorien und Modelle und der Welt der konkreten Objekte und Ereignisse,
- dem Wechsel zwischen unterschiedlichen Repräsentationsformen (schematisch und sprachlich) sowie
- im kommunikativen Austausch (hier mit einem Mitschüler), durch den sich der Lernende mit einer anderen Sichtweise auseinandersetzt.

10 Original Titel: „What Does “In the Infinite” Mean?: The Difficulties with Dealing with the Representation of the “Infinite” in a Teaching Sequence on Optics” (zit. n. Mortimer & Buty, 2009, S. 225).

Auch diese Studie weist darauf hin, dass ein enger Zusammenhang zwischen der Förderung des Umgangs mit fachgebunden Repräsentationen und dem konzeptuellen Verständnis der Schüler besteht.

1.3.2 Fazit und weiterer Forschungsbedarf

Als Fazit der unterschiedlichen Studien können folgende Trends festgehalten werden:

- Das explizite Arbeiten mit Repräsentationen unterstützt offenbar den Aufbau des konzeptuellen Verständnisses von Schülern unterschiedlicher Altersbereiche (vgl. Wilhelm, 2005; Prain & Waldrip, 2006; Botzer & Reiner, 2007; Waldrip et al., 2010; Mortimer & Buty, 2009).

Förderlich ist hierbei offenbar:

- das Thematisieren der Verbindungen zwischen unterschiedlichen Repräsentationen, die in unterschiedlichen Formaten gegeben waren (vgl. Prain & Waldrip, 2006; Waldrip et al., 2010; Mortimer & Buty, 2009) oder auf unterschiedlichen Abstraktionsebenen liegen (vgl. Wilhelm, 2005; Mortimer & Buty, 2009)
- sowie das Generieren eigener Repräsentationen und ihre Diskussion im Unterricht (vgl. Prain & Waldrip, 2006; Botzer & Reiner, 2005; Waldrip et al., 2010; Hubber et al., 2010; Mortimer & Buty, 2009).

Die vorgestellten Studien befassen sich mit dem Aufbau von Wissen und konzeptuellem Verständnis. Einen Einblick in die kognitiven Mechanismen, der auch den Umgang mit Experimenten reflektiert, bietet die Studie von Botzer und Reiner (2005). Die vorliegenden Ergebnisse basieren auf einer Fallstudie mit einer kleinen Anzahl an Probanden. Unklar ist daher, ob der vorgestellte Ansatz zu statistisch signifikantem Lernfortschritt führt und welche Effektstärke hiermit verbunden wäre.

Forschungsbedarf besteht zudem in der Analyse von Prozessen der kognitiven Aktivierung im Kontext der Förderung des kompetenten Umgangs mit Repräsentationen und des Aufbaus eines fundierten konzeptuellen Verständnisses, so dass die Schüler in der Lage sind, eine Verbindung zwischen physikalischen Experimenten, beobachtbaren Phänomenen und wissenschaftlichen Erklärungen,

welche depiktional schematisch oder deskriptiv verbal bzw. abstrakt mathematisch dargestellt werden, herzustellen. Für die Unterrichtspraxis gilt es zu klären, wie Aufgaben so gestellt werden können, dass die Lernenden gefordert werden, ihre kognitiven Fähigkeiten und ihre Ressourcen im Umgang mit Repräsentationen einzusetzen, insbesondere, wenn es darum geht, weitverbreitete Lernschwierigkeiten, die sich in naiven Schülervorstellungen manifestieren, zu überwinden.

Auch Interaktion zwischen Lernen mit und über Experimente und der Verwendung multipler Repräsentationen stand in den vorliegenden Studien nicht im Zentrum der wissenschaftlichen Analyse.

Sowohl zum Thema kognitive Aktivierung als auch zum Thema Physiklernen aus Experimenten existieren jedoch zahlreiche Forschungsergebnisse (ohne speziellen Fokus auf das Lernen mit multiplen Repräsentationen in Naturwissenschaften), die in den folgenden Kapiteln im Überblick dargestellt werden.

1.3.3 Kognitive Aktivierung

Bezugstheorien kognitiver Aktivierung sind zum einen Theorien der Gedächtnispsychologie und der kognitiven Verarbeitung von (Lern-)inhalten (siehe Kapitel 1.1.2 Gedächtnispsychologische Grundlagen und Kapitel 1.1.3 Modelle der kognitiven Verarbeitung von Texten, Bildern und Sprache) und zum anderen kognitiv-konstruktivistische Lerntheorien, welche die Bedeutung der fachgebundenen Interaktion für den Aufbau von neuem Wissen betonen (siehe Wygotski, 1986/1964; Piaget 1979) und die Konfrontation mit anderen Sichtweisen und Standpunkten (siehe diSessa 1983, 1988, 1993; Vosniadou 1994) hervorheben.

Die Förderung eines Konzeptwechsels, wie sie in Kapitel 1.2 *Die Rolle von Konzeptwechseln in Lehr-Lernprozessen* diskutiert wurde, kann als ein wesentlicher Aspekt kognitiver Aktivierung gesehen werden: durch die Auseinandersetzung mit Informationen, Erfahrungen oder Phänomenen, die bisherigen Konzepten widersprechen, werden die Lernenden angeregt, eigene Schemata und Strategien zu verändern.

Im Rahmen der Unterrichtsforschung wurde der Begriff der kognitiven Aktivierung von Baumert und Klieme (vgl. Baumert et al., 2004, zit. n. Lipowsky, 2009, S. 93) in Abgrenzung zu anderen zentralen Faktoren der Unterrichtsqualität, wie Schülerorientierung und Klassenführung, etabliert.

Da die Konfrontation mit anderen Sichtweisen bereits ausführlich diskutiert wurde, wird der Begriff der kognitiven Aktivierung im Folgenden aus Sicht der Gedächtnispsychologie und Unterrichtsforschung beleuchtet.

Aus dem Levels-of-Processing-Ansatz der Gedächtnispsychologie ist bekannt, dass auf Seiten der Lernenden für das langfristige Behalten eines Sachverhalts die Tiefe der kognitiven Verarbeitung entscheidend ist (vgl. Craik & Lockhart, 1972). Die erweiterte Variante des Levels-of-Processing-Ansatzes ergänzt das Konzept der Verarbeitungstiefe um das Konzept der Verarbeitungsbreite, die auf die Elaboration der Lerninhalte zielt. Eine tiefe und breite Verarbeitung des Lerngegenstandes fördert neben dem Behalten auch die Flexibilität des Wissens, so dass dieses leichter in Transferaufgaben angewandt werden kann.

Die Initiierung kognitiver Aktivierung in der pädagogischen Praxis steht somit in enger Beziehung zu empirischen Ergebnissen der Lernstrategieforschung. Zum Einsatz effektiver Lernstrategien besteht mittlerweile eine große Zahl an Veröffentlichungen, deren Diskussion über das Thema dieser Arbeit hinaus führen würde. Überblicksdarstellungen finden sich unter anderen in Mandl und Friedrich (1992); Pintrich und Garcia (1994); Schnotz (2006a), S. 177 ff. sowie Levin und Arnold (2006).

Schrader und Helmke (2006, S. 638 f.) unterscheiden drei Formen anspruchsvoller Lernstrategien, die auf eine tiefe und breite Verarbeitung zielen: kognitive Lernstrategien, metakognitive Lernstrategien und ressourcenorientierte Strategien.

1. *Kognitive Lernstrategien* zielen darauf, die Prozesse der Aufnahme, Verarbeitung und Speicherung von Information zu optimieren. Diesbezügliche Lernaktivitäten umfassen das Wiederholen des gelernten Materials bzw. das erneute mentale Durchspielen von Zusammenhängen, die Organisation und Strukturierung des Lernmaterials und die Elaboration, bei der Zusammenhänge zwischen verschiedenen Inhalten hergestellt werden.
2. *Metakognitive Prozesse* betreffen die Planung des Lernvorgangs, die Prüfung, ob das Gelernte hinreichend verstanden und behalten wurde und die Anpassung des Vorgehens, wenn das bisherige Vorgehen nicht zielführend war oder Schwierigkeiten auftraten (Regulation).
3. *Ressourcenorientierte Strategien* schließlich zielen darauf, die Rahmenbedingungen des Lernens zu optimieren, wobei sowohl äußere Bedingungen wie die Gestaltung des Arbeitsplatzes und die Verwendung von Lernmaterialien als auch innere Faktoren wie Motivation und Aufmerksamkeit berücksichtigt werden.

Vor diesem Hintergrund kann kognitive Aktivierung als Einsatz anspruchsvoller Lernstrategien anstelle einer mechanischen Wiederholung von Lerninhalten definiert werden (vgl. Klauer & Leutner, 2007, S. 108 f.).

Im Rahmen von Unterrichtsforschung wurde kognitive Aktivierung als Kriterium für Unterrichtsqualität vorwiegend im Mathematikunterricht untersucht.

Baumert und Kunter (2011, S. 13 f.) arbeiteten auf Basis von Übersichtsartikeln und Metaanalysen folgende Instruktionsdimensionen heraus, die kognitive Aktivierung fördern:

- Die Lernenden erhalten kognitiv herausfordernde Lerngelegenheiten.
- Der Lernprozess der Lernenden wird sorgfältig unterstützt und überprüft, die Lernenden erhalten individuelle Rückmeldungen, der Unterricht adaptiert an den Kenntnis- und Fähigkeitsstand der Lernenden.
- Der Unterrichtsverlauf zeichnet sich durch eine effektives Klassen- und Zeitmanagement aus.

Einen Überblick zu empirischen Befunden zur Wirkung kognitiver Aktivierung im Unterricht bietet Lipowsky (2009, S. 93 f.). Lipowsky (2009) weist darauf hin, dass positive, wenn auch schwache Effekte zwischen kognitiver Aktivierung und Lernzuwachs im Bereich Mathematik in folgenden Studien berichtet wurden: Studien von Hiebert und Wearne (1993), im QUASAR-Projekt (Stein & Lane, 1996), zufolge von Ergebnissen des britischen CAME-Projekts (Shayer & Adhami, 2007) sowie Ergebnissen aus der TIMSS-Videostudie (vgl. hierzu Baumert & Kunter 2011, im Überblick).

Im Bereich des Schriftspracherwerbs bezieht sich Lipowsky (2009, S. 93 f.) auf eine Studie von Taylor, Pearson, Peterson und Rodriguez (2003), in der positive Effekte eines kognitiv anspruchsvolleren Unterrichts auf den Lernzuwachs benachteiligter Grundschüler ermittelt wurde.

Ein positiver Zusammenhang zwischen kognitiver Aktivierung und Lernzuwachs bei der Bearbeitung von Hausaufgaben wurden zudem u.a. von Lipowsky selbst in der Pythagoras Studie gefunden (Klieme, Lipowsky, Rakoczy & Ratzka, 2006).

Im folgenden Abschnitt wird auf die zitierten Studien nun näher eingegangen:

Hiebert und Wearne (1993) führten eine Studie mit sechs „zweiten“ Klassen einer Schule in ländlicher Gegend, mit breitem Spektrum an Leistungsniveaus durch. Vier Klassen, in denen nach einer konventionellen Lehrbuchmethode unterrichtet wurde, wurden mit zwei Klassen verglichen, in denen ein alternativer Zugang implementiert wurde. Lerninhalte waren der Zahlenraum 1 bis 100, darauf aufbauend der Zahlenraum 1 bis 1000 sowie die Addition und Subtraktion von zwei- und dreistelligen Zahlen. Die Zuordnung zu den Bedingungen erfolgte aufgrund eines Tests am Anfang des Schuljahres, der grundlegende Kenntnisse und

Fähigkeiten erfasste. Die 60 besten Schüler wurden zufällig auf zwei Klassen mit hohem Leistungsniveau und die übrigen 87 Schüler auf vier Klassen mit mittlerem bis geringerem Leistungsniveau verteilt. Insgesamt nahmen von den 147 Schülern der sechs Klassen 135 an der Studie teil (vgl. ebd., S. 397). Eine überdurchschnittliche und eine unterdurchschnittliche Klasse wurden der Bedingung mit der alternative Unterrichtsmethoden zugewiesen. Der Unterricht aller teilnehmenden Klassen wurde in einem Zeitraum von 12 Wochen einmal pro Woche beobachtet und per Audiotape aufgezeichnet.

Die alternative Methode (im Folgenden Treatmentbedingung genannt) folgte zwar den Themen des Schulbuchs, es wurden aber keine Aufgaben oder Lehrbuchtexte bzw. sonstige Materialien aus dem Buch verwendet. Die zugrunde liegende Idee bestand darin, das Verständnis der Schüler zu fördern, indem die Schüler aktiviert werden, Beziehungen zwischen ihrem derzeitigen Wissen und neuen Lerninhalten, Beziehungen zwischen verschiedenen Repräsentationsformen herzustellen sowie Zusammenhänge zwischen verschiedenen Lösungswegen zu erkennen. Diese Prinzipien wurden operationalisiert, indem die Schüler mit konzeptualisierten Problemstellungen konfrontiert wurden; eigene Darstellungen von Mengen durch geschriebene Zahlen und „Bauklötzchen“ entwickelten; Lösungswege in beiden Formen darstellen und diese mit der Klasse diskutierten. Im konventionellen Unterricht konnten die Lehrer uneingeschränkt in ihrem Sinn agieren, der Unterricht in den Klassen mit mittleren und geringerem Leistungsniveau folgte hier meist dem Lehrbuch, die Lehrer verwendeten nur wenige zusätzliche Aktivitäten und Aufgaben, Lösungswege wurden mittels Papier und Bleistift gefunden; in der überdurchschnittlichen Klasse brachte der Lehrer häufiger zusätzliche Problemstellungen mit ein.

Ausgewertet wurde neben dem Lernzuwachs der Schüler, die Anzahl an Problemen bzw. Aufgaben, die bearbeitet wurden, die Zeit, die pro Problem / Aufgabe aufgewendet wurde, der Problem-Kontext bzw. die Strategien, welche verwendet wurden, um die Probleme/Aufgaben zu bearbeiten und sowie Kriterien zur Beschreibung der Diskussionen im Klassenraum (Redeanteil und Form der Diskussion der Beteiligten, z.B. Art der Fragestellung durch die Lehrer) (vgl. ebd., S. 405 - 415).

Klassen, welche der Treatmentbedingung zugeordnet waren, bearbeiteten weniger Aufgaben, verwendeten jedoch mehr Zeit auf die Bearbeitung der jeweiligen Problemstellung als Klassen in der Kontrollbedingung. Der Redeanteil der Schüler in der Treatmentbedingung war höher als in der Kontrollbedingung. Die Lehrer in der Treatmentbedingung stellten weniger Fragen, die es erforderten Fakten oder schematische Vorgehensweisen zu wiederholen. Die

Schülerantworten in der Treatmentbedingung zeichneten sich zudem durch ausführlichere Beschreibungen und Erklärungen aus.

In den abschließenden Tests zeigte sich in Mittelwertvergleichen (unter Berücksichtigung des Vorwissens), dass Schüler in der Treatmentbedingung bessere Leistungen erzielten und am Ende des Jahres bei den meisten Items höhere Punkte als Schüler in der Kontrollbedingung erreichten. Schüler in der Treatmentbedingung konnten zudem längere, umfassendere Antworten geben, wenn sie nach alternativen Lösungsstrategien gefragt wurden. Den höchsten Lernfortschritt und die besten Gesamtleistungen erreichte die überdurchschnittliche Klasse, welche der Treatment-Bedingung zugeordnet war. Schüler dieser Klasse bearbeiten weniger Aufgaben, verwendeten pro Aufgabe jedoch deutlich mehr Zeit, engagierten sich stärker an Diskussionen und teilten ihre Lösungsstrategien und ihre Erklärungen in den Diskussionen häufiger mit ihren Klassenkameraden. Sie erzielten bei allen Aufgaben eine deutliche Verbesserung (vgl. ebd., S. 419).

Stein und Lane (1996) führten von 1990 bis 1993 eine Studie an vier Projekt-schulen im Mathematikunterricht der Klassenstufen 6 bis 8 durch. Untersucht wurde u.a. der Zusammenhang zwischen der Art der Aufgabenstellungen durch die Lehrer und der Aufgabebearbeitung durch die Schüler. Dabei wurden zur Kategorisierung der Aufgabenstellungen sogenannte Reformmerkmale ausgearbeitet (Aufgaben mit verschiedenen Lösungswegen und unterschiedlichen Darstellungsweisen). Aufgaben, welche Reformmerkmale aufweisen, sollten zu einem höheren Lernzuwachs führen, erfasst durch Tests zu verschiedenen Messzeitpunkten und zu einer besseren generellen Unterrichtsqualität, erfasst durch Beobachtungen und Videoanalysen. Im Ergebnis zeigte sich, dass Schüler die größten Lernerfolge erzielen, wenn die verwendeten Aufgaben sie zu hochwertigen Denkprozessen und zum gegenseitigen Austausch durch Fachdiskussionen anregen. Moderate Lernerfolge wurden erzielt mit Aufgaben von kognitiv ansprechender Gestaltung, welche die Schüler jedoch nicht in hochwertige mathematische Argumentations- und Problemlöseprozesse verwickelten. Am wenigsten profitierten die Schüler von Aufgaben, welche mit einer einfachen und klar offen zu legenden Strategie bewältigt werden konnten, wobei die Suche des Lösungswegs wenig gegenseitige Kommunikation erforderte.

Shayer und Adhami (2007) untersuchten im Rahmen des CAME-Projekts in einer Längsschnittuntersuchung, ob es möglich ist, durch reflektierenden Mathematikunterricht die kognitive Entwicklung der Schüler in domänenspezifischer versus domänenübergreifender Hinsicht zu steigern. 2500 Schüler im Alter von 11 bis 13 Jahren erhielten über einen Zeitraum von zwei Jahren alle zehn Tage eine „Thinking Maths“ Unterrichtsstunde. Solche „Thinking Maths“ Unterrichtsstunden bestanden aus drei Sequenzen:

1. Vorbereitung: Es wurde eine Ausgangssituation geschaffen, in der sich die Schüler gegenseitig erklären, worum es in dem Arbeitsblatt ging.
2. Partner- bzw. Gruppenarbeit: Die Schüler bearbeiteten das Arbeitsblatt mit der Aussicht, ihre Lösungen später der Klasse vorzustellen. Der Lehrer gab in dieser Zeit keine Hilfestellungen, sondern beobachtete die Schüler und notierte sich die Vorgehensweise.
3. Im Unterrichtsgespräch stellten die einzelnen Gruppen ihre Lösungsstrategie vor, während die zuhörenden Schüler Fragen stellten. Der Lehrer agierte als Moderator und leitete den Diskussionsprozess.

Daran anschließend bearbeiteten die Schüler weitere Aufgabenblätter, wobei die Phasen 1 bis 3 durchlaufen wurden, die Argumentationen jedoch weniger ausführlich und dadurch weniger zeitaufwendig besprochen wurden.

Tatsächlich konnte durch das Projekt immerhin ein domänenübergreifender Effekt in der Steigerung der allgemeinen Intelligenz bei Schülern erzielt werden, welche von vornherein durch ein hohes kognitives Niveau gute Lernvoraussetzungen aufweisen. Die Steigerung war jedoch an die Voraussetzung geknüpft, dass die Lehrkraft die Rolle eines Moderators einnahm und sich insgesamt zurückhaltend verhielt. Als weiteres Ergebnis der Studie ist festzuhalten, dass sich die Herangehensweise der teilnehmenden Schüler auch auf andere Fächer übertrug.

Klieme et al. (2006) untersuchten im Rahmen des Projekts „Pythagoras“ als eine Qualitätsdimension von Unterricht die kognitive Aktivierung bezüglich des Umgangs mit Hausaufgaben. Das Projekt startete mit einer repräsentativen Lehrerbefragung ($N = 40$) in Deutschland und der Schweiz. Die Ergebnisse der Befragung wurden in Relation zu Erhebungen der Unterrichtspraxis gesetzt, welche videografiert wurden (Lerninhalt war, wie der Projektname schon andeutet, vorwiegend der Satz des Pythagoras). Die Stichprobe der Videografie bestand aus 20 Klassen der Jahrgangsstufe 9 in Deutschland sowie 20 Klassen der Jahrgangsstufe 8 in der Schweiz. Um Aussagen über die Wirkungsrichtung der angenommenen Qualitätsmerkmale machen zu können, wählten die Autoren ein Längsschnittdesign (vgl. ebd., S. 133 f.). Als Indikator für kognitiv aktivierenden Unterricht wurde eine Fragebogenskala eingesetzt, welche die Wahrnehmung des Umgangs mit Hausaufgaben der Lehrer aus Schülersicht erhob. Die Skala umfasste u.a. Fragen zum Umgang mit Fehlern und der Anregung zum Nachdenken, zur Wertschätzung von Lösungsprozessen und dem Finden neuer Lösungswege sowie zur Wertschätzung der Anstrengung der Schüler unabhängig von der Richtigkeit der Lösungen (vgl. ebd., S. 13). Mittels Mehrebenenanalysen konnten die Autoren einen signifikant positiven Einfluss eines prozessorientierten Umgangs mit Hausaufgaben auf die Leistungs-

entwicklung im Schuljahr belegen. In weiteren Ergebnissen des Projekts zeigte sich auf mikrogenetischer Ebene zudem ein Interaktionseffekt zwischen kognitiven und motivationalen Faktoren: Von kognitiv aktivierendem Unterricht profitierten v.a. Schüler, welche sich stärker für das Fach interessierten (vgl. ebd., S. 140).

Abschließend zum Thema kognitiver Aktivierung im Mathematikunterricht werden noch zentrale Ergebnisse der TIMSS-Videostudie hervorgehoben:

So stellte sich heraus, dass sich das Niveau der kognitiven Herausforderung im Mathematikunterricht in erster Linie durch die Aufgabenauswahl und die Behandlung der Aufgaben ergibt (vgl. Baumert & Kunter, 2011, S. 13 f).

Lehrende können Prozesse der kognitiven Aktivierung u.a. fördern, indem sie

- die Lernenden mit kognitiv herausfordernden Aufgaben konfrontieren und kognitive Konflikte initiieren. So können kognitiv aktivierende Aufgaben an das Vorwissen der Schüler anknüpfen, indem sie Schüler ermutigen, eigene Vorstellungen und Überzeugungen der Schüler auf die Probe zu stellen (vgl. Lipowsky, 2009, S. 94; Baumert & Kunter, 2011, S. 13),
- Lernende ermutigen, eigene Gedanken, Konzepte, Ideen und Lösungswege zu erläutern, die Lernenden anregen durchdachte Fragen zu stellen und
- insgesamt auf eine Lernatmosphäre achten, in der die Lernenden inhaltliche Konzepte und Ideen intensiv diskutieren.

Für den sprachlichen Bereich liegt eine Studie von Taylor, Pearson und Rodriguez (2003) vor. Die Autoren untersuchten unter Beteiligung von 88 Lehrkräften bei 792 Schülern aus neun Schulen in sozialen Brennpunkten in den USA den Lernzuwachs im Schriftspracherwerb. Pro Klasse (der Klassenstufen 1 bis 5) wurden neun Schüler per Zufall aus einer Klasse eines teilnehmenden Lehrers ausgewählt (88 Klassen mit je 9 Schüler: $N = 9 \times 88 = 792$). Die teilnehmenden Lehrkräfte wurden zu drei Zeitpunkten im Frühjahr, Sommer und im Winter während eines Schuljahres im Leseunterricht für je eine Schulstunde von geschulten Beobachtern nach einem festgelegten System (CIEARA Classroom Observation Scheme, vgl. ebd., S. 9) beurteilt. Die Kompetenzen im Schriftspracherwerb der Schüler wurden ebenfalls zu drei Messzeitpunkten im Frühjahr, Sommer und Winter erhoben (vgl. ebd., S. 8 ff.): neben der Leseleistung der Schüler, welche pro Schüler individuell beurteilt wurde, erhob man auch das Leseverständnis und die schriftlichen Fähigkeiten der Schüler. Es handelte sich um Aufgaben mit angepasstem Schwierigkeitsniveau an das jeweilige Niveau der Klassenstufe (1 bis 5). Eine Mehrebenenanalyse ergab, dass Lehrer, welche den Schwerpunkt auf anspruchsvolles Denken bei der Form der Fragestellungen

und den verwendeten Aufgaben legten, in höherem Maß eine Verbesserung des Schriftspracherwerbs bezüglich des Leseverständnis und der schriftlichen Fähigkeiten erzielten. In den Beobachtungen zeigte sich zudem, dass Lehrer, welche anspruchsvollere Fragen stellten, die Schüler förderten, indem sie von den Schülern forderten darüber nachzudenken, welche Inhalte sie gelesen hatten (vgl. ebd., S. 22).

Im Rückblick auf die vorgestellten Studien im Bereich der Unterrichtsforschung wird deutlich, dass der Begriff der kognitiven Aktivierung, wenn auch mit inhaltlichen Überschneidungen, durchaus unterschiedlich aufgefasst wurde. Ein weiteres Problem in der Begriffseingrenzung ergibt sich nach Lipowsky (2009, S. 93 f.), daraus, dass es oft nur schwer möglich ist, zu beobachten, inwieweit die pädagogische Maßnahmen, welche auf kognitive Aktivierung zielen, tatsächlich auch in der intendierten Weise wirken.

Als Ergebnis des Kapitels sollen folgende Aspekte kognitiver Aktivierung festgehalten werden:

Kognitive Aktivierung definiert sich als Umsetzung anspruchsvoller Lernstrategien (vgl. Taylor, Pearson & Rodriguez 2003; Klauer & Leutner, 2007, Baumert & Kunter, 2011). Lipowsky (2009, S. 93) folgend kann kognitive Aktivierung auf Seiten der Lehrenden unter dem Gesichtspunkt analysiert werden, wie Lehrende Prozesse der kognitiven Aktivierung initiieren und fördern oder auf Seiten der Lernenden, wie Inhalte verarbeitet und Lernangebote genutzt werden.

Anspruchsvolle Lernstrategien beinhalten:

- Beziehungen zwischen dem Vorwissen bzw. dem derzeitigem Lernstand und neuen Lerninhalten herzustellen (Hiebert & Wearne, 1993; Baumert & Kunter, 2011),
- Lernende mit kognitiv herausfordernden Aufgaben zu konfrontieren, herausfordernde Fragen bzw. Aufgaben zu stellen (Taylore et al., 2003; Baumert & Kunter, 2011),
- kognitive Konflikte zu initiieren (Baumert & Kunter, 2011),
- Lernende zu aktivieren, Beziehungen zwischen verschiedenen Repräsentationsformen herzustellen (Hiebert & Wearne, 1993; Stein & Lane, 1996),
- Lernende zu fördern, verschiedenen Lösungswege zu kennen (Stein & Lane, 1996) und Zusammenhänge zwischen diesen Wegen herzustellen (Hiebert & Wearne, 1993),
- Lernende zu ermutigen, eigene Gedanken, Konzepte, Ideen und Lösungswege zu erläutern (Hiebert & Wearne, 1993; Shayer & Adhami, 2007) und

- einen prozessorientierten Umgang mit Problemen zu fördern, bei dem Lernende angeregt werden inhaltliche Konzepte und Ideen intensiv zu diskutieren (Hiebert & Wearne, 1993; Klieme et al., 2006; Shayer & Adhmi, 2007; Baumert & Kunter, 2011).

Die Förderung des kompetenten Umgangs mit verschiedenen Darstellungsformen (Repräsentationskompetenz) greift v.a. auf der Ebene der Angebotsgestaltung an: Die hierzu entwickelten Lernmaterialien zielen auf die Elaboration der Lerninhalte. Die Lernenden sollen durch entsprechende Aufgaben dazu angeregt werden, die erlernten physikalischen Vorgänge mental durchzuspielen. Der Aufbau solcher adäquaten mentalen Modelle sollte, sofern dieser Schritt gelingt, zum einen dazu führen, naive Schülervorstellungen zu überwinden, da Inkonsistenzen mit einer adäquaten Modellvorstellung inkompatibel sind und sich zum anderen auch in externen Repräsentationen niederschlagen, die wiederum dazu genutzt werden können Transfer- und Problemlöseaufgaben zu bearbeiten.

Die bewusste Anwendung von Lernstrategien beinhaltet auch eine metakognitive Komponente (vgl. Schnotz, 2006a, S. 186), weil sie damit verbunden ist, den eigenen kognitiven Zustand zu erkennen und die eigenen Verarbeitungsschritte an den aktuellen Lernstand anzupassen (vgl. ebd.). Bezüglich des Lernens mit Repräsentationen bedeutet dies, die Lernenden anzuregen,

- den physikalischen Sachverhalt zeichnerisch und schriftlich darzustellen, also das eigene Verständnis der physikalischen Zusammenhänge explizit zu machen,
- selbst erstellte Repräsentationen und das dahinterliegende eigene Verständnis physikalischer Konzepte kritisch zu prüfen sowie
- mit erstellten Repräsentationen kognitiv zu operieren.

Die kritische Prüfung steht dabei in enger Verbindung mit dem weiterführenden Umgang mit Repräsentationen: Die Operation mit den erstellten externen Repräsentationen beinhaltet, deren Zweckmäßigkeit zur Lösung von physikalischen Problemen zu prüfen und die eigenen Verarbeitungsschritte anzupassen.

1.3.4 Lernen aus Experimenten

Lernen aus Experimenten ist ein eigenes Forschungsgebiet in der Fachdidaktik, das hier weder umfassend behandelt werden kann noch soll. Da die konkrete Ebene der

beobachtbaren Phänomene jedoch eine zentrale Ebene des Physiklernens darstellt und auch einen Ausgangspunkt des Umgangs mit Repräsentationen bildet, wird das Lernen aus Experimenten unter der Perspektive des Umgangs mit internen und externen Repräsentationen kurz allgemein thematisiert.

Beim Lernen aus Experimenten lassen sich prinzipiell zwei didaktische Herangehensweisen unterscheiden: Das Schülerexperiment und das Demonstrationsexperiment. Der lernwirksame Einsatz von Schülerexperimenten ist eines der grundlegenden Themen der Physikdidaktik (vgl. Kircher, Girwidz & Häußler, 2000; Séré, 2002; Hopf, 2007 im Überblick). Die Idee des selbsttätigen Experimentierens lässt sich mit dem Credo einer konstruktivistischen Lehr-Lernauffassung verbinden, nach dem Lernen besonders dann erfolgreich ist, wenn die Lernenden zu Selbsttätigkeit aufgefordert werden. Konsens besteht über folgende Zieldimensionen von Schülerexperimenten (vgl. Séré, 2002):

- in der Verbindung von Theorie und Praxis,
- im Erwerb experimenteller Fähigkeiten und
- in der wissenschaftstheoretisch motivierten Zielsetzung des Kennenlernens naturwissenschaftlichen Denkens.

Die hohen Erwartungen, die mit dem Lernen aus eigenen Experimentiererfahrungen verbunden werden, konnten in empirischen Studien nicht eindeutig gestützt werden. Tatsächlich zeigt sich ein gemischtes Bild. Zahlreiche empirische Studien zur Wirksamkeit von Schülerexperimenten belegen, dass der Einsatz von Schülerexperimenten nicht automatisch zur Verbesserung des Unterrichts führt (vgl. Hopf, 2007, S. 35). Harlen (1999) analysiert die Lernwirksamkeit von Schülerexperimenten aus angloamerikanische Sicht und kommt zu dem Ergebnis, dass die bisherigen empirischen Untersuchungen weder einen positiven Einfluss auf die Motivation, die Fertigkeiten, den Wissenserwerb oder das Verständnis naturwissenschaftlicher Vorgehensweisen belegen. Den Grund für diese Befundlage sieht sie im Wesentlichen in Lernumgebungen, die den Schülern den Sinn des Experimentierens nicht vermitteln konnten. Faktoren, die hierzu beitragen, bestehen nach Harlen (1999) darin, dass Lernende sich nicht mit den eigentlichen Inhalten befassen, sondern damit, die Durchführung des Experiments zu organisieren.

Dabei überfordere die hohe Komplexität der Experimente die Lernenden. Durch die kognitive Überlastung konzentrierten sich die Lernenden auf das Befolgen der Experimentieranweisung, arbeiteten mit Rezepten oder fokussierten auf einen Teilaspekt. Zudem falle es den Schülern schwer, die Ergebnisse des Versuchs mit den wissenschaftlichen Konzepten zu verknüpfen. Oft komme es dabei zu folgender

Situation: Die Schüler wissen, auf welche Beobachtung das Experiment zielt und interpretieren diese Beobachtung dann auf Basis ihrer eigenen Vorstellungen.

Einen umfassenden Überblick zu Problemen beim Einsatz von Schülerexperimenten, der sich teilweise mit Harlens Sicht deckt, gibt Hopf (2007, S. 13-18). Dieser Überblick gibt Hinweise, warum die empirische Befundlage wenig eindeutig ist:

- Viele Schüler experimentieren ohne nachzudenken, d.h. die Schüler greifen nur wenig auf höhere kognitive Strategien zurück, da sie sich entweder strikt an die Experimentieranweisung halten ohne nachzudenken oder das Vorgehen ihrer Mitschüler kopieren.
- Lehrende und Lernende verbinden unterschiedliche Ziele mit dem Experimentieren. Während die Lehrenden die Schüler dazu anregen möchten, auf einer konzeptuellen Ebene wissenschaftliche Ideen zu überprüfen und zu manipulieren, sehen die Schüler das Ziel in der Manipulation der konkreten Versuchsmaterialien, dem Befolgen der Experimentieranweisung und dem Erhalten der korrekten Antwort, wodurch sie die Experimentiersituation nicht effektiv nutzen.
- Hintergrund für eine solche ineffektive Nutzung besteht auch in einem Mangel an Vorkenntnissen, was bedeutet, dass die Schüler oft überhaupt den wissenschaftlichen Hintergrund und das Ziel des jeweiligen Experiments nicht kennen.
- Die Schüler haben zudem große Schwierigkeiten, Theorie und Praxis miteinander zu verbinden, also die erworbenen theoretischen Konzepte und Begriffe auf die Experimentiersituation und die beobachteten Phänomene anzuwenden.
- Der Mangel an Verbindung zwischen Theorie und Praxis kann darin begründet sein, dass die Vorstellungen der Lernenden dem Lernen aus Experimenten entgegen stehen. Der Einfluss der Schülervorstellungen kann sogar so weit reichen, dass sie zu völlig unterschiedlichen Beobachtungen gelangen.
- Oft fehlt es den Schülern auch an den praktischen experimentellen Fähigkeiten, z.B. der Handhabung von Messinstrumenten, was die Durchführung des Experiments beeinträchtigt.
- Um ein Gelingen der experimentellen Durchführung zu garantieren, enthalten die Experimente oft kochbuchartige Anweisungen, was die gedankliche Eigenleistung der Schüler reduziert und den eigentlichen Vorteil, die Anregung zur Selbsttätigkeit konterkariert.

Lernen aus Experimenten unter dem Fokus der Repräsentationskompetenz zielt darauf, die Schüler zu einer intensiven kognitiven Auseinandersetzung mit den dahinterliegenden physikalischen Konzepten zu aktivieren, welche auf verschiedene Weisen extern repräsentiert werden können und die Schüler in der Konstruktion adäquater und relevanter mentaler Modelle zu unterstützen. Die Experimentiersituation soll konsequent zur Anregung des Denkens und Lernens genutzt werden (vgl. Hopf, 2007, S. 32). So schlagen Gunstone und Champagne (1990) vor:

„Teaching strategies must, at least in part, aim to help students understand the nature of appropriateness of the links being generated. [...] This involves time for interaction and reflection. [...] Less time is spent in interacting with apparatus, worksheets and instructions, an (in extreme cases) recipes. More time is spent in interacting with ideas“ (S. 179).

Diese Zielsetzung ist, wenn sie adäquat in die Praxis umgesetzt wurde, kein Allheilmittel für die verschiedenen Probleme, die dazu führen, dass das Lernen aus Schülerexperimenten nicht zu einem höheren Lernerfolg führt. Vielversprechend ist ein solcher Ansatz jedoch im Hinblick auf die Förderung eines adäquaten Grundverständnisses physikalischer Konzepte und somit für die Überwindung von naiven Schülervorstellungen.

Eine besondere Form der kognitiven Aktivierung beim Lernen aus Experimenten in der Naturwissenschaft besteht in der Predict-Observe-Explain Sequenz (POE-Sequenz) von White und Gunstone (1992). White und Gunstone (1992) entwickelten die Predict-Observe-Explain Sequenz, um individuelle Vorhersagen von Lernenden aufzudecken und die Diskussion der Lernenden über diese Ideen zu fördern. Die POE-Sequenz basiert auf dem klassischen Forschungsmodell bei dem eine Hypothese aufgestellt wird und die Annahme der Hypothese begründet, relevante Daten gesammelt und die Ergebnisse bewertet werden.

Zunächst werden die Lernenden gebeten, eine Vorhersage über den Ausgang eines Experiments zu machen. Diese Vorhersage halten die Lernenden schriftlich fest und geben Gründe für ihre Vorhersage an. Die Begründung liefert dem Lehrenden wertvolle Hinweise über die Theorien der Lernenden, hilft ihm, Schülervorstellungen aufzudecken und das Denken der Lernenden über den Sachverhalt zu verstehen.

Indem die Lernenden ihre eigenen Vorhersagen erklären und bewerten, beginnen sie ihre eigenen Lernprozesse zu bewerten sowie neue Deutungen zu

entwickeln. Angeregt und unterstützt wird dieser Prozess zusätzlich dadurch, dass auch die Vorhersagen und Begründungen der anderen Lernenden gehört werden. Die POE-Sequenz gliedert sich in folgende Schritte:

1. Im ersten Schritt der Vorhersage (predict) werden die Lernenden gebeten, unabhängig voneinander ihre Vorhersagen schriftlich festzuhalten und Gründe für ihre Vorhersage zu nennen.
2. Im zweiten Schritt (observe) wird das Experiment durchgeführt und die Lernenden erhalten die Möglichkeit, sich auf die Beobachtung zu konzentrieren, anschließend werden die Lernenden gebeten aufzuschreiben, was sie gesehen haben.
3. Im dritten Schritt (explain) ergänzen und überarbeiten die Lernenden ihre Erklärung unter Berücksichtigung der Beobachtungen. Nachdem die Lernenden ihre Erklärungen zu Papier gebracht haben, werden die überarbeiteten Erklärungen der Lernenden gemeinsam diskutiert.

In zahlreichen Studien wurde die POE-Sequenz verwendet und erwies sich vielfach als effektiv (Kearney & Treagust, 2001; Kearney, Treagust, Yeo & Zadnik, 2001; Crouch, Fagen, Callan & Mazur, 2004; Botzer & Reiner, 2005).

Eine Studie von Crouch et al. (2004) verglich drei verschiedene Formen der Darbietung von Demonstrationsversuchen untereinander und mit einer Kontrollgruppe, der überhaupt kein Versuch demonstriert wurde. Die Stichprobe bestand aus 133 College-Studenten, die einen Anfänger – Physikkurs vor Beginn ihres Medizinstudiums absolvierten. Die drei Bedingungen waren wie folgt konzipiert:

- In der Beobachtungsbedingung wurde den Lernenden der Versuch vorgeführt, parallel hörten die Lernenden die Erklärung des Lehrenden
- In der Vorhersage-Bedingung zeichneten die Lernenden ihre Vorhersage auf, beobachteten den Versuch und erhielten danach die Erklärung des Lehrenden
- In den Diskussionsbedingungen sagten die Lernenden die Versuchsaufgaben schriftlich vorher, beobachteten den Versuch, diskutierten ihre Beobachtungen mit den anderen und hörten abschließend die Erklärung des Lehrenden.

Die Studenten der Vorhersage- und Diskussionsbedingung schnitten jeweils signifikant besser im anschließenden Leistungstest ab als Studenten in der Beobachtungs- sowie in der Kontrollbedingung. Der Vorteil der Diskussionsbedingung gegenüber der Vorhersagebedingung war jedoch marginal. Dabei zeigte sich im Abschlusstest, dass Studenten, welche Demonstrationsversuche

passiv beobachten, die zugrundeliegenden Konzepte nicht signifikant besser verstanden als Lernende, welchen der Versuch überhaupt nicht demonstriert wurde.

Eine Studie von Botzer und Reiner (2005), in welcher die POE-Sequenz ebenfalls zum Einsatz kam, nahm die repräsentationalen Anforderungen beim Erwerb wissenschaftlicher Konzepte in den Blick. Die Autoren arbeiteten mit 16 Schülern aus Israel der neunten Klassenstufe zu einer Lerneinheit über Magnetfeldern. Die Teilnehmer verfügten über Grundkenntnisse in Mechanik, Wärmelehre und Elektrizitätslehre, hatten jedoch zuvor keinen Unterricht in Magnetismus erhalten. Ziel des Lernexperiments war es, Einblick in die mentalen Prozesse von Lernen mit geringem Vorwissen zu erhalten, während diese spontan visuelle Repräsentationen magnetischer Phänomene konstruierten. Die Schüler bearbeiteten in Zweiergruppen eine Reihe von Predict-Observe- und Explain-Sequenzen, wobei sie Versuchsmaterialien wie Magnete, Kompass und Nägel verwendeten. Dabei wurden sie aufgefordert, Phänomene des Magnetismus verbal zu beschreiben und mit einer Zeichnung zu erklären. Sie wurden ermutigt, frei zu sprechen und sich keine Sorgen über die Richtigkeit ihrer Antworten zu machen. In der Vorhersagephase wurden die Schüler gebeten, ein Modell zu antizipieren und zeichnerisch darzustellen. In der Beobachtungsphase wurden sie aufgefordert, ein deskriptives Modell zu entwickeln, das beschreibt, was unter den verschiedenen Versuchsbedingungen passiert und in der Erklärungsphase sollten sie in Zeichnungen ein Erklärungsmodell entwickeln.

Die Interaktionen der Schüler wurden videografiert und es wurden Feldnotizen erstellt. Zudem wurden die schriftlichen Antworten und die Zeichnungen einschließlich der Entwürfe eingesammelt. Die Lernaktivität umfasste drei Lehrabschnitte:

- Während der ersten Sitzung arbeiteten die Teilnehmer mit konkreten Objekten wie Magneten, Stahlnägeln, Kompassen und Eisenspänen, z.B. zu der Fragestellung „How can we distinguish between two steel bars that look identical, but one is magnetized and the other not“ (zit. n. Botzer & Reiner, 2005, S. 156).
- Im zweiten Abschnitt setzten sich die Teilnehmer mit Repräsentationen auf der mikroskopischen Ebene auseinander, z.B. durch die Bearbeitung der Fragestellung „If we break a magnet, each of the pieces will still have two poles. Suggest a structure that will explain this phenomenon“ (zit. n. ebd.).
- Im letzten Abschnitt lernten die Schüler mit formalen Repräsentationen zu operieren. So erforschten sie Muster, die durch Eisenspäne gebildet werden, um die Form von Feldlinien zu beobachten.

Untersucht wurden die Fragestellungen,

- in welchen Kategorien oder Modi die Repräsentationen magnetischer Phänomene der Schüler eingeordnet werden können,
- die Beziehung zwischen bildlicher Repräsentation und dem physikalischen Kontext sowie
- systematische Muster in der Entwicklung der bildlichen Repräsentationen während der Lernaktivität.

Botzer und Reiner (2005) entdeckten drei übergeordnete Klassen bildlicher Repräsentationen:

1. *Sensorisch basierte Repräsentationen*: Hier können geordnet nach dem Abstraktionsgrad drei Modi unterscheiden werden:
 - a) Repräsentationen, in denen Situationen auf einem oberflächlichen Niveau dargestellt werden und die auch Elemente enthalten, die für das Verständnis des physikalischen Sachverhalts irrelevant sind (von den Autoren als fotografisch sensorische Erfahrungen bezeichnet),
 - b) die Übertragung vorheriger sensorischer Erfahrungen auf eine neue sensorische Situation, also Projektionen früherer sensorischer Erfahrungen auf neue sensorische Situationen. Solche Repräsentationen enthalten Überlegungen, welche die Lernenden nicht direkt vor Augen hatten und
 - c) Transformationen konkreter Bilder. In diesem Repräsentationsmodus werden visuell wahrnehmbare Informationen deutlich verändert. Die konstruierten Repräsentationen enthalten visuelle Elemente, welche die Lernenden zuvor noch nie gesehen hatten.
2. *Rein bildliche Repräsentationen*: Rein bildliche Repräsentationen beziehen sich auf die Konstruktion mentaler Modelle mikroskopischer Strukturen, die nicht sichtbar sind. Sie basieren entweder auf vorheriger sensorischer Erfahrungen, die nun auf rein mental vorgestellten Situationen übertragen werden (Transformationen konkreter Bilder) oder es handelt sich um Transformationen mikroskopischer Repräsentationen. In letztem Fall werden mikroskopische Repräsentationen, welche die Schüler aus dem Bereich der Mechanik oder der Elektrizität kennen, verändert. Sie werden z.B. gedreht, ausgedehnt oder verlagert, um das mentale Bild der Zeichnung des physikalischen Problems zuzuordnen.
3. *Repräsentationen, die auf Formalismen beruhen, also aus formalen Regeln abgeleitet wurden*. Repräsentationen, die diesem Modus zugeordnet wurden,

enthalten formale Symbole, wie das Plus- oder das Minuszeichen und werden durch formale Regeln begründet (Ableitungen von Repräsentationen aus formalen Regeln).

Bei der Analyse der Beziehung zwischen physikalischem Kontext und Repräsentationsmodus zeigte sich, dass ein enger Zusammenhang zwischen Modus und Kontext besteht. Die Lernumgebung beeinflusst also maßgeblich, welche Form von Repräsentationen zum Lernen verwendet wird. Während im ersten Lehrabschnitt, in dem die Teilnehmer mit konkreten Objekten wie Magneten, Stahlnägeln, Kompassen und Eisenspänen arbeiteten, am häufigsten sensorisch-basierte Repräsentationen konstruiert wurden (Anteil von 70 %), wurden in der Sitzung zum Umgang mit formalen Repräsentationen vorwiegend abstrakte regelbasierte Repräsentationen entwickelt.

Die Analyse übergeordneter Entwicklungsverläufe der Repräsentationen (Entwicklungsmuster) ergab, dass sich die Vorstellungen der Schüler über die Zeit und die Interaktionen hinweg verändern. Während in einer Lernumgebung zur Auseinandersetzung mit konkreten Objekten die Konstruktion sensorisch basierter (photographischer) Repräsentationen erfordert wird, erfordern Lernumgebungen zur Auseinandersetzung mit formalen Regeln eine Kombination aus bedeutungsvollen Symbolen und fotografisch bildhaften Repräsentationen. Dabei werden situative mentale Modelle als Denkwerkzeuge eingesetzt und auf mikroskopisch relevante neue Situationen angewandt. Bildliche Repräsentationen bestimmter Situationen werden im Zug der Bearbeitung der Aufgaben als Werkzeuge für Schlussfolgerungen eingesetzt. Bei der Konstruktion mental visueller Repräsentationen sind zwei grundlegende kognitive Prozesse hervorzuheben: Projektion und Transformation. Das Übertragen früherer Erfahrungen auf neue Situationen zur Interpretation der aktuellen Situation kann erklären, wie es zur Bildung wissenschaftlich inadäquater Vorstellungen kommen kann. Die Transformation mentaler Bilder ermöglicht es, Repräsentationen an die spezifischen Erfordernisse der Aufgabenstellung oder des physikalischen Problems anzupassen.

Obgleich die Generalisierbarkeit der Ergebnisse durch die kleine Stichprobe (8 Fallstudien zur Interaktionen zwischen Schülern) limitiert ist, kommen die Autoren zum folgenden Schluss: Die Integration unterschiedlicher visueller Repräsentationen erleichtert das qualitative Verständnis physikalischer Phänomene. Der graduelle Fortschritt von der Konstruktion konkreter über die Bildung mikroskopischer bis hin zur Entwicklung mündlich sprachlicher Repräsentationen könne die Visualisierungsstrategien der Schüler fördern. Die Verwendung der Predict-Observe-Explain-Strategie könne die Schüler darin unterstützen, ihre

individuellen mentalen Modelle zu wissenschaftlich angemessenen Modellen weiterzuentwickeln. Dabei ist zu erwarten, dass die Vorhersagen der Schüler „Präkonzepte“ reflektieren, die im Widerspruch zu wissenschaftlichen Konzepten stehen. Die Autoren vermuten, dass ein angemessenes Design der Lernumgebung den Schülern sensorische Erfahrungen ermöglicht, welche einen Konzeptwechsel unterstützen. Die gemeinsame Konstruktion visueller Repräsentationen ermutigt zudem fruchtbare Fachdiskussionen über physikalische Phänomene. Abschließend weisen die Autoren darauf hin, dass die von den Schülern generierten Repräsentationen den Lehrenden einen Einblick in die Vorstellungen der Schüler bieten, somit nützliche Bewertungswerkzeuge darstellen und die Lehrenden darin unterstützen mit den Schülern über die Lerninhalte zu sprechen.

1.4 Integrierte Betrachtung: Physiklernen mit multiplen Repräsentationen aus Experimenten zur Förderung des konzeptuellen Verständnisses

Die Fähigkeit, Phänomene auf unterschiedliche Weise zu repräsentieren und zwischen verschiedenen Repräsentationsformen hin- und herzuwechseln (vgl. Dolin, 2007) ist eine Kernkompetenz in den Naturwissenschaften.

Der Wissens- und Kompetenzerwerb in den Naturwissenschaften beim Lernen mit multiplen Repräsentationen basiert auf allgemeinen kognitionspsychologischen Grundlagen der menschlichen Informationsverarbeitung.

Gemäß des *integrativen theoretischen Rahmenmodells des Text- und Bildverstehens* nach Schnotz und Bannert (2003) stellt sich der Lernprozess wie folgt dar: Wenn Lernende einen Text über für sie neue physikalische Inhalte lesen, die Erklärung des Lehrers hören, oder sich mit einer bildhaften Repräsentation befassen, tritt die visuelle verbale bzw. auditive verbale Information in das jeweilige Register ein.

Die Informationen werden dann durch den visuellen bzw. auditiven Kanal in das visuelle bzw. auditive Arbeitsgedächtnis weitergeleitet. Im Fall von verbalen Informationen wird nach selektiven Prozessen im Arbeitsgedächtnis eine propositionale Repräsentation gebildet, welche Informationen des Gelesenen bzw. Gehörten auf einer konzeptuellen Ebene enthält. Diese Propositionen wiederum führen zur Bildung und Elaboration eines mentalen Modells. Beim Bildverstehen kommt es zur visuellen perzeptuellen Repräsentation des Bildes im Arbeitsgedächtnis, wo sie nach Prozessen der Selektion von Informationen zur Konstruktion und Elaboration eines mentalen Modells führt. Während die aus-

gewählten verbalen Informationen zunächst ins propositionale Arbeitsgedächtnis gelangen und anschließend zur Konstruktion eines mentalen Modells im Arbeitsgedächtnis führen, wird aus den ausgewählten bildlichen Informationen also direkt ein mentales Modell im Arbeitsgedächtnis gebildet. Die Bildung eines aufgabenorientierten mentalen Modells, indem der physikalische Sachverhalt anschaulich visualisiert werden kann, spielt eine Schlüsselrolle im Verstehensprozess. Als Modelle, die strukturelle oder funktionelle Analogie zu den repräsentierten Objekten bzw. zu den repräsentierten Abläufen aufweisen, ermöglichen sie es den Lernenden, physikalische Zusammenhänge intern dynamisch zu simulieren und logische Schlüsse zu ziehen, um physikalische Probleme zu lösen. Beim Bearbeiten physikalischer Probleme fungieren mentale Modelle oft als interne Repräsentation der Ausgangslage.

Eine „gute“ Repräsentation der Ausgangslage zeichnet sich dadurch aus, dass mit ihr leicht diejenigen Operationen durchgeführt werden können, die zum Erreichen des Ziels führen. Die Konstruktion eines solchen mentalen Modells wird entscheidend durch die kognitiven Schemata mitbestimmt, die im Langzeitgedächtnis abgespeichert wurden, also durch das Vorwissen und das Vorverständnis der Lernenden. So kann ein mentales Modell von einem physikalischen Versuch Eigenschaften und Beziehungen enthalten, die in der Beschreibung des physikalischen Prozesses oder in der Abbildung des Versuchsaufbaus nicht enthalten sind. Umgekehrt werden an mentalen Modellen auch neue Informationen abgelesen und zu den bereits bestehenden propositionalen Repräsentationen im Langzeitgedächtnis hinzugefügt (vgl. Schnotz, 2006b). Modelle, die sich in einem bestimmten Kontext als nützlich erwiesen haben, sind nur schwer zu verändern. Ihre Korrektur ist mit erheblichem kognitivem Aufwand verbunden. Dies könnte ein weiteres Erklärungsmodell sein, warum Schülervorstellungen sehr hartnäckig sind.

Die internen Vorstellungen und mentalen Modelle der Lernenden schlagen sich dabei auch in den externen Repräsentationen nieder, welche die Lernenden bei der Aufgabenbearbeitung erstellen. Insbesondere in der Physik erfordert Problemlösen, dass eine adäquate und zweckmäßige physikalische Repräsentation gefunden werden muss (diese ist oft bildlich). Externe (bildliche) Repräsentationen haben die Funktion, die Lernenden dadurch zu entlasten, dass nicht alle relevanten Informationen innerhalb des Arbeitsgedächtnisses präsent gehalten werden müssen. Auf diese externen Repräsentationen können dann beispielsweise mathematische Operationen angewandt werden.

Gemäß des Konzeptwechsellansatzes der Reorganisation von Wissensfragmenten nach diSessa (1983, 1988, 1993) beginnen Lernende ohne systematische

physikalische Vorbildung¹¹ mit einem naiven physikalischen Verständnis, das aus einer Ansammlung von Phänomenen besteht, über das, was sie in der Welt beobachten. Diese p-Prims werden in dem Moment aktiviert, indem eine Situation eintritt, in der sie Relevanz aufweisen. Sie sind jedoch nicht hierarchisch geordnet und nur lose miteinander verbunden. Entsprechend fehlt es den Lernenden an den metakognitiven Fähigkeiten, zwischen theoretischen Annahmen und Beobachtung zu unterscheiden sowie die Kohärenz der verschiedenen Teilbereiche des Wissens kritisch zu durchdenken.

Das Verstehen eines wissenschaftlichen Konzepts beinhaltet: die Verwendung wissenschaftlich angemessener Repräsentationen,

- das Verständnis der Verbindungen zwischen den physikalischen Phänomenen, den entsprechenden physikalischen Konzepten und den unterschiedlichen Repräsentationsformen (wie verbale Beschreibungen, geometrischen Diagrammen, Tabellen und Gleichungen), welche sich auf das physikalische Konzept beziehen sowie
- die Kompetenz verschiedene Repräsentationsformen mit äquivalenter Bedeutung ineinander zu übersetzen.

Infolgedessen sollten die Lernenden darin unterstützt werden, selbst generierte Repräsentationen auf ihre interne Kohärenz zu prüfen sowie abzugleichen, ob unterschiedliche Repräsentationsformen einander inhaltlich entsprechen oder wechselseitig ergänzen.

In manchen Fällen ist das Vorverständnis der Lernenden durch inadäquate Rahmentheorien (vgl. Vosniadou & Brewer, 1992) geprägt. Dies kann auch dazu führen, dass die Schüler im Unterricht selbst Vorstellungen zu physikalischen Sachverhalten bilden, die inadäquat sind. Ein solch inadäquates Verständnis schlägt sich auch in den externen Repräsentationen nieder (vgl. Cox, 1999), welche die Lernenden erstellen. Im Lernprozess nehmen externe Repräsentationen u.a. durch die Entlastung des Arbeitsgedächtnisses eine Schlüsselrolle ein. Voraussetzung für diese unterstützende Funktion externer Repräsentationen ist jedoch, dass die jeweilige externe Repräsentation wissenschaftlich angemessen und für die Bearbeitung der Aufgabe relevant ist. Ist diese Voraussetzung nicht erfüllt, ist es sinnvoll, die bisherigen Sichtweisen der Lernenden zu reflektieren und die Lernenden darauf aufmerksam zu machen, dass ihre Repräsentation inkonsistent

¹¹ Dies trifft auf Mittelstufenschüler der Klassenstufe 7 und 8 zu. Zumal die Schüler in diesen Jahrgangsstufen erstmalig Physikunterricht erhalten.

zu dem physikalischen Phänomen oder zu anderen Repräsentationen ist. Anhand externer Repräsentationen können die Vorstellungen der Lernenden dann thematisiert und im Licht der neuen Annahmen reinterpretiert werden, so dass die Lernenden ihr Denken über einen Sachverhalt von inadäquaten Rahmenannahmen befreien können.

Beide Strategien, sowohl die Reorganisation von Wissensfragmenten als auch die Überwindung von inadäquaten Rahmentheorien, implizieren, dass die Bestrebungen der Lernenden wissenschaftliche Konzepte in den Naturwissenschaften zu verstehen, die aktive Auseinandersetzung mit Repräsentationen erfordern, bei denen die Lernende ihre kognitiven Fähigkeiten und ihre Ressourcen im Umgang mit Repräsentationen einsetzen.

Lernen mit multiplen Repräsentationen beinhaltet die Auseinandersetzung mit deskriptiven und depiktionalen Repräsentationen unterschiedlicher Abstraktionsgrade. Hierbei kann zwischen der Phänomen-, der Modell- und der Theorieebene (vgl. Leisen, 1998) unterschieden werden. Oft dienen die unterschiedlichen fachbezogenen Repräsentationen der Erklärung eines physikalischen Phänomens, das in einem physikalischen Experiment beobachtet werden kann. Diese unterschiedlichen Repräsentationsformen wie die „Verbalisierung der Beobachtung“, die „schematische Darstellung des zugehörigen physikalischen Prozesses“ oder die „deskriptiv-abstrakte Repräsentation einer zugehörigen allgemeinen mathematischen Gesetzmäßigkeit“, werden dabei oft auf die Phänomene, die im Experiment beobachtet werden können, rückbezogen. Lernen aus Experimenten unter dem Fokus der Repräsentationskompetenz beinhaltet, die Schüler zu einer intensiven kognitiven Auseinandersetzung mit den dahinterliegenden physikalischen Konzepten zu aktivieren, die auf verschiedene Weisen extern repräsentiert werden können.

Die der Predict-Observe-Explain-Sequenz nach White und Gunstone (1992) ist hierbei eine vielversprechende Strategie der kognitiven Aktivierung. Ihre Verwendung hat das Potential, die Schüler darin zu unterstützen, ihre individuellen mentalen Modelle zu wissenschaftlich angemessenen Modellen weiterzuentwickeln. Dabei ist zu erwarten, dass die Vorhersagen der Schüler, Schülervorstellungen reflektieren, die im Widerspruch zu wissenschaftlichen Konzepten stehen.

2 Empirische Untersuchungen

2.1 Theoretische Vorüberlegungen zum Lerngegenstand

2.1.1 Gründe für die Wahl des Lerngegenstands

Die im zweiten Teil dieser Arbeit vorgestellten empirischen Studien (Pilot- und Hauptstudie) fanden im Bereich der Strahlenoptik statt. Mehrere Gründe sprachen für die Entscheidung, Strahlenoptik als Lerngegenstand zu wählen.

1. Viele Studien im Bereich des Umgangs mit multiplen Repräsentationen im Hinblick auf die Förderung des konzeptuellen Verständnisses fanden im Bereich der Mechanik / Teilchenmodelle (Plötzner & Spada, 1998; Waldrip et al., 2010; Hubber et al., 2010) bzw. der Kinematik (Wilhelm, 2005) statt. Im Vergleich hierzu ist das Gebiet der Strahlenoptik weniger stark erforscht. Die in dieser Arbeit vorgestellten Studien im Bereich der Strahlenoptik wurden mit Oberstufenschülern (vgl. Mortimer & Buty, 2009) oder College-Studenten (vgl. Goldberg & McDermott, 1987) durchgeführt.
2. Das Gebiet der Strahlenoptik weist für die Untersuchung den „technischen“ Vorteil auf, dass die Unterschiede in den Vorkenntnissen der Schüler kaum oder in geringem Maße durch den vorherigen Unterricht bedingt sind. Zum einen liegt dies daran, dass das Themengebiet der Strahlenoptik entsprechend dem Lehrplan für das Land Rheinland-Pfalz (vgl. Ministerium für Bildung, Wissenschaft, Jugend und Kultur. Rheinland-Pfalz. Lehrplan-Entwürfe Lernbereich Naturwissenschaften Biologie Physik Chemie S. 175 f., 190) das erste Gebiet ist, in dem die Schüler im Fach Physik an Gymnasien und Realschulen unterrichtet werden. Selbst wenn ein anderes Gebiet zeitlich vorgezogen wurde, fällt dies wenig ins Gewicht, da „das Gebiet der Optik [...] wenig weitergehende Bezüge zu den nachfolgenden Lerninhalten aufweist“ (zit. n. ebd., S. 190).
3. Die Strahlenoptik stellt entsprechend ein relativ in sich geschlossenes Gebiet dar, in dem die wesentlichen zentralen Grundstrukturen der Physik enthalten sind (zit. n. ebd.):

„Ausgehend von der Beobachtung von Phänomenen der Alltagswelt sollen die Kinder einen ersten Einblick in die Untersuchung optischer Erscheinungen gewinnen. [...] Gerade das Gebiet der Optik, das wenig weitergehende Bezüge zu den nachfolgenden Lerninhalten aufweist, [...] ist gut geeignet, die Kinder behutsam in die Arbeitsweisen der Physik einzuführen. Elemente der physikalischen Fachsprache und Begriffswelt werden eingeübt. Am Beispiel des Lichts wird zum ersten Mal eine sinnvolle Modellbildung aufgezeigt. Das Experiment steht im Mittelpunkt des Unterrichts. Wo es nur angeht, sollen die Kinder selbst tätig werden“ (S. 190).

4. Die Strahlenoptik enthält somit die grundlegenden wissenschaftstheoretischen Aspekte und Methoden der Physik als wissenschaftliche Disziplin: Experimentieren, Beobachten und Modellbildung. Die Bildentstehung durch die Sammellinse wird hierbei explizit im Lehrplan für Realschulen und Gymnasien als ein zentraler Themenbereich genannt. Die folgenden Lerninhalte sind gemäß Lehrplan mit der Bildentstehung verbunden: Lichtausbreitung, Lichtquellen, Lichtbündel und Lichtstrahl, Lichtstrahl als Modellvorstellung, Verlauf spezieller Strahlen oder Lichtbündel an der Sammellinse, Brechungsverhalten paralleler Strahlen, Vereinfachung der Brechung an der Linsenmitte, Konstruktion der Bilder; Übersicht über Art und Lage der Bilder (vgl. ebd., 190 f.).
5. Die Bildentstehung bei der Sammellinse umfasst unterschiedliche Repräsentationen auf verschiedenen Abstraktionsebenen.
6. Nicht zuletzt sind für den Bereich der Strahlenoptik eine Reihe von domänen-spezifischen Schülervorstellungen dokumentiert, die sich auch gerade auf das repräsentationale Verständnis der Schüler auswirken.
7. Um die kognitiven Anforderungen der experimentellen Durchführung des physikalischen Experiments zur Bildentstehung am Hohlspiegel bzw. der Sammellinse sowie der Datenauswertung und Interpretation zu erfassen, wurde eine kognitive Aufgabenanalyse durchgeführt (vgl. Gagné, Briggs & Wager, 1988). Im Zuge der Aufgabenanalyse wurden auch die verschiedenen Repräsentationsformen identifiziert, die bei der Durchführung, Auswertung und Interpretation des Schülerexperiments zur Bildentstehung am Hohlspiegel bzw. der Sammellinse zum Tragen kommen:
 - Die Ebene der verbalen Beschreibung der Phänomene (deskriptive Repräsentation, die sich an der konkreten Beobachtung orientiert).
 - Die Ebene der Messwerte. Die Schüler können Gegenstandsweite, Bildweite Gegenstandsgröße und Bildgröße abmessen und in tabellarischer Form festhalten (deskriptive Repräsentation, die durch die räumliche Anordnung eine depiktionale Komponente enthält) (vgl. Abbildung 8).

- Die Ebene der Modellbildung: Ausgehend vom Modell des Lichts als Strahl, kann die Entstehung der Bilder in einer Strahlenkonstruktion (depiktional-schematische Repräsentation) dargestellt werden (vgl. Abbildung 9).
- Die Ebene der generalisierenden Beschreibung der Verhältnisse von Gegenstandsweite, Bildweite und Abbildungsmaßstab. Diese verbale abstrakte Beschreibung stellt eine verbale Repräsentation dar (vgl. Abbildung 10), die auf die mathematische Repräsentation der Abbildungsgleichung hinführt (vgl. Abbildung 11).
- Die Ebene der Mathematisierung: hier das Abbildungsgesetz bzw. die Linsengleichung. Da der Lehrplan einen „sparsamen“ Umgang mit der Mathematik nahelegt (vgl. ebd., S. 175), beschränkte sich die hier thematisierten mathematischen Zusammenhänge auf die Abbildungsgleichung (vgl. Abbildung 11). Ein Vorteil der Abbildungsgleichung besteht darin, dass ihre Herleitung geometrisch gut veranschaulicht werden kann, was ebenfalls mit den Zielen des Lehrplans (vgl. ebd., S. 190) konform ist (vgl. Abbildung 12).

Auf Basis der Aufgabenanalyse wurden die folgenden zehn kognitiven Schritte identifiziert. Die vollständige Aufgabenanalyse befindet sich in Anhang A (auf der Produktseite zu diesem Buch unter www.springer.com).

1. Aktivierung von Vorwissen und Nachdenken anregen. Die Aktivierung von Vorwissen bezieht sich in der 8. Klasse auf das Suchen von Alltagsbeispielen und technischen Anwendungen, wie etwa bei Autoscheinwerfern oder beim Kosmetikspiegel. Auf der repräsentationalen Ebene kommen hier Fotografien oder realistische Zeichnungen von Gegenständen ins Spiel.
2. Erste Fragestellungen identifizieren und Erklärungsversuche starten. Die Schüler können hier an bekannte Phänomene der Optik erinnert werden, wie das Reflexionsgesetz (Einfallswinkel = Ausfallswinkel). Sie operieren also mit abstrakten verbalen und schematischen Repräsentationen wie der geometrischen Anwendung des Reflexionsgesetzes.
3. Versuchsplanung der Schüler: Hypothesen aufstellen, Zusammenhänge auf Basis des Vorwissens und der Beobachtung vermuten und zu testende Einflussgrößen identifizieren.
4. Versuchsplan erarbeiten und einen Versuchsaufbau (vgl. Abbildung 7) festlegen: dabei die technische Umsetzung bedenken, den Versuchsaufbau sprachlich beschreiben. Zudem müssen die Schüler erkennen, was variiert, was beobachtet und was gemessen wird, d.h. abhängige und unabhängige Variablen müssen identifiziert und unterschieden werden.

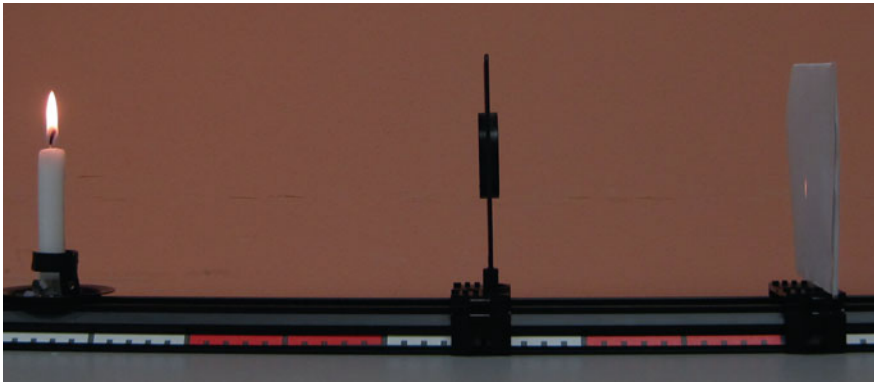


Abbildung 7: Experiment zur Bildentstehung bei der Sammellinse

5. Messwertetabelle planen (vgl. Abbildung 8): Festhalten, welche Größen gemessen werden soll und mit welche Ergebnissen zur Bestätigung oder dem Verwurf der Hypothesen in eine mathematisches Modell überführt werden kann.

Bildfall	Gegenstandsweite g / cm	Bildweite b / cm	Gegenstandsgröße G / cm	Bildgröße B / cm
gleich großes Bild	$g = 2f$	$b = 2f$	$G = B$	$B = G$
	10 cm	10 cm	4 cm	4 cm
ver- kleinertes Bild	$g > 2f$	$f < b < 2f$	$G > B$	$B < G$
	15 cm	7,5 cm	4 cm	2 cm
ver- größertes Bild	$f < g < 2f$	$b > 2f$	$G < B$	$B > G$
	8 cm	13,3 cm	4 cm	6,8 cm

Abbildung 8: Im Unterricht verwendete Messwertetabelle zur Bildentstehung bei der Sammellinse

6. Justierung des experimentellen Aufbaus und Messung der Werte unter Berücksichtigung des Versuchsplans. Hier wird erfordert, die konkrete Beobachtung zu quantifizieren und entsprechend in das Auswertungsschema (hier eine Tabelle, vgl. Abbildung 8) einzuordnen.
7. Auswertung und Abgleich mit Hypothese: Ggf. Fehlerquellen notieren und deren mögliche Ursachen erklären.
8. Modellbildung: Erstellung der Abbildungskonstruktion für die 3 Fälle. Vergrößertes Bild, verkleinertes Bild und gleichgroßes Bild. Die Modellbildung betrifft das Verstehen der schematischen Konstruktion des Strahlendiagramms (vgl. Abbildung 9).

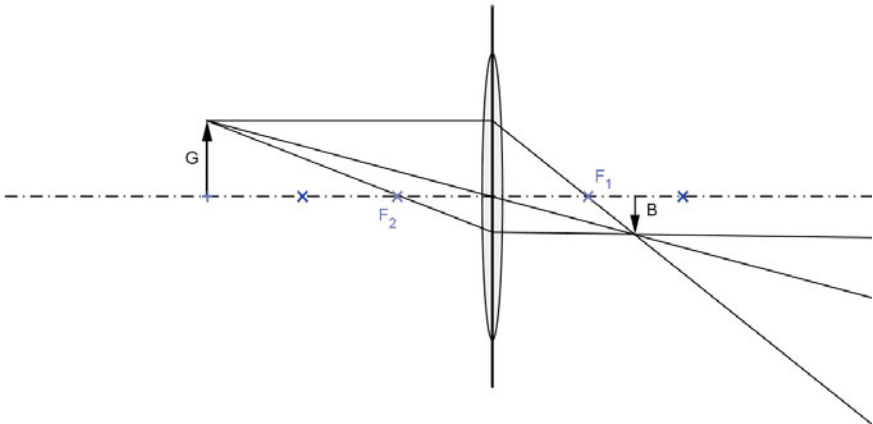


Abbildung 9: Strahlenkonstruktion zur Bildentstehung bei der Sammellinse

9. Bestätigung des Modells durch eine mathematische Erklärung: Dieser Schritt erfordert schließlich die schematische Darstellung des Strahlengangs (vermittelt über eine verbale Beschreibung) in eine mathematische Repräsentation zu übertragen, in diesem Fall in das Abbildungsgesetz.

Das Verhältnis von Bildgröße zu Gegenstandsgröße entspricht dem Verhältnis von Bildweite zu Gegenstandsweite.

Abbildung 10: Generalisierende Beschreibung der Verhältnisse von Gegenstandsweite, Bildweite und Abbildungsmaßstab

$$A = \frac{B}{G} = \frac{b}{g}$$

Abbildung 11: Abbildungsgleichung

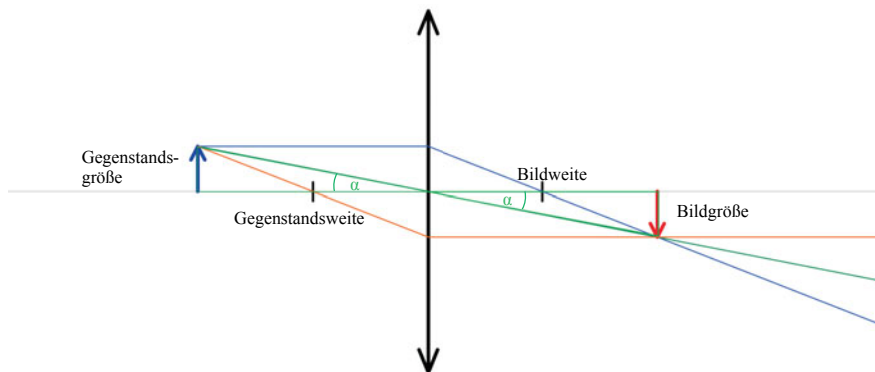


Abbildung 12: Geometrische Veranschaulichung der Abbildungsgleichung

10. Reflexion des Gelernten: Zusammenhänge der einzelnen Schritte reflektieren: Was war die Ausgangsfragestellung? In welche Teilziele konnte die Fragestellung zergliedert werden? Welche Operationen mussten angewandt werden, um diese Etappen zu erreichen, welche Repräsentationsformen wurden dabei genutzt? Wie gut wurden die Zwischenziele erreicht? In welcher Relation steht das Erreichte zur Ausgangsfragestellung? Konnte diese letztlich beantwortet werden?

2.1.2 Schülervorstellungen im Kontext der Bildentstehung bei der Sammellinse

Naive Vorstellungen im Bereich der Optik wurden insbesondere in der Fachdidaktik in den letzten Jahrzehnten genauer untersucht (vgl. Kärrqvist & Andersson, 1983; Guesne, 1985; Wiesner, 1986, 1992a, 1992b, 1994; Reiner et al., 2000). In der Regel wurden die Teilnehmer, Schüler oder Studenten, zu Alltagsphänomenen (z.B. „Was verstehst du unter Licht?“ zit. n. Guesne, 1985, S. 80) oder zu physikalischen Experimenten (vgl. Goldberg & McDermott, 1987, S. 109) befragt.

Generell ergaben die wissenschaftlichen Untersuchungen, dass die Teilnehmer die Konzepte nicht systematisch verwenden, dass viele Vorstellungen unter wissen-

schaftlichen Gesichtspunkten nicht akzeptabel waren und dass es schwierig ist, diese Vorstellungen durch übliche Instruktionen zu verändern. Solche naiven und wissenschaftlich unangemessenen Vorstellungen wurden dabei in den folgenden Bereichen gefunden: Entstehung von Licht und Schatten, Entstehung von Farben, physikalische Beschreibung des Sehvorgangs, Entstehung von Spiegelbildern und Entstehung reeller und virtueller Bilder bei der Sammellinse sowie ontologische Grundannahmen zum Thema Licht.

In dem folgenden Literaturrückblick werden naive Vorstellungen zu den Bereichen der Strahlenoptik vorgestellt, die in einem engeren inhaltlichen Zusammenhang des gewählten Lerninhalts der Bildentstehung bei der Sammellinse stehen. Diese betreffen alle genannten Bereiche außer der Entstehung von Licht und Schatten und der Entstehung von Farben.

Eine der ersten und maßgeblichen Studien zu naiven Vorstellungen über Licht von Jugendlichen hat Guesne (1985) durchgeführt. Sie interviewte 30 Jugendliche im Alter von 13 und 14 Jahren, die zuvor keinen Optikunterricht in der Schule erhalten hatten, in standardisierten Interviews.

Zu den wesentlichen Vorstellungen zählte, dass die Befragten Licht mit seiner Quelle, seinen Wirkungen oder einem Zustand gleichsetzten.

- Gleichsetzung mit der Quelle: Manche der interviewten Jugendliche lokalisierten das Licht ausschließlich in der Lichtquelle z.B. in der Glühbirne.
- Gleichsetzung mit den Wirkungen: Einige Kinder gaben an, Licht könne man nur an den Stellen sehen, auf denen das Licht auf der Wand helle Flecken erzeugt (z.B. durch Sonnenlicht oder durch die Reflexion eines Spiegels).
- Gleichsetzung mit einem Zustand: Jugendliche, die Licht mit einem Zustand gleichsetzen, bezeichnen Licht als Helligkeit, die sich z.B. mit dem Wetter ändere.

Als wichtigste Tatsache stellte sich hierbei heraus, dass kaum einer der befragten Jugendlichen annahm, gewöhnliche Gegenstände werfen Licht zurück. Guesne (1985) betont, dass diese Erkenntnis aus zwei Gründen von prinzipieller Bedeutung für den gesamten Bereich der Optik sei:

- Erstens sei es praktisch unmöglich, die Entstehung des Bildes von irgendeinem Gegenstand, der nicht selbst leuchtet z.B. in der Fotografie, zu verstehen, solange man diese Vorstellung nicht begriffen habe (vgl. Guesne, 1985, S. 85).
- Zweitens sei diese Annahme entscheidend dafür, den Sehvorgang zu begreifen. Da Kinder im Alltag Licht nur erkennen, wenn es einen deutlich wahrnehmbaren

Effekt hervorbringe, glaubten die befragten Kinder nicht, dass auch bei nicht-selbst-leuchtenden Körpern Licht ins Auge gelange (vgl. Guesne, 1985, S. 91).

Diese Beobachtung und die zugehörige Deutung greift Wiesner (1992a, S. 16) nun wie folgt auf: Der Autor stellt ausgehend von den grundlegenden Lernschwierigkeiten einer unzureichenden physikalischen Sehvorstellung weitere zentrale Lernschwierigkeiten in der Strahlenoptik zusammen. Wiesner (1992a) bestätigte die Beobachtungen Guesnes (1985) und zeigte, dass Schüler auch nach dem Optikunterrichtangaben, beleuchtete Gegenstände sehen zu können, ohne dass dazu Licht von dem wahrgenommenen Gegenstand ins Auge falle (vgl. ebd.).

Wiesner (1992a) sieht nun einen engen Zusammenhang zwischen einer fehlerhaften physikalischen Sehvorstellung und fehlerhaften Konzepten zur Streuung. Gerade das Konzept der Streuung betrifft den Kern einer solchen physikalischen Sehvorstellung, zu der der wichtige Aspekt zählt, dass beleuchtete Gegenstände selbst Licht abstrahlen (Wiesner, 1994, S. 7). So zeigte Wiesner (1986, S. 26), dass viele Schüler davon ausgehen, beleuchtete Gegenstände wie Tische, Bücher oder Bilder strahlten kein Licht ab. Auch hier fehlt nach Wiesner (1986) die Verbindung zwischen wahrgenommenen Gegenständen und den Augen des Betrachters. Daran schließt sich das Konzept an, das auftreffende Licht mache die Gegenstände hell, bleibe auf diesen liegen oder verschwinde allmählich (vgl. ebd.). Ebenfalls damit verbunden ist die Vorstellung, dass bei Lichtquellen mit geringer Intensität, wie Räucherstäbchen oder weit entfernte beleuchtete Fenster, kein Licht mehr ins Auge gelange (vgl. ebd.).

Guesne (1985) beschreibt im Zusammenhang mit der Vorstellung, beleuchtete Gegenstände strahlten kein Licht ab, eine Interpretation des Sehens, bei der dem Auge ein aktiver Part zugeschrieben wird, während dem Gegenstand nur eine passive Rolle zukommt. Dabei bleibe die Bewegung, die vom Auge zum Gegenstand hingehe, abstrakt. Das Subjekt werde dabei als Ursprung des Prozesses gesehen und nicht als Empfänger von Licht.

Die Schwierigkeiten mit der physikalischen Sehvorstellung schlagen sich nach Wiesner ebenfalls auf Lernschwierigkeiten mit dem Spiegelbild nieder. Auch hier erkennen viele Schüler nicht, dass das Licht aus der Richtung des Spiegels ins Auge fallen muss, damit das Spiegelbild wahrgenommen werden kann.

Nach Wiesner (1986, 1992a, 1992b) bereite Schülern aber nicht nur die Erklärung der Wahrnehmung, sondern auch die Lage des Spiegelbildes große Probleme: so werden die strahlengeometrische Konstruktion des Spiegelbildortes von den meisten Schülern als nicht überzeugend eingestuft. Die Schüler gingen oft davon aus, dass das Spiegelbild auf der Spiegeloberfläche liege (vgl. Wiesner,

1986, S. 26, 27, 1992a, 1992b, S. 288). Der Spiegel werde auch oft als ein Gegenstand aufgefasst, der das Spiegelbild zum Betrachter zurückwerfe. Wiesner beschreibt hierzu folgende Situation (zit. n. Wiesner, 1992a):

„Fragt man z.B. einen Schüler (oder auch einen Erwachsenen), was man tun kann, um in einem kleinen Taschenspiegel mehr vom eigenen Gesicht sehen zu können, hält die überwiegende Mehrzahl den Spiegel weiter weg vom Gesicht. [...] Je weiter entfernt vom Gegenstand der Spiegel ist, desto kleiner erschienen nach dieser Meinung in dem Spiegel die Gegenstände, und dieses verkleinerte Bild wirft er zum Betrachter zurück“ (S. 16).

Auch mit der Entstehung von reellen und virtuellen Bildern durch die Sammellinse ist eine Reihe von Fehlvorstellungen verbunden:

Schwierigkeiten bereiten Schülern insbesondere die physikalischen Vorstellungen des Abbildungsvorgangs. Viele Schüler nutzen zur Erklärung der Entstehung des reellen Bildes bei der Sammellinse nicht das Konzept einer Punkt-zu-Punkt-Abbildung.

Zu den gängigen Vorstellungen zählt: Das Bild ginge als Ganzes durch die Linse zum Schirm und werde dabei in der Linse umgedreht (Wiesner, 1994, S. 8). Der Autor bezeichnet dieses weitverbreitete Konzept als holistische Erklärung des Abbildungsvorgangs. Dass Lernende auf eine solche holistische Erklärung des Abbildungsvorgangs zurückgreifen, wird insbesondere bei Abdeckaufgaben deutlich. Unter der Annahme, das Bild werde als Ganzes vom Gegenstand aus durch die Linse auf den Schirm transportiert, ist es nur konsequent anzunehmen, dass ein Teil des Bildes abgeschnitten werde, wenn man eine Blende vor die Linse hält (vgl. Wiesner, 1992b, S. 288). Hält man eine ringförmige Blende vor die Linse, glauben viele Lernende entsprechend, das Bild werde ringförmig am äußeren Rand abgeschnitten. Wird die Linse zur Hälfte abgedeckt, gehen viele Schüler und auch Studenten davon aus, dass auch das reelle Bild zur Hälfte abgeschnitten werde, einige überlegen sich sogar, welche Hälfte des Bildes (obere versus untere Hälfte) betroffen sei (vgl. Goldberg & McDermott, 1987, S. 112; Wiesner, 1994, S. 8).

Goldberg und McDermott (1987) berichten in einer Studie über das Verständnis der Entstehung reeller Bilder durch die Sammellinse und den Hohlspiegel von der folgenden weiteren Verständnisschwierigkeit, die sie bei jungen Erwachsenen beobachten konnten. Die Autoren interviewten 80 College Studenten, die einen Einführungskurs in Physik besuchten, ca. die Hälfte aller Studenten hatten kein Vorwissen in Strahlenoptik durch Schule oder Universität erworben, die andere Hälfte hatte bereits ein experimentelles Praktikum in Optik absolviert. Den

Studenten wurde ein Versuchsaufbau gezeigt, bei dem eine Glühbirne, eine Linse und ein Schirm hintereinander auf einer optischen Bank montiert sind. Im Verlauf des Interviews wurden die Studenten gefragt, wo das Bild wäre, wenn man den Schirm entfernt und sie frei um den Versuchsaufbau im Raum herumgehen können. Nur wenige Studenten waren in der Lage zu erkennen, dass sich das Bild an der gleichen Position befindet wie der Schirm. Die übrigen Studenten gaben eine Erklärung ab wie etwa, das Bild sei auf oder in der Linse. Insbesondere war die Vorstellung verbreitet, dass ein Bild nur mit Hilfe eines Schirms gesehen werden kann und dass die Linse das Bild quasi einrahme (vgl. Goldberg & McDermott, 1987, S. 114).

Das virtuelle Bild bei der Sammellinse bereitet Schülern häufig ähnliche Schwierigkeiten wie die Entstehung des Spiegelbildes. Wiesner (1986) berichtet diesbezüglich zwei wesentliche Ansichten „(a) man schaut durch die Linse (quasi wie durch eine Gardine) hindurch auf einen Gegenstand und (b) das Bild liegt wie ein Spiegelbild auf der Linsenoberfläche“ (zit. n. Wiesner, 1986, S. 28).

Weitere Erklärungen, die ebenfalls das Konzept der Punkt-zu-Punkt-Abbildung außer Acht ließen, bestehen in der Idee, die Linse konzentriere das Licht oder hinter der Linse sei mehr Licht bzw. seien mehr Strahlen vorhanden als vor der Linse (vgl. ebd., S. 158). Häufig werde die Entstehung reeller Bilder durch Spiegelung und Reflexion erklärt, dabei werde einem Gegenstandspunkt in der Regel nur ein Strahl zugeordnet und nicht ein divergierendes Strahlenbündel (vgl. ebd., S. 16).

Guesne (1985) befragte in ihrer Studie zu „Vorstellungen von Kindern über Licht“ auch zur Rolle der Sammellinse in der Funktion als Lupe und Brennglas. Dabei teilten sich die Kinder in zwei Antworttypen auf: Die erste Gruppe war der Ansicht, das Vergrößerungsglas mache das Licht größer, während die andere Hälfte davon überzeugt war, die Sammellinse konzentriere das Licht.

Kinder, welche das Konzept der Lichtkonzentration vertraten, waren der Ansicht, die gesamte Lichtmenge, die durch das Vergrößerungsglas hindurchgehe, bleibe hinter der Linse erhalten, was wissenschaftlich korrekt ist. Dass auch diese Gruppe von Kindern nicht notwendigerweise eine physikalisch angemessene Vorstellung von der Funktionsweise haben, zeigte sich in den angefertigten Zeichnungen der Kinder, mit denen die Kinder verdeutlichen sollten, wie die Linse das Licht konzentriere (vgl. Guesne, 1985, S. 87). So gab einer der befragten Jugendlichen an, ein einziger Strahl verlasse nach der Bündelung die Linse.

Kinder, die der Ansicht waren, das Vergrößerungsglas mache das Licht größer, gaben entweder an, hinter der Lupe sei mehr Licht als vor der Lupe oder das Licht werde hinter der Lupe verstärkt bzw. vermehrt. So stellte sich eines der befragten

Kinder vor, Licht könne nicht nur verstärkt werden, sondern auch verloren gehen oder verschwinden.

Gerade letztere Vorstellung zeigt, dass Kinder und auch Erwachsene mit wenig physikalischer Vorbildung wissenschaftstheoretisch grundlegend verschiedene Vorstellungen von Licht besitzen. Dies zeigt sich insbesondere auch in naiven Vorstellungen, die eine materialistische Konzeption von Licht annehmen. Reiner et al. (2000) beschreiben eine solche Konzeption unter Berufung auf Forschungsergebnisse verschiedener Studien (vgl. Andersson & Kärrqvist, 1983; Guesne, Séré & Tieberghien, 1983; Reiner, 1987; Smith, 1987; zit. n. Reiner et al., 2000, S. 14, 15) wie folgt: Wenn Lernende gefragt werden, wie das Sehen funktioniert, gaben sie an, dass Moleküle zwischen dem gesehenen Gegenstand z.B. einem „Buch“ und dem „Auge“ vorhanden seien. Der Sehvorgang wird demgemäß als das Ergebnis sich bewegender Lichtpartikel interpretiert. Hierbei wird Licht oft als Flüssigkeitsstrom beschrieben, der in Bewegung ist, sich aber auch in Ruhelage befinden kann.

Der eben beschriebene Literaturreisblick zeigt, dass naive und wissenschaftlich unangemessene Konzepte in allen Inhaltsbereichen, zu finden sind, welche für das Verständnis der Bildentstehung durch die Sammellinse relevant sind. Der nächste Abschnitt zielt darauf, diese Vorstellungen erstens auch vor dem Hintergrund der in Kapitel 1.3.2 beschriebenen Konzeptwechselansätzen einzuordnen und zweitens, mögliche Lernschwierigkeiten, die aus diesen Schülervorstellungen resultieren, beim Umgang mit (multiplen) Repräsentationen aufzuzeigen.

Zusammenfassend lassen sich die Schülervorstellungen diesen drei übergeordneten wissenschaftlich inadäquaten Konzepten zuordnen:

1. *Licht wird als eine Art Gegenstand oder Zustand aufgefasst.* Unter dieses Konzept lassen sich folgende Schülervorstellungen subsumieren:
 - Licht wird mit seiner Quelle gleichgesetzt.
 - Licht wird als Zustand („Helligkeit“) verstanden.
 - Licht wird als Substanz bzw. Materie (z.B. als eine Art Flüssigkeit) betrachtet und
 - Auftreffendes Licht macht die Gegenstände hell und bleibt auf den Gegenständen liegen.
2. *Reelles und virtuelles Bild werden als Gegenstände verstanden.* Diesem Konzept entsprechen die Vorstellungen:
 - Das Spiegelbild liegt auf der Spiegeloberfläche bzw. auf die Linsenoberfläche.
 - Das Bild geht als Ganzes durch die Sammellinse.

- Bei einer Lupe schaut man durch die Linse hindurch auf das Bild.
 - Wird ein Teil der Linse abgedeckt, so wird auch ein Teil des Bildes abgeschnitten.
3. *Optische Geräte und auch das menschlichen Auge spielen eine aktive Rolle beim Sehvorgang oder bei der Bildentstehung.* Diesem Konzept lassen sich folgende Vorstellungen zuordnen:
- Die Interpretation des Sehens, bei der dem Auge ein aktiver Part zugeschrieben wird, während dem Gegenstand nur eine passive Rolle zukommt.
 - Je weiter entfernt vom Gegenstand der Spiegel ist, desto kleiner erscheinen in dem Spiegel die Gegenstände und dieses verkleinerte Bild wirft der Spiegel zum Betrachter zurück.
 - Die Linse empfängt das Bild als Ganzes und dreht es aktiv um.
 - Die Linse konzentriert oder vergrößert das Licht.

Im folgenden Abschnitt wird der Versuch gestartet, die aufgeführten Schülervorstellungen vor dem Hintergrund des Rahmentheorieansatzes von Vosniadou (1992, 1994) und der Perspektive des fragmentarischen Wissens nach diSessa (1983, 1988, 1993) einzuordnen. An dieser Stelle soll ausdrücklich betont werden, dass die hier vorgenommene Einordnung ein hypothetisches Gedankenspiel darstellt. Ziel ist es, die Vorstellungen zu strukturieren und mögliche Erklärungen für das Zustandekommen der Vorstellungen zu finden. Zudem sollen Konsequenzen für das Lernen mit multiplen Repräsentationen aufgezeigt werden. Der Autorin liegen keine empirischen Daten vor, ob die hier skizzierten „möglichen“ Rahmentheorien oder p-Prims in der formulierten Weise bei Lernenden beobachtet werden können. Die Frage, welcher Ansatz sich besser eignet, Schülervorstellungen in der Strahlenoptik zu erklären und ihre Überwindung zu fördern, ist ebenfalls kein Gegenstand dieser wissenschaftlichen Arbeit.

Fasst man bspw. das erste Konzept in Anlehnung an Vosniadou und Brewer (1992, 1994) bzw. Vosniadou (1992) als implizite epistemologische Rahmentheorien auf, könnten Probleme mit dem Konzept der Streuung und dem physikalischen Sehvorgang wie folgt erklärt werden:

1. Ist Licht eine Art Substanz, dann strahlen beleuchtete Gegenstände kein Licht ab, weil das Licht auf diesen Gegenständen liegen bleibt und diese erhellt. Das Auge als aktiver Part würde diese Helligkeit erfassen.
2. Eine andere epistemologische Rahmentheorie (möglicherweise ein etwas weiter fortgeschrittener Ansatz) könnte in der Interpretation des Sehvorgangs

als das Ergebnis sich bewegender Lichtpartikel bestehen, wobei Licht oft als Flüssigkeitsstrom beschrieben wird, der zum Auge gelangt (hier käme dem Auge kein aktiver Part zu). Konsistent wäre diese Theorie jedoch nur dann, wenn der Strom von beleuchteten Gegenständen reflektiert würde.

3. Auch das Konzept, in welchem die Sammellinse eine aktive Funktion übernimmt, indem sie das Licht zu einem Strahl konzentriert, könnte als eine epistemologische Rahmenannahme aufgefasst werden. In diesem Fall würde das Konzept der „Punkt-zu-Punkt-Abbildung“ keinen Sinn ergeben und einem Gegenstandspunkt würde dann (hinter der Linse) nur ein Strahl zugeordnet und nicht ein divergierendes Strahlenbündel.
4. Eine (möglicherweise etwas fortgeschrittenere) epistemologische Rahmentheorie zum Abbildungsvorgang könnte in der von Wiesner (1994) angesprochenen holistischen Konzeption des Abbildungsvorgangs bestehen, nach der das Bild als Ganzes durch die Linse zum Schirm geht und dabei in der Linse umgedreht wird (vgl. ebd.). Unter dieser Annahme ergibt sich bei einer partiellen Abdeckung, dass auch ein Teil des Bildes abgeschnitten werde. Ist die Blende ringförmig, wird das Bild entsprechend ringförmig am äußeren Rand abgeschnitten. Wird die Linse zur Hälfte abgedeckt, wird die obere bzw. die Hälfte des Bildes abgeschnitten.

Die Vermittlung wissenschaftlich adäquater Vorstellungen sollte darauf zielen, die Lernenden darin zu unterstützen ein internes mentales Modell der Lichtausbreitung und des Abbildungsvorgangs zu generieren. Auf diese Weise können die Lernenden die Zusammenhänge in kohärenter Weise intern dynamisch simulieren, z.B. indem sie sich vorstellen, was passiert, wenn man den Gegenstand von der Linse weg oder auf die Linse zuschiebt. Ist den Lernenden bewusst, dass von jedem Gegenstandspunkt eines leuchtenden Gegenstands Licht ausgeht, sollten sie erkennen können, dass bei einer Abdeckung der Linse das Bild lediglich schwächer bzw. blasser, aber dennoch vollständig abgebildet wird. Das mentale Modell der Lernenden würde als ein wesentliches Element die geometrische Strahlenkonstruktion beinhalten. Voraussetzung für ein wissenschaftlich angemessenes mentales Modell des Abbildungsvorgangs ist die physikalische Sehvorstellung. Die Schüler sollten entsprechend kognitiv dazu aktiviert werden, die Lichtstrahl ausbreitung nachzuvollziehen und mögliche epistemologischen Rahmentheorien, wie das Konzept, Licht sei eine Art Substanz, die auf Gegenständen liege, zu revidieren. Auf Basis eines adäquaten mentalen Modells der Lichtausbreitung, sollten die Schüler in der Lage sein, die geometrischen Strahlenkonstruktionen in der Optik flexibel anzuwenden.

Fraglich ist jedoch, ob Lernende, die von Schülervorstellungen ausgehen, überhaupt über eine konsistente Erklärung des physikalischen Sehvorgangs oder der Bildentstehung verfügen. In Anlehnung an diSessa (1983, 1988, 1993) könnten die Lernschwierigkeiten auf bruchstückhafte intuitive Vorstellungen (sogenannte p-Prims) zurückgeführt werden. Zur Erinnerung, den Lernenden fehlt es gemäß dieser Theorie an den metakognitiven Fähigkeiten, die Inkonsistenz ihres Wissens zu erkennen.

Folgende mögliche p-Prims könnten auf Basis der erläuterten Schülervorstellungen formuliert werden.

1. Das Auge sieht, wenn es hell ist und kein Hindernis im Weg steht.
 - Dieses mögliche p-Prim beinhaltet die Vorstellung beleuchtete Gegenstände oder auch ein Spiegelbild könne man sehen, ohne dass dazu Licht ins Auge fallen müsse; bzw. beleuchtete Gegenstände strahlen kein Licht ab.
 - Auch die weitverbreitete Vorstellung, dass bei einer partiellen Abdeckung der Linse nur ein Teil des Bildes entsteht, lässt sich hier zuordnen. Da die Abdeckung ein „Hindernis“ darstellt, folgern einige Lernende, dass auch das Bild entsprechend abgeschnitten wird.
 - Auch die eher selten beschriebenen Vorstellungen einer aktiven Rolle des Auges könnten diesem möglichen p-Prim zugeordnet werden. Hierzu zählen die Vorstellungen von Sehstrahlen, die vom Auge ausgehend zum Gegenstand verlaufen oder die Vorstellung einiger Lernender, eine Katze könnte selbst bei völliger Dunkelheit noch Gegenstände sehen.
2. Licht erhellt Gegenstände.
 - Licht wird in folgender Schülervorstellung mit Helligkeit gleichgesetzt: Auftreffendes Licht mache die Gegenstände hell und bleibe auf ihnen liegen (wie eine Substanz). Von dieser Vorstellung gibt es zwei Varianten, die sich widersprechen: die Gleichsetzung von Licht mit seinen Wirkungen oder mit seiner Quelle.
3. Optische Geräte wie Spiegel oder Linsen erzeugen Bilder (als Ganzes), die den Gegenständen inhärent sind (z.B. auf der Linsen- oder Spiegeloberfläche liegen).
 - Dieses mögliche p-Prim könnte auf Alltagsbeobachtungen von Spiegelungen basieren. Die meisten dieser Spiegelungen sind nicht perfekt (wie z.B. bei Spiegelungen durch getönte Fensterscheiben an Gebäuden oder Autos oder Spiegelungen an Wasseroberflächen). Der Beobachter sieht ein Spiegelbild zusammen mit der sich spiegelnden Oberfläche und erhält den Eindruck, das Bild läge auf der Oberfläche. Nur bei sehr hochwertigen und großen

Spiegeln ist die Illusion eines Raumes hinter dem Spiegel glaubhaft. Liegt das Bild „auf dem Spiegel“, dann ist es auch nur folgerichtig, dass ein Befragter den Spiegel weiter weg halten will, wenn er oder sie einen größeren Bildausschnitt sehen möchte.

4. Je weiter entfernt ein Objekt ist, desto kleiner sieht man es (und einen desto größeren Ausschnitt kann man sehen).

- Auch dieses mögliche p-Prim könnte direkt der Alltagserfahrung entspringen. So entfernt sich ein Fotograf, der eine Ganzkörper-Ansicht einer Person erstellen möchte, von der Person und sieht im Sucher seine Annahme bestätigt. Probleme ergeben sich nun, wenn diese Erfahrung auf Spiegelbilder übertragen wird. Dann liegt die Annahme nahe, den Spiegel weiter weg zu halten, um einen größeren Bildausschnitt sehen zu können.

Vorteile in der Einordnung der Schülervorstellungen in den Rahmentheorieansatz (vgl. Vosniadou & Brewer, 1992) liegen in der Erklärung der Fortschritte der Lernenden. Der Bezug der Vorstellungen zu impliziten ontologischen Hintergrundannahmen zeigt auf, dass die Lernschwierigkeiten der Schüler aus impliziten Annahmen resultieren, die Lernenden in der Regel nicht explizit zugänglich sind.

Mit diSessas Ansatz (1983, 1988, 1993) der Reorganisation von Wissensfragmenten lässt sich jedoch ebenfalls sehr gut erklären, dass die Schwierigkeiten der Lernenden aus Vorstellungen resultieren, die sie nicht explizit formulieren. Die Ursachen hierfür gründen sich nicht darauf, dass den Lernenden ihre epistemologischen Rahmentheorie nicht bewusst zugänglich sind (und sie diese daher auch nicht gezielt überprüfen können), sondern vielmehr darauf, dass die Vorstellungen fragmentarisch und inkonsistent sind und nicht explizit formuliert werden. Die Stärke dieses Ansatzes besteht darin, den diffusen Vorstellungen der Schüler keine vermeintlich höhere Konsistenz zu unterstellen, als sie tatsächlich aufweisen.

Geeignete Strategien für das Lernen mit multiplen Repräsentationen können darin bestehen, die Lernenden zu aktivieren, die eigene Vorstellungen, die extern repräsentiert wurden, mit Lösungen im Unterricht abzugleichen, wobei die Lernenden angehalten werden, die Lösungen nicht einfach zu übernehmen sondern auch kritisch zu prüfen. Indem die Lernenden eigene Repräsentationen, z.B. eine selbst erstellte Strahlenkonstruktion mit anderen Repräsentationen und den beobachteten Phänomenen abgleichen, können sie auf Inkonsistenzen aufmerksam gemacht werden, so z.B. bei der Darstellung der Entstehung von reellen Bildern bei Abdeckaufgaben. Auch diese Strategie zielt darauf, die Lernenden in der Konstruktion eines adäquaten mentalen Modells der Lichtausbreitung, des

Sehvorgangs und der Entstehung reeller Bilder zu unterstützen. Der Schwerpunkt unter der Perspektive des Lernens mit multiplen Repräsentationen bestünde darin, die unterschiedlichen Repräsentationen in konsistenter und kohärenter Weise zu verwenden. Ziel wäre es, ein Wissensnetz zu schaffen, in dem die unterschiedlichen Repräsentationen (wie die Messwertetabelle, die Strahlenkonstruktion und die Abbildungsgleichung) miteinander verknüpft werden.

2.2 Pilotstudie

2.2.1 Zielsetzung der Pilotstudie

Zahlreiche Forschungsergebnisse belegen, dass (naive) Schülervorstellungen sehr hartnäckig sind und sich gegenüber Veränderungsversuchen als widerständig erweisen (vgl. Goldberg & McDermott, 1987; Wiesner, 1992a; Vosniadou & Brewer, 1992; Tyson et al., 1997; Özdemir & Clark, 2007). Die naiven Vorstellungen der Lernenden beeinflussen das formale Erlernen wissenschaftlicher Konzepte und schlagen sich in externen Repräsentationen nieder (vgl. Cox, 1999), welche die Lernenden bei der Aufgabenbearbeitung erstellen. Der im ersten Teil der Arbeit dargestellte Forschungsstand zum Konzeptwechsel zeigt, dass die Veränderung des konzeptuellen Grundverständnisses der Schüler jeweils domänenspezifisch zu erreichen ist. Entsprechend gilt es, die Instruktionen auf die Schülervorstellungen zu beziehen, die in Relation zur Bildentstehung in der Strahlenoptik stehen. Die entwickelten Instruktionen zielen darauf ab, den Schülern zu zeigen, warum gegebenenfalls alternative oder qualitativ neue Repräsentationen zweckmäßiger zur Erklärung und Problemlösung sind (vgl. Tyson et al., 1997; Özdemir & Clark, 2007). Auf diese Weise sollten die Schülervorstellungen im Licht der neuen theoretischen Annahmen reinterpretiert werden. Die Pilotstudie verfolgte vorrangig zwei Hauptziele: erstens die Validierung von Erhebungsinstrumenten und zweitens die Validierung des Unterrichtsmaterials.

Die Treatmentbedingung beinhaltete Aufgabenstellungen, welche die Überwindung von gängigen Schülervorstellungen erfordern, die in einschlägiger Fachliteratur berichtet werden (Goldberg & McDermott, 1987; Wiesner, 1992; Reiner et al., 2000). In der Kontrollbedingung hingegen bearbeiteten die Schüler Aufgabenstellungen, die keine aktive Auseinandersetzung mit Schülervorstellungen erfordern. In beiden Bedingungen wurden die Schüler kognitiv aktiviert, sich mit verschiedenen Repräsentationsformen auseinander zu setzen.

2.2.2 Fragestellung und Hypothesen

Die folgenden Forschungsfragen wurden im Rahmen der Pilotstudie untersucht:

1. Führen die entwickelten kognitiv aktivierenden repräsentationalen Aufgaben zu einem Lernzuwachs in Bezug auf den Lerninhalt: die Bildentstehung in der Strahlenoptik?
2. Fördern kognitiv aktivierende Aufgaben, die darauf zielen, weitverbreitete Schülervorstellungen durch die Auseinandersetzung mit fachbezogenen Repräsentationen zu überwinden, Wissen, Problemlösen und den Umgang mit fachspezifischen Repräsentationen von Schülern in der Strahlenoptik?
3. Fördern diese Aufgaben das konzeptuelle Verständnis der Schüler?

Hypothese 1: Zur Beantwortung der ersten Forschungsfrage wurde analysiert, ob sowohl die Treatment- als auch die Kontrollgruppe Wissen und Problemlösen bei repräsentationsbezogenen Aufgaben und ihr konzeptuelles Verständnis in der Strahlenoptik nach der Intervention verbesserten.

Hypothesen 2 und 3: Zur Untersuchung der beiden Fragestellungen wurde geprüft, ob die Treatmentgruppe die Kontrollgruppe nach der Intervention in zwei Aspekten einen höheren Lernzuwachs erzielt.

- Erstens im Wissen und Problemlösen beim Umgang mit fachspezifischen Repräsentationen im Bereich der Strahlenoptik (erfasst in einem inhaltspezifischen Leistungstest),
- Zweitens im konzeptuellen Verständnis im Bereich der Strahlenoptik (erfasst in einem Konzepttest zum Thema Strahlenoptik).

2.2.3 Stichprobe und Design

Die Stichprobe bestand aus 57 Schülern aus zwei Parallelklassen, welche die achte Klasse eines Gymnasiums besuchten. Eine Klasse wurde der Treatment- und eine Klasse der Kontrollbedingung zugeordnet. Alle Schüler waren im Alter von 13 und 14 Jahren. Der Stichprobe gehörten 24 Jungen und 33 Mädchen an. Die Teilnahme an der Studie war freiwillig und mit Schülern, Eltern und Lehrern im Vorhinein abgesprochen. Fünf Schüler fehlten entweder beim Prä- oder beim Posttest und wurden daher in der Analyse nicht berücksichtigt.

Bei der Pilotstudie handelte es sich um ein einfaktorielles quasi-experimentelles Prä-Posttest-Design. Die Treatmentbedingung beinhaltete Aufgabenstellungen, welche die Überwindung von gängigen Schülervorstellungen erfordern, die in einschlägiger Fachliteratur berichtet werden (Goldberg & McDermott, 1987; Wiesner, 1992a, 1992b; Reiner et al., 2000). In der Kontrollbedingung hingegen bearbeiteten die Schüler Aufgabenstellungen, die keine aktive Auseinandersetzung mit Schülervorstellungen erfordern. In beiden Bedingungen wurden die Schüler kognitiv aktiviert, sich mit verschiedenen Repräsentationsformen auseinander zu setzen. Es handelt sich um eine quasi-experimentelle Studie, da die Schüler im Klassenverbund als Gruppe den Bedingungen zugeteilt wurden. Grund für die Wahl eines quasi-experimentellen Designs bestand in der Einbindung in den regulären Schulalltag. Um den Stundenplan einzuhalten, konnten die Schüler nicht individuell den Bedingungen zugeordnet werden. In der Folge müssen Leistungsunterschiede, die von vornherein bestanden, in Kauf genommen werden. Verfahren der Parallelisierung auf Basis von Vorleistungen und / oder Intelligenz waren somit ausgeschlossen. Beide Klassen, Treatment- und Kontrollgruppe, wurden vom gleichen Lehrer im gleichen Zeitumfang von 135 min, also drei Schulstunden, unterrichtet. Die Prä- und Posttests fanden jeweils in der Physikstunde vor und nach der Unterrichtseinheit statt und erforderten jeweils insgesamt 45 min. (25 min. + 15 min.).

2.2.4 Durchführung und Unterrichtsmaterial

Um die Hauptstudie im Schuljahr 2010/2011 durchführen zu können, wurde die Pilotstudie Ende April und Anfang Mai 2010 umgesetzt. Zu diesem Zeitpunkt hatten die Schüler seit August 2009 Unterricht im Fach Physik. Das Thema Strahlenoptik war zwei Monate zuvor abgeschlossen worden, so dass die Schüler bereits Vorkenntnisse zum Thema Bildentstehung mit der Sammellinse besaßen. Aus diesem Grund fiel die Wahl des Lerngegenstands auf das Thema der Bildentstehung beim Hohlspiegel: Mit der Bildentstehung am Hohlspiegel wurde ein Lerninhalt gewählt, der in den repräsentationalen Anforderungen äquivalent zur Bildentstehung bei der Sammellinse ist und zugleich als Wiederauffrischung und Vertiefung früherer Lerninhalte in das Jahresscurriculum eingebunden werden konnte (Unterrichtsthema unmittelbar vor Beginn der Studie war Mechanik). Die Unterrichtsreihe umfasste drei Schulstunden zur Bildentstehung beim Hohlspiegel:

1. In der ersten Stunde erfolgte zunächst eine kurze Einführung mittels eines Demonstrationsexperimentes mit einer optischen Scheibe, bei der Brechungsvorgänge am Hohlspiegel demonstriert wurden. Nach der Einführung führten die Schüler ein Schülerexperiment zur Entstehung reeller Bilder am Hohlspiegel durch.
2. In der zweiten Stunde wurden die Strahlenkonstruktion und die Abbildungsgleichung behandelt.
3. In der dritten Stunde bearbeiteten die Schüler Übungsaufgaben zum Umgang mit verschiedenen Repräsentationsformen, wie der Strahlenkonstruktion und der Abbildungsgleichung, die es erforderten, diese Repräsentationsformen aufeinander und auf das Schülerexperiment zu beziehen

Die Treatmentgruppe bearbeitete kognitiv aktivierende Aufgaben, die verlangten, sich vertieft mit Repräsentationen auseinanderzusetzen und die in der Strahlenoptik bekannten weitverbreiteten Schülervorstellungen ansprechen. Im Gegensatz zur Treatmentgruppe bearbeitete die Kontrollgruppe ebenfalls kognitiv aktivierende Aufgaben, die es erforderten, sich vertieft mit Repräsentationen auseinanderzusetzen. Weit verbreitete Schülervorstellungen wurden jedoch nicht thematisiert. Ausführliche Informationen und weitere Beispiele zu den verwendeten Unterrichtsmaterialien finden sich in Hettmannsperger, Schnotz, Müller und Scheid (in Druck). Zur Verdeutlichung der Unterschiede zwischen Treatment- und Kontrollgruppe wird hier folgendes Aufgabenbeispiel genannt (vgl. Abbildung 13), indem die weit verbreitete Lernschwierigkeit thematisiert wird, dass sich die Lichtstrahlen, die ein reelles Bild formen, auch im freien Raum treffen können (vgl. Goldberg McDermott, 1987, S. 114).

Die Treatmentgruppe wurde gefragt, an welcher Stelle der Beobachter A, B, oder C ein scharfes Bild einer Kerze sehen kann, wenn man

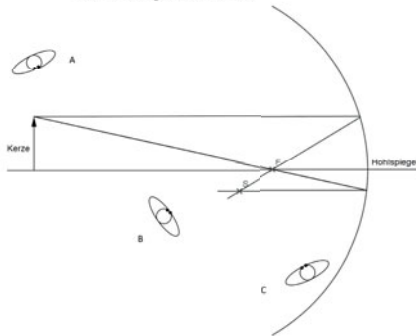
- a) einen intransparenten Schirm an der Position S aufstellt,
- b) den intransparenten Schirm gegen einen transparenten Schirm austauscht oder
- c) den Schirm entfernt.

Parallel zu dieser Aufgabe operierte die Kontrollgruppe mit derselben Repräsentation und wurde gefragt, an welcher Position ein Beobachter ein scharfes Bild der Kerze erkennen kann. Zudem wurde sie aufgefordert, den dargestellten Bildpunkt H der Kerzenflamme in der gegebenen Abbildung zu konstruieren. Die Kontrollgruppe wurde also kognitiv aktiviert, mit der gleichen Repräsentation zu operieren, die erwähnte Schülervorstellung wurde jedoch nicht thematisiert. Das

vollständige Unterrichtsmaterial und die zugehörige Unterrichtsplanung findet sich in Anhang B1 und B2 auf der Produktseite zu diesem Buch unter www.springer.com.

A1

- i. Welcher Beobachter (A, B, C) kann das Bild der Kerze sehen, wenn Max ein weißes Stück Pappe an der Position S aufstellt?
- ii. Welcher Beobachter (A, B, C) kann das Bild der Kerze sehen, wenn Maren die Pappe gegen ein transparentes weißes Papier vertauscht?
- iii. Welcher Beobachter (A, B, C) kann das Bild der Kerze bei einer Versuchsanordnung ohne Schirm sehen?

**A1**

- a) An welcher Position kann das Bild der Kerze (hier dargestellt als Pfeil) auf einem Schirm aufgefangen werden? (Zeichne das Bild der Kerze ein)
- b) Konstruiere den Bildpunkt zu Punkt H des Gegenstandes (Kerze).

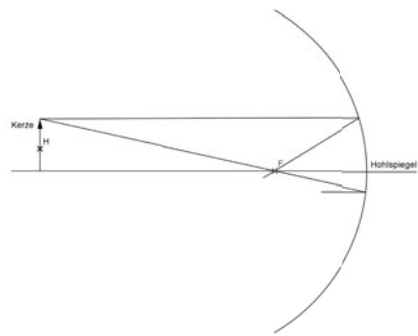


Abbildung 13: Aufgabenblatt 4: Übungen zur Strahlenkonstruktion und zur Abbildungsgleichung am Hohlspiegel, Aufgabe 1 [Material 7-TG] bzw. [Material 7-KG], links Aufgaben der Treatmentgruppe (TG), rechts Aufgabe der Kontrollgruppe (KG)

2.2.5 Variablen und Erhebungsinstrumente

2.2.5.1 Leistungstest

Die in einem Leistungstest erhobenen abhängigen Variablen Wissen und Problemlösen beim Umgang mit fachspezifischen Repräsentationen orientierten sich am Lehrplan für Gymnasien in Rheinland-Pfalz der Jahrgangsstufe 8. Der Test wurde von Jochen Scheid in Zusammenarbeit mit der Autorin eigens für die Studie entwickelt, da zum Zeitpunkt der Studie kein Instrument vorlag, welches das Themengebiet der Bildentstehung in Klassenstufe 8 hinsichtlich des Umgangs mit (multiplen) Repräsentationen erfasste. Der Test bestand aus neun Aufgaben und zielte darauf, Wissen und Problemlösen im Bereich der Strahlenoptik durch Aufgaben zu erfassen, die den Umgang mit fachspezifischen Repräsentationen erfordern. Die Testdauer betrug sowohl im Prä- als auch im Posttest 25 Minuten. Maximal konnten jeweils 17 Punkte je Test erreicht werden. Prä- und Posttest

enthielten die gleichen Aufgabenstellungen, wobei die Aufgaben des Prätests die Bildentstehung an der Sammellinse zum Thema hatten und die Aufgaben des Posttests die Bildentstehung am Hohlspiegel. Der vollständige Test findet sich in Anhang B3 auf der Produktseite zu diesem Buch unter www.springer.com.

Der Test umfasste folgende Aufgabentypen:

- Typ 1: drei Aufgaben, welche das Operieren mit Repräsentationen in unterschiedlichen Formaten erforderten (depiktionale und deskriptive Repräsentationen). Die repräsentationsbezogenen Aufgaben konnten nur korrekt gelöst werden, wenn das Denken der Schüler nicht durch spezifische Schülervorstellungen geleitet wurde (Aufgaben 1, 2a und 2b¹²). Aufgaben dieses Typs hatten zum Ziel, Wissen und Problemlösen beim Umgang mit Repräsentationen in unterschiedlichen Formaten im Hinblick auf die Überwindung spezifischer Schülervorstellungen zu erfassen.
- Typ 2: zwei Aufgaben, in denen die Schüler Repräsentationen in depiktional schematischer Form in eine deskriptiv verbale Repräsentationsform übersetzen sollten (Aufgaben 3 und 4) und Fehler in gegebenen depiktionalen Repräsentationen zu erkennen. Dieser Aufgabentyp zielte darauf, das fachliche Wissen im Umgang mit Repräsentationen und die repräsentationale Kohärenz zu erfassen.
- Typ 3: zwei Leistungstestaufgaben zur Abbildungsgleichung (Aufgaben 5a und 5b). Aufgaben des dritten Typs erfassten die Lernleistung der Schüler und prüften, ob die Schüler mit der mathematischen Repräsentation des Abbildungsvorgangs umgehen konnten. Mit Aufgabenstellungen, die prüften, ob die Schüler die Zweckmäßigkeit der mathematischen Repräsentation der Abbildungsgleichung erkennen, sollte auch eine metakognitive Komponente von Repräsentationskompetenz erhoben werden.

Aufgaben des ersten Typs werden exemplarisch anhand der Aufgaben 2a und 2b erläutert (vgl. Abbildung 14 und Abbildung 15). Aufgabe 2 gliedert sich in zwei Teilaufgaben, die aufeinander aufbauen. Das Anwenden der Strahlenkonstruktion (Operieren mit der gegebenen Repräsentation) in Aufgabe 2a unterscheidet sich von einer Standardaufgabenstellung dadurch, dass der Schirm nicht an die richtige Stelle gerückt ist.

12 Die Aufgaben 2a und 2b wurden nur von der Autorin eingesetzt, da sie in spezifischer Weise darauf zielen, den Umgang mit Repräsentationen im Hinblick auf die Überwindung weitverbreiteter Schülervorstellungen zu erfassen.

2. Aufgabe (4P)

Die Skizze zeigt den Gegenstand G und sein Bild B, das durch den Hohlspiegel entsteht. Der Bildpunkt Q zum Gegenstandspunkt P wurde richtig konstruiert. Die Strahlenkonstruktion ist nicht dargestellt. Der Schirm ist jedoch nicht an der richtigen Stelle aufgestellt, sondern ein Stück zum Spiegel hingertickt.

- a) Zeige durch eine Zeichnung, dass auf dem Schirm anstatt des Bildpunktes Q ein unscharfer Bildfleck entsteht. Verwende hierfür die gegebene Abbildung (unten).

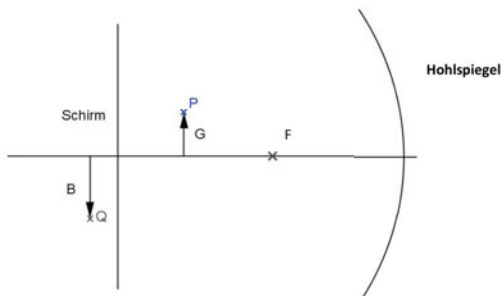


Abbildung 14: Aufgabenbeispiel aus dem Leistungstest: Anwenden der Strahlenkonstruktion

Diese Aufgabe prüft, ob die Schüler ein adäquates mentales Modell der geradlinigen Lichtausbreitung besitzen. Um die korrekte Größe des Lichtflecks zu bestimmen, müssen die Schüler neben den ausgezeichneten Strahlen auch den oberen Randstrahl einzeichnen, der gerade noch auf den Hohlspiegel trifft. Voraussetzung hierfür ist das Verständnis, das sich Licht in alle Richtungen ausbreitet. Dies berührt Schülervorstellungen zu den Konzepten der Lichtausbreitung und der Punkt-zu-Punkt-Abbildung. Aufbauend auf die Situation in Aufgabe 2a wurde im zweiten Aufgabenteil eine Blende vor den Hohlspiegel gesetzt.

Auch Aufgabe 2b zielt darauf, das Konzept der Punkt-zu-Punkt-Abbildung zu prüfen und spricht die weitverbreitete Schülervorstellung an, eine Lochblende schneide das Bild ringförmig ab. Die Lösung bestand darin, auch hier diejenigen Strahlen einzuzichnen, welche am Rand der Lochblende gerade noch auf den Hohlspiegel treffen. Über eine Transferleistung beim Operieren mit der bildschematischen Repräsentation der Strahlenkonstruktion hinaus verlangt diese Aufgabe zudem, die bildliche Repräsentation in eine verbale Repräsentation zu übersetzen. Zudem prüft die Aufgabe die Kohärenz der verbalen und bildlichen Repräsentation. Eine Übersicht über die repräsentationalen Anforderungen der Aufgaben des Typs 2 und 3 finden sich in der Dissertationsschrift von Scheid (2013).

- b) Wie ändert sich das Bild, wenn man jetzt noch eine Lochblende vor den Hohlspiegel stellt?

Begründe Deine Antwort durch eine Zeichnung. Verwende hierfür die gegebene Abbildung (unten).

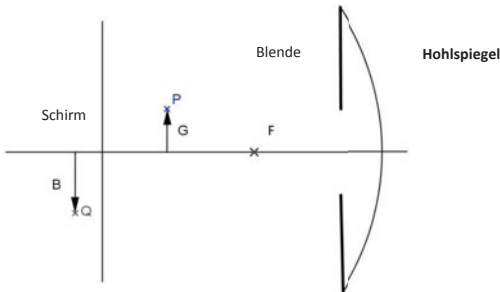


Abbildung 15: Aufgabenbeispiel aus dem Leistungsposttest: Lochblende vor Hohlspiegel

2.2.5.2 Konzepttest

Das konzeptuelle Verständnis der Schüler wurde in einem Konzepttest erfasst. Der Test wurde von der Autorin in Zusammenarbeit mit Jochen Scheid eigens für die Studie entwickelt, da zum Zeitpunkt der Studie kein Konzepttest zur Strahlenoptik vorhanden war. Ziel des Tests war die Erfassung physikalischer Grundkonzepte, welche allesamt für das Verständnis der Bildentstehung an der Sammellinse und am Hohlspiegel relevant sind.

Der Test beinhaltete Fragen zu folgenden Konzepten: Lichtausbreitung, Streuung, physikalische Sehvorstellung, zu ontologischen Konzepten, was unter „Licht“ verstanden werden kann, zur gerichtete Reflexion (Planspiegel) sowie zur Entstehung reeller Bilder bei der Sammellinse und am Hohlspiegel. Der als Multiple-Choice-Test konzipierte Konzepttest bestand aus 30 Items. Pro Aufgabe konnten maximal 2 Punkte erzielt werden, die maximal erreichbare Gesamtpunktzahl betrug somit 60. Die Fragen sollten erfassen, wie die Schüler mit bekannten Lernschwierigkeiten umgehen. Die Distraktoren bezogen sich hierbei auf bekannte Schülervorstellungen. Die Anzahl an Distraktoren orientierte sich je Aufgabe an der Anzahl alternativer Antwortmöglichkeiten, die auf Basis bekannter Schülervorstellungen gefunden werden konnten. Sie variiert zwischen 3 bis 6 Antwortmöglichkeiten.

Das folgende Aufgabenbeispiel (vgl. Abbildung 16) erfasst Konzepte zur Bildentstehung bei der Sammellinse. Korrekt war in diesem Fall die dritte Antwortmöglichkeit: „Lichtstrahlen eines Gegenstandspunktes werden durch die Sammellinse abgelenkt und treffen sich im Bildpunkt“, die Distraktoren bezogen sich auf die Schülervorstellungen:

- Das reelle Bild bei der Sammellinse entstehe durch Spiegelung oder Reflexion.
 - Der holistischen Konzeption des Abbildungsvorgangs: Das Bild gehe als Ganzes durch die Linse zum Schirm und werde dabei in der Linse umgedreht.
9. Wie entsteht durch Verwendung einer Sammellinse ein Bild, das auf einem Schirm aufgefangen werden kann?
- Das reelle Bild durch eine Sammellinse entsteht durch Spiegelung der Lichtstrahlen an der Linse nach dem Reflexionsgesetz.
 - Eine Sammellinse hat den Effekt, die Lichtstrahlen aufzuhellen.
 - Lichtstrahlen eines Gegenstandspunktes werden durch die Sammellinse abgelenkt und treffen sich im Bildpunkt.
 - Das Bild geht als Ganzes durch die Linse zum Schirm, dabei wird es in der Linse unter Einhaltung der Linsengesetze umgedreht (siehe Skizze).

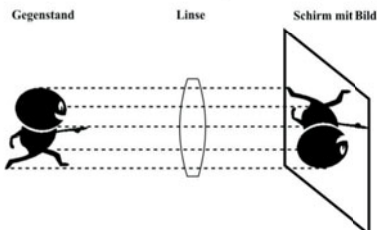


Abbildung 16: Aufgabenbeispiel (Item 9) aus dem Konzepttest

Pro Aufgabe war in der Regel eine Antwortmöglichkeit korrekt, mit Ausnahme der Aufgaben 4, 5, 7a, 8 und 14, in denen jeweils zwei Kreuze gesetzt werden mussten.

Die Antworten der Teilnehmer wurden mit zwei Punkten bewertet, wenn ausschließlich die korrekte Alternative gewählt wurde, ein Punkt wurde vergeben, wenn zusätzlich zu der richtigen Antwort auch ein Distraktor angekreuzt wurde (zwei Punkte für die Lösung weniger ein Punkt Abzug). In allen anderen Fällen wurde die Entscheidung mit null Punkten bewertet.

Im obigen Beispiel wurden also zwei Punkte vergeben, wenn die dritte Antwortmöglichkeit angekreuzt wurde. Die Schülerantwort wurde mit einem Punkt bewertet, wenn zusätzlich noch ein weiteres Kreuz gesetzt wurde. Eine Bewertung mit null Punkten erfolgte bei allen anderen Antwortverteilungen.

Bei Aufgaben, welche die Wahl von zwei Antwortalternativen erforderten, wurden zwei Punkte vergeben, wenn beide Lösungen gewählt wurden. Ein Punkt wurde

vergeben, wenn zusätzlich zu den beiden korrekten Antworten ein Distraktor für richtig befunden wurden (zwei Punkte für die Lösung weniger ein Punkt Abzug). Eine Bewertung mit null Punkten erfolgte bei allen anderen Antwortverteilungen. Folgende Grundkonzepte wurden in dem Test erfasst. Der vollständige Test befindet sich in Anhang B4 (auf der Produktseite zu diesem Buch unter www.springer.com).

2.2.6 Auswertung und Ergebnisse

2.2.6.1 Itemstatistiken zum Leistungstest

Um eine tragfähige Testfassung zu erstellen, wurde zunächst eine Itemanalyse des Leistungstests durchgeführt. Da sich die Werte von Treatment- und Kontrollgruppe weder im Prätest noch im Posttest wesentlich unterscheiden, werden an dieser Stelle die Werte für die Gesamtstichprobe (Treatment- und Kontrollgruppe) berichtet.

Die Analyse der Itemschwierigkeiten ergab, dass alle Items mit Ausnahme der Item 3 und 4 innerhalb des Toleranzbereichs von $0.20 \leq P_i \leq 0.80$ liegen (vgl. Anhang B5, Tabelle 1, auf der Produktseite zu diesem Buch unter www.springer.com). Die Aufgaben 3 und 4 wurden offenbar von vielen Schülern gut beantwortet, daher eignen sie sich offenbar wenig, die Leistung beim Umgang mit Repräsentationen differenziert zu messen. Item 3 und 4 bedürfen daher einer Überarbeitung.

Die Analyse der Trennschärfe zeigte, dass die korrigierten Trennschärfen der Items zwischen $.01 \leq r_{it} \leq .45$ für den Prätest und zwischen $-.11 \leq r_{ip} \leq .35$ für den Posttest rangieren (vgl. Anhang B5, Tabelle 2, auf der Produktseite zu diesem Buch unter www.springer.com). Die Analyse belegt, dass Item 5a (Multiple Choice Item: Wozu kann man die Abbildungsgleichung gebrauchen?) durch eine negative Trennschärfe im Posttest auffiel. Dieses Item wird tendenziell sogar häufiger von Personen mit geringer Kompetenz in repräsentationsbezogenen Aufgaben korrekt beantwortet als von Personen mit hoher Kompetenz. Möglicherweise kommt dieses Ergebnis teilweise durch Raten zustande, da es sich um eine Multiple-Choice Frage handelt. Alternativ könnten auch die Distraktoren ungeeignet sein. Da dieses Item im Rahmen der Erfassung der Leistungen nicht als invertiert aufgefasst werden kann, wurde es von der Analyse ausgeschlossen.

Cronbachs Alpha erhöht sich unter Ausschluss des Items von $\alpha = .55$ auf $\alpha = .59$ im Posttest. Im Prätest sind die Werte nochmals niedriger ($\alpha = .49$ - Wert für alle Items, $\alpha = .50$ - unter Ausschluss von Item 5a). Für die Retest-Reliabili-

tät ergibt sich für die verbleibenden Items ein Wert von $r_{\text{Test-Retest}} = .50$. Ohne Berücksichtigung von Aufgabe 5a konnten nunmehr maximal 15.50 Punkte erzielt werden. Generell weisen die Items Trennschärfen unter $r_{it} = .40$ auf und auch die interne Konsistenz für den Posttest liegt mit knapp $\alpha = .59$ im unteren Bereich, was für eine notwendige und grundlegende Überarbeitung des Tests spricht.

Eine ausführlichere Diskussion der statistischen Kennwerte und der Struktur des Tests findet sich in der Dissertationsschrift von Scheid (2013)

2.2.6.2 Itemstatistiken zum Konzepttest

Um eine tragfähige Testfassung zu erstellen, wurde auch für den Konzepttest zunächst eine Itemanalyse durchgeführt.

Die Analyse der Itemschwierigkeit ergab, dass alle Items mit Ausnahme von Item 17a innerhalb des Toleranzbereichs von $0.20 \leq P_i \leq 0.80$ liegen (vgl. Anhang B5, Tabellen 3 und 4, auf der Produktseite zu diesem Buch unter www.springer.com).

Die Analyse der korrigierten Trennschärfen zeigt, dass die Mehrheit der Items in einem Trennschärfenbereich von $0.11 \leq r_{it} \leq 0.4$ im Posttest liegen. Bei den Items 3, 4, 5, 7a, 12a, 17a und 17b zeigen sich Werte nahe 0 oder sogar negative Werte (vgl. Anhang B5, Tabelle 5, auf der Produktseite zu diesem Buch unter www.springer.com).

Die Items 5, 7a und 17a und b, 12a wurden wegen der deutlich negativen Werte im Posttest ausgeschlossen. Eine Invertierung der Items lässt sich theoretisch nicht begründen: Schüler, welche die Bildentstehung bei der Sammellinse gut verstanden haben, sollten prinzipiell besser in der Lage sein diese Items zu beantworten als Schüler mit geringem Wissen.

Die Item 3 und 4, welche Trennschärfen nahe 0 aufweisen, wurden im Test belassen, müssen jedoch für die künftige Verwendung überarbeitet werden.

Für die korrekte Beantwortung der verbleibenden 26 Items konnten maximal 52 Punkte erzielt werden. Cronbachs Alpha erhöht sich unter Ausschluss der Items 5, 7b, 12a, und 17a von $\alpha = .72$ auf $\alpha = .78$ im Posttest. Im Prätest fällt Cronbachs Alpha entsprechend niedriger aus. Das geringere Wissen der Schüler im Vortest führt offenbar zur heterogeneren Beantwortung, was zu einem geringeren Wert der internen Konsistenz führt ($\alpha = .45$).¹³ Für die Retest-Reliabilität ergibt sich für die verbleibenden 26 Items ein Wert von $r_{\text{Test-Retest}} = .39$.

¹³ Unter Ausschluss der Items 5, 7b, 12a und 17a ergibt sich der identische Wert von $\alpha = .45$.

2.2.6.3 Exkurs: Berechnung einer polychorischen Korrelationsmatrix für ordinalskalierte Daten

Da die Aufgaben des Konzepttests auf Itemebene mit den Abstufung 0, 1 oder 2 Punkte ordinalskaliert sind, wurde der Berechnung der Trennschärfen und Cronbachs Alpha in der Pilotstudie und in der später dargestellten Hauptstudie eine polychorische Korrelationsmatrix für ordinalskalierte Daten zugrunde gelegt.

Das Modell der polychorischen bzw. tetrachorischen Korrelation basiert auf der Annahme, dass eine latente kontinuierliche Variable durch Schwellenparameter in zwei bzw. drei oder mehr Kategorien eingeteilt wird. Je nachdem, wo der Wert einer Versuchsperson liegt, wählt die jeweilige Person eine der Kategorien. Die Schwellenparameter werden hierbei auf Basis der Häufigkeiten der Kategorien bestimmt. Unter der Annahme, dass nicht nur jede der latenten Variablen einer Normalverteilung folgt, sondern auch beide latente Variablen bivariat normalverteilt sind, lässt sich nun die Produkt-Moment-Korrelation der beiden latenten Variablen schätzen (vgl. Eid, Gollwitzer & Schmitt, 2011, S. 516). Die Berechnung basiert im Fall der tetrachorischen Korrelation, die ein Sonderfall der polychorischen Korrelation darstellt, darauf, dass sich der tetrachorische Koeffizient r_{tet} als Kosinus eines Winkels beschreiben lässt (vgl. ebd., S. 529):

$$r_{tet} = \cos \left(\frac{180^\circ}{1 + \sqrt{\frac{n_{11}n_{22}}{n_{12}n_{21}}}} \right)$$

$n_{11}, n_{12}, n_{21}, n_{22} =$ Anzahlen in den jeweiligen Merkmalskombinationen, welche man durch eine Kreuztabelle darstellen kann

Der Berechnung liegt allgemein folgender Ansatz zugrunde: Gegeben seien zwei beobachtete ordinalskalierte Variablen A und B (mit in unserem Falle jeweils drei möglichen Werten). Es wird davon ausgegangen, dass A und B nach einem System von Grenzen x_1, x_2, y_1, y_2 wie folgt von unbeobachteten, metrischen Variablen X und Y abgeleitet werden können:

$$A = \begin{cases} 0 & \text{falls } X < x_1 \\ 1 & \text{falls } x_1 < X < x_2 \\ 2 & \text{falls } x_2 < X \end{cases}$$

$$B = \begin{cases} 0 & \text{falls } Y < y_1 \\ 1 & \text{falls } y_1 < Y < y_2 \\ 2 & \text{falls } y_2 < Y \end{cases}$$

Weiterhin wird angenommen, dass X und Y multivariat normalverteilt sind.

Ziel ist es, aus der Kontingenztafel von A und B die Korrelation der zugrundeliegenden Variablen X und Y zu schätzen. Hierzu wird die Maximum-Likelihood Methode angewendet. Freie Parameter sind die Korrelation (ρ_{XY}) und die Grenzen (x_1, x_2, y_1, y_2). Mittelwerte und Standardabweichungen können $(\mu_X, \sigma_X) = (0,1)$ und $(\mu_Y, \sigma_Y) = (0,1)$ gewählt werden.

Die Berechnung der polychorischen Korrelationen erfolgte mit dem R-Befehl „polychor“ des Paketes „polycor“. Nach Olsson (1979) bieten sich folgende Methoden zur Lösung des Problems, welche beide im R-Befehl `polychor` aufrufbar sind:

1. Es wird eine vollständige Maximum-Likelihood Maximierung durchgeführt, bei welcher ρ_{XY} und die Grenzen gleichzeitig geschätzt werden.
2. Es werden zuerst die Grenzen aus den marginalen relativen Häufigkeiten mittels der Inversen der Normalverteilung berechnet. Danach wird ρ_{XY} mittels Maximum Likelihood Schätzung bestimmt. Diese „Two-Step-Method“ ist schneller und numerisch einfacher als die erste Methode, wenngleich erstere formal korrekter ist.

Im Folgenden ist dies beispielhaft dargestellt. Zu sehen ist die gemeinsame Dichte zweier multivariat normalverteilter Zufallsvariablen, welche jeweils eine Standardnormalverteilung als Randverteilung haben und mit $\rho = 0.7$ korreliert sind. Die Farben repräsentieren die 3 x 3 Felder der Kontingenztafel der ordinalen Variablen. Das Integral über die einzelnen Farben entspricht der relativen Häufigkeit der ordinalen Kombination.

Da für die erste Methode Konvergenz nicht immer gegeben war, basieren die im Rahmen dieser Arbeit vorgestellten Ergebnisse (Maße der internen Konsistenz als Schätzer für die Realibilität, Korrelationsmatrizen) auf der zweiten Methode.

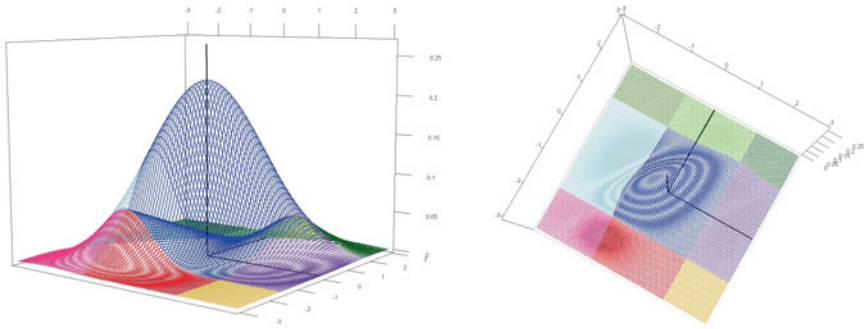


Abbildung 17 und Abbildung 18: Gemeinsame Dichte zweier multivariat normalverteilter Zufallsvariablen aus unterschiedlichen Perspektiven

2.2.6.4 Erste Hypothese: Veränderungen durch die Intervention

Zunächst wurde untersucht, ob die vorliegenden kognitiv aktivierenden repräsentationalen Aufgaben dazu geeignet sind, den Lerninhalt der Bildentstehung am Hohlspiegel inklusive der hiermit verbundenen physikalischen Grundkonzepte zu vermitteln.

Hierzu wurde analysiert, ob durch die eingesetzten Aufgaben überhaupt eine Änderung des Lernzuwachses prä – post erzielt werden kann. Wie Tabelle 1 zeigt, schneiden alle teilnehmenden Schüler sowohl im Leistungs- als auch im Konzeptposttest besser ab als in den jeweiligen Prätests.

Tabelle 1 Deskriptive Ergebnisse Mittelwerte und Standardabweichungen für Treatment- (TG) und Kontrollgruppe (KG)

AV ^a	Zeitpunkt	TG (n = 27)		KG (n = 25)		Alle (N = 52)	
		M	SD	M	SD	M	SD
LT ^b	prä	5.00	1.81	7.64	2.11	6.26	2.35
	post	9.50	3.63	11.40	1.72	10.41	3.01
KT ^c	prä	15.41	4.98	17.08	5.75	16.21	5.38
	post	27.74	10.34	22.44	5.98	25.19	8.86

^aAV: abhängige Variable

^bLT: Leistung bei repräsentationsbezogenen Aufgaben, maximal erreichbare Punkte: 15.50

^cKT: Konzeptuelles Verständnis; maximal erreichbare Punkte: 52

Zur Überprüfung, ob diese Unterschiede auch signifikant sind, wurde ein t-Test für abhängige Stichproben durchgeführt, wobei jeweils die Ergebnisse des Prätests mit den Ergebnissen des Posttests verglichen wurden.

$$\begin{aligned} H_0 : & \quad \mu_{d(\text{post-prä})} \leq 0 \\ H_1 : & \quad \mu_{d(\text{post-prä})} > 0 \end{aligned}$$

Voraussetzung hierfür ist die Normalverteilung für die jeweiligen Mittelwertdifferenzen (im Leistungstest und im Konzepttest). Diese Voraussetzung wurde per QQ-Plot und per Shapiro-Wilk-Test überprüft (vgl. Nachtigall & Wirtz, 2009, S. 171). Sie war für beide Tests, in der Treatmentgruppe, in der Kontrollgruppe wie auch in der Gesamtstichprobe gegeben.

Die Ergebnisse in Gemäß Cohen (1988, zitiert nach ebd.) handelt es sich in beiden Fällen von d'' um einen großen Effekt.

Tabelle 2 bestätigen, dass der Mittelwertevergleich vor und nach der Unterrichtseinheit mit Ausnahme des Unterschieds im Konzepttest für die Kontrollgruppe auf dem $\alpha = 0.001$ Niveau signifikant sind und zwar sowohl was die Ergebnisse im Leistungstest $t(51) = 10.87, p < .001, d'' = 1.51$ als auch, was das Abschneiden im Konzepttest angeht, $t(51) = 6.21, p < .001, d'' = 1.08$. Die Ergebnisse unterstützen somit die erste Hypothese. Bei d'' handelt es sich um die aus den Daten geschätzte Effektstärke für den Effekt d'' auf Ebene der Population. Die Berechnung von d'' errechnet sich nach Eid et al. (2011, S. 353) wie folgt:

$$d'' = \frac{\mu_D}{\sigma_D}$$

μ_D : Mittelwert der Differenz in der Population
 σ_D : Standardabweichung der Differenz in der Population

Gemäß Cohen (1988, zitiert nach ebd.) handelt es sich in beiden Fällen von d'' um einen großen Effekt.

Tabelle 2 Ergebnisse Vergleich des Mittelwertevergleichs für abhängige Stichproben prä und post

AV ^c	TG ^a (n = 27)			KG ^b (n = 25)			Alle (N = 52)		
	M (SD)	t(26)	p	M (SD)	t(24)	p	M (SD)	t(51)	p
LT ^d	4.50 (3.25)	7.19	<.001	3.76 (2.07)	9.07	<.001	4.14 (2.75)	10.87	<.001
KT ^e	12.33 (8.85)	7.24	<.001	5.36 (6.21)	4.32	<.001	8.98 (8.28)	6.21	<.001

^aTG: Treatmentgruppe

^bKG: Kontrollgruppe

^cAV: abhängige Variable

^dLT: Leistung bei repräsentationsbezogenen Aufgaben

^eKT: Konzeptuelles Verständnis

2.2.6.5 Zweite Hypothese: Wirkung des Treatments auf die Physikleistung

Zur Überprüfung von Hypothese 2 wurde analysiert, ob die Treatmentgruppe die Kontrollgruppe hinsichtlich Wissen und Problemlösen beim Umgang mit fachspezifischen Repräsentationen (erfasst durch den Leistungstest) übertrifft.

Die Aufgaben des Präleistungstests unterscheiden sich von den Aufgaben des Postleistungstests in einem Punkt: zwar bearbeiteten die Schüler parallele Fragestellungen, doch im Prätest bezogen sich die Aufgaben mit den gleichen Formulierungen der Fragestellungen jedoch auf die Bildentstehung bei der Sammellinse, während im Posttest die Bildentstehung beim Hohlspiegel geprüft wurde. Diese Entscheidung begründete sich durch die Zielsetzung, das Vorwissen der Schüler möglichst genau einzuschätzen. Die Verwendung von Aufgaben zur Bildentstehung bei der Sammellinse bot sich an, weil die Schüler bereits Vorkenntnisse hierzu im Unterricht erworben hatten. Die Verwendung einer Varianzanalyse mit Messwiederholung war hiermit jedoch ausgeschlossen.

Zur Prüfung der zweiten Hypothese wurde eine ANCOVA als Analysemethode gewählt, da ein t-Test für unabhängige Stichproben im Leistungsvortest ergab, dass die Kontrollgruppe von einem höheren Niveau aus startete als die Treatmentgruppe $t(50) = 3.96, p < .001$. Formal kann das getestete Hypothesenpaar wie folgt dargestellt werden.

- H_0 : $\mu_{\text{postTG}} - \mu_{\text{postKG}} \leq 0$
- H_1 : $\mu_{\text{postTG}} - \mu_{\text{postKG}} > 0$

Die Kovarianzanalyse erfordert die Erfüllung folgender Voraussetzungen (Bortz, 2005, S. 369): erstens Intervallskaliertheit der Daten, zweitens Normalverteilung der Daten, drittens Varianzhomogenität und viertens die Korrelation der Kovariate mit der abhängigen Variablen.¹⁴

Zu 1): Auf Basis der Testkonstruktion wurde von der Intervallskaliertheit der Daten für die Bildung der Gesamtpunktzahl ausgegangen.

Zu 2): Zur Prüfung der Normalverteilungsannahme wurde jeweils für den Prä- und für den Posttest ein Shapiro-Wilk-Test durchgeführt und ein QQ-Plot erstellt (vgl. Nachtigall & Wirtz, 2004, S. 171).¹⁵ Im Prätest zeigte sowohl der QQ-Plot als auch der Shapiro-Wilk-Test, dass die Daten hinreichend einer Normalverteilung ähneln. Dies gilt jedoch nicht für den Posttest: Hier handelt es sich um eine rechtssteile und linksschiefe Verteilung. Gemäß Bortz (2005b, S. 287) können Verletzungen der Normalverteilungsannahme bei schiefen Verteilungen und Stichproben $N > 10$ in Kauf genommen werden.

Zu 3): Der Levene-Test auf Gleichheit der Fehlervarianzen ergab ein signifikantes Ergebnis, so dass auch die Voraussetzung der Homogenität der Varianzen verletzt ist ($F(1, 50) = 16.63, p < .001$). Bei gleich großen Stichproben wird der F-Test durch heterogene Varianzen jedoch nur unerheblich beeinflusst (vgl. Bortz, 2005, S. 287). Bei der Interpretation der Ergebnisse wurden die entsprechenden Korrekturen, die SPSS ausgibt, berücksichtigt.

Zu 4): Abhängige Variable und Kovariate korrelieren signifikant miteinander ($r(50) = .50, p < .001$). Beim Vergleich der Ergebnisse von ANOVA und ANCOVA zeigt sich, dass die Berücksichtigung der Kovariate zugunsten der Treatmentbedingung ausfällt. Es ergibt sich ein positives Regressionsgewicht für die Kovariate, $b = 0.60, SD = 0.19; t = 3.10, p = .003$.

Insgesamt zeigte sich in der ANCOVA kein signifikanter Unterschied zwischen Treatment- und Kontrollgruppe im Posttest, $F(1, 52) = 1.36, MSE = 0.96, p = .714$ n.s.¹⁶.

14 Kovariaten mit geringer Reliabilität reduzieren die Teststärke und können bei nicht randomisierten Untersuchungen (z.B. in quasi-experimentellen Studie wie dieser) zu Verzerrungen führen (vgl. ebd.).

15 Dieser Test bietet sich bei kleinen Stichproben an.

16 n.s. = nicht signifikant

2.2.6.6 Dritte Hypothese: Wirkung des Treatments auf das konzeptuelle Verständnis

Schließlich wurde geprüft, ob die Treatmentgruppe im Konzepttest einen höheren Lernzuwachs verzeichnen kann als die Kontrollgruppe. Da der Kenntnisstand von Teilnehmern unter verschiedenen Bedingungen wiederholt mit dem gleichen Testinstrument erhoben wurde, bot sich als Analysemethode eine Varianzanalyse mit Messwiederholung an.

Zur Analyse der Ausgangsbedingungen ergab ein t-Test für unabhängige Stichproben im Konzeptvortest keine signifikanten Mittelwertunterschiede zwischen den beiden Gruppen ($t(50) = 1.12, p = .267, n.s.$).

An die Durchführung einer Varianzanalyse mit Messwiederholung sind zunächst die gleichen Voraussetzungen (1 bis 3) wie an eine ANOVA gebunden. Eine weitere Voraussetzung für die Durchführung einer zweifaktoriellen Varianzanalyse mit Messwiederholung ist Sphärizität. Sphärizität ist dann gegeben, wenn die Varianz der Differenzen (Kovarianz) für jedes beliebige Paar von Bedingungen gleich ist, oder anders ausgedrückt, wenn keine beliebigen zwei Bedingungen mehr oder weniger abhängig voneinander sind als irgendwelche anderen beiden. Sphärizität wird also erst ab drei Vergleichen zu einem möglichen Problem.

Zu 1) Auf Basis der Testkonstruktion wurde auch hier von der Intervallskaliertheit der Daten für die Bildung der Gesamtpunktzahl ausgegangen.

Zu 2) Die Prüfung der Normalverteilungsannahme erfolgte jeweils für den Prä- und für den Posttest per Shapiro-Wilk-Test und per QQ-Plot (vgl. Nachtigall & Wirtz, 2004, S. 171).¹⁷ Im Prä- und im Posttest zeigte sich sowohl im QQ-Plot wie auch im Shapiro-Wilk-Test, dass die Daten hinreichend einer Normalverteilung ähneln.

Zu 3) Der Levene-Test auf Gleichheit der Fehlervarianzen ergab kein signifikantes Ergebnis, so dass von der Homogenität der Varianzen ausgegangen werden kann.

Zu 4) Die im Mauchly-Test auf Sphärizität angegebene Signifikanz von $p < .001$ ist kleiner als das Signifikanzniveau von $p < .05$, was darauf hinweist, dass für die Daten keine Sphärizität vorliegt. Da jedoch nur zwei Gruppen miteinander verglichen werden, sind die Ergebnisse unter der Annahme, Sphärizität sei gegeben, identisch mit den Ergebnissen welche auch unter Berücksichtigung der Korrekturen von Box (zit. n. Rasch, Friese & Naumann, 2010, S. 111), von Greenhouse & Geisser (1959) (zit. n. ebd.) sowie von Huynh & Feldt (1976) (zit.n.

¹⁷ Dieser Test bietet sich bei kleinen Stichproben an.

Field, 2009, S. 461; vgl. auch Rasch et al., 2006, S. 111) ausgegeben werden ($\varepsilon_{\text{box}} = 1$; $\varepsilon_{\text{Huynh-Feldt}} = 1$; $\varepsilon_{\text{Greenhouse-Geisser}} = 1$).

Die Analyse wurde, wie folgt, umgesetzt: Die Quadratsummenanteile zwischen den Personen bezogen sich auf die Unterscheidung der Bedingung Treatment- versus Kontrollgruppe und die Quadratsummenanteile innerhalb der Personen auf den Faktor Zeit (Prä- versus Posttest). Für die Klärung der Fragestellung war insbesondere der Interaktionseffekt: Bedingung x Zeit relevant. Zur Analyse bot sich ein „einfacher“ linearer Kontrast an, bei dem der Effekt jeder Faktorstufe mit dem Effekt einer Referenzfaktorstufe verglichen wird (vgl. Rudolf & Müller, 2004, S. 100). Für diese Analyse wurde der Faktor Zeit (Prämessung) mit dem Effekt der folgenden Faktorstufe (Postmessung) verglichen. Formal kann das getestete Hypothesenpaar wie folgt dargestellt werden.

$$\begin{aligned} H_0 : & \quad (\mu_{\text{postTG}} - \mu_{\text{präTG}}) - (\mu_{\text{postKG}} - \mu_{\text{präKG}}) \leq 0 \\ H_1 : & \quad (\mu_{\text{postTG}} - \mu_{\text{präTG}}) - (\mu_{\text{postKG}} - \mu_{\text{präKG}}) > 0 \end{aligned}$$

Die ANOVA mit Messwiederholung belegt einen signifikanten Interaktionseffekt zwischen Zeit und Bedingung $F(1, 50) = 10.65, MSE = 315.60, p = .002, \eta_p^2 = .18$.

Abbildung 19 zeigt, dass die Treatmentgruppe einen signifikant höheren Lernzuwachs erreichte als die Kontrollgruppe. Die Effektstärke wurde mit dem Programm: OMEGA 3.1 (Fischer, 2001) berechnet. Auf Ebene der Population errechnet sich wie folgt (Rasch et al., 2010, S. 37):

$\sigma_{\text{systematisch}}^2$	systematische Varianz auf Ebene der Population
σ_{Gesamt}^2	Gesamtvarianz auf Ebene der Population

Der auf Basis der zugrundeliegenden Stichprobe *geschätzte* Effekt ω^2 für den Effekt in der Population betrug: $\omega^2 = .16$. Gemäß Cohen (1988) handelt es sich um einen starken Effekt: 16 % der erklärten Varianz lassen sich auf das Treatment zurückführen. Die dritte Hypothese konnte daher bekräftigt werden.

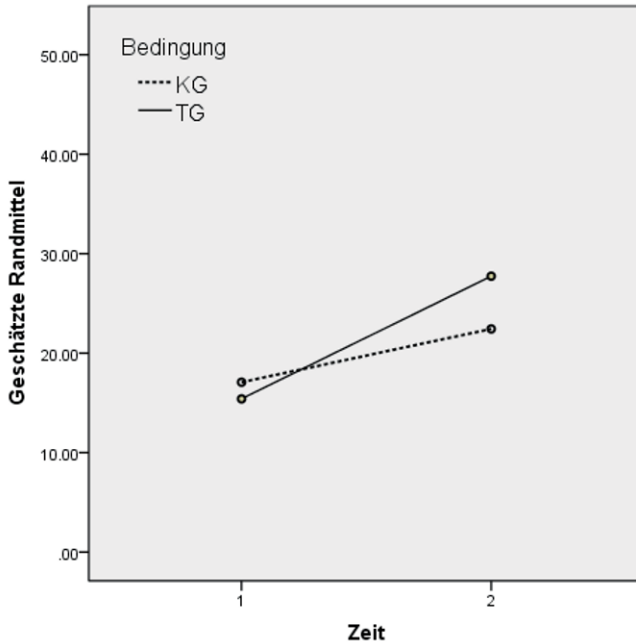


Abbildung 19: Interaktion zwischen Bedingung Treatment- (TG) versus Kontrollgruppe (KG) und Zeit zur Veranschaulichung des höheren Lernzuwachses der TG in Bezug auf das konzeptuelle Verständnis

2.2.7 Diskussion der Pilotstudie und Konsequenzen für die Hauptstudie

Das Hauptziel der Pilotstudie bestand darin, die entwickelten Lernmaterialien und Tests in der Praxis zu erproben. Die Ergebnisse weisen darauf hin, dass die entwickelten Materialien sowohl in der Treatment- als auch in der Kontrollgruppe lernwirksam sind.

Die Pilotstudie zeigt zudem, dass eine relativ kurze Intervention (135 min.), welche auf die Überwindung von weitverbreiteten Schülervorstellungen abzielt, zu einer signifikanten und praktisch bedeutsamen Verbesserung des konzeptuellen Verständnisses in der Strahlenoptik führt. Durch die Thematisierung von Schülervorstellungen durch Aufgaben, welche Schüler kognitiv aktivieren, sich mit multiplen Repräsentationen auseinanderzusetzen, kann offenbar das konzeptuelle Verständnis der Schüler gefördert werden. Der höhere Lernzuwachs in der Treatmentgruppe weist darauf hin, dass sich die Schüler signifikant häufiger

für korrekte Antworten und gegen Antwortalternativen entschieden haben, welche weitverbreitete Schülervorstellungen beinhalten. Angesichts der Hartnäckigkeit von Schülervorstellungen lässt sich dies durchaus als Erfolg werten.

Im Leistungstest bestehen Unterschiede im Vorwissen zwischen beiden Gruppen: So startete die Kontrollgruppe mit besseren Lernvoraussetzungen als die Treatmentgruppe. Unter Berücksichtigung der Gesamtpunktzahl im Leistungsvortest als Kovariate in der Analyse konnte kein signifikant höherer Lernzuwachs in der Treatment- im Vergleich zur Kontrollgruppe festgestellt werden.

Die Ergebnisse sind aus fünf Gründen jedoch mit Vorsicht zu interpretieren.

1. Für die Ergebnisse des Leistungstests besteht eine maßgebliche Einschränkung in der geringen Reliabilität des Messinstruments. Der Test sollte daher überarbeitet werden, bevor er erneut eingesetzt wird.
2. Bei der Beobachtung der Unterrichtsreihe zeigte sich, dass insbesondere in der letzten Stunde mehr Zeit für die Diskussion der Aufgaben erforderlich gewesen wäre. So konnten Rückfragen der Schüler nicht in der gewünschten Tiefe behandelt werden und die Zeit auf individuelle Lernschwierigkeiten einzugehen, musste zu Gunsten der Ergebnissicherung stark begrenzt werden.
3. Eine weitere Einschränkung der Interpretierbarkeit der Ergebnisse des Leistungstests ergibt sich aus den Verletzungen der Voraussetzungen: so waren die Daten nicht normalverteilt und auch die Homogenität der Varianzen war nicht gegeben. Die Verwendung der ANCOVA erweist sich gegenüber der Verletzung der Voraussetzungen jedoch als robust, wenn die verglichenen Gruppen annähernd gleich groß sind und den kritischen Wert von $N = 10$ pro Gruppe nicht unterschreiten.
4. Es liegen keine Daten zur längerfristigen Entwicklung vor. So kann z.B. keine Aussage darüber gemacht werden, ob der Lernzuwachs beider Gruppen im Leistungs- und im Konzepttest mittelfristig und längerfristig erhalten bleibt und ob der höhere Lernzuwachs der Treatmentgruppe im Konzepttest mittel- und langfristig stabil ist.
5. Nicht zuletzt handelt es sich um eine recht kleine Stichprobe, die aus einer Schule und einem Schultyp stammt. Die Ergebnisse lassen sich somit nicht auf den Lernerfolg von Schülern an Realschulen plus und Gesamtschulen generalisieren. Die Teststärke ($1-\beta$), also die Wahrscheinlichkeit die H_0 zu widerlegen, wenn sie tatsächlich falsch ist und die H_1 zutrifft, ist für geringe Effekte bei kleinen Stichproben eingeschränkt: So ist die Schätzung der Effektstärke für kleine Stichproben mit großer Unsicherheit behaftet (große Konfidenzintervalle). In der Konsequenz lassen sich für die Ergebnisse des

Leistungstests, der zudem auch nur über eine eingeschränkte Realibilität verfügt, keine zuverlässigen Schlüsse zur Wirkung des Treatments ziehen.

Für den Konzepttest sind die Ergebnisse eindeutiger. Der selbst in der kleinen Stichprobe nachweisbare gefundene starke Effekt ($\omega^2 = .16$) für den höheren Lernzuwachs der Treatmentgruppe weist auf die positive Wirkung der gezielten Thematisierung von wissenschaftlichen Konzepten hin. Schüler, welche hierzu ein gezieltes Treatment erhielten, entschieden sich signifikant häufiger für die korrekten Konzepte und wählten seltener Alternativen, welche weitverbreitete Schülervorstellungen beschreiben.

In der Folge wurden nachstehende Änderungen bei der Durchführung der Hauptstudie vorgenommen:

- Der Leistungstest zur Erfassung von Wissen und Problemlösen bei repräsentationsbezogenen Aufgaben wurde grundlegend überarbeitet. Auch der Konzepttest wurde optimiert.
- Um den längerfristigen Lernerfolg erheben zu können, wurde das konzeptuelle Verständnis und die Physikleistung bei repräsentationsbezogenen Aufgaben zu einem dritten Messzeitpunkt erfasst. Aus organisatorischen Gründen wurde ein Zeitabstand von acht Wochen gewählt: acht Wochen betrug der maximale zeitliche Abstand, der es noch ermöglichte, die Studie im Schuljahr 2010/2011 durchzuführen.
- Der Stichprobenumfang wurde deutlich erhöht: So nahmen an der Hauptstudie 21 Klassen aus 10 Schulen teil. Zusätzlich zu acht Gymnasien konnten auch zwei Gesamtschulen für die Teilnahme gewonnen werden.
- Angesichts des klar feststellbaren Effekts des Treatments der Pilotstudie auf die Förderung des konzeptuellen Verständnisses wurde im Anschluss an dieses Ergebnis das Treatment weiterentwickelt. Da sich die entwickelten Aufgaben, welche darauf zielen, weitverbreitete Schülervorstellungen durch die Auseinandersetzung mit fachbezogenen Repräsentationen zu überwinden, offenbar dazu eignen das konzeptuelle Verständnis zu fördern, wurde nun die Art der Aufgabengestaltung in den Blick genommen. Zur Untersuchung dieser Frage wurden zwei Versuchsbedingungen gewählt, die sich im Ausmaß der kognitiven Aktivierung im Umgang mit (multiplen) Repräsentationen unterscheiden. Im Gegensatz zur Pilotstudie, in der sich beide Bedingungen darin unterscheiden, dass Schülervorstellungen thematisiert wurden oder nicht, wurde in der Hauptstudie die Art und Weise mit Repräsentationen umzugehen und das Ausmaß

der kognitiven Aktivierung variiert. Ziel der Studie war es, tiefergehend zu analysieren, wie der Umgang mit Repräsentationen gestaltet werden kann, um das konzeptuelle Verständnis und die Leistung im Umgang mit Repräsentationen zu fördern, anstelle zu prüfen, ob dies überhaupt lernwirksam ist.

- Vor diesem Hintergrund wurde die Unterrichtsreihe von drei auf sechs Unterrichtsstunden erweitert. Die höhere Lernzeit zielte erstens auf einen nachhaltigeren Aufbau von Kompetenzen als dies in der Hälfte der Zeit möglich wäre. Zweitens wurde der Beobachtung der Zeitknappheit in der dritten Unterrichtsstunde in der Pilotstudie Rechnung getragen und drittens wurden zusätzliche Übungszeiten eingeplant, die sich aus der Weiterentwicklung des Treatments ergaben.

2.3 Hauptstudie

2.3.1 Forschungsfragen und Hypothesen

1. Fördern Aufgaben, welche Schüler kognitiv aktivieren, kohärente interne Repräsentationen (mentale Modelle) und externe Repräsentationen zu bilden sowie mit diesen zu operieren, Wissen und Problemlösen in Strahlenoptik (und zwar unmittelbar nach dem Unterricht und mittelfristig)?
2. Fördern Aufgaben, welche Schüler kognitiv aktivieren, kohärente interne Repräsentationen (mentale Modelle) und externe Repräsentationen zu bilden sowie mit diesen zu operieren, unmittelbar nach dem Unterricht und mittelfristig das konzeptuelle Grundverständnis in Strahlenoptik?
3. Steigert der Einsatz der Predict-Observe-Explain-Strategie unter der Perspektive des Lernens mit multiplen Repräsentationen das konzeptuelle Verständnis der Schüler?
4. Führen Aufgaben, welche Schüler kognitiv aktivieren, kohärente interne Repräsentationen (mentale Modelle) und externe Repräsentationen zu bilden sowie mit diesen zu operieren, unmittelbar nach dem Unterricht und mittelfristig zu einer Veränderung der Motivation im Vergleich zu einer Kontrollgruppe?
5. Wirkt sich die Thematisierung von Schülervorstellungen positiv auf die Entwicklung des konzeptuellen Verständnisses aus (im Vergleich zu Physikunterricht zum gleichen Lerninhalt ohne Thematisierung von Schülervorstellungen). Oder anders gefragt: Lassen sich die Ergebnisse der Pilotstudie unter leicht veränderten Bedingungen replizieren?

Hypothese 1 Die Treatmentgruppe erreicht im Wissen und Problemlösen beim Umgang mit fachspezifischen Repräsentationen im Bereich der Strahlenoptik unmittelbar nach der Intervention und ca. zwei Monate später einen höheren Lernzuwachs als die Kontrollgruppe.

Hypothese 2: Die Treatmentgruppe erreicht im konzeptuellen Verständnis im Bereich der Strahlenoptik unmittelbar nach der Intervention und ca. zwei Monate später einen höheren Lernzuwachs als die Kontrollgruppe.

Für die Untersuchung der beiden ersten Hypothesen wurden zwei Versuchsbedingungen gewählt, die sich im Ausmaß der kognitiven Aktivierung im Umgang mit (multiplen) Repräsentationen unterscheiden.

Die Treatmentbedingung beinhaltete kognitiv aktivierende Aufgabenstellungen, welche die Überwindung von gängigen Schülervorstellungen erfordern. In der Kontrollbedingung wurden die gleichen Schülervorstellungen thematisiert und prinzipiell auch vergleichbare Repräsentationsformen verwendet. Schüler in der Kontrollbedingung wurden jedoch nicht gezielt kognitiv aktiviert, eigene Repräsentationen zu erstellen, zu reflektieren und mit gegebenen und selbst erstellten Repräsentationen zu operieren.

Wissen und Problemlösen beim Umgang mit fachspezifischen Repräsentationen im Bereich der Strahlenoptik wurde in einem eigens entwickelten inhaltspezifischen Leistungstest zu drei Messzeitpunkten erfasst: zum Messzeitpunkt prä (vor der Intervention), zum Messzeitpunkt post (unmittelbar nach der Intervention) und zum Messzeitpunkt follow-up (ca. zwei Monate später).

Das konzeptuelle Verständnis im Bereich der Strahlenoptik wurde ebenfalls in einem eigens entwickelten Konzepttest zu den drei genannten Messzeitpunkten erhoben.

Hypothese 3: Ein Teil der Treatmentgruppe, die ein variiertes Treatment erhielt, welches die Predict-Observe-Explain-Strategie (POE-Sequenz) unter der Perspektive des Lernens mit multiplen Repräsentationen beinhaltet, erreicht im Vergleich zu dem regulären Treatment und im Vergleich zur Kontrollgruppe einen höheren Lernzuwachs im konzeptuellen Verständnis und unmittelbar nach der Intervention und ca. zwei Monate später.

Hypothese 4: Die Entwicklung der Motivation der Schüler in der Treatmentgruppe unterscheidet sich von der Entwicklung in der Kontrollgruppe.

Diese Hypothese wurde als ungerichtete Hypothese formuliert: einerseits könnte sich die intendierte höhere kognitive Aktivierung in der Treatmentgruppe positiv auf die Motivation der Schüler auswirken, andererseits könnte die höhere kognitive Aktivierung in der Treatmentgruppe auch zu erhöhten Belastungen und Anforderungen führen, was die Motivation mindert. Insgesamt steht die

Motivation nicht im Mittelpunkt dieser Studie, da das Treatment vorwiegend auf die Optimierung der kognitiven Lernbedingungen zielt. Als wichtiger Faktor für Lernprozesse wurde sie jedoch mit untersucht.

Hypothese 5: Schüler, die ein Treatment zur Förderung des konzeptuellen Verständnisses erhielten (Treatment- und Kontrollgruppe dieser Studie) erzielten einen höheren Lernzuwachs in Bezug auf das konzeptuelle Verständnis als Schüler, die ein Treatment zur Förderung der Kohärenz (Scheid, 2013) erhielten bzw. der Kontrollgruppe der entsprechenden Studie angehörten.

Zur Beantwortung der fünften Forschungsfrage wurde innerhalb des Projekts die Studie von Scheid (2013) mit den hier berichteten Daten verglichen. Unterschiede zwischen Treatment- und Kontrollgruppe innerhalb der Einzelstudien werden in die Analyse einbezogen, sofern signifikante Unterschiede vorlagen.

Des Weiteren wurden im Rahmen des oben dargelegten Forschungsprogramms (Fragen 1-5) die beiden folgenden Forschungsziele verfolgt.¹⁸

1. Vor dem Hintergrund, dass für die Domäne der Strahlenoptik im Gegensatz zu anderen Bereichen der Physik, wie der Mechanik (vgl. Hestenes et al., 1992) bislang kein Konzepttest vorliegt, sollte eine tragfähige Fassung eines Konzepttests in der Strahlenoptik erstellt werden. Damit verbundene offene Forschungsfragen betreffen vor allem die Beurteilung, inwiefern die entwickelte Testfassung die Gütekriterien der Objektivität, der Reliabilität und der Validität erfüllt.
2. Im Rahmen der Exploration der Lernprozesse, während der Intervention, wurde der Frage nachgegangen, ob der erfolgreiche Umgang mit zentralen Repräsentationen ein geeigneter Prädiktor für den Lernerfolg unmittelbar nach dem Unterricht darstellt.

Zur Umsetzung der ersten Zielsetzung wurde neben einer ausführlichen Analyse der Itemkennwerte, der Test auf Rasch-Skalierbarkeit geprüft, Korrelationen mit relevanten Außenkriterien (z.B. Physik- und Mathematiknote, figural-räumliches Schlussfolgern etc.) berechnet (Analyse der Kriteriumsvalidität) sowie eine Analyse der Dimensionalität des Tests mittels einer faktorenanalytischen Kreuzvalidierung vorgenommen (Konstruktvalidität).

Zur Umsetzung der zweiten Zielsetzung wurde eine per Zufall ausgewählte Teilstichprobe von Arbeitsblättern, die im Rahmen der Intervention eingesammelt wurden, vertiefend analysiert.

¹⁸ Fragen, welche mit den genannten Forschungszielen einhergehen, werden im Rahmen dieser Arbeit exploriert, jedoch nicht vollumfänglich beantwortet.

2.3.2 Stichprobe und Design

Die Stichprobe bestand aus $N = 525$ Schülern aus $N = 21$ Schulklassen von 10 Schulen, die von $N = 10$ unterschiedlichen Lehrern unterrichtet wurden. Von den 525 Schülern gingen die Daten von mindestens $N = 443$ Schülern in die Analysen ein. Die übrigen 82 Schüler fehlten an den Tagen, an denen die abhängigen Variablen oder die kognitiven Fähigkeiten erfasst wurden oder es lagen aus diversen Gründen (Schulwechsel zum zweiten Halbjahr) keine vollständigen Informationen zu den Fachnoten vor. In späteren Modellen weicht die zugrundeliegende Datenbasis von $N = 443$ nach oben ab, wenn zu den einzelnen Messzeitpunkten mehr Daten zur Verfügung standen oder aus inhaltlichen Gründen weniger Kovariaten in das Modell eingingen, was entsprechend die Chance erhöht, dass weniger Daten fehlen. Die 525 Schüler waren im Alter von 12 bis 15 Jahren ($M = 13.60$, $SD = 8$ [Monate]) und besuchten entweder ein Gymnasium oder eine Integrierte Gesamtschule (IGS) im Bundesland Rheinland-Pfalz. Zur Stichprobe zählten dabei 17 Gymnasialklassen, $n = 433$ Schüler ($\approx 82\%$ der Gesamtstichprobe) und 4 Klassen an Gesamtschulen, $n = 92$ Schüler ($\approx 18\%$ der Gesamtstichprobe). Der Stichprobe gehörten $n = 256$ Jungen und $n = 269$ Mädchen an (vgl. Tabellen Tabelle 3 und 4). Unter den 10 teilnehmenden Physik-Lehrern waren drei Frauen und sieben Männer, die Berufserfahrung der Lehrkräfte variierte zwischen 2 und 30 Jahren. Detaillierte Informationen zur Stichprobe der Lehrer und Schüler können Tabelle 6 im Anhang C9 (auf der Produktseite zu diesem Buch unter www.springer.com), entnommen werden.

Die Umsetzung der Studie war im Vorfeld durch den Datenschutzbeauftragten und die Aufsichts- und Dienstleistungsdirektion des Landes Rheinland-Pfalz geprüft und genehmigt worden. Die Teilnahme an der Studie war freiwillig. Alle Schülerleistungen wurden anonym erfasst. Jeder Schüler erhielt eine Teilnehmernummer, welche er oder sie durchgehend in der Studie verwendete, um die Zuordnung der Testleistungen über unterschiedliche Erhebungsinstrumente und Messzeitpunkte anonym, aber dennoch individuell, auswerten zu können. Die Zuordnung der Teilnehmernummer zu den Namen erfolgte durch die Lehrer und ist der Versuchsleiterin und den beteiligten Mitgliedern der Arbeitsgruppe nicht bekannt. Voraussetzung für den Start der Studie in der jeweiligen Schule war die schriftliche Erlaubnis der Eltern. Nichtteilnehmende Schüler beteiligten sich am Unterricht, ihre Testleistungen wurden jedoch nicht erfasst.

Bei der Hauptstudie handelte es sich um ein einfaktorielles quasi-experimentelles Design. Die Leistung der Schüler wurde zu drei Messzeitpunkten erhoben: unmittelbar vor Beginn der Unterrichtsreihe, direkt nach der Unterrichtsreihe sowie zwei Monate später.

Tabelle 3 und *Tabelle 4* Informationen zur Stichprobe nach Bedingung (Treatment-versus Kontrollgruppe) Klassenstufe, Geschlecht und Schultyp

		Bedingung			Klassenstufe				
		KG ^a	TG ^b		7	8			
Geschlecht	m	121	135	256	Schulart	^a GY	171	262	433
	w	129	140	269		^b IGS	51	41	92
		250	275	525			222	303	525

^aKontrollgruppe, ^bTreatmentgruppe

^aGymnasium, ^bIGS

Auch diese Studie wurde in den regulären Schulalltag eingebunden, was die externe Validität durch die realitätsnahen Bedingungen erhöht. Um den Stundenplan einzuhalten, wurden die Schüler im Klassenverbund als Gruppe den Bedingungen zugeteilt.

An den 10 teilnehmenden Schulen unterrichtete jeweils eine Lehrkraft eine Treatment- und eine Kontrollgruppe derselben Jahrgangsstufe (Parallelklassen), eine Lehrkraft an einem Gymnasium unterrichtete drei Klassen der siebten Jahrgangsstufe, an dieser Schule wurden zwei Klassen der Treatment und eine Klasse der Kontrollbedingung zugeordnet.

Dadurch dass jede Lehrkraft jeweils sowohl in der Treatment- als auch in der Kontrollbedingung unterrichtete, sollte der Faktor der Lehrerpersönlichkeit soweit wie möglich kontrolliert werden.

Auf Basis der Vornoten (Mathematik, Deutsch und Physik) wurden die Klassen abwechselnd der Treatment- und der Kontrollgruppe zugeordnet, um insgesamt möglichst vergleichbare Stichproben zu erhalten. Durch dieses Verfahren kann nur eine Parallelisierung auf Gruppenebene erzielt werden und nicht auf Individual-ebene. Merkmalsunterschiede zwischen Treatment- und Kontrollbedingung, die von vornherein bestanden, können somit nicht vollständig ausgeschlossen werden.

Zur Prüfung der dritten Hypothese wurde mit einem Teil der Stichprobe (6 Klassen)¹⁹ ein variiertes Treatment durchgeführt, in welchem in der ersten und in der vierten Unterrichtsstunde POE-Sequenzen (White & Gunstone, 1992; Palmer, 1995; Kearney, Treagust, Yeo & Zadnik, 2001; Crouch et al., 2004) eingebaut waren (vgl. Tabelle 5). Die Implementierung der Treatmentvariation zielte darauf ab, zu

¹⁹ Diese Teilstichprobe enthielt zwei 8. Klassen eines Gymnasiums, zwei 7. Klassen eines Gymnasiums und zwei 7. Klassen einer Integrierten Gesamtschule.

untersuchen, ob das Verständnis der Bildentstehung vertieft werden kann, wenn die Schüler mit Hilfe der POE-Sequenz kognitiv dazu aktiviert werden, ihre eigenen Vorhersagen, welche sie auch mittels schematischer Repräsentationen darstellen sollten, mit den beobachteten Ergebnissen des Versuchs und der repräsentationalen Deutung der beobachteten Ergebnisse zu kontrastieren. In Parallelklassen, in denen das variierte Treatment umgesetzt wurde, erhielt die Kontrollgruppe den gleichen Unterricht wie die übrigen Kontrollgruppen an den anderen Schulen.

Tabelle 5 Informationen zur Stichprobe je Bedingung und Treatmentform

		Bedingung		
		KG	TG	
Treatmentform	regulär	179	192	371
	variiert	71	83	154
		250	275	525

Beide Klassen, Treatment- und Kontrollgruppe, wurden jeweils von der gleichen Lehrkraft im gleichen Zeitumfang unterrichtet. Die Unterrichtsreihe war ursprünglich auf sechs Schulstunden ausgerichtet. Bei der Umsetzung zeigte sich jedoch, dass alle Lehrer durch unvorhergesehene Unterbrechungen im Schulalltag etwas mehr Zeit benötigten, so dass der Lernstoff auf sieben Unterrichtsstunden (315 min.) verteilt wurde. Dabei wurde die absolute Lernzeit für Treatment- und Kontrollgruppe parallel gehalten. Die Prä- und Posttests fanden jeweils in der Physikstunde vor und nach der Unterrichtseinheit statt und erforderten jeweils 45 min. Die Follow-up Erhebung zwei Monate später erforderte ebenfalls eine Unterrichtsstunde. Zusätzlich wurde noch eine Unterrichts- oder Vertretungsstunde für die Erhebung der kognitiven Fähigkeiten der Schüler aufgewandt.

Die Umsetzung der Studie fand Ende des Schuljahres 2010/2011 statt. Zu diesem Zeitpunkt wurden die Schüler entweder seit August 2010 oder seit Januar 2011²⁰ erstmals im Fach Physik unterrichtet. Die Unterrichtsreihe wurde begonnen, sobald das Thema der Bildentstehung bei der Sammellinse auch im regulären Curriculum vorgesehen war. Entsprechend hatten bereits alle Schüler einige Vor-

²⁰ Manche Schulen starteten mit dem Physikunterricht erst zum zweiten Halbjahr (Epochalunterricht).

kenntnisse zu den Konzepten der geradlinigen Lichtausbreitung, zur Streuung, zur Brechung an Grenzflächen und zur physikalischen Sehvorstellung erworben.

2.3.3 Operationalisierung des Treatments

2.3.3.1 Genereller Überblick

Für die Intervention wurden zwei Versuchsbedingungen gewählt, die sich im Ausmaß der kognitiven Aktivierung im Umgang mit (multiplen) Repräsentationen unterscheiden. Die Unterrichtsreihe umfasste in der Planung sechs Schulstunden:

1. eine Einführungsstunde zum Thema Brechungsvorgänge an der Sammellinse. In dieser Stunde zeigte die jeweilige Lehrkraft den Schülern ein Demonstrationsexperiment, bei dem parallel einfallende Lichtstrahlen auf eine Sammellinse treffen und eine optische Tafel streifen; auf Basis des Demonstrationsexperiments befassten sich die Schüler mit den Grundlagen der Konstruktion von Strahlengängen,
2. der Durchführung des Schülerexperiments zur Entstehung reeller Bilder bei der Sammellinse inklusive der Auswertung des Experiments in der zweiten Stunde,
3. die Konstruktion der unterschiedlichen Bildfälle (vergrößertes, verkleinertes und gleich großes Bild) in der dritten Stunde,
4. die Thematisierung und Konstruktion des virtuellen Bildes in der vierten Stunde,
5. der Einführung des Abbildungsgesetzes mit verbaler und geometrischer Veranschaulichung der Zusammenhänge in der fünften Stunde und
6. der Abschlussstunde, in der die Bildkonstruktion und das Abbildungsgesetz vertiefend geübt wurden.

Kognitiv aktivierende Lernstrategien zielen darauf, die Lernenden durch anspruchsvolle Aufgaben dazu anzuregen, die erlernten physikalischen Vorgänge mental durchzuspielen, also geeignete mentale Modelle aufzubauen. Der Aufbau solcher adäquaten mentaler Modelle sollte, sofern dieser Schritt gelingt, zum einen dazu führen, Schülervorstellungen zu überwinden, da Inkonsistenzen mit einer adäquaten Modellvorstellung inkompatibel sind (vgl. Gentner & Gentner, 1983; diSessa, 1983, 1988, 1993; Botzer & Reiner, 2005; Mortimer & Buty, 2009) und sich zum anderen auch in externen Repräsentationen niederschlagen (vgl. Cox, 1999), die wiederum dazu genutzt werden können, Problemlöseaufgaben zu bearbeiten (vgl. Gentner & Gentner, 1983; Ohlsson, 1992; Zhang, 1997).

Tabelle 6 Überblick über Formen der kognitiven Aktivierung in der Treatmentbedingung

Art der kognitiven Aktivierung in der TG	Theoretischer / empirischer Bezug:	Aufgaben ^a siehe Anhang C1 ^b	Std.
Kognitive Aktivierung durch das Erstellen eigener Repräsentationen	Eigene Gedanken, Konzepte, Ideen und Lösungswege erläutern (Hiebert & Wearne, 1993; Shayer & Adhmi 2007).	Aufgabenblatt 1 [Mat2_K1] A1	1
		Experimentieranleitung [Mat5_K1]	2
		Aufgabenblatt 4 [Mat7_K1]	3
		Aufgabenblatt 5 [Mat8_K1]	4
		Aufgabenblatt 7 [Mat10_K1] A2 und A3	5
Abgleich der eigenen Vorstellung mit der Lösung. Der Abgleich mit der Lösung bezweckt, die eigene Sichtweise zu reflektieren und die eigenen Verarbeitungsschritte entsprechend anzupassen	Beziehungen zwischen Vorwissen bzw. dem derzeitigen Lernstand und neuen Lerninhalten herstellen (Hiebert & Wearne, 1993; Baumert & Kunter, 2011).	Aufgabenblatt 8 [Mat12_K1]	5
		Experimentieranleitung [Mat5_K1]	2
		Aufgabenblatt 4 [Mat7_K1]	3
		Aufgabenblatt 7 [Mat10_K1] A2 und A3	5
		Aufgabenblatt 3 [Mat6_K1]	2
Kognitive Aktivierung durch die Anforderung, Repräsentationen aufeinander zu beziehen bzw. ineinander zu übersetzen, die in unterschiedlichen Formaten vorliegen oder auf unterschiedlichen Abstraktionsebenen liegen	Beziehungen zwischen verschiedenen Repräsentationsformen herstellen (Hiebert & Wearne, 1993; Stein & Lane 1996).	Aufgabenblatt 4 [Mat7_K1]	2
		Aufgabenblatt 6 [Mat9_K1]	3
		Aufgabenblatt 7 [Mat10_K1], A2 und A3	4
		Aufgabenblatt 8 [Mat12_K1]	5
		Aufgabenblatt 3 [Mat6_K1]	2
Repräsentationales Fading-out	Beziehungen zwischen dem derzeitigen Lernstand und neuen Lerninhalten herstellen (Hiebert & Wearne, 1993; Baumert & Kunter, 2011).	Aufgabenblatt 4 [Mat7_K1] (Konstruktion der Bildfälle)	3
		Aufgabenblatt 7: [Mat10_K1], A2 und A3	5
Förderung der kognitiven Flexibilität im Umgang mit Repräsentationen	Beziehungen zwischen verschiedenen Repräsentationsformen herstellen (Hiebert & Wearne, 1993; Stein & Lane 1996).	Aufgabenblatt 9 [Mat13_K1], A1 und A2	6
		Aufgabenblatt 3 [Mat6_K1]	2
Einsatz von Aufgaben, die darauf angelegt sind, ein internes mentales Modell physikalischer Abläufe oder	Beziehungen zwischen verschiedenen Repräsentationsformen herstellen (Hiebert & Wearne, 1993; Stein & Lane 1996).	Aufgabenblatt 6 [Mat9_K1]	4
		Aufgabenblatt 3 [Mat6_K1]	5

Zusammenhänge auf Basis depliktionaler schematischer Repräsentationen zu bilden, welche die Schüler die physikalischen Abläufe bzw. Zusammenhänge in kohärenter Weise intern dynamisch simulieren lassen.	Aufgabenblatt 7 [Mat10_K1], A2 und A3
Treatmentvariation	
POE-Sequenz, Vorhersage unter Verwendung von Repräsentationen in unterschiedlichen Formaten, sowie Abgleich der eigenen Vorstellung mit der Lösung.	Demonstrationsexperiment an der optischen Scheibe [Zusatz-1-K1] TG und [Zusatz-2-K1] TG
Kognitive Konflikte initiieren (Baumert & Kunter, 2011); Beziehungen zwischen verschiedenen Repräsentationsformen herstellen (Hiebert & Wearne, 1993; Stein & Lane 1996); einen prozessorientierten Umgang mit Problemen fördern, Kliceme et al., 2006)	

^aMehrfachnennungen möglich

^bauf der Produktseite zu diesem Buch unter www.springer.com

Die Art der im Unterricht verwendeten Repräsentationen und die angesprochenen Schülervorstellungen sowie der Einsatz eines Demonstrations- und eines Schülerexperiments waren in beiden Versuchsbedingungen konstant gehalten. Folgende Schülervorstellungen wurden sowohl in der Treatment- als auch in der Kontrollgruppe behandelt:

Tabelle 7 Überblick über thematisierte Konzepte und damit verbundene Schüler- vorstellungen in der Treatment- (TG) und der Kontrollbedingung (KG)

Thematisierte Grundkonzepte	Adressierte Schülervorstellungen (TG und KG)	Aufgaben ^a	Std
Wissenschaftstheoretisch adäquates Lichtverständnis; Physikalische Selbvorstellung und geradlinige Lichtausbreitung	Verwechslung Lichtbündel und Lichtstrahl Verständnisprobleme mit der geradlinigen Lichtausbreitung: Fokussierung auf ausgezeichnete Strahlen	siehe Anhang C1 ^b Aufgabenblatt 1 [Mat2_K1] A1 Aufgabenblatt 3 [Mat6_K1] TG	1 2
Entstehung reeller Bilder bei der Sammellinse, Konzept der Punk-zu-Punkt-Abbildung	Fehlerhaftes Verständnis der Funktionsweise von Sammellinsen, fehlerhafte bildliche Repräsentation, Vorbeugung einer holistischen Konzeption des Abbildungsvorgangs durch Einüben der Strahlenkonstruktion	Aufgabenblatt 3 [Mat6_K1] TG Aufgabenblatt 4 [Mat7_K1] (Konstruktion der Bildfälle)	2 3
Virtuelle Bilder bei der Sammellinse Entstehung reeller Bilder bei der Sammellinse	Probleme beim Verständnis der Entstehung (und Lage) virtueller Bilder Bei Abdeckung der oberen bzw. unteren Hälfte der Linse wird entsprechend die obere bzw. untere Hälfte des Bildes angeschnitten	Aufgabenblatt 5 [Mat8_K1] Aufgabenblatt 7: [Mat10_K1], A2 (Hinführung) und A3 Aufgabenblatt 9 [Mat13_K1], A1	4 5 6
Entstehung reeller Bilder bei der Sammellinse: Aspekt Funktion des Schirms / Ort des Bildes	Schwierigkeiten beim Verständnis des Bildortes: Vorstellung, das Bild sei auf der Linse; nur wenige Lernende erkennen, dass sich das Bild ohne Schirm an der gleichen Position befindet wie der Schirm.		
Treatmentvariation Entstehung reeller Bilder bei der Sammellinse	Bei Abdeckung der oberen bzw. unteren Hälfte der Linse wird entsprechend die obere bzw. untere Hälfte des Bildes angeschnitten	[Zusatz-1-K1] TG und [Zusatz-2-K1] TG (Demonstrationsversuch „abgedeckte Sammellinse“)	4

^aMehrfachnennungen möglich

^b auf der Produktseite zu diesem Buch unter www.springer.com

Solche kognitiven Lernstrategien beinhalten auch eine metakognitive Komponente (vgl. Schrader & Helmke 2006, S. 638 f.). Durch die Aufforderung, den physikalischen Sachverhalt zeichnerisch und schriftlich darzustellen, verlangen sie von den Lernenden, ihre Vorstellungen zu reflektieren sowie mit den erstellten Repräsentationen kognitiv zu operieren. Die Operation mit den erstellten externen Repräsentationen beinhaltet, deren Zweckmäßigkeit zur Lösung von physikalischen Problemen zu prüfen, und zielt somit darauf, die eigenen Verarbeitungsschritte anzupassen.

Das vollständige Unterrichtsmaterial ist in Anhang C1, die zugehörige Unterrichtsplanung in Anhang C2 (auf der Produktseite zu diesem Buch unter www.springer.com) zu finden. Zusammenfassungen zur Operationalisierung finden sich in Tabelle 6 und Tabelle 7. Zur guten Lesbarkeit der Aufgabentexte sei ebenfalls auf Anhang C1 unter www.springer.com hingewiesen.

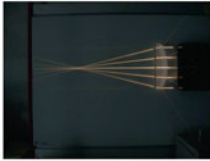
2.3.3.2 Operationalisierung des Treatments in der ersten Unterrichtsstunde

Unterschiede in der kognitiven Aktivierung im Umgang mit Repräsentationen bezogen sich in der ersten Stunde auf eine unterschiedliche Thematisierung der Begriffe „Lichtbündel“ und „Lichtstrahl“. Beiden Gruppen wurde folgendes Demonstrationsexperiment gezeigt: auf einer optischen Tafel ist eine Sammellinse und eine Lampe mit einer Schlitzblende montiert. Durch die Schlitzblende werden schmale parallel einfallende Lichtbündel erzeugt, welche die optische Tafel streifen (die Lichtbündel sind durch die Streuung an der optischen Tafel zu sehen). Die Sammellinse vereinigt das achsenparallel einfallende Licht im Brennpunkt. Nachdem sich die parallel einfallenden Lichtbündel im Brennpunkt getroffen haben, divergieren sie. Bezogen auf das Demonstrationsexperiment sollten sich die Schüler den Unterschied zwischen Lichtbündel und Lichtstrahl klar machen.

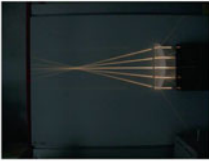
Während man im Demonstrationsexperiment Lichtbündel sehen kann, die eine optische Tafel streifen, handelt es sich bei Lichtstrahlen um ein gedankliches Konstrukt, das zur Beschreibung und zeichnerischen Darstellung von Lichtbündeln dient. Die hiermit verbundene Schülervorstellung betrifft eine materialistische Konzeption von Licht, in welcher der Begriff (Licht-)„Strahl“ nicht im mathematischen Sinn aufgefasst wird, sondern als konkreter Strom, analog einem Flüssigkeitsstrom verstanden wird (vgl. Andersson & Kärrqvist, 1983; Reiner et al., 2000, S. 14 f.).

Gerade letztere Vorstellung zeigt, dass Lernende mit wenig physikalischer Vorbildung ontologisch grundlegend verschiedene Vorstellungen von Licht besitzen.

In beiden Gruppen wird diese Schülervorstellung durch eine verbale Repräsentation angesprochen, die sich auf das Demonstrationsexperiment bezieht (vgl. Abbildung 20). Durch das offene Frageformat in der Treatmentgruppe wird diese jedoch kognitiv aktiviert, eine eigene verbale Repräsentation zu erstellen und zu prüfen, während die Kontrollgruppe aus zwei gegebenen verbalen Repräsentationen die zutreffende Repräsentation wählen muss.



b) Was versteht man unter dem Brennpunkt einer Sammellinse?



b) Was versteht man unter dem Brennpunkt einer Sammellinse?

c) Beschreibe den Unterschied zwischen einem Lichtbündel und einem Lichtstrahl

c) Welcher Satz ist richtig?

- Lichtstrahlen gibt es nur in der gedanklichen Vorstellung. Sie dienen zur Beschreibung und zeichnerischen Darstellung von Lichtbündeln.
- Lichtbündel gibt es nur in der gedanklichen Vorstellung. Sie dienen zur Beschreibung und zeichnerischen Darstellung von Lichtstrahlen.

Abbildung 20: Aufgabenblatt 1 - Die Sammellinse[Mat 2_K1] TG bzw. [Mat 2_K2] KG links Aufgabe der Treatmentgruppe (TG), rechts Aufgabe der Kontrollgruppe (KG)

2.3.3.3 Treatmentvariation in der ersten Unterrichtsstunde

In der Treatmentgruppe wurden nun zwei kurze POE-Sequenzen eingebaut. Vor der Vorführung des Demonstrationsexperimentes wurden die Schüler gefragt, wie das Licht verläuft, wenn man die Lampe anschaltet und eine Linse davor setzt. Die Schüler sollten den Verlauf des Lichts vorhersagen und ihre Vermutungen zum Strahlenverlauf in einer schematischen Repräsentation an der Tafel zeichnen (predict).

Anschließend wurde den Schülern das Experiment demonstriert und die Beobachtung der Schüler festgehalten (observe). Abschließend wurde die Funktionsweise der Linse erklärt und die zentralen Begriffe wie „der Brennpunkt F “ und „die Brennweite f “ sowie der Begriff der optischen Achse erarbeitet (explain). Das Unterrichtsmaterial zur Treatmentvariation kann in Anhang C3 und die Unterrichtsplanung zum variierten Treatment in Anhang C4 jeweils auf der Produktseite zu diesem Buch unter www.springer.com eingesehen werden.

In einer zweiten Sequenz wurde die Auswirkung verschiedener Brennweiten thematisiert. So sollten die Schüler vorhersagen, was passiert, wenn man eine stärker gekrümmte Linse verwendet. Auf Basis der Schülerantworten erstellte die

Lehrkraft entsprechende Tafelbilder (predict). Anschließend führte die Lehrkraft den Schülern das Experiment vor, wobei die Beobachtungen der Schüler festgehalten wurden (observe). Abschließend klärten Schüler und Lehrer im Dialog das Ergebnis (explain).

In der Kontrollgruppe wurde der Unterrichtslauf belassen: Die Schüler erarbeiteten auf Basis des Demonstrationsexperiments die zentralen Begriffe ohne POE-Sequenz, in der die Schüler ihre Vorhersagen verbal und zeichnerisch darstellten und anhand des Experiments nachbereiteten.

2.3.3.4 Operationalisierung des Treatments in der zweiten Unterrichtsstunde

In der zweiten Unterrichtsstunde führten sowohl die Treatment- als auch die Kontrollgruppe ein Schülerexperiment zur Entstehung reeller Bilder bei der Sammellinse durch. Analog zu dem Versuch zur Bildentstehung am Hohlspiegel der Pilotstudie montierten sie zunächst eine Kerze (auf einem Kerzenteller), eine Sammellinse und einen Schirm auf einer optischen Bank. Durch das Verschieben von Kerze, Linse und Schirm konnten die Abstände zwischen Kerze und Linse sowie zwischen Linse und Schirm eingestellt werden. Durch die Wahl geeigneter Abstände war es für die Schüler nun möglich, ein reelles, verkleinertes, vergrößertes oder gleichgroßes Bild der Kerzenflamme auf dem Schirm aufzufangen. Die Aufgabe bestand darin, diese Bilder zu finden und die Flammengröße von Kerze und Bild zu messen.

Da das Experiment die Unterrichtszeit von 45 min. nahezu ausfüllte, konzentrierte sich das Herausarbeiten des Unterschieds zwischen Treatment- und Kontrollgruppe auf das Nacharbeiten in der Hausaufgabe.

Beide Gruppen befassten sich zur Wiederholung mit dem Versuchsaufbau und den hiermit verbundenen physikalischen Begriffen: „Bildweite“, „Gegenstandsweite“, „Bildgröße“ und „Gegenstandsgröße“. Eine Schülervorstellung wurde in diesem Kontext nicht adressiert.

In beiden Gruppen erforderte es die Aufgabe, zwei unterschiedliche Ebenen miteinander zu verknüpfen:

- das reale Experiment (Ebene der konkreten Gegenstände), auf welche die Skizze des Versuchsaufbaus (Repräsentation des Versuchsaufbaus) referiert,
- sowie die Beschriftung des Versuchsaufbaus, bei der die physikalischen Größen (deskriptiv abstrakte Repräsentationen) auf die Ebene des realen Experiments und ihre Repräsentation in der Skizze bezogen werden. In den Unterrichts-

beobachtungen zeigte sich, dass diese Aufgabe für die Schüler keineswegs so trivial war, wie sie erscheinen mag: Probleme ergaben sich v.a. in der Zuordnung von Gegenstandsweite und Bildweite. Hier kam es des Öfteren zu Verwechslungen.

Die kognitive Aktivierung in der Treatmentgruppe bestand darin, die Skizze selbst zu erstellen, während der Kontrollgruppe die Skizze vorgegeben wurde. Das eigene Erstellen der Skizze stellt nicht nur eine kognitive Aktivierung im Sinn eigener Tätigkeit dar. Die kognitive Aktivierung der Treatmentgruppe sollte als Ergebnis das aktive Abrufen der Lerninhalte zeigen. Aus der kognitionspsychologischen Forschung ist bekannt, dass selbstgenerierte Elaborationen die elaborative Verarbeitung und damit die Behaltensleistung der Lernenden fördern (vgl. Stein & Bransford, 1979, zit. n. Anderson, 1996, S. 187). Auf metakognitiver Ebene verlangt die Fragestellung in der Treatmentgruppe von den Schülern, ihr Verständnis auch von der Funktion des Versuchsaufbaus im Abgleich mit der Lösung zu reflektieren (vgl. Abbildung 21).

Erstelle eine Skizze des Versuchsaufbaus und beschrifte sie mit den Dir bekannten Größen.



Beschrifte den Versuchsaufbau:

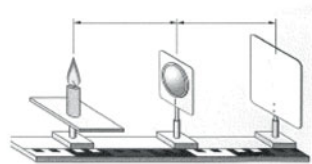


Abbildung 21: [Experimentieranleitung Mat. 5_ K1] bzw. [Mat5 1], Abbildung rechts entnommen aus Backhaus et al. (2008). Fokus Physik Gymnasium Rheinland-Pfalz, S. 35

Die zweite Hausaufgabe bestand in der Bearbeitung eines Arbeitsblatts (Aufgabenblatt 3), das auf die Bildkonstruktion hinführte.

In beiden Gruppen arbeiteten die Schüler mit einer schematischen Repräsentation, in der eine Kerze vor einer Sammellinse dargestellt war. Anhand dieser Repräsentation wurde die physikalische Sehvorstellung thematisiert.

Zum Kern der physikalischen Sehvorstellung zählt, dass beleuchtete Gegenstände durch Streuung Licht abstrahlen (Wiesner, 1994, S. 7) und dieses Licht

ins Auge fällt. Fehlt eine solche physikalische Sehvorstellung, ergeben auch die Strahlenkonstruktionen in der Strahlenoptik wenig oder überhaupt keinen Sinn (Wiesner, 1986, 1992b).

Aus der Fokussierung auf die Konstruktion der ausgezeichneten Strahlen ergibt sich zudem häufig ein im Unterricht selbst erzeugtes Problem. Viele Schüler beachten folgendes nicht: Die ausgezeichneten Strahlen werden zur Bildkonstruktion verwendet; bei dem physikalischen Phänomen der Bildentstehung im konkreten Experiment hingegen trägt alles Licht, das auf die Sammellinse fällt, zur Bildentstehung bei. An dieser Stelle bestehen daher Verständnisschwierigkeiten auf zwei Ebenen:

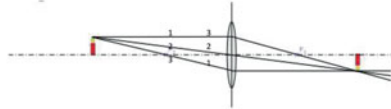
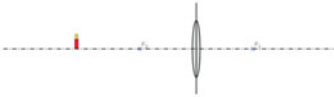
- der Unterscheidung Lichtbündel versus Lichtstrahl und
- einer adäquaten Vorstellung der Lichtausbreitung.

Hieraus können spätere Verständnisschwierigkeiten resultieren, die sich in Schülervorstellungen äußern, wie etwa in der Vorstellung, wenn man eine Blende vor die Linse hält, werde ein Teil des Bildes abgeschnitten (Wiesner, 1992b ,S. 288). Verfügen die Schüler hingegen über ein adäquates Modell der Lichtausbreitung und kennen den Unterschied zwischen einem Lichtbündel und dem idealisierten Modell eines Lichtstrahls, könnte diesen Verständnisschwierigkeiten möglicherweise vorgebeugt werden. Die im Folgenden entwickelte Instruktion sollte auf ein solches adäquates mentales Modell hinführen.

1. Die Treatmentgruppe wurde in einem ersten Schritt aufgefordert, vom obersten Punkt der Kerze Strahlen in all diejenigen Richtungen einzuzeichnen, in denen die Kerze gemäß ihrer Vorstellung Licht abstrahlt.
2. Im zweiten Schritt sollten die Schüler die ausgezeichneten Strahlen, Brennpunktstrahl, Mittelpunktstrahl und Parallelstrahl, einzeichnen. Die Durchführung beider Schritte in der gegebenen Reihenfolge sollte den Schülern helfen, durch eigenes Operieren mit der Repräsentation diesen Zusammenhang zu erkennen.
3. In einem dritten Schritt wurden die Schüler aufgefordert, explizit zeichnerisch zu verdeutlichen, dass auch andere Strahlen außer den ausgezeichneten Strahlen zur Bildentstehung beitragen.

Die Kontrollgruppe erhielt parallel hierzu die gleiche Repräsentation. Anstelle sich ein adäquates Modell der Lichtausbreitung durch das Operieren mit der schematischen Repräsentation der Kerze vor der Sammellinse zu erarbeiten, wurde die Kontrollgruppe direkt durch einen kurzen Text instruiert (vgl. Abbildung 22).

- b) Zeichne von dem obersten Punkt der Kerze Strahlen in alle Richtungen in die Abbildung unten ein, von denen du glaubst, dass es Strahlen gibt.
- c) Es gibt einige Strahlen, die auf die Sammellinse fallen und deren Verlauf du leicht angeben kannst. Zeichne diese Strahlen in die Abbildung unten ein und benenne sie! Welche Lichtstrahlen tragen ebenfalls zur Bildentstehung bei? Zeichne ein.



Der oberste Punkt einer Kerzenflamme sendet, wie jeder Punkt der Flamme, Licht nach allen Seiten aus.
 Ein Teil des Lichts fällt auf die Sammellinse.
 Alles Licht, das von der Kerze auf die Sammellinse fällt, trägt zur Bildentstehung bei. Von bestimmten Strahlen kannst du den Verlauf.

b) Benenne dies Strahlen (siehe Abbildung oben)!

1 _____
 2 _____
 3 _____

Neben achsenparallel einfallenden Lichtbündeln und Lichtbündeln durch den optischen Mittelpunkt, trägt auch alles andere Licht, das auf die Sammellinse trifft, zur Bildentstehung bei.

Abbildung 22: Aufgabenblatt 3: Bildkonstruktion für die Sammellinse [Mat.6_K1] bzw. [Mat. 6_K2] links Aufgaben der TG, rechts Aufgaben der KG 1

Auch in dieser Gruppe wurde den entsprechenden Verständnisschwierigkeiten also vorgebeugt.

2.3.3.5 Operationalisierung des Treatments in der dritten Unterrichtsstunde

Basierend auf der Hinführung auf die Strahlenkonstruktion durch die Hausaufgabe und die Einführung der ausgezeichneten Strahlen in Aufgabenblatt 2, lernten die Schüler in der dritten Stunde unterschiedliche Bildfälle (vergrößerte, verkleinerte und gleich große Bilder) zu konstruieren, die den Beobachtungen im Schülerexperiment entsprachen²¹ (siehe Aufgabenblatt 4: Die Konstruktion der Bildfälle [Mat. 7_K1] bzw. [Mat.7_K2] in Anhang 6 auf der Produktseite zu diesem Buch unter www.springer.com).

In der Treatmentgruppe setzten sich die Schüler in einem dreischrittigen Verfahren mit der Konstruktion auseinander. Dabei erhielten die Schüler ein repräsentationales „Fading-out“²² mit dem Ziel, die Schüler sowohl beim Erstellen

21 Die Schüler verwendeten für die Strahlenkonstruktionen die Hauptebene der Linsen. Der tatsächliche Strahlenverlauf infolge der Brechung an den beiden Außenflächen der Linse wurde aus Zeitgründen nicht thematisiert und ist auch nicht im Lehrplan des Landes Rheinland-Pfalz für das Fach Physik vorgesehen (vgl. Ministerium für Bildung, Wissenschaft, Jugend und Kultur. Rheinland-Pfalz. Lehrplan-Entwürfe Lernbereich Naturwissenschaften Biologie Physik Chemie).

22 Beim Fading-out (Englisch „fade out“ = abblenden, ausblenden) werden anfänglich gegebene Hilfestellungen allmählich verringert (Prinzip abnehmender Lernhilfen), auf diese Weise wird den wachsenden Fähigkeiten der Lernenden Rechnung getragen und eigenständiges Arbeiten ermöglicht und eingefordert (vgl. Reinmann-Rothmeier & Mandl, 1994, S. 632; Schnotz, 2006b, S. 126).

der Konstruktion als auch beim Verknüpfen der Konstruktion (schematische Repräsentation) mit dem Merksatz zu unterstützen. Der Merksatz, der als deskriptiv verbale Repräsentation einzuordnen ist, repräsentiert den Zusammenhang zwischen Bildweite, Gegenstandsweite, Brennweite und Bildgröße in einer verallgemeinernden und abstrakten Form.

1. In einem ersten Schritt erhielten die Schüler eine Repräsentation, die analog zur schematischen Repräsentation in Aufgabenblatt 3 war. Sie wurden aufgefordert, das Bild der dargestellten Kerze vor einer Sammellinse zu konstruieren und zu benennen, um welchen Fall es sich handelt (vergrößertes, verkleinertes oder gleich großes Bild). Zu Zwecken der Datenerhebung wurde das erste Arbeitsblatt nach einer vorgegeben Bearbeitungszeit von 5 min. eingesammelt.
2. In einem zweiten Schritt erhielten die Schüler eine Teillösung der Strahlenkonstruktion und sollten, die gefragten Größen, Gegenstands- und Bildweite sowie Gegenstands- und Bildgröße, abmessen und den Merksatz ergänzen. Die Ergänzung des Merksatzes erforderte von den Schülern, die schematisch-bildhafte Repräsentation der Strahlenkonstruktion im Hinblick auf die allgemeinen Verhältnisse von Bildweite, Gegenstandsweite, Brennweite und Bildgröße zueinander zu abstrahieren und in eine verbale Repräsentation zu übersetzen. Auch das zweite Arbeitsblatt wurde nach 5 Minuten eingesammelt.
3. In einem dritten Schritt bekamen die Schüler zur Ergebnissicherung für ihre persönlichen Unterlagen eine vollständige Lösung der Aufgaben mit einem Lösungsblatt, welches die Konstruktion, die Messergebnisse der Zeichnung und den vollständigen Merksatz enthielt.

Das Fading-out fand zeitlich auf zwei Ebenen statt:

- Zum einen erhielten die Schüler mit der Teillösung eine Rückmeldung, in der sie abgleichen konnten, inwiefern sie das erste Aufgabenblatt korrekt bearbeitet hatten. Schüler, denen die Strahlenkonstruktion Probleme bereitet hatte, erhielten durch die Teillösung sowohl eine Hilfestellung bei der Konstruktion als auch die Chance, die Konstruktion im zweiten Anlauf zu vervollständigen und auf Basis der Lösung die zweite Aufgabe, den Merksatz, zu ergänzen.
- Zum anderen wurden die Teillösungen des zweiten Aufgabenblatts zunehmend unvollständiger sowie Aufgaben und die Ergänzung des Merksatzes von Mal zu Mal offener gestaltet: So war im ersten Durchgang bei dem gleichgroßen Bild der Parallelstrahl vollständig und Brennpunktstrahl sowie Mittelpunktstrahl bis zur Hauptebene (Linsenmitte) konstruiert. Im Merksatz musste

lediglich der Zusammenhang hergestellt werden, dass Gegenstandsweite und Bildweite einander und beide jeweils der doppelten Brennweite entsprechen. Um das Erkennen dieses Zusammenhangs aus der Konstruktion (depiktionale Repräsentation) zu erleichtern, war die doppelte Brennweite eingezeichnet und benannt.

- Im zweiten Durchgang (verkleinertes Bild) waren in der Teillösung - die ausgezeichneten Strahlen - nur bis zur Hauptebene konstruiert. Im Merksatz mussten neben den gefragten Größen auch Texte ergänzt werden.
- Im dritten Durchgang (vergrößertes Bild) waren in der in der Teillösung die ausgezeichneten Strahlen ebenfalls nur bis zur Hauptebene konstruiert. Den Merksatz sollten die Schüler entsprechend der Lösungsmodelle, die sie zuvor erhalten hatten, selbst formulieren.

Die Treatmentgruppe erhielt parallel hierzu einen ausführlichen Lehrervortrag, in dem die Strahlenkonstruktion und die Merksätze erläutert wurden. Die Aufgabe der Schüler bestand darin, die dargestellten Zusammenhänge zu verstehen und die gefragten Größen (Bildweite, Gegenstandsweite, Gegenstandsgröße und Bildgröße) auf den ausgeteilten Arbeitsblättern zu ergänzen.

Prinzipiell arbeiteten also beide Gruppen mit den gleichen Repräsentationen: der Strahlenkonstruktion, der Ergänzung der konkreten Zahlenwerte gefragter physikalischer Größen und dem Merksatz, in dem allgemeinen Zusammenhänge zwischen den involvierten physikalischen Größen abstrakt verbal repräsentiert werden.

Die Treatmentgruppe wurde jedoch kognitiv aktiviert, die entsprechenden Repräsentationen teilweise selbst zu erstellen, Repräsentationen zu ergänzen und eigenständig unterschiedliche Repräsentationen aufeinander zu beziehen bzw. die schematische Repräsentation (über den Zwischenschritt des Abmessens der konkreten Zahlenwerte) in eine abstrakte generalisierende verbale Repräsentation zu überführen.

In der Hausaufgabe wurde die Idee der Vervollständigung eines repräsentationalen „Lückentextes“ aufgegriffen. Während die Treatmentgruppe die Ergebnisse in einer Tabelle über die Bildfälle zusammenfassen sollte, bestand die Aufgabe der Kontrollgruppe darin, die gegebenen Bildfälle mündlich erläutern zu können. Die tabellarische Übersicht war hier als Lernmaterial gedacht.

2.3.3.6 Operationalisierung des Treatments in der vierten Unterrichtsstunde

Thema der vierten Unterrichtsstunde war die Konstruktion des virtuellen Bildes. Mit dem virtuellen Bild wurde die Bildentstehung bei der Sammellinse vervollständigt. Die Konstruktion des virtuellen Bildes wurde nicht im Leistungstest abgefragt und spielt daher für die Operationalisierung des Gesamttreatments im Vergleich zu Aufgaben, die auf die Vertiefung der Konstruktion der reellen Bilder zielen, eine untergeordnete Rolle.

Da es sich bei dem Konzept des virtuellen Bildes um einen relativ anspruchsvollen und abstrakten Lerninhalt handelt, der vielen Schülern Lernschwierigkeiten bereitet (vgl. Wiesner, 1986), wurde sowohl die Kontrollgruppe als auch die Treatmentgruppe schrittweise auf die Konstruktion des virtuellen Bildes hingeführt.

Während der Kontrollgruppe die Konstruktion weitgehend vorgegeben war, sollte die Treatmentgruppe die Strahlenkonstruktion in mehreren Schritten bearbeiten und die gegebenen Konstruktionen ergänzen. Aufgabenblatt 5: Das virtuelle Bild [Mat. 8 _ K1] bzw. [Mat. 8 _ K2]. Auch hier wurde die Treatmentgruppe kognitiv dazu aktiviert, sich intensiver mit der Konstruktion auseinanderzusetzen. Während die Treatmentgruppe unter Berücksichtigung der selbst erstellten Konstruktionen sich die Begriffe „virtuelles Bild“ im Gegensatz zu einem „reellen Bild“ selbst erarbeiten mussten, wurde der Kontrollgruppe dieser Unterschied vom Lehrer erklärt. Die Erklärung der Unterscheidung forderte dabei, Schlüsse aus der unter Anleitung selbst erstellten Konstruktion zu ziehen und diese in Worte zu fassen (also zwei Repräsentationen, die in unterschiedlichen Formaten vorliegen, aufeinander zu beziehen). In der Kontrollgruppe wurden diese Schritte vom Lehrer übernommen. Zum Abschluss des Themas „virtuelle Bilder“ bearbeiteten beide Gruppen eine Aufgabe zu einer schematischen Abbildung des Versuchsaufbaus des virtuellen Bildes. Während die Treatmentgruppe gebeten wurde, einen Versuchsaufbau zum virtuellen Bild mit eigenen Worten zu erklären und die vorliegende Versuchsskizze zu erläutern, sollte die Kontrollgruppe die Versuchsskizze lediglich beschriften.

Nachdem nun alle Fälle der Bildentstehung vollständig thematisiert waren und auch in der tabellarischen Übersicht vervollständigt werden konnten, wurde die übrige Unterrichtszeit genutzt, um die Konstruktion des reellen Bildes zu vertiefen. Die Treatmentgruppe bearbeitete eine Aufgabe, mit dem Ziel, ein internes mentales Modell der Bildentstehung auf Basis der Strahlenkonstruktion zu bilden, das den Abbildungsvorgang in kohärenter Weise intern dynamisch simuliert.

A1

Ein Gegenstand - dargestellt durch einen Pfeil - steht in einem Abstand $g = 2f$ vor einer Sammellinse

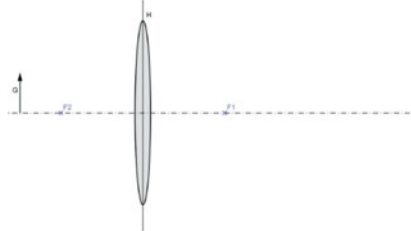


- a) Wie verändern sich Bildgröße und Lage des Bildes, wenn man den Gegenstand auf die Linse zuschiebt, d.h. g nähert sich f ? Gib die ungefähre Lage und Größe des Bildes in der Abbildung an (ohne Konstruktion) und kennzeichne den Vorgang durch Pfeile.
- b) Welche Abbildungseigenschaften liegen vor, wenn g innerhalb der einfachen Brennweite f liegt (siehe Tabelle)?
- c) Wie verändern sich Bildgröße und Lage des Bildes, wenn man den Gegenstand weiter von der Linse weg schiebt ($g > 2f$)? Gib die ungefähre Lage und Größe des Bildes in der Abbildung an (ohne Konstruktion) und kennzeichne den Vorgang durch Pfeile.
- d) Was passiert, wenn man den Gegenstand sehr weit von der Linse entfernt? Welchem Wert nähert sich die Bildweite?

A1

a) 6 cm vor einer Sammellinse steht ein Gegenstand mit der Größe von 2 cm. Die Linse hat eine Brennweite von 4 cm. Bestimme durch die Zeichnung Ort und Größe des Bildes (Abb. 2).

b) Nadine möchte den Brennpunkt ihrer Lupe (Sammellinse) bestimmen.



Ein brennendes Teelicht (Gegenstandshöhe $G = 2$ cm) befindet sich 10 cm vor der Linse (Abb. 1). Mit einer weißen Pappe kann sie das Bild 7 cm hinter der Linse auffangen. Es hat eine Bildgröße von 1,4 cm.

Bestimme den Brennpunkt mit einer Strahlenkonstruktion!

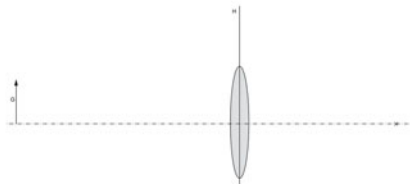


Abbildung 23: Aufgabenblatt 6: Übungen zur Bildkonstruktion [Mat. 9_K1] bzw. [Mat. 9_K2] links Aufgaben der TG, rechts Aufgabe der KG

Hierzu operierte die Treatmentgruppe wie zuvor mit der schematischen Darstellung eines leuchtenden Gegenstands, der im Abstand $g = 2f$ vor einer Sammellinse steht (vgl. Abbildung 23). Die Schüler wurden nun kognitiv aktiviert, sich vorzustellen, was mit der Größe und Lage des Bildes passiert, wenn man einen Gegenstand auf die Linse zuschiebt bzw. weiter von der Linse wegschiebt. Das Ergebnis dieser Überlegungen sollten die Schüler jeweils in der gegebenen Abbildung durch eine Skizze darstellen (siehe Abbildung 23).

Schließlich wurden die Schüler gebeten, sich den extremen Fall vorzustellen, in dem der Gegenstand sehr weit von der Linse entfernt wurde und Konsequenzen für die Bildweite aufzuzeigen.

Die Kontrollgruppe bearbeitete parallel dazu an zwei Strahlenkonstruktionen. In der ersten Aufgabe konstruierte die Kontrollgruppe das Bild eines vergrößerten selbstleuchtenden Gegenstands; die zweiten Aufgabe enthielt ein Transferproblem: hier waren die Werte der physikalischen Größen des Bildes gegeben. Mit Hilfe der Strahlenkonstruktion sollte der Brennpunkt bestimmt werden (vgl. Abbildung 23).

Auch die Kontrollgruppe operierte also mit der schematischen Repräsentation der Strahlenkonstruktion und löste hierzu nicht nur Standardfragestellungen, so dass der Unterschied in den Instruktionen der beiden Bedingungen nicht auf die Bearbeitung von Standardaufgaben versus Problemlöseaufgaben reduziert werden kann. Im Gegensatz zur Treatmentgruppe wurde die Kontrollgruppe jedoch nicht kognitiv aktiviert, die Bildentstehung durch das Operieren mit der Strahlenkonstruktion intern mental zu simulieren.

2.3.3.7 Treatmentvariation in der vierten Unterrichtsstunde

Die Treatmentvariation zielte darauf, die weitverbreitete Schülervorstellung zu überwinden, dass bei der Entstehung reeller Bilder das Bild abgeschnitten werde, wenn man eine Blende vor eine Sammellinse halte. Zur Förderung eines angemessenen Verständnisses dieses physikalischen Settings wurde folgende POE-Sequenz umgesetzt, bei der sich die Schüler mit multiplen Repräsentationen auseinandersetzten (zum Unterrichtsmaterial der Treatmentvariation sei auf Anhang C3 und zur Unterrichtsplanung der Treatmentvariation auf Anhang C4 jeweils auf der Produktseite zu diesem Buch unter www.springer.com hingewiesen).

Die Schüler wurden in der variierten Treatmentbedingung kognitiv aktiviert, sich mit dem konkreten Versuch und unterschiedlichen Repräsentationen des physikalischen Phänomens, das in dem Versuch beobachtet werden kann, auseinander zu setzen (Fotografie des Versuchsaufbaus, verbale Erklärungen des Phänomens, schematische Repräsentation des Strahlengangs).

Im ersten Schritt (predict) sollten die Schüler vorhersagen, was ein Beobachter auf dem Schirm sieht, wenn man die Kerze anzündet und entsprechend der Abbildungen die obere Hälfte der Sammellinse abdeckt. Der fertige Versuchsaufbau wurde den Schülern vom Lehrer gezeigt. Unterstützend erhielten die Schüler zur Bearbeitung ihrer Vorhersage auf einem Arbeitsblatt folgende Fotografien des Versuchsaufbaus aus unterschiedlichen Perspektiven (vgl. Abbildung 24):

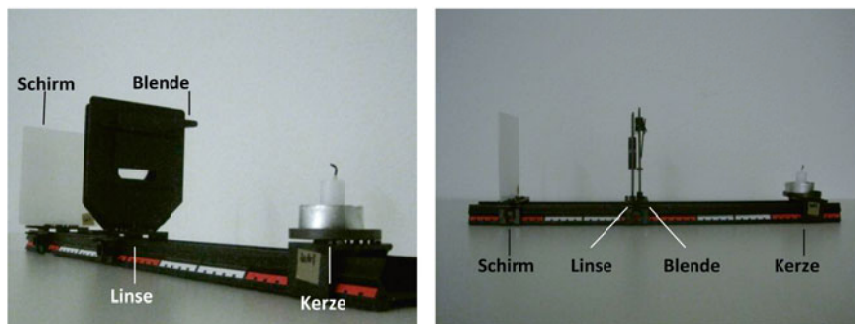


Abbildung 24: Demonstrationsversuch „Abgedeckte Sammellinse“ [Zusatz-1-K1]

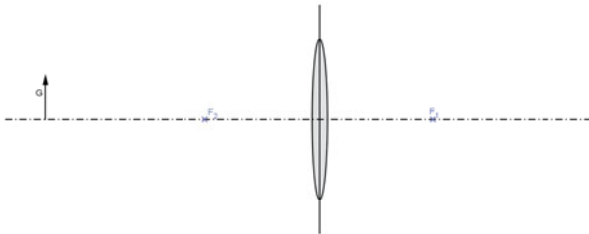
Ihre Vermutungen sollten die Schüler bezogen auf die Abbildungen schriftlich auf dem Arbeitsblatt formulieren. Im nächsten Schritt (observe) wurde den Schülern der Versuch demonstriert (vgl. Abbildung 25). Ihre Beobachtungen wurden in der Klasse diskutiert (observe).



Abbildung 25: Ausgang des Demonstrationsversuchs „Abgedeckte Sammellinse“

Im dritten Schritt (explain) bearbeiteten die Schüler ein Arbeitsblatt (vgl. Abbildung 26), in dem die Erklärung des Phänomens in einer schematischen Strahlenkonstruktion dargestellt wurde. Die grafische Lösung ist in Abbildung 27 dargestellt.

Deutung des Versuchsergebnisses



- Konstruiere das Bild der Kerze (Abb. oben).
- Zeichne den Lichtstrahl ein, der vom obersten Punkt des Gegenstands ausgeht und gerade noch den oberen Rand der Linse erreicht.
- Wie ändert sich das Bild, wenn man nur den Bereich des oberen Randes der Linse abdeckt?

- Warum sieht man ein vollständiges Bild, wenn man die komplette obere Hälfte des Bilde abgedeckt hat (siehe Demonstrationsversuch)? Stelle grafisch dar, was passiert?
- Grenze durch Einzeichnen ein, welche Lichtstrahlen nach der Abdeckung der Sammellinse noch zur Bildentstehung beitragen.

Abbildung 26: Demonstrationsversuch „Abgedeckte Sammellinse“ [Zusatz-2-K1]

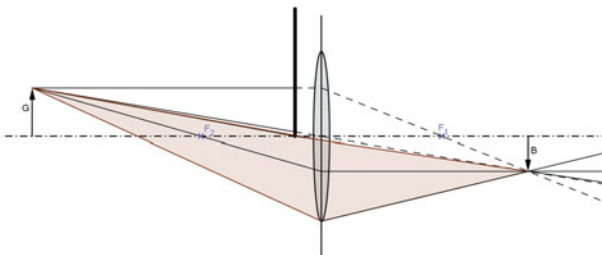


Abbildung 27: Konstruktion des reellen Bildes einer Kerze, welches durch eine Sammellinse entsteht, wobei die obere Hälfte der Linse abgedeckt wurde. Die hervorgehobene Fläche zeigt den Bereich des Lichts, welcher nach Abdeckung der Linse zur Entstehung des reellen Bildes beiträgt.

Die zusätzlich für die Durchführung der POE-Sequenz in der Treatmentgruppe aufgewendete Zeit wurde bei der Erarbeitung der Entstehung virtueller Bilder eingespart. Die Entstehung virtueller Bilder war kein Gegenstand des Leistungstests. Die Kontrollgruppe setzte sich entsprechend wie in Abschnitt 2.3.3.6 Operationalisierung des Treatments in der vierten Unterrichtsstunde dargestellt, ausführlicher mit der Konstruktion des virtuellen Bildes auseinander.

2.3.3.8 Operationalisierung des Treatments in der fünften Unterrichtsstunde

Bevor sich die Schüler mit dem Abbildungsgesetz befassten, bearbeiteten sie ein weiteres Arbeitsblatt zur Vertiefung der Strahlenkonstruktion, in dem auch weitverbreitete Schülervorstellungen thematisiert wurden. Dieses Arbeitsblatt (Arbeitsblatt 7) wurde zum Zweck der Datenerhebung in beiden Gruppen eingesammelt. Zur Ergebnissicherung erhielten die Schüler die Lösung der Aufgaben für ihre Lernunterlagen.

In beiden Gruppen befassten sich die Schüler mit der Strahlenkonstruktion unter den Bedingungen, dass die Linse erstens kleiner als der leuchtende Gegenstand ist und zweitens die obere Hälfte der Linse abgedeckt wird. Die hier adressierten Lernschwierigkeiten betreffen die physikalischen Vorstellungen des Abbildungsvorgangs. Viele Schüler nutzen zur Erklärung der Entstehung des reellen Bildes bei der Sammellinse nicht das Konzept einer „Punkt-zu-Punkt-Abbildung“ (vgl. Wiesner, 1994). Eine der gängigen Vorstellung besteht in diesem Kontext in einer holistischen Erklärung des Abbildungsvorgangs, nach der das Bild als Ganzes durch die Linse zum Schirm geht und dabei in der Linse umgedreht wird (Wiesner, 1994, S. 8). Eine solche holistische Erklärung des Abbildungsvorgangs zeigt sich insbesondere bei den hier verwendeten Abdeckaufgaben.

Unter der Annahme, das Bild werde als Ganzes vom Gegenstand aus durch die Linse auf den Schirm transportiert, ist es nur konsequent anzunehmen, dass ein Teil des Bildes abgeschnitten werde, wenn man eine Blende vor die Linse hält, (Wiesner, 1992b). Hält man eine ringförmige Blende vor die Linse, glauben viele Lernende entsprechend, das Bild werde ringförmig am äußeren Rand abgeschnitten. Wird die Linse zur Hälfte abgedeckt, gehen viele Schüler und auch Studenten davon aus, dass auch das reelle Bild zur Hälfte abgeschnitten werde, einige überlegen sich sogar, welche Hälfte des Bildes (obere versus untere Hälfte) betroffen sei (Goldberg & McDermott, 1987, S. 112; Wiesner, 1994, S. 8). Schüler, die bereits an dieser Stelle über ein adäquates mentales Modell der Lichtausbreitung und der Bildentstehung verfügen, sollten bereits hier erkennen, dass auch

in diesem Fall ein vollständiges Bild entsteht, das in seiner Intensität jedoch abgeschwächt ist. Das Problem in der repräsentationalen Darstellung besteht darin, dass – sowohl im Fall einer kleineren Linse als auch im Fall der abgedeckten Linse – Brennpunktstrahl und Parallelstrahl nicht mehr auf die Linse treffen. Die Lösung besteht darin, zunächst die Hauptebene der Linse zur Strahlenkonstruktion zu verwenden, um herauszufinden, wo das Bild entsteht. Anschließend können diejenigen Strahlen eingezeichnet werden, die auf den oberen und unteren Rand der Linse treffen. Auf diese Weise kann der Bereich angegeben werden, in dem das Licht, welches von dem leuchtenden Gegenstand ausgeht, zur Entstehung des reellen Bildes beiträgt.

Die Treatmentgruppe bearbeitete beide Aufgaben (Bildkonstruktion, wenn Linse kleiner als der leuchtender Gegenstand; Bildkonstruktion, wenn die oberen Linsenhälfte abgedeckt ist) in zwei Schritten (siehe Abbildungen 28 und 29).

- Im ersten Schritt bestand die Aufgabe darin, das Bild unter den gegebenen Bedingungen zu konstruieren. Diese Aufgabenstellung bezweckte, die Schüler ihre Vorstellung des Abbildungsvorgangs in einer externen Repräsentation reflektieren zu lassen. Da die Arbeitsblätter der Schüler eingesammelt wurden, ergibt sich die Möglichkeit, den Lernstand der Schüler auszuwerten, um zu sehen, ob sich die Instruktionen als fruchtbar erweisen, welche diesen Lernschwierigkeiten vorbeugen sollen. So war etwa in Aufgabenblatt 3 bereits durch das Arbeiten mit der Strahlenkonstruktion thematisiert worden, dass alles Licht zur Bildentstehung beiträgt, das auf die Sammellinse fällt.
- In einem zweiten Schritt erhielten die Schüler eine graphische Teillösung. Sie wurden nun aufgefordert, die Teillösung zu ergänzen und anhand der Abbildung zu erläutern, warum es möglich ist, auch dann noch ein Bild zu konstruieren, wenn der Gegenstand kleiner als die Linse ist bzw. wenn die obere Hälfte der Linse abgedeckt worden war (siehe Abbildung 28 und Abbildung 29).

Die Schüler sollten in mehreren Hinsichten durch diese Aufgabenstellung kognitiv aktiviert werden:

- Durch den Abgleich der eigenen Vorstellung mit der Teillösung sollten die Schüler ihre Sichtweise reflektieren. Die Aufgabe zielte darauf, ihre Lösung des physikalischen Problems zu prüfen und die eigenen Verarbeitungsschritte entsprechend anzupassen.
- Die Vervollständigung der Strahlenkonstruktion war als repräsentationales Fading-out gedacht, bei dem die Schüler erneut die Bildkonstruktion durch-

denken sollten, sofern sie mit der Lösung der Aufgabe Schwierigkeiten hatten.

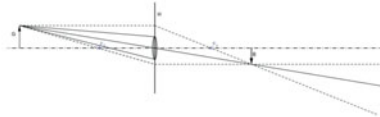
- Die schriftliche Erklärung, warum es möglich ist, ein Bild unter den gegebenen Bedingungen (Linse < Gegenstand) zu konstruieren, erforderte das Erstellen von zwei Repräsentationen in unterschiedlichen Formaten (Bild und Text), die aufeinander bezogen werden mussten. Bei der Abdeckaufgabe wurde zudem gefragt, wie sich das Abdecken auf das Bild auswirkt. Hier galt es aus der Konstruktion und der Beschreibung eine Schlussfolgerung über Auswirkungen auf das reelle Bild (Ebene des Phänomens) zu beziehen.

A2
 Kannst du auch ein Bild konstruieren, wenn man anstelle einer großen Sammellinse eine deutlich kleinere Linse mit gleicher Brennweite verwendet (siehe Abbildung unten)? Verdeutliche deine Überlegungen mit einer Zeichnung.



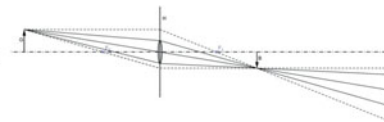
Seite 2

b) Vervollständige die Konstruktion.



Erläutere anhand Abbildung:

A2
 Warum kannst du auch ein Bild konstruieren, wenn man anstelle einer großen Sammellinse eine deutlich kleinere Linse mit gleicher Brennweite verwendet (Abb. 3)? Begründe:



Die Konstruktionsmethode führt also auch dann noch zum Ziel, wenn der Gegenstand, der abgebildet werden soll, viel größer ist als die Linse.

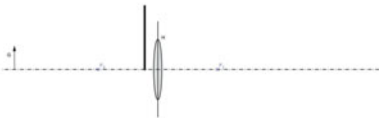
Abbildung 28: Aufgabenblatt 7 Wiederholung zur Bildkonstruktion [Mat. 10_K1], A2 bzw. [Mat 10_K2], A2 links Aufgabe der TG, rechts Aufgabe der KG

In der Kontrollgruppe wurde die gleiche Schülervorstellung adressiert.

Im Gegensatz zur Treatmentgruppe erhielt die Kontrollgruppe die Abbildung inklusive vollständiger Strahlenkonstruktion und sollte die Frage beantworten, warum es jeweils möglich ist, unter den gegebenen Bedingungen (Linse < Gegenstand; Abdeckung der oberen Linsenhälfte), das Bild zu konstruieren (vgl. Abbildungen 29). Im Gegensatz zur Treatmentgruppe wurde die Kontrollgruppe nicht aufgefordert, mit der Konstruktion selbst zu operieren. Die Aufgabe, beide Repräsentationen aufeinander zu beziehen, bestand zwar auch in der Kontrollgruppe; die zentrale graphische Repräsentation war der Kontrollgruppe jedoch vorgegeben. Somit wurde die Kontrollgruppe nicht aktiviert, ihre eigenen Vorstellungen extern darzustellen, diese auf ihre Zweckmäßigkeit hin zu prüfen oder die eigenen Verarbeitungsschritte entsprechend anzupassen.

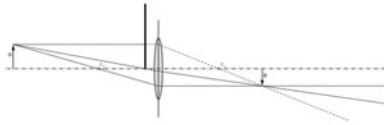
A3

a) Kannst Du noch das Bild einer Kerze (hier als Pfeil dargestellt) konstruieren, wenn man entsprechend der Abbildung die obere Hälfte der Sammellinse abdeckt? Verdeutliche Deine Überlegungen mit einer Zeichnung (Abb. 5).

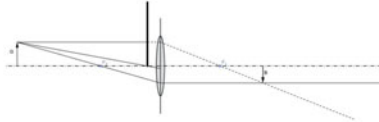


A3

Was passiert mit dem Bild auf dem Schirm, wenn man entsprechend der Abbildung die obere Hälfte der Sammellinse abdeckt? (Abb. 4)?



b) Vervollständige die Konstruktion.



Was passiert mit dem Bild, wenn man wenn man entsprechend der Abbildung die obere Hälfte der Sammellinse abdeckt? Erläutere anhand der Abbildung oben:

Abbildung 29: Aufgabenblatt 7: Wiederholung zur Bildkonstruktion [Mat_10K1], A3 bzw. [Mat. 10_K2] A3, links Aufgabe TG, rechts Aufgabe der KG

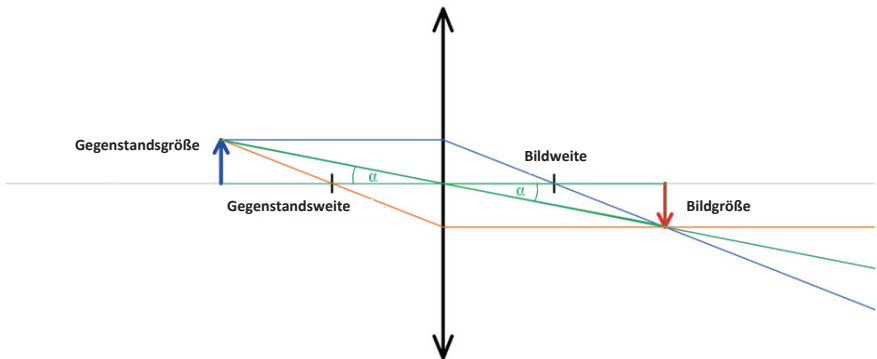


Abbildung 30: Aufgabenblatt 8: Die Abbildungsgleichung bei der Sammellinse [Mat.12_K1] bzw. [Mat. 12_K2] geometrische Veranschaulichung zur Herleitung der Abbildungsgleichung

Im zweiten Teil der Stunde wurde die Abbildungsgleichung eingeführt und verbal sowie geometrisch veranschaulicht (vgl. Abbildung 30). Das hier eingesetzte Arbeitsblatt war zuvor in der Pilotstudie von Scheid (2013) verwendet worden. Schülervorstellungen wurden in diesem Kontext nicht thematisiert.

Betrachtet man die beiden Dreiecke, welche die Gegenstandsgröße bzw. Bildgröße, die Gegenstandsweite bzw. die Bildweite und den Mittelpunktstrahl

als Seiten haben, können hieraus unter Anwendung des geometrischen Satzes zur Ähnlichkeit von Dreiecken eine allgemeine Formel über die Seitenverhältnisse der Dreiecke abgeleitet werden. Da die Ähnlichkeit von Dreiecken jedoch gemäß Lehrplan des Landes Rheinland-Pfalz im Fach Mathematik nicht zuvor im Mathematikunterricht behandelt worden war (vgl. Ministerium für Bildung, Wissenschaft, Jugend und Kultur. Rheinland-Pfalz. Rahmenlehrplan Mathematik, 2007), wurden die Schüler in beiden Gruppen schrittweise zu diesem Zusammenhang hingeführt.

Die Treatmentgruppe erhielt hierzu mehrere aufeinander aufbauende Aufgaben:

- In einem ersten Schritt sollten die Schüler, wie zuvor geübt, den Strahlengang in der gegebenen Abbildung durch das Einzeichnen des Mittelpunktstrahls ergänzen. In der folgenden Abbildung waren alle irrelevanten Informationen ausgeblendet, so dass nur die für die Herleitung der Gleichung relevanten Dreiecke zu sehen sind. Diese sollten die Schüler mit den entsprechenden physikalischen Größen korrekt beschriften und sich Gedanken über die Winkelverhältnisse der Dreiecke machen.
- Im nächsten Schritt wurden die Schüler aufgefordert sich zu veranschaulichen, dass beide Dreiecke in ihren Seitenverhältnissen übereinstimmen; ihre Schlussfolgerung über die Seitenverhältnisse sollten sie in einem kurzen Satz schriftlich festhalten.
- In den letzten Schritten wurde von den Schülern gefordert, diesen Zusammenhang über den Zwischenschritt einer Wortgleichung als Formel auszudrücken.

Der Kontrollgruppe wurde parallel dazu der gleiche Zusammenhang unter Verwendung der gleichen Repräsentationen in einem Lehrervortrag erläutert. Sie erhielt somit die gleichen Informationen, jedoch ohne gezielte kognitive Aktivierung zu den jeweiligen Bearbeitungsschritten.

Folgende Aspekte der Aufgabenstellung zielten auf eine kognitive Aktivierung beim Operieren mit den Repräsentationen in der Treatmentgruppe im Vergleich zur Kontrollgruppe:

- Die Treatmentgruppe war gefordert, die relevanten Informationen selbst in den gegebenen graphischen Repräsentationen zu ergänzen.
- Die Aufgabenstellung, sich die Übereinstimmung der Seitenverhältnissen mental zu veranschaulichen und die Schlussfolgerung zu verbalisieren, zielte auf die Bildung eines mentalen Modells des mathematischen Zusammenhangs;

aus diesem mentalen Modell können die relevanten Schlussfolgerung im Idealfall abgelesen werden.

- Durch das schriftliche Festhalten der Schlussfolgerung und die Übersetzung der Schlussfolgerung in die Sprache abstrakter Formeln sollten die Schüler lernen, die Schlussfolgerungen aus der grafischen Repräsentation sowohl in eine sprachliche als auch in eine mathematische Repräsentation zu übersetzen.

Im Anschluss wandten beide Gruppen die Abbildungsgleichung in einer Übungsaufgabe an. Die Aufgabenformulierung und der Rechenaufwand waren in beiden Gruppen parallel gehalten. Die Treatmentgruppe wurde jedoch aufgefordert, das Ergebnis ihrer Rechnung anhand einer gegebenen Strahlenkonstruktion, die im Originalmaßstab ausgeteilt wurde, zu beziehen, um die Verbindung zwischen Abbildungsgleichung und Strahlenkonstruktion zu festigen.

2.3.3.9 Operationalisierung des Treatments in der sechsten Unterrichtsstunde

Die letzte Unterrichtsstunde steuerte darauf hin, die gelernten Inhalte durch Übungen zur Bildkonstruktion und zum Abbildungsgesetz zu vertiefen (Aufgabenblatt 9). Hierzu bearbeiteten die Schüler zwei Aufgaben mit Teilfragestellungen. In der ersten Aufgabe wurde der Treatment- und auch der Kontrollgruppe die graphische Repräsentation einer Kerze (dargestellt durch einen Pfeil) vor einer Sammellinse präsentiert. Beide Gruppen sollten sich überlegen, ob das Bild der Kerze auch bei einer Versuchsanordnung ohne Schirm gesehen werden kann. Die hier adressierte Schülervorstellung wurde in der (bereits in vorigen Kapiteln erläuterten) Studie von Goldberg und McDermott (1987) beschrieben.

Zur Erinnerung: Goldberg und McDermott (1987) zeigten Studenten einen Versuchsaufbau, bei dem eine Glühbirne, eine Linse und ein Schirm hintereinander auf einer optischen Bank montiert waren. Die Studenten wurden gefragt, wo das Bild wäre, wenn man den Schirm entfernt und sie frei um den Versuchsaufbau im Raum herumgehen könnten. Nur wenige Studenten waren in der Lage zu erkennen, dass sich das Bild an der gleichen Position befindet wie der Schirm. Die übrigen Studenten gaben unter anderem die Erklärungen ab, das Bild sei auf oder in der Linse. Verbreitet war auch die Vorstellung, dass ein Bild nur mit Hilfe eines Schirms gesehen werden kann, wobei die Linse das Bild quasi einrahme (vgl. ebd., S. 114).

Die Herausforderung bei dieser Aufgabe besteht darin, zu erkennen, dass sich die Lichtstrahlen, die ein reelles Bild formen, im freien Raum treffen. Unter

der Voraussetzung, dass die Schüler ein kohärentes mentales Modell des Abbildungsvorgangs gebildet haben, sollten sie in der Lage sein, auch in dieser Aufgabe die korrekte Schlussfolgerung zu ziehen. Die Aufgabenstellung in Treatment- und Kontrollgruppe unterschied sich nun wie folgt (vgl. Abbildung 31): In der schematischen Repräsentation des Strahlengangs, welcher der Treatmentgruppe vorgelegt wurde, waren zusätzlich drei Beobachter an unterschiedlichen Positionen platziert.

Die Treatmentgruppe wurde gefragt, an welcher Stelle der Beobachter A, B, oder C ein scharfes Bild einer Kerze sehen kann, wenn man ...

1. einen intransparenten Schirm an der Position S aufstellt,
2. den intransparenten Schirm gegen einen transparenten Schirm austauscht oder
3. den Schirm entfernt.

Die Schüler wurden gebeten, ihre Antworten zu begründen. Durch die Fragestellung wurde die Treatmentgruppe kognitiv aktiviert, sich mental in die Position der verschiedenen Beobachter hineinzusetzen. Die Überlegungen, die zur korrekten Lösung führen, können zudem grafisch durch eine entsprechende Skizze veranschaulicht werden.

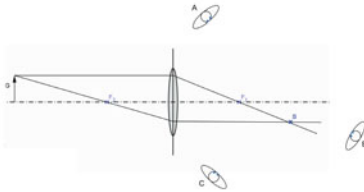
Im Fall des intransparenten Schirms können nur die Beobachter A und B das Bild sehen, bei Verwendung eines transparenten Schirms sehen alle Beobachter das Bild, da das Licht auf dem Schirm in alle Richtungen streut und im letzten Fall (kein Schirm) kann nur Beobachter A das Bild sehen, da nur er im Strahlengang steht. Mit den ersten beiden Fällen (intransparenter und transparenter Schirm) wurde auch implizit nochmals das Konzept der Streuung und die physikalische Sehtheorie (vgl. Wiesner, 1986) thematisiert, weil die Schüler erkennen mussten, dass beleuchtete Gegenstände (hier der Schirm) Licht abstrahlen und dieses Licht in die Augen des Beobachters fallen muss, wenn der Beobachter das Bild sehen kann.

Die Kontrollgruppe wurde im Gegensatz zur Treatmentgruppe direkt gefragt, ob es möglich sei, das Bild der Kerze bei einer Versuchsanordnung ohne Schirm zu sehen und ihre Antwort zu begründen (vgl. Abbildung 8). Die Kontrollgruppe erhielt über die Adressierung der Schülervorstellung hinaus, die an sich kognitiv anspruchsvoll ist, also keine zusätzliche kognitive Aktivierung, mit der gegebenen schematischen Repräsentation der Strahlenkonstruktion mental und zeichnerisch zu operieren.

A1

Begründe jeweils deine Antwort:

- Welcher Beobachter (A, B, C) kann das Bild der Kerze sehen, wenn Max einen undurchsichtigen Schirm, z.B. ein weißes Stück Pappe, an der Position S aufstellt?
- Welcher Beobachter (A, B, C) kann das Bild der Kerze sehen, wenn Max die Pappe gegen einen transparenten weißen Schirm vertauscht?
- Welcher Beobachter (A, B, C) kann das Bild der Kerze bei einer Versuchsanordnung ohne Schirm sehen?

**A1**

a) Wo entsteht das Bild der Kerze?

- b) Ist es möglich, das Bild der Kerze (hier dargestellt durch einen Pfeil) bei einer Versuchsanordnung ohne Schirm zu sehen?

Wenn nein, warum nicht?

Wenn ja, warum?

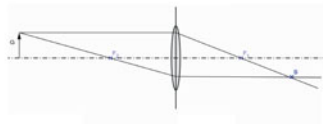


Abbildung 31: Aufgabenblatt 9 Übungen zur Abbildungsgleichung und zur Bildkonstruktion [Mat. 13_K1], A1 bzw. [Mat. 13_K2], A1 links Aufgabe der TG, rechts Aufgabe der KG

Die letzte Übungsaufgabe der Unterrichtseinheit war auf eine Vertiefung der Strahlenkonstruktion angelegt.

Beide Gruppen erhielten zunächst die gleiche graphische Repräsentation, bei der eine Kerze (dargestellt durch einen Pfeil) vor einer Sammellinse steht. Beide Gruppen sollten eine Strahlenkonstruktion durchführen, die von der üblichen Vorgehensweise abweicht.

In der Treatmentgruppe waren hierzu zwei Punkte P und Q auf der Hauptebene (Linsenmitte) angegeben. Die Aufgabe bestand darin, jeweils einen Strahl zu konstruieren, der von der Kerzenspitze ausgeht und in dem Punkt P bzw. in dem Q auf der Linsenmitte auftrifft (vgl. Abbildung 32). Die Lösung der Aufgabe umfasst zwei Bearbeitungsschritte:

- Im ersten Schritt wird wie üblich das reelle Bild unter Verwendung der ausgezeichneten Strahlen eingezeichnet.
- In einem zweiten Schritt kann dann ausgehend von der Konstruktion des Bildes der Verlauf der Strahlen angegeben werden, die am Punkt P bzw. Q auf der Linsenmitte auftreffen und ebenfalls zur Entstehung des Bildes beitragen.

Diese Aufgabe zielte darauf, die Schüler kognitiv zu aktivieren, die Strahlenkonstruktion erneut zu durchdenken und sich zu überlegen, dass außer den ausgezeichneten Strahlen noch mehr Licht zur Bildentstehung beiträgt, was sich wiederum zeichnerisch darstellen lässt.

Die Kontrollgruppe konstruierte parallel hierzu zwei Gegenstandspunkte P und Q, die in der gegebenen Zeichnung eingetragen waren (vgl. Abbildung 32). Sie wurde jedoch nicht kognitiv aktiviert, die Strahlenkonstruktion unter der Perspektive zu durchdenken, dass außer den ausgezeichneten Strahlen auch andere Lichtbündel im zeichnerischen Modell dargestellt werden können.

A2

Konstruiere das Bild der Kerze (hier dargestellt durch einen Pfeil) und bestimme die Bildgröße durch Konstruktion (Abb. 2)!
Hinweis: Die Linse ist durch die Linsenmitte (H) dargestellt.

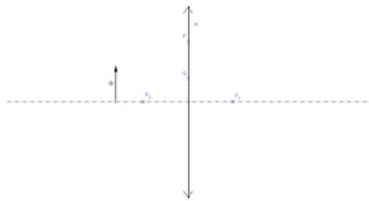


Abb. 2

b) Bestimme den Abbildungsmaßstab

c) Welche Abbildungseigenschaften treffen zu:

- seitenrichtig seitenverkehrt
 vergrößert verkleinert

d) Zeichne folgende Strahlen:

- Strahl, der von der Kerzenstipitze ausgeht und in dem Punkt P auf der Linsenmitte auftrifft
- und
- Strahl, der er von der Kerzenstipitze ausgeht und in dem Punkt Q auf der Linsenmitte auftrifft

A2

a) Konstruiere das Bild des Pfeils und bestimme die Bildgröße durch Konstruktion (Abb. 2)!
Hinweis: Die Linse ist durch die Linsenmitte (H) dargestellt.

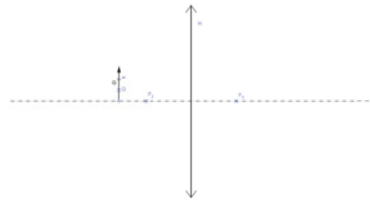


Abb. 2

b) Bestimme den Abbildungsmaßstab

c) Welche Abbildungseigenschaften treffen zu:

- seitenrichtig seitenverkehrt
 vergrößert verkleinert

d) Konstruiere die Bildpunkte zu den Punkten P und Q des Gegenstandes

Abbildung 32: Aufgabenblatt 9 Übungen zur Abbildungsgleichung und zur Bildkonstruktion [Mat. 13_K1] bzw. [Mat. 12_K2], A2 links Aufgabe der TG, rechts Aufgabe KG

2.3.4 Variablen und Erhebungsinstrumente

2.3.4.1 Überblick

Tabelle 8 zeigt einen Überblick über die in der Hauptstudie erhobenen abhängigen Variablen und Kovariaten und ihre Operationalisierung.

Tabelle 8 Überblick über Variablen und die Erhebungsinstrumente

Variable	Erhebungsinstrument
abhängige Variablen	
Wissen und Problemlösen beim Umgang mit Repräsentationen in der Strahlenoptik	Leistungstest (25 min.): Im Rahmen der Studie entwickeltes Testinstrument, drei Messzeitpunkte
Konzeptuelles Grundverständnis in Strahlenoptik	Konzepttest (15 min.): Im Rahmen der Studie entwickeltes Testinstrument, drei Messzeitpunkte
Motivation der Schüler	Fragebogen (5 min.), Basierend auf Hoffmann, Häußler & Peters-Haft, 1997; entnommen aus Kuhn (2008, S. 305-308)
Kovariaten	
Relevante kognitive Fähigkeiten	Intelligenz-Struktur-Test (I-S-T) 2000 R: drei Subskalen (45 min.), Satzergänzung, Würfelaufgaben und Matrizen, einmalige Erfassung
Verbale Fähigkeiten	
Räumliches Denken (figural)	
Logisches Schlussfolgern (figural)	
Vorleistung in Mathematik, Physik, Deutsch	Jeweilige letzte Zeugnisnote
Weitere mögliche Einflussfaktoren	
Geschlechtszugehörigkeit	Dummy – Kodierung (0 = weiblich, 1 = männlich)
Schultyp	Dummy – Kodierung (0 = IGS, 1 = Gymnasium)
Klassengröße	Anzahl teilnehmender Schüler je Schulklasse
Einflussfaktoren bei der Anwendung des Lehrmaterials	
Ablauf der Unterrichtsstunden	Lehrernotizheft (Items entnommen aus Helmke et al., 2010), einmalige Erhebung nach jeder Unterrichtsstunde
Engagement des Lehrers aus Schülersicht	Teilfragen des obigen Motivationsfragebogens, Skala aus: Seidel, Prenzel, Duit & Lehrke (2003, S. 366), drei Messzeitpunkte

2.3.4.2 Leistungstest

Zur Erhebung der abhängigen Variablen Wissen und Problemlösen beim Umgang mit fachspezifischen Repräsentationen wurde der Leistungstest der Pilotstudie

von Jochen Scheid in Zusammenarbeit mit der Autorin für die Studie grundlegend überarbeitet: acht Aufgaben kamen neu hinzu.

Das in dieser Studie verwendete Testinstrument bestand aus 14 Aufgaben und erfasste Wissen und Problemlösen im Bereich der Strahlenoptik durch Aufgaben, die den Umgang mit fachspezifischen Repräsentationen erfordern (siehe Tabelle 9 zum Überblick über die repräsentationalen Anforderungen). Die Testdauer betrug sowohl im Prä- als auch im Posttest 25 Minuten. Maximal konnten jeweils 42 Punkte je Test erreicht werden. Prä-, Post- und Follow-up Test waren identisch. Der Test umfasste folgende Aufgabentypen:

- Repräsentationsbezogene Aufgaben, die nur korrekt gelöst werden konnten, wenn das Denken der Schüler nicht durch spezifische Schülervorstellungen geleitet wurde (Aufgaben 1, 4a und 4b).²³ Aufgaben dieses Typs zielten darauf, Wissen und Problemlösen beim Umgang mit Repräsentationen in unterschiedlichen Formaten im Hinblick auf die Überwindung spezifischer Schülervorstellungen zu erfassen. Sie waren mit den Aufgaben 1, 2a und 2b des Pilotstudienvortests identisch. Beispiel: „Wie ändert sich das Bild des Gegenstandes, wenn man eine Lochblende vor die Linse stellt? Begründe Deine Antwort durch eine Zeichnung. Verwende hierfür die gegebene Abbildung (unten).“
- Aufgaben, bei denen die Schüler mit der schematischen Repräsentationsform des Strahlengangs operierten (Aufgabe 2 Teilbereich „Zeichnung“, Aufgabe 4a und 4b - Teilaufgabe Zeichnung)²³. Beispiel: „Zeichne den Strahlengang des folgenden Versuchsaufbaus“.
- Aufgaben, in denen die Schüler Repräsentationen in depiktional schematischer Form in eine deskriptiv verbale Repräsentationsform übersetzen (2a, b; 4b - Teilaufgabe-Erklärung, 6)²³ oder mit einer deskriptiven Repräsentation operieren sollten (3, 7). Beispiel: „Ein Gegenstand befindet sich vor einer Linse. Die Gegenstandsweite liegt zwischen einfacher und doppelter Brennweite der Linse. Wie groß ist das Bild verglichen mit der Gegenstandsgröße? Wie groß ist die Bildweite verglichen mit der Brennweite?“
- Aufgaben, in denen die Schüler Repräsentationen auf Basis einer depiktional realistischen Repräsentation Schlussfolgerungen ziehen und diese verbal formulieren sollten. Beispiel: „Ein gewöhnliches Zimmerfenster wird durch eine Sammellinse auf einer Wand in einem Raum abgebildet. Um welchen Bildfall handelt es sich?“

²³ Mehrfachzuordnung der Aufgaben 4a und 4b: Teilaufgaben Zeichnung und Erklärung

Die beiden letztgenannten Aufgabentypen waren daraufhin angelegt, das fachliche Wissen im Umgang mit Repräsentationen und die repräsentationale Kohärenz zu erfassen.

Tabelle 9 Übersicht über die repräsentationalen Anforderungen der Aufgabentypen im Leistungstest

Aufgabe	Thematisierte Schülervorstellung	Repräsentationale Anforderung	Punkt
A 1	Abdeckaufgabe	Ein mentales Modell der Situation auf Basis von Hinweisen in depiktionaler Form entwickeln, aus diesem Modell Schlüsse ziehen und diese verbal begründen (identisch mit Aufgabe 1 der Pilotstudie).	4
A 2	-	Entwicklung einer sachlich korrekten und vollständigen schematischen Repräsentation (Konstruktion des Strahlengangs mit vollständiger Beschriftung).	7
Teil a)	-	Auf Basis der angefertigten Konstruktion logische Schlüsse ziehen und diese verbal ausdrücken (Operation mit der selbst entwickelten Repräsentation: mentale Simulation der Bildentstehung).	4
Teil b)	-	Auf Basis der angefertigten Konstruktion logische Schlüsse ziehen und diese verbal ausdrücken.	4
A 3	-	Ein mentales Bild der Situation aufbauen oder eine Skizze anfertigen und dann auf Basis dieser bildlichen internalen oder externalen Repräsentation Schlussfolgerungen ziehen und diese verbal bzw. unter Verwendung der Abkürzung der entsprechenden physikalischen Formel ausdrücken.	2
A 4	Abdeckaufgabe	Eine schematische Repräsentation entwickeln und mit dieser operieren und physikalische Schlussfolgerungen (Konsequenzen auf der Ebene der Phänomene) ziehen (identisch mit Aufgabe 2a der Pilotstudie).	5
Teil b) mit Erklärung	Abdeckaufgabe	Eine schematische Repräsentation entwickeln, mit dieser operieren und physikalische Schlussfolgerungen (Konsequenzen auf der Ebene der Phänomene) ziehen und diese begründen (identisch mit Aufgabe 2b der Pilotstudie).	3 + 1,5
A 5	-	Auf Basis einer realistisch depiktionalen Abbildung ein Phänomen kategorisieren.	1
Teil a)	-	Auf Basis einer realistisch depiktionalen Abbildung ein Phänomen kategorisieren.	1
Teil b)	-	Auf Basis einer realistisch depiktionalen Abbildung ein Phänomen kategorisieren.	1
Teil c)-1	-	Die Entscheidungen der Kategorisierung 5a) und 5b) verbal begründen.	1
Teil c)-2	-	Die Entscheidungen der Kategorisierung 5a) und 5b) verbal begründen.	1
A 6	-	Eine gegebene Repräsentation korrigieren und an die Anforderungen der Fragestellung anpassen. Den Zweck einer bildlich-schematischen Repräsentationen erkennen (metakognitiver Aspekt). Aus einer schematisch bildlichen Repräsentation mathematisch (logische) Schlussfolgerungen ziehen. Bildlich-schematische Repräsentationen in verbale Repräsentationen übersetzen, eigene Schlussfolgerungen begründen.	3
A 7	-	Mit einer mathematischen Repräsentation operieren (identisch mit Aufgabe 7 der Pilotstudie).	4,5
			Σ 42

Ebenso wie der Test der Pilotstudie orientierte sich auch das Erhebungsinstrument der Hauptstudie am Lehrplan für Gymnasien in Rheinland-Pfalz der Jahrgangsstufe 8. Eine ausführlichere Beschreibung des Tests kann in der Dissertationsschrift von Scheid (2013, S. 139 ff.) nachgelesen werden. Der vollständige in dieser Arbeit verwendete Test ist in Anhang C5 auf der Produktseite zu diesem Buch unter www.springer.com enthalten.

2.3.4.3 Konzepttest

Der Konzepttest wurde für die Hauptstudie wie folgt überarbeitet, eine tabellarische Übersicht zur Überarbeitung findet sich in Tabelle 7 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com:

- Items, die Konzepte zur Bildentstehung beim Hohlspiegel erfragten, wurden weggelassen, da die Bildentstehung an der Sammellinse Thema der Untersuchung war.
- Hinzu kamen drei Items zum Thema Streuung und physikalische Sehvorstellung, weil das Verständnis der Strahlenkonstruktion auf diesen Konzepten basiert sowie zwei Items zum virtuellen Bild, dessen physikalische Erklärung im Gegensatz zur Pilotstudie in der Hauptstudie mit aufgenommen wurden.²⁴
- Der Test wurde homogener gestaltet. Jedes Item enthielt drei Distraktoren und eine korrekte Antwort, mit Ausnahme von Item 5 und 8, in denen aus sachlichen Gründen zwei korrekte Antworten bestanden.
- Items mit negativer Trennschärfe im Posttest der Pilotstudie wurden entfernt oder überarbeitet.

Die Punktevergabe der Items erfolgte analog zur Pilotstudie. Entsprechend konnten je Item maximal 0, 1 bis 2 Punkte erzielt werden. Die maximal erreichbare Punktzahl lag bei 21×2 Punkten = 42 Punkte. Der vollständige Test befindet sich in Anhang C6 auf der Produktseite zu diesem Buch unter www.springer.com. Folgende Grundkonzepte wurden in dem Test erfasst (vgl. Tabelle 10)

²⁴ Aus zeitlichen Gründen wurde in der Pilotstudie auf die Erklärung virtueller Bilder am Hohlspiegel verzichtet. Da in der Hauptstudie die Bildentstehung umfassend gemäß des Lehrplans eingeführt werden konnte und die Thematisierung virtueller Bilder auch in Hinblick auf die Förderung des konzeptuellen Verständnisses reizvoll ist, wurde dieses Thema (wenn auch nicht schwerpunktmäßig) mit aufgenommen.

Tabelle 10 Erfasste Grundkonzepte und zugeordnete Items des Konzepttests

korrekte Lösung: wiss. Konzept	Distraktoren: (wiss. inadäquate) Schülervorstellungen	Item
Lichtausbreitung	Licht wird mit seiner Quelle / seinen Wirkungen / mit einem Zustand gleichgesetzt.	1, 9, 10
Sekundäre Lichtquellen und Physikalische Sehvorstellung: Streuung	Beleuchtete Gegenstände wie Tische, Bücher oder Bilder strahlen kein Licht ab / beleuchtete Gegenstände werfen kein Licht zurück.	3, 9, 10
Physikalische Sehvorstellung	Bei Lichtquellen mit geringer Intensität gelangt kein Licht mehr ins Auge.	2
Wissenschaftstheoretisch adäquates Modellverständnis	Licht wird als Flüssigkeitsstrom beschrieben, der in Bewegung ist, sich aber auch in Ruhelage befinden kann. Verwechslung von Lichtbündel und Lichtstrahl.	4, 4
Entstehung reeller Bilder bei der Sammellinse	Eine Sammellinse macht das Licht größer Das reelle Bild bei der Sammellinse entsteht durch Spiegelung oder Reflexion. Holistische Konzeption des Abbildungsvorgangs: das Bild geht als Ganzes durch die Linse zum Schirm und wird dabei in der Linse umgedreht. Bei Abdeckung der oberen bzw. unteren Hälfte der Linse wird entsprechend die obere bzw. untere Hälfte des Bildes angeschnitten.	5, 11, 15a),
Entstehung reeller Bilder bei der Sammellinse:	Bei Verwendung einer ringförmigen Blende werden die Bildränder kreisförmig abgeschnitten.	15b), 15c)
Entstehung reeller Bilder bei der Sammellinse:	Fehlerhafte bildliche Repräsentation der Strahlenkonstruktion. Fehlerhaftes Verständnis der Funktionsweise von Sammellinsen. Schwierigkeiten beim Verständnis des Bildortes und der Funktion des Schirms: Vorstellung, das Bild sei auf der Linse; nur wenige Lernende erkennen, dass sich das Bild ohne Schirm an der gleichen Position befindet wie der Schirm.	13, 8, 15d), 15e), 15f)
Ort und Lage des Spiegelbilds	Das Spiegelbild liegt auf der Spiegeloberfläche.	6,
Gerichtete Reflexion (Planspiegel)	Der Spiegel werde als ein Gegenstand aufgefasst, der das Spiegelbild zum Betrachter zurückwirft. (Licht muss also nicht aus Richtung des Spiegels ins Auge fallen, damit das Spiegelbild wahrgenommen werden kann.)	7
Virtuelle Bilder bei der Sammellinse	Probleme beim Verständnis der Entstehung (und Lage) virtueller Bilder.	12, 16
Punkte je Item: 2, Maximalpunktzahl 42		\sum_{Items}^{21}

2.3.4.4 Motivationsfragebogen

Der Motivationsfragebogen bestand aus Items validierter Erhebungsinstrumente und umfasste die Skalen: Selbstkonzept, intrinsische Motivation, (wahrgenommenes) Lehrer-Engagement aus Schülersicht und die Skala „Folgen und Folgenanreize – Gute Noten“. Die Skala „Lehrer-Engagement aus Schülersicht“ (LES) wurde eingesetzt, um möglich Unterschiede zwischen Treatment- und Kontrollgruppe zu erfassen, welche die Lernleistung beeinflussen, bei der Entwicklung des Treatments jedoch nicht intendiert waren. Für den ersten Messzeitpunkt zielt die Skala auf die Erfassung des Engagements ihres Lehrers aus Sicht der Schüler. Hier sollten zwischen Treatment und Kontrollgruppe möglichst keine Unterschiede bestehen. Zum zweiten Messzeitpunkt interagiert die Bewertung mit den eingesetzten Materialien, welche sich in Treatment- und Kontrollgruppe unterscheiden. Wünschenswert wäre, dass Schüler der Treatmentgruppe in der Postmessung zu einer ähnlichen Einschätzung kommen, wie Schüler der Kontrollgruppe.

Insgesamt bestand der Fragebogen aus 28 Items, die auf einer sechsstufigen Skala mit den Polen „trifft gar nicht zu“ (1) bis „trifft voll und ganz zu“ (6) bewertet werden konnten.

Der Posttest wurde um 6 selbstformulierte Items ergänzt, welche das Selbstkonzept in Bezug auf die erworbenen Lerninhalte und den Umgang mit zentralen Repräsentationen erfassten. Diese Items wurden mit dem Kürzel FI (Forschungssitem zur künftigen Auswertung) versehen und im Rahmen dieser Arbeit nicht ausgewertet. Die verwendeten Fragebögen sind in Anhang C7 auf der Produktseite zu diesem Buch unter www.springer.com zu finden. Der Fragebogen setzte sich wie folgt zusammen (vgl. Tabelle 11).

Tabelle 11 Skalen des Motivationsfragebogens

Skala	Items-Anzahl	Beispiel	Quelle der Items
Selbstkonzept (SK)	10	„Der Unterrichtsstoff in Physik ist für mich verständlich.“	Basierend auf Hoffmann, Häußler & Peters-Haft, 1997; entnommen aus Kuhn (2008, S. 305-308)
Intrinsische Motivation bzw. Engagement (IE)	8	„In meiner Freizeit beschäftige ich mich auch über die Hausaufgaben hinaus mit Themen, die mit Physik zu tun haben.“	Basierend auf Hoffmann, Häußler & Peters-Haft, 1997; entnommen aus Kuhn (2008, S. 305-308)
Lehrer-Engagement aus Schülersicht (LES)	4	„Mein Physiklehrer wirkt begeistert im Physikunterricht.“	Aus Seidel et al. (2003, S. 366)
Folgen und Folgenanreize – Gute Noten (GN)	6	„In Physik viel zu können und gut zu sein ist für mich wichtig, weil ich gute Noten bekommen möchte.“	Rheinberg & Wendland (2003, 2004)

2.3.4.5 Erfassung der Kovariaten

Zur Erfassung relevanter kognitive Fähigkeiten wurden Teilskalen des I-S-T 2000 R eingesetzt (Lipmann, Beauducel, Brocke & Amthauer, 2007). Ziel war es, schlussfolgerndes Denken im verbalen und figuralen Bereich zu erfassen (siehe auch Tabelle 12 und Abbildung 33). Aus Zeit- und Kostengründen musste ein Instrument gewählt werden, das eine Gruppentestung ermöglicht. Für ein tiefgehendes Verständnis der Bildentstehung bei der Sammellinse sollten gerade diese Fähigkeiten eine zentrale Rolle spielen. Wie zuvor ausgeführt, besteht eines der Lernziele in der Förderung des kompetenten Umgangs mit multiplen Repräsentationen, wobei vor allem schematische Repräsentationen und verbale Repräsentationen wechselseitig verknüpft werden. Die mathematische Ebene spielt beim Verständnis der Abbildungsgleichung eine Rolle. Der Schwerpunkt wurde in den entwickelten Lernmaterialien auf die geometrische Herleitung gelegt. Zur Bildung eines adäquaten mentalen Modells der Bildentstehung ist zudem die Verknüpfung mit der realen Experimentiersituation und damit das Verständnis der physikalischen Zusammenhänge im dreidimensionalen Raum wichtig. Daher fiel die Wahl auf Skalen zur Erfassung des figural-räumlichen und figurallogischen Schlussfolgerns. Um die Schüler und auch die teilnehmende Schulen

nicht übermäßig zu belasten, wurde die Zeit für die Durchführung des Tests inklusive Instruktion auf 45 Minuten begrenzt. Wäre den teilnehmenden Schülern und Lehrkräften eine längere Testzeit zumutbar gewesen, wäre der Teilbereich der numerischen Intelligenz ebenfalls erfasst worden.

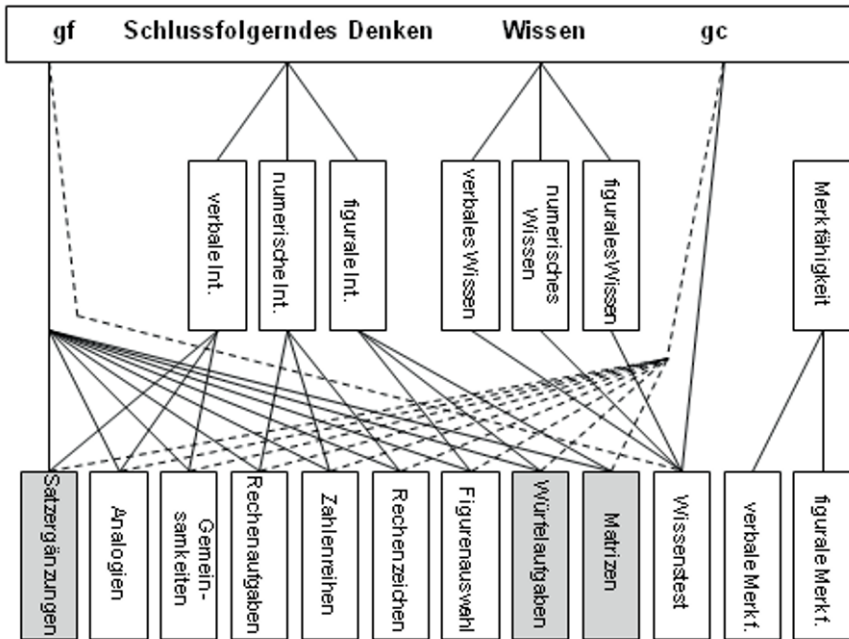


Abbildung 33: Die mit dem I-S-T 2000 R erfasste Fähigkeitsstruktur, (entnommen aus Liepmann et al., 2007, S. 13). In der Studie eingesetzte Subskalen wurden von der Autorin grau hervorgehoben, gf = schlussfolgerndes Denken, gc = Wissen (ebd., S. 12)

Durch die Verwendung einer Auswahl an Subskalen lässt sich zwar keine umfassende Diagnose der Intelligenz auf Individualebene erstellen. Dies ist jedoch für die Umsetzung der Studie auch nicht nötig, da lediglich der Einfluss besonders relevanter kognitiver Fähigkeiten auf die beiden abhängigen Variablen „Wissen und Problemlösen beim Umgang mit (multiplen) Repräsentationen“ sowie auf das „konzeptuelle Verständnis“ statistisch kontrolliert werden sollte.

Tabelle 12 Auswahl an Subskalen zur Erfassung relevanter kognitiver Fähigkeiten

Subskala	Teilbereich	Aufgabenart	Dauer
Satz-ergänzung	verbale Intelligenz	„Jede Aufgabe besteht aus einem Satz, in dem ein Wort fehlt. Aus fünf vorgegebenen Wörtern soll jenes ausgewählt werden, das den Satz richtig vervollständigt“ (zit. n. Liepmann et al., 2007, S. 18).	6 min.
Würfel-aufgaben:	figurale Intelligenz (räumlich)	„In der Aufgabe werden Würfel vorgegeben, auf denen jeweils sechs verschiedene Muster abgebildet sind, drei davon sichtbar. Die auszuwählenden Würfel zeigen einen der vorgegebenen Würfel in veränderter Lage. Es soll herausgefunden werden, um welche Würfel es sich jeweils handelt“ (zit. n. ebd., S. 18).	9 min.
Matrizen	figurale Intelligenz (logisch)	„Es werden Anordnungen von Figuren vorgegeben, die nach einer bestimmten Regel aufgebaut sind. Aus vorgegebenen Auswahlfiguren soll jeweils die regelkonforme Lösung herausgefunden werden“ (zit. n. ebd., S. 19).	10 min.

Vor jedem Aufgabentyp erhielten die Schüler eine kurze Instruktion anhand einer Overheadfolie, auf welcher an zwei Beispielen die Aufgabe erklärt wurde (vgl. ebd., im Testheft, S. 7, S. 17, S. 19). Parallel dazu konnten sie die Instruktion im Testheft mitlesen. Das erste Aufgabenbeispiel wurde von der Versuchsleiterin erläutert, das zweite wurde von einem der Schüler erklärt. Der Test startete gemeinsam, wenn die Teilnehmer keine Rückfragen zur Aufgabenstellung mehr stellten und signalisierten, dass sie die Fragestellung verstanden hatten. Die Durchführung des Tests benötigte inklusive Instruktion eine volle Schulstunde (45 Minuten).

Jede Aufgabengruppe umfasst 20 Fragen. Pro Aufgabe standen jeweils fünf Antwortmöglichkeiten zur Auswahl, aus denen die Schüler eine korrekte Antwort wählen sollten. Pro Aufgabe konnte ein Punkt erzielt werden, je Aufgabengruppe also maximal 20 Punkte. Die jeweils erzielten Punktzahl pro Aufgabengruppe wurden entsprechend der individuell erzielten Gesamtpunktzahl in der Aufgabengruppe dem jeweiligen Normwert der Altersgruppe (vgl. ebd., S. 96) zugeordnet. Die teilnehmenden Schüler durften keine Hilfsmittel verwenden.

Des Weiteren wurde vor Beginn der Unterrichtsreihe von allen teilnehmenden Schülern die Vorleistung von den Lehrern erfragt; verwendet wurden die jeweils aktuellsten Zeugnisnoten in den Fächern, Mathematik, Deutsch und Physik. In

Klassen, die zuvor keinen Physikunterricht erhalten hatten, wurde ersatzweise anstelle der Physiknote die Note im Fach Naturwissenschaften (aus Klassenstufe 6) erfasst.

Als weitere Einflussfaktoren wurden zudem die Klassengröße, die Geschlechtszugehörigkeit und der Schultyp (Gymnasium versus Integrierte Gesamtschule) berücksichtigt.

2.3.4.6 Einflussfaktoren bei der Anwendung des Lehrmaterials

Um Einblick in die Umsetzung der Studie zu erlangen und zu sehen, wie die teilnehmenden Lehrkräfte und Schüler mit den zur Verfügung gestellten Materialien umgehen, wurden die Unterrichtsstunden so oft wie möglich vor Ort beobachtet.

Des Weiteren wurden die Schüler im Rahmen des Motivationsfragebogens zu allen drei Messzeitpunkten um eine Einschätzung des Engagements ihres Lehrers gebeten (siehe Erläuterungen zum Motivationsfragebogen).

Zudem wurden die Lehrer in einer offenen Frage gebeten, in einem zur Verfügung gestellten Notizheft besondere Vorkommnisse im Unterricht in eigenen Worten zu notieren, was zumindest einen groben Einblick in die Umsetzung der Stunden erlaubt, in denen aus organisatorischen Gründen (z.B. zeitliche Überschneidung von Hospitation und Datenerhebung an unterschiedlichen Schulen) keine Hospitation möglich war. Die hier verwendeten Items basierten auf Items zur Erfassung von Unterrichtsqualität von Helmke et al. (2010). Pro Stunde wurden acht kurze Fragen zu folgenden Themen gestellt (siehe Anhang C7 auf der Produktseite zu diesem Buch unter www.springer.com):

Einschätzung der Aufgabenschwierigkeit (Bewertung auf einer Skala von 0 zu einfach bis 3 zu schwierig)

- Einschätzung des Aufgabenumfangs (Bewertung auf einer Skala von 0 zu wenige Aufgaben bis 3 zu umfangreich)
- Fragen zum Unterrichtsablauf (Zustimmung zu drei Aussagen, Skala von 1 stimme nicht zu bis 4 stimme zu)
- Bewertung der Lernbilanz (Zustimmung zu drei Aussagen, Skala von 1 stimme nicht zu bis 4 stimme zu)

2.3.5 Ergebnisse zur Messung der abhängigen Variablen

2.3.5.1 Itemstatistiken zum Leistungstest

Um zu untersuchen, wie sich die Änderungen der Aufgaben im Vergleich zur Pilotstudie statistisch auswirken, wurde zunächst eine Itemanalyse des Leistungstests durchgeführt. Eine Übersicht über die Itemmittelwerte und Standardabweichungen findet sich in Tabelle 8 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com.

Die Analyse der Itemschwierigkeit ergab, dass für den Posttest alle Items mit Ausnahme von Item 2b und Item 6 innerhalb des Toleranzbereichs von $0.20 \leq P_i \leq 0.80$ liegen (vgl. Abbildung 34: Itemschwierigkeiten des Leistungstests prä, post und follow-up, 2Z: Aufgabenteil Zeichnung, 4b - E: Aufgabenteil Erklärung, 5c-E1: Aufgabenteil Erklärung 1, 5 c-E2: Aufgabenteil-Erklärung 2).

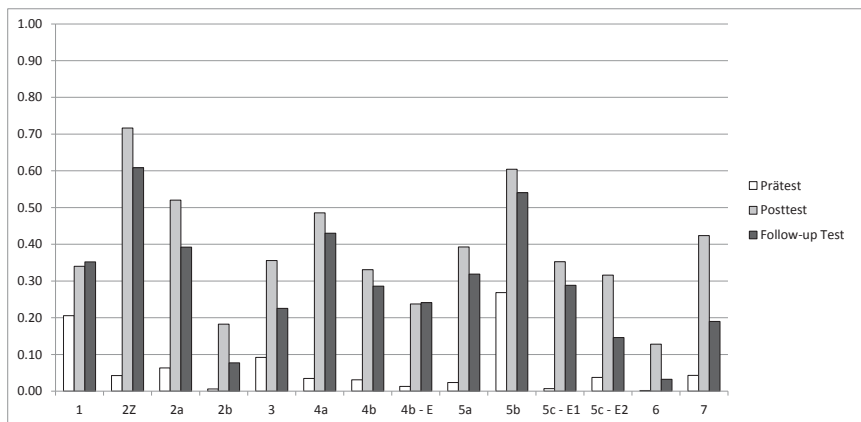


Abbildung 34: Itemschwierigkeiten des Leistungstests prä, post und follow-up, 2Z: Aufgabenteil Zeichnung, 4b - E: Aufgabenteil Erklärung, 5c-E1: Aufgabenteil Erklärung 1, 5 c-E2: Aufgabenteil-Erklärung 2

Für den Prätest zeigen sich durchweg deutlich geringere Werte. In der Follow-up Messung fällt die Lösungswahrscheinlichkeit der Items 6 und 2b Items unter die kritische Grenze von 20%. Zudem sinkt die Lösungswahrscheinlichkeit für

Item 5c (Teil 2)²⁵ auf einen Wert von 15% (vgl. Tabelle 9 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com). Bemerkenswert ist, dass die Lösungswahrscheinlichkeit für die Wahl der korrekten Gegenstandsweite (5c - Erklärung1) deutlich weniger stark sinkt als die Lösungswahrscheinlichkeit für die Erklärung dieser Wahl (5c - Erklärung2), siehe auch Abbildung 35.

In der Analyse der Trennschärfe zeigt sich, dass die Werte aller Items in der Post-Messung in einem Bereich von $0.29 \leq r_{ii} \leq 0.56$ und in der Follow-up Messung in einem Bereich von $0.18 \leq r_{ii} \leq 0.59$ liegen (vgl. Abbildung 35). Negative Werte in der Trennschärfe bestanden lediglich in Aufgabe 7 (Anwendung der Abbildungsgleichung) im Vortest (vgl. Tabelle 9 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com).

Item 6 zeichnet sich sowohl in der Post- als auch in der Follow-up Messung durch die geringste Trennschärfe aus. Die Analyse der Verteilung des Items zeigt, dass das Item rechtssteil und linksschief verteilt ist. Konsistent mit diesem Befund ist die geringe Lösungswahrscheinlichkeit des Items. Daher kann in Erwägung gezogen werden Item 6 auszuschließen.

Cronbachs Alpha als Maß für die interne Konsistenz und Schätzer für die Reliabilität kann mit $\alpha = .77$ im Post- und $\alpha = .75$ im Follow-up Test als zufriedenstellend und mit $\alpha = .29$ als gering im Prätest bewertet werden. Unter Ausschluss von Item 6 ergeben sich die gleichen Werte. Der niedrige Wert im Vortest ist auf das geringe Vorwissen der Schüler zurückzuführen, die zuvor keinen Unterricht in Strahlenoptik erhalten hatten. Die Test-Retest-Reliabilität liegt mit $r_{\text{Test-Retest}}(482) = .44$, $p < .01$ im akzeptablen Bereich. Da die zeitliche Stabilität des Merkmals aufgrund des geringen Vorwissens der Schüler nicht vorausgesetzt werden kann, wurde davon abgesehen die Retest-Reliabilität prä – post und prä – follow-up zu berechnen.

25 Wortlaut des Items: (Erklärung – Teil 2: Kreuze an, wie groß die Gegenstandsweite g etwa sein muss! Erkläre, wie du auf die Lösung des Aufgabenteils b) gekommen bist!)

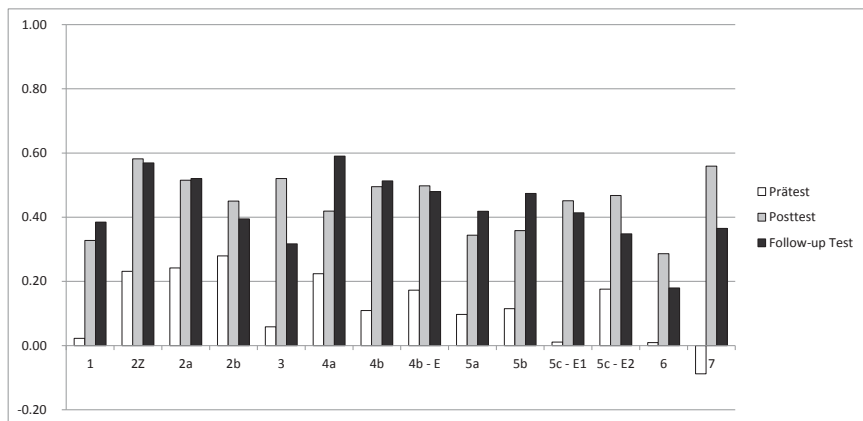


Abbildung 35: Korrigierte Trennschärpen der Items des Leistungstests, prä, post, und follow-up, 2Z: Aufgabenteil Zeichnung, 4b - E: Aufgabenteil Erklärung, 5c - E1: Aufgabenteil Erklärung 1, 5c - E2: Aufgabenteil-Erklärung 2

2.3.5.2 Faktorenanalyse des Leistungstests

Um die Dimensionalität des Instruments (die Struktur des Tests) zu analysieren, wurde eine konfirmatorische Faktorenanalyse durchgeführt. Ergebnisse einer exploratorischen Faktorenanalyse des Instruments auf Basis einer von den hier berichteten Daten unabhängigen Stichprobe, finden sich in der Dissertation von Scheid (2013).

Mittels der konfirmatorischen Faktorenanalyse wurde zunächst untersucht, ob es sich um ein eindimensionales Testinstrument handelt. Hierzu wurde für jeden der drei Messzeitpunkte getrennt ein Generalfaktorenmodell aufgestellt, bei dem alle Variablen auf einen Faktor laden. In diesem Modell wird angenommen, dass alle 13 Variablen nur ein einziges latentes Merkmal (hier das physikalische Verständnis beim Umgang mit (multiplen) Repräsentationen) erfassen und darüber hinaus nur zufällige Messfehler und indikatorspezifische Varianz. Im Ergebnis zeigt sich, dass das Generalfaktorenmodell für jeden der drei Messzeitpunkte verworfen werden muss, da das Modell außerhalb der in der Literatur angegebenen Grenzen für Cut-off-Werte liegen: CFI und $TLI \geq 0.95$ – guter Bereich bzw. $CFI \geq 0.90$ – akzeptabler Bereich; $RMSEA < 0.06$ ($N > 250$), $SRMR < 0.11$ (vgl. Bühner, 2011, S. 428). Exemplarisch werden an dieser Stelle die Fit-Indizes des Generalfaktorenmodells für den zweiten Messzeitpunkt aufgeführt, um die Ablehnung des Generalfaktorenmodells zu belegen ($\chi^2_{(65)} = 308.35, p < .001$; $CFI = 0.83$; $TLI =$

0.80; $RMSEA = 0.09$; $SRMR = 0.06$). Die Fit-Indizes des dritten Messzeitpunktes weichen hierbei nur geringfügig von den Ergebnissen des zweiten Messzeitpunktes ab. Für den ersten Messzeitpunkt ergeben sich die deutlichsten Abweichungen von den Untergrenzen der Cut-Off-Werte. Offenbar handelt es sich um ein mehrdimensionales Testinstrument, das unterschiedliche Fähigkeitsfacetten oder möglicherweise sogar voneinander unabhängige Fähigkeiten erfasst.

Untersucht wurde hierbei zum einen die inhaltliche Struktur des Tests, d.h. Klärung der Frage, welche Items inhaltlich zusammenhängen und ein latentes Merkmal abbilden und zum anderen die Frage, wie stark die verschiedenen Facetten inhaltlich miteinander verbunden sind. Bei der Durchführung der Analyse zeigte sich, dass sich die Interpretierbarkeit der Modelle unter Ausschluss von Item 13 (Aufgabe 6) deutlich erhöhte, welches bereits bei der Analyse der Itemstatistiken aufgefallen war. Zur Analyse der inhaltlichen Struktur wurde ein 2-Faktorenmodell einem 3-Faktorenmodell gegenübergestellt (siehe Tabelle 14).

1. Im 2-Faktorenmodell wurde folgende Unterscheidung getroffen: Ein Faktor, welcher das konzeptuelle Verständnis im Umgang mit Repräsentationen und ein Faktor, welcher nur den Umgang mit Repräsentationen (ohne Thematisierung von Schülervorstellungen) erfasst.

- Faktor 1 („Schülervorstellungen“) = y_1, y_6, y_7, y_8
- Faktor 2 („Umgang mit Repräsentationen“) = $y_2, y_3, y_4, y_5, y_9, y_{10}, y_{11}, y_{12}, y_{14}$

2. Im 3-Faktorenmodell wurden drei Faktoren angenommen, welche auf den Umgang mit verschiedenen Repräsentationsformen zielen: Umgang mit deskriptiven Repräsentationen, Umgang mit depiktional realistischen Repräsentationen, die sich eng am Experiment orientieren und Umgang mit depiktional schematischen Repräsentationen.

- Faktor 1 („Umgang mit deskriptiven Repräsentationen“) = $y_1, y_3, y_4, y_5, y_8, y_{14}$ ²⁶. Der erste Faktor bezieht sich auf das Ziehen logischer Schlüsse auf Basis des Operierens mit deskriptiven Repräsentationen. Die deskriptiven Repräsentationen schließen hierbei sowohl verbale Begründungen als auch die Verwendung der Abbildungsgleichung (mathematische Repräsentation) ein.
- Faktor 2 („Bezüge zu einem realen Experiment herstellen“) = $y_9, y_{10}, y_{11}, y_{12}$ ²⁷. Der zweite Faktor bezieht sich ausschließlich auf alle Teilfragen

26 entspricht den Aufgaben: 1, 2a, 2b, 3, 4b - E (4b - Aufgabenteil Erklärung), 7

27 entspricht den Aufgaben: 5a, 5b, 5c - E1 (Aufgabenteil c, Erklärung 1), 5c - E2 (Aufgabenteil c, Erklärung 2)

von Aufgabe 5, in der es darum ging, aus einer depiktional realistischen Repräsentation, in der ein reales Experiment dargestellt ist, Informationen abzulesen.

- Faktor 3 („Umgang mit schematisch depiktionalen Repräsentationen“) = y_2 , y_6 , y_7^{28} . Der dritte Faktor bezieht sich auf das eigene Erstellen von schematisch deskriptiven Repräsentationen der Strahlenkonstruktion und dem Operieren mit denselben.

Um zu analysieren, wie stark die verschiedenen Facetten miteinander zusammenhängen wurde jedes der beiden Modelle (2-Faktorenmodell versus 3-Faktorenmodell), nun auf unterschiedliche Weise spezifiziert:

- Basismodell mit Korrelationen: Zunächst wurde jeweils ein Modell aufgestellt, bei dem die jeweiligen Items auf zwei bzw. drei Faktoren laden und das Programm die Korrelationen zwischen den Faktoren schätzt.
- Bifaktormodell: In einem nächsten Schritt wurde ein bifaktorielles Modell aufgestellt. Gemäß diesem Modell bestehen ebenso wie im Basismodell zwei bzw. drei Faktoren, die für die Erfassung voneinander unabhängiger Merkmale stehen. Korrelationen zwischen den Faktoren wurden daher nicht zugelassen. Über diese zwei bzw. drei Faktoren hinaus wird angenommen, dass ein Generalfaktor existiert, welcher die Bearbeitung aller Items beeinflusst.
- Second-Order-Modell: Im letzten Schritt wurde ein Second-Order-Modell aufgestellt. In diesem Modell wird angenommen, dass ein Faktor erster Ordnung besteht, der sich auf zwei bzw. drei Faktoren zweiter Ordnung auswirkt, welche für voneinander unabhängige latente Merkmale stehen. Entsprechend sind die Faktoren zweiter Ordnung untereinander unkorreliert.
- Insgesamt wurden zusätzlich zum Generalfaktorenmodell also sechs Modelle zu drei Zeitpunkten verglichen. Die Berechnung der Modelle erfolgte unter Verwendung der freien Statistiksoftware R und der Pakete lavaan (Rosseel, 2012), qgraph (Epskamp et al., 2012) und psych (Revelle, 2013).

Im Hinblick auf die Fit-Indizes zeigt sich, dass von den getesteten Modellen das Bifaktormodell „Umgang mit unterschiedlichen Repräsentationsformen“ innerhalb der in der Literatur angegebenen Grenzen der Cut-off-Werte liegt (vgl. Tabelle 13 und Tabelle 14). Das Modell (vgl. Abbildung 37) kann inhaltlich wie folgt interpretiert werden: Es existiert eine generelle Kompetenz, welche die Be-

28 entspricht den Aufgaben: 2Z (Zeichnung), 4a und 4b

arbeitung aller gestellten Aufgaben beeinflusst. Dass diesbezüglich jeder der Pfadkoeffizienten signifikant wurde, stützt die Gültigkeit dieser Annahme. Des Weiteren existieren drei Subskalen, die über diese Fähigkeit hinaus den „Umgang mit deskriptiven Repräsentationen“, die Fähigkeit Bezüge zu einem realen Experiment herzustellen“ und „den Umgang mit schematisch depiktionalen Repräsentationen“ erfassen (vgl. Abbildung 36). Jede dieser Kompetenzen ist voneinander unabhängig.

Tabelle 13 Fit-Indizes: Konfirmatorische Faktorenanalysen Leistungsposttest, ($N = 484$)

	2-Faktorenmodell (inklusive Skala zum Konzeptuellen Verständnis)			3-Faktorenmodell (Umgang mit unterschiedlichen Repräsentationsformen)		
	^a BMK	^b BFM	^c SOM	^a BMK	^b BFM	^c SOM
$\chi^2_{(df)}$	257.47 ₍₆₄₎ ***	152.95 ₍₅₂₎ ***	257.47 ₍₆₃₎ ***	177.72 ₍₆₂₎ ***	112.78 ₍₅₂₎ ***	169.85 ₍₆₁₎ ***
CFI	0.87	0.93	0.87	0.92	0.96	0.92
TLI	0.84	0.90	0.83	0.90	0.94	0.89
RMSEA	0.08	0.06	0.08	0.06	0.05	0.06
SRMR	0.06	0.04	0.06	0.05	0.04	0.05
AIC	14184.38	14103.86	14186.38	14108.63	14063.69	14112.77
BIC	14297.30	14266.96	14303.48	14229.91	14226.79	14238.23

*** $p < .001$

^aBMK: Basismodell mit Korrelationen zwischen Faktoren, ^bBFM: Bifaktormodell, ^cSOM: 2nd Order Modell

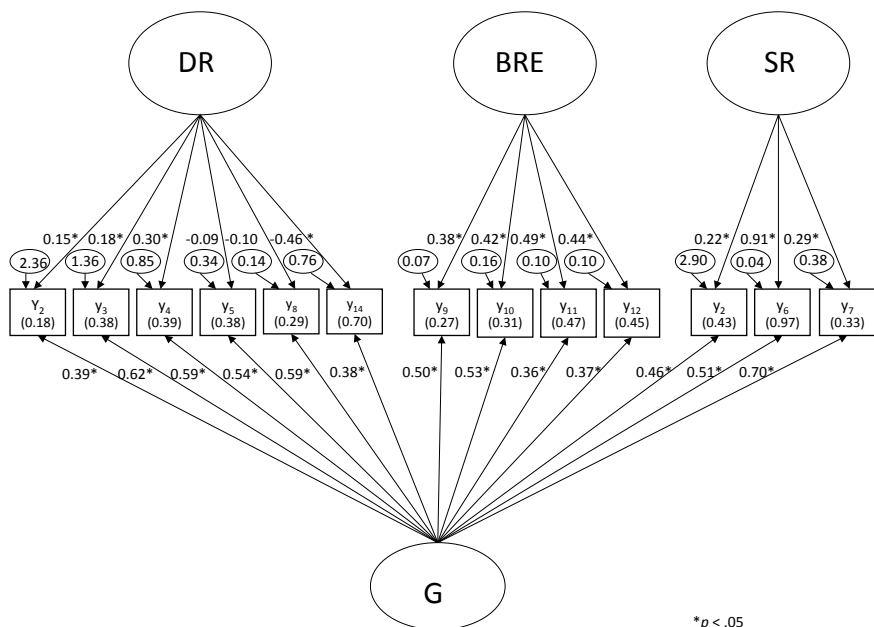


Abbildung 36: Bifaktormodell „Umgang mit unterschiedlichen Repräsentationsformen“, DR = Umgang mit deskriptiven Repräsentationen, BRE = Bezüge zu einem realen Experiment herstellen, SR = Umgang mit schematisch depiktionalen Repräsentationen, G = Generalfaktor

Da jedoch nicht alle Pfadkoeffizienten für die Pfade zwischen den Items und denjenigen Faktoren, welche die Subskalen abbilden, signifikant sind, ist auch dieses Modell mit Vorsicht zu interpretieren. Dies betrifft die Items: y_5 und y_8 . Bei diesen Items wurden die betreffenden Pfadkoeffizienten nicht signifikant. Geringere, noch akzeptable globale Fitwerte ($\chi^2_{(62)} = 177.72, p < .001; CFI = 0.92; RMSEA = 0.06; SRMR = 0.05$), jedoch bessere lokale Fitwerte, weist das Basismodell mit Korrelationen zwischen den drei Faktoren auf (vgl. Tabelle 13).

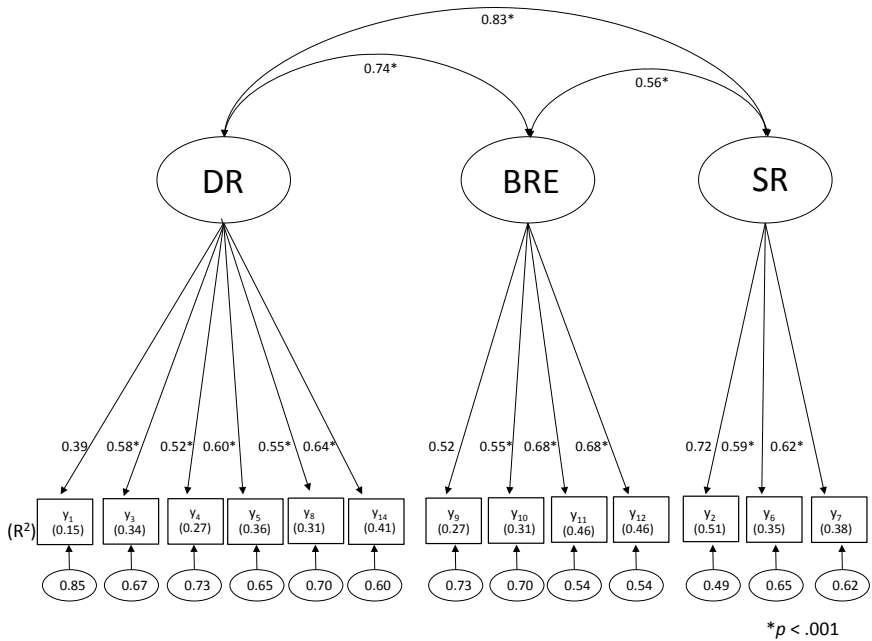


Abbildung 37: BMK: Basismodell mit Korrelationen zwischen Faktoren „Umgang mit unterschiedlichen Repräsentationsformen“, DR = Umgang mit deskriptiven Repräsentationen, BRE = Bezüge zu einem realen Experiment herstellen, SR = Umgang mit Umgang mit schematisch depiktionalen Repräsentationen

Zusammenfassend lässt sich festhalten, dass sich die Physikleistung durch drei Kernkompetenzen beschreiben lässt, die vermutlich jedoch durch eine generelle „Repräsentations“-Kompetenz beeinflusst werden.

Tabelle 14 Zuordnung der Items des Leistungstests zu den Skalen „Umgang mit verschiedenen Repräsentationsformen, Reliabilität der Skalen und lokale Gütemaße, ($N = 484$)

<i>F</i>	Item / Aufgabe	Indikatorreliabilität	Faktor-reliabilität	<i>DEV</i>	Formel Larcker Kriterium	α
DR ^a	Y ₁ 1	0.15	.68	0.29	.69 (nicht erfüllt)	.66
	Y ₃ 2a	0.34				
	Y ₄ 2b	0.27				
	Y ₅ 3	0.36				
	Y ₈ 4b -E	0.30				
	Y ₁₄ 7	0.41				
BRE ^b	Y ₉ 5a	0.27	.70	0.38	>.54	.69
	Y ₁₀ 5b	0.31				
	Y ₁₁ 5C -E1	0.46				
	Y ₁₂ 5C -E2	0.46				
SR ^c	Y ₂ 2Z	0.15	.67	0.47	>.31	.57
	Y ₆ 4a	0.31				
	Y ₇ 4b	0.16				

Erläuterungen

Y _i	Item	Faktor
Y ₁ 1	Deskriptive R. / Verbale Begründung/ Schülervorstellung: Tobias möchte an einem sonnigen Tag mit Hilfe einer Sammellinse ein Streichholz entzünden. Nadine schlägt Tobias vor, eine verstellbare Lochblende vor die Sammellinse zu setzen. Kann Tobias dadurch das Streichholz besser entzünden?	DR ^a
Y ₃ 2a	Deskriptive R. / Verbale Begründung: Wie würde sich die Bildgröße und Bildweite verändern, wenn man den Gegenstand weiter von der Linse entfernt?	
Y ₄ 2b	Deskriptive R. / Verbale Begründung: Begründe mit den Strahlen in der Abbildung oben: Warum verändert sich die Bildgröße so wie oben beschrieben, wenn der Gegenstand weiter von der Linse entfernt wird?	
Y ₅ 3	Deskriptive R. / Deskriptive Darstellung: Ein Gegenstand befindet sich vor einer Linse. Die Gegenstandsweite liegt zwischen einfacher und doppelter Brennweite der Linse. Wie groß ist das Bild verglichen mit der Gegenstandsgröße? Wie groß ist die Bildweite verglichen mit der Brennweite?	
Y ₈ 4b - E	4b Aufgabenteil Erklärung Schülervorstellung/Lochblende/Strahlenkonstruktion/verbale Beschreibung: Erkläre in Worten, was mit dem Bild passiert.	
Y ₁₄ 7	Deskriptive R. / Berechnung: Berechne <i>B</i> bei einer scharfen Abbildung eines Gegenstandes mit einer Sammellinse.	
Y ₉ 5a	Bezug reales Experiment / Fotografie Ein gewöhnliches Zimmerfenster wird durch eine Sammellinse auf einer Wand in einem Raum abgebildet. Um welchen Bildfall handelt es sich?	BRE ^b
Y ₁₀ 5b	Bezug reales Experiment / Fotografie Kreuze an, wie groß die Gegenstandsweite <i>g</i> etwa sein muss!	
Y ₁₁ 5C - E1	Aufgabenteil Erklärung 1 Bezug reales Experiment / Fotografie: Erkläre, wie du auf die Lösung des Aufgabenteils a) gekommen bist!	

Y ₁₂	5c - E2	Aufgabenteil Erklärung 2 Bezug reales Experiment / Fotografie: Erkläre, wie du auf die Lösung des Aufgabenteils b) gekommen bist!	
Y ₂	2Z	Aufgabe 2 Zeichnung Schematisch depiktionale R. / Strahlenkonstruktion: Zeichne den Strahlengang des folgenden Versuchsaufbaus.	SR ^c
Y ₆	4a	Schematisch depiktionale R. / Problemlöseaufgabe / Strahlenkonstruktion: Abbildung (unten) zeigt einen Gegenstand (hier als Pfeil dargestellt) mit der Gegenstandsgröße G und sein Bild, das durch die Sammellinse entsteht, mit der Bildgröße B. Der Bildpunkt Q zum Gegenstandspunkt P wurde richtig konstruiert. Die Strahlenkonstruktion ist nicht dargestellt. Der Schirm ist jedoch nicht an der richtigen Stelle aufgestellt, sondern ein Stück zur Linse hingerückt. a) Zeige durch Einzeichnen, dass auf dem Schirm anstatt des Bildpunktes Q ein unscharfer Bildfleck entsteht. Verwende hierfür für die gegebene Abbildung (unten).	
Y ₇	4b	Schematisch depiktionale R. / Schülervorstellung / Lochblende / Strahlenkonstruktion: Wie ändert sich das Bild des Gegenstandes, wenn man eine Lochblende vor die Linse stellt? Begründe Deine Antwort durch eine Zeichnung. Verwende hierfür die gegebene Abbildung (unten).	

^a DR = Umgang mit deskriptiven Repräsentationen

^b BRE = Bezüge zu einem realen Experiment herstellen

^c SR = Umgang mit Umgang mit schematisch depiktionalen Repräsentationen

2.3.5.3 Itemstatistiken zum Konzepttest

Zur Analyse, wie sich die Überarbeitung der Aufgaben statistisch auswirkt, wurde für den Konzepttest ebenfalls eine Itemanalyse durchgeführt (vgl. Tabellen 11-13 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com).

Die zugrunde liegende Stichprobe basiert neben den Daten der hier diskutierten Studie auf den Daten von Jochen Scheid. Bei den Daten von Jochen Scheid handelt es sich um eine Stichprobe von $N = 444$ Schülern, die ebenfalls der 7. und 8. Klassenstufe angehörten, die allerdings ausschließlich das Gymnasium besuchen. Detailliertere Angaben zur Stichprobe finden sich in der Dissertationsschrift von Jochen Scheid (vgl. Scheid, 2013). Ziel der Studie von Jochen Scheid war es, Effekte der Förderung der Kohärenz im Umgang mit Repräsentationen auf den Lernzuwachs zu untersuchen. In der Studie von Jochen Scheid wurde der Konzepttest ebenfalls zu drei Messzeitpunkten verwendet: vor der Intervention, nach der Intervention, welche in dieser Studie einen vergleichbaren Umfang von sechs Stunden hatte, sowie sechs Wochen später. Durchführung und Auswertung des Tests erfolgten in gleicher Weise wie in der hier dargestellten Studie. Die Testfassung war exakt identisch.

In einem ersten Schritt wurden die Mittelwerte der beiden Stichproben zum ersten Messzeitpunkt verglichen, als die Schüler in beiden Stichproben noch keinen Unterricht zur Bildentstehung bei der Sammellinse erhalten hatten.

Die Stichproben werden wie folgt unterschieden:

- Die Stichprobe von Jochen Scheid wird im Folgenden mit Stichprobe „Kohärenz“ abgekürzt „Ko“ bezeichnet, da Schüler in dieser Studie in der Treatmentgruppe Unterricht zur Förderung der Kohärenz von Repräsentationen erhielten, während Schüler der Kontrollgruppe herkömmliche Aufgaben bearbeiteten.
- Die Stichprobe, welche der hier vorgestellten Studie zu Grunde liegt, wird Stichprobe „Schülervorstellungen“ genannt und im Folgenden mit „SV“ abgekürzt, da der Unterricht in beiden Bedingungen (Treatment- und Kontrollgruppe) auf die Förderung des konzeptuellen Verständnisses und damit einhergehend darauf zielte, Schülervorstellungen zu überwinden.

Da in der Studie zur Förderung der Kohärenz kein Unterschied zwischen Treatment- und Kontrollgruppe gefunden wurde (vgl. Scheid, 2013), wurde darauf verzichtet, die Itemstatistiken separat je Bedingung (Treatment- versus Kontrollgruppe) zu berichten. Detailliertere Angaben zu den Ergebnissen des Konzepttests für die Stichprobe „Ko“ finden sich in Scheid (2013, S. 135 ff.). Analog dazu wurden auch in der Studie zu Schülervorstellungen so verfahren, dass Treatment- und Kontrollgruppe gesamt betrachtet wurden. Der Vergleich zielt also darauf, Unterschiede zwischen den Stichproben bzw. den Studien auf Itemebene darzustellen, welche darauf beruhen, dass in der Studie „SV“ explizit auf die Förderung des konzeptuellen Verständnisses eingegangen wurde, während in der Studie „Ko“ andere Lernziele im Vordergrund standen.

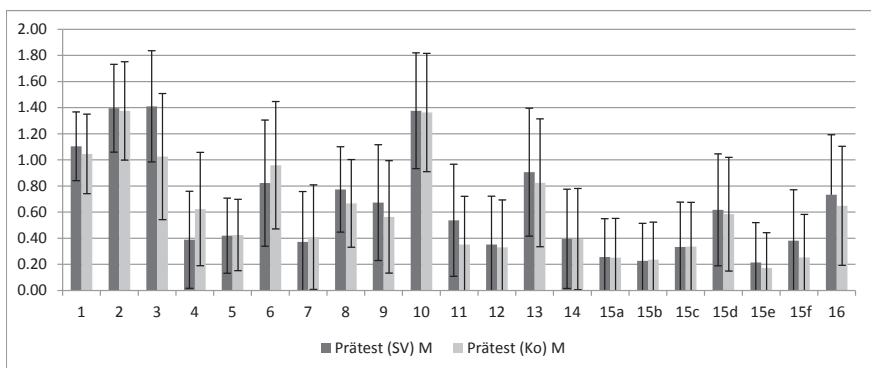


Abbildung 38: Mittelwerte und Standardabweichungen des Konzeptprätest der Stichproben „SV“ (= Schülervorstellungen) und „Ko“ (= Kohärenz) im Vergleich (nur Gymnasiasten)

Im ersten Schritt wurde analysiert, ob sich die Stichproben „SV“ und „Ko“ bereits zu Beginn systematisch unterscheiden, was zum ersten Messzeitpunkt erwartungsgemäß nicht der Fall sein sollte. Hierzu wurden in der Stichprobe „SV“ ausschließlich Gymnasiasten einbezogen, um größtmögliche Vergleichbarkeit zwischen den Stichproben zu erreichen (vgl. Tabelle 10 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com und Abbildung 38). Bezüglich der Gesamtpunktzahl im Konzeptprätest lassen sich keine Unterschiede zwischen den Gruppen feststellen ($t(860) = .64, n.s.$).

In den folgenden Itemstatistiken wurden die Daten aller Schüler der Stichproben „SV“ und „Ko“ verglichen, also auch die Daten der Gesamtschüler in der Stichprobe „SV“ einbezogen. Erwartungsgemäß erzielten die teilnehmenden Schüler nach dem Unterricht zur Bildentstehung bei der Sammellinse im Schnitt durchweg höhere Punktzahlen als zum ersten Messzeitpunkt, unabhängig davon, ob gezielt auf die Förderung des konzeptuellen Verständnisses eingegangen wurde oder nicht (siehe Tabelle 11 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com).

Die Analyse der Itemschwierigkeit ergab, dass für den Posttest der Stichprobe „SV“ alle Items mit Ausnahme von Item 15e innerhalb des Toleranzbereichs von $0.20 \leq P_i \leq 0.80$ liegen (vgl. auch Abbildung 39). Die Lösungswahrscheinlichkeit beträgt für Item 15e 15 % (vgl. Tabelle 12 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com).

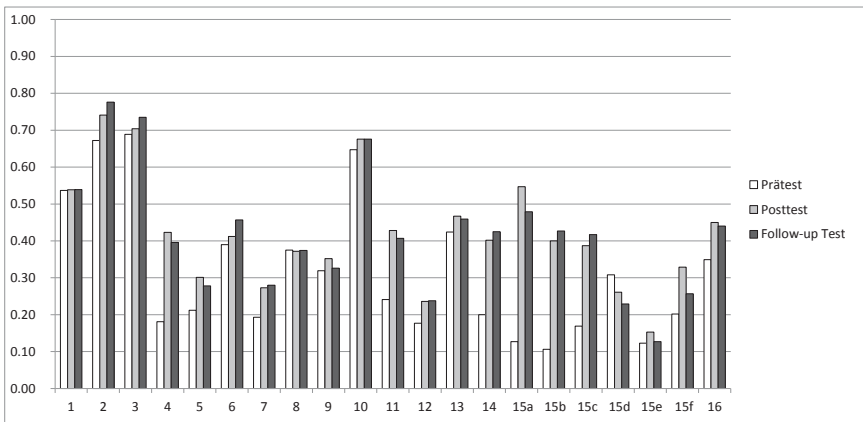


Abbildung 39: Abbildung: Itemschwierigkeiten des Konzepttests für die Stichprobe „SV“ (= Schülervorstellungen) je Messzeitpunkt

In Bezug auf die Stichprobe „SV“ bestätigt sich beim Blick auf die Lösungswahrscheinlichkeiten, dass sich der Unterricht stärker auf Bearbeitung von Aufgaben auswirkt, die spezielle Konzepte zur Bildentstehung thematisieren, wie etwa den Strahlengang, die Bildentstehung unter Verwendung einer Blende oder Fragen zum virtuellen Bild (Item 11 bis 16), als auf die Bearbeitung von Aufgaben, in denen es allgemein um Konzepte der Lichtausbreitung, gerichteten Reflexion, Streuung und der physikalischen Sehvorstellung geht (Items 1-10), siehe auch Abbildung 39. Item 7 und 15a fallen zudem ebenfalls mit einer geringen Lösungswahrscheinlichkeit im Posttest und Follow-up Test der Stichprobe „Ko“ auf (vgl. Tabelle 13 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com).

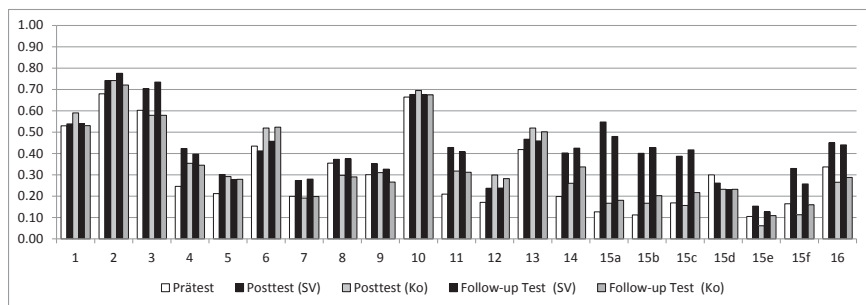


Abbildung 40: Itemschwierigkeit für den Konzeptprätest: Prätest: Gesamt, Post- und Follow-up Test je Stichprobe („SV“ = Schülervorstellungen versus „Ko“ = Kohärenz)

In der Stichprobe SV weist Item 15a jedoch „gute“ Werte der Lösungswahrscheinlichkeit zum zweiten und dritten Messzeitpunkt auf, was darauf hindeutet, dass dieses Item sich dafür eignet das Wissen von Schülern zu differenzieren, nachdem die Schüler Unterricht zur Förderung des konzeptuellen Verständnisses erhalten hatten (vgl. Abbildung 40 sowie Tabelle 12 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com).

Eine abschließende Beurteilung kann hier ein Blick auf die Trennschärfe geben. Die Analyse der Trennschärfe zeigt, dass die Items 7, 15d, 15e eine geringe Trennschärfe aufweisen. Die Items 7 und 15e waren bereits zuvor negativ aufgefallen (vgl. Abbildung 41 sowie Tabelle 13 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com).

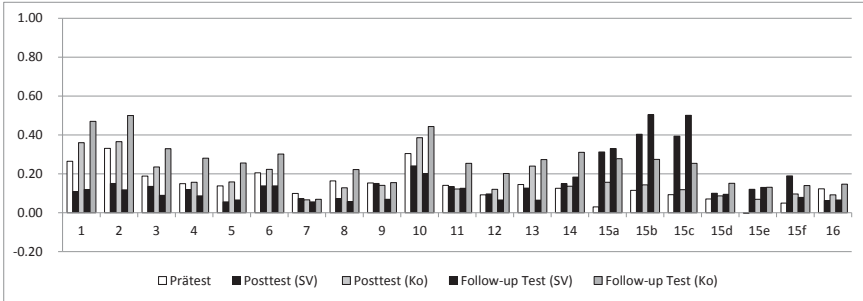


Abbildung 41: Korrigierte Trennschärfe für den Konzeptprätest: Prätest: Gesamt, Post- und Follow-up Test je Stichprobe „SV“ (Schülvorstellungen) versus „Ko“ (Kohärenz)

Des Weiteren geht die Trennschärfe bei einigen Items im Vergleich der Gesamtstichprobe („SV“ und „Ko“) und der Postmessung („SV“) zurück (vgl. Tabelle 13 in Anhang C9). Um diesen Rückgang genauer zu analysieren, wurde die Trennschärfe zu allen drei Messzeitpunkten gesondert nochmals für die Stichprobe „SV“ betrachtet (vgl. Abbildung 42).

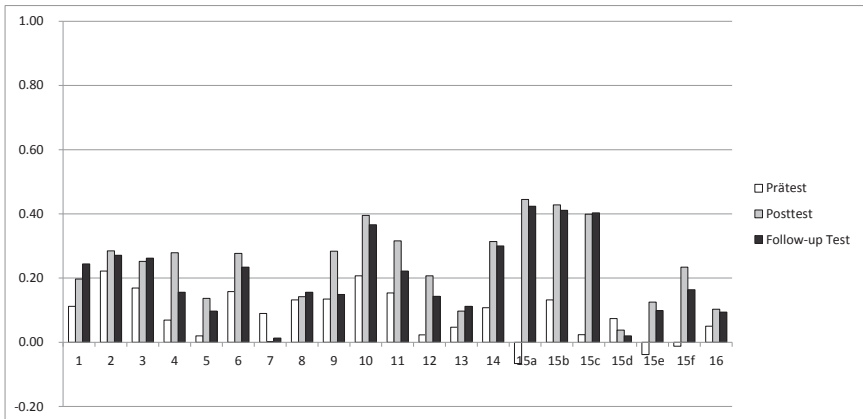


Abbildung 42: korrigierte Trennschärfen des Konzepttests für die Stichprobe „SV“ (Schülvorstellungen) je Messzeitpunkt

Ein Blick auf Abbildung 42 zeigt, dass der Effekt des Rückgangs von Messzeitpunkt 1 zu Messzeitpunkt 2, mit Ausnahme von Item 15d, verschwindet, wenn lediglich die Stichprobe „SV“ zu Grunde gelegt wird.

Die Analyse der Trennschärfe in der Stichprobe „SV“ ergibt zudem, dass die Werte in der Post-Messung zwischen $0.10 \leq r_{it} \leq 0.45$ und in der Follow-up Messung in einem Bereich von $0.10 \leq r_{it} \leq 0.40$ liegen (Tabelle 13 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com). Negative Werte in der Trennschärfe bestanden lediglich im Vortest und zwar in den Items: 15a, 15e und 15f.

Diese negativen Werte liegen jedoch nahe am Bereich 0, was darauf hinweist, dass diese Items spezifisches Wissen zur Bildentstehung bei der Sammellinse erfassen und daher wenig dazu geeignet sind das konzeptuelle Verständnis von Schülern ohne Vorkenntnisse dieser Lerninhalte einzuschätzen. Im Posttest erweisen sie sich jedoch als geeignet.

Zwischen-Fazit: Ein Vergleich der Lösungswahrscheinlichkeiten und der Trennschärfe zwischen den Stichproben von Jochen Scheid (Förderung der repräsentationalen Kohärenz ohne gezielte Berücksichtigung von Schülervorstellungen) und dem Treatment der hier vorgestellten Studie legt nahe, dass die Items 7, 15d und 15e einer Überarbeitung bedürfen. Folgende Schwierigkeiten sind mit den Items verbunden:

- Die ursprüngliche Fassung von Item 7 enthielt eine ungenaue Formulierung, welche zu einer mehrdeutigen Interpretation führte: „In einem großen dunklen Raum stehen Corinna und Nadine nebeneinander vor einem Spiegel. Corinna beleuchtet den Spiegel schräg mit einer Taschenlampe, die ein schmales Lichtbündel erzeugt. Wer kann das Licht der Taschenlampe sehen?“ Je nachdem wie „schmal“ der jeweilige Schüler das Lichtbündel interpretierte, fiel das Licht auf die Beobachterin Nadine oder ging knapp an ihr vorbei.
- Bei Item 15d „Was passiert mit dem Bild, wenn der Schirm entfernt wird?“ und 15e „In einer Versuchsanordnung sind eine Glühlampe, eine Sammellinse und ein Schirm so montiert, dass ein vergrößertes, umgekehrtes, scharfes Bild des Glühfadens entsteht: Wo ist das Bild, nachdem der Schirm weggenommen wurde?“ könnte der Begriff des Bildes problematisch gewesen sein. Möglicherweise wurde der Ort an dem sich die Strahlen gemäß der Modellvorstellung vom Lichtstrahl treffen nicht als „Bild“ interpretiert, da der Beobachter bzw. der Schirm fehlt, weshalb Schüler, denen klar ist, dass sich die Strahlen auch im Raum treffen, diesen Sachverhalt nicht unter dem Begriff „Bild“ subsumieren, sodass zwischen diesen Schülern und Schülern, welche diesen Sachverhalt nicht verstanden haben, nicht ausreichend differenziert werden kann.

In der Diskussion mit einer externen Arbeitsgruppe ergab sich zudem folgendes Problem bezüglich des Item 1 („Du siehst hier auf dem Bild eine brennende Kerze. Wo ist das Licht? Schraffiere den Bereich / die Bereiche, in dem das Licht deiner Meinung nach ist, mit einem Stift“) wurde lediglich mit der vollen Punktzahl bewertet wenn der Schattenraum unterhalb der Kerze ausgespart wurde. Hier sollten die Auswertungskriterien überarbeitet werden, da durch die Rückstreuung auch der Schattenraum erhellt ist, weshalb der Körper der Kerze zu sehen ist.

Aus den genannten Gründen wurden die Items 1, 7 und 15d und 15e von der Gesamtwertung ausgeschlossen. Cronbachs Alpha²⁹ als Maß für die interne Konsistenz und Schätzer für die Reliabilität verbessert sich unter Ausschluss der Items 1, 7, 15d und 15e systematisch in der Post- und der Follow-up Messung (vgl. Tabelle 15). Der Ausschluss weiterer Items auf Basis der Ergebnisse der Faktorenanalyse, welche in dem folgenden Kapitel vorgestellt wird, führt zu einer weiteren Verbesserung der internen Konsistenz. Insgesamt wurden 11 Items beibehalten, welche in die Errechnung der Gesamtpunktzahl eingingen (maximal erreichbare Höchstpunktzahl: 22 Punkte). Insgesamt kann die interne Konsistenz mit .79 zum zweiten Messzeitpunkt unter Ausschluss der problematischen Items in der Stichprobe „SV“ als gut und in der Gesamtstichprobe mit .75 als zufriedenstellend eingestuft werden. Die Test-Retest Reliabilität kann ebenfalls als zufriedenstellend bewertet werden. Da aufgrund des geringen Vorwissens der Schüler keine zeitliche Stabilität des Merkmals gegeben war, wurde davon abgesehen, die Retest-Reliabilität prä – post und prä – follow-up zu berechnen.

29 Da es sich um ordinalskalierte Daten handelt, wurde Cronbachs Alpha basierend auf einer polychorischen Korrelationsmatrix (vgl. Eid, Gollwitzer & Schmitt, 2010, S. 922) berechnet. Erläuterungen zur Anwendung dieses Verfahrens finden sich in Kapitel 2.2.6.3 Exkurs: Berechnung einer polychorischen Korrelationsmatrix für ordinalskalierte Daten.

Tabelle 15 Reliabilitätsstatistiken für den Konzepttest

	Gesamtstichprobe			Stichprobe „SV“		
	Prätest (N = 935)	Posttest (N = 869)	Follow-up (N = 922)	Prätest (n = 491)	Posttest (n = 480)	Follow-up (n = 486)
α (21 Items)	.37	.69	.64	.42	.75	.71
α (17 Items) ^a	.40	.75	.66	.42	.77	.72
α (11 Items) ^b	.45	.75	.70	.51	.79	.76
<i>r</i> _{Test-Retest} (21 Items)		post – follow-up			post – follow-up	
		.62 ***			.60 ***	
(17 Items) ^a		.62 ***			.62 ***	
(11 Items) ^b		.64 ***			.63 ***	

*** $p < .001$

^a unter Ausschluss der Items: 1, 7, 15d, 15e

^b Auf Basis späterer Analysen wurden weitere sechs Items ausgeschlossen (5, 8, 12, 13, 15f, 16), siehe Kapitel 2.3.5.5 Kreuzvalidierung des Konzepttests.

2.3.5.4 Raschanalyse des Konzepttests

Zur weiteren Analyse des Testinstruments wurde der Test auf Raschskalierbarkeit geprüft. Die zugrundeliegenden Daten basieren auf der Gesamtstichprobe. Hierzu wurde pro Messzeitpunkt ein „Ein-Parameter-Logistisches-Modell für dichotome Items“ aufgestellt. Die Analyse wurde mit dem freien Statistikprogramm R unter Verwendung des Softwarepackage eRm von Poinstingl, Mair und Hatzinger (2007) durchgeführt. Da in den Items 0, 1 oder 2 Punkte erreicht werden konnten, wurden die Items nachträglich wie folgt dichotomisiert:

- Vollständig richtig gelösten Antworten wurde der Wert 1 zugeordnet.
- Antworten, die mit einem Punkt bewertet worden waren, weil die teilnehmende Person sowohl die richtige Antwort als auch einen Distraktor gewählt hatte, wurden mit 0 Punkten bewertet.
- Antworten, die zuvor mit 0 Punkten bewertet wurden, wurden mit 0 Punkten belassen.

Zur Modellprüfung wurde ein Likelihood-Quotienten-Test (Andersen, 1973, zit. n. Bühner 2011, S. 531). Die χ^2 -Prüfgröße ergibt sich daraus, dass der Wert des Likelihood-Quotienten-Tests in eine approximative Prüfgröße überführt wird, welche einer χ^2 -Verteilung folgt (vgl. Bühner, 2011, S. 532). Die Stichprobe wird zunächst in zwei Gruppen geteilt. Pro Gruppe werden sodann erneut die spezifischen Item- und Personenparameter geschätzt. Aus den Parameterschätzungen werden

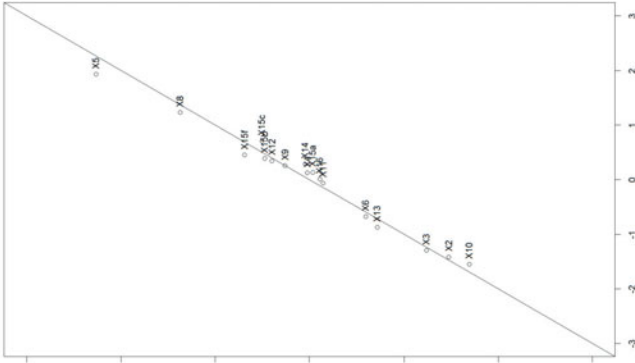
pro Gruppe die Likelihoods bestimmt und der entsprechende Quotient berechnet und in einen empirischen χ^2 -Wert überführt. Im Vergleich mit einem kritischen χ^2 -Wert, welcher die jeweilige Anzahl an Freiheitsgraden aufweist, kann man testen, ob das Modell abzulehnen ist oder nicht (vgl. ebd.). Das Modell ist zu verwerfen, wenn bei einer Teilung des Datensatzes signifikante Unterschiede zwischen den beiden Teildatensätzen auftreten. Als Teilungskriterium wurden zum ersten ein Median-Split, zum zweiten der arithmetische Mittelwert und zum dritten eine zufällige Teilung herangezogen. Die Ergebnisse des Andersen Tests zeigen, dass das Rasch-Modell im Hinblick auf die ersten beiden Betrachtungen (Mittelwert- und Median-Split) abgelehnt werden muss (siehe Tabelle 16).

Tabelle 16 Ergebnisse des Likelihood-Quotienten-Test

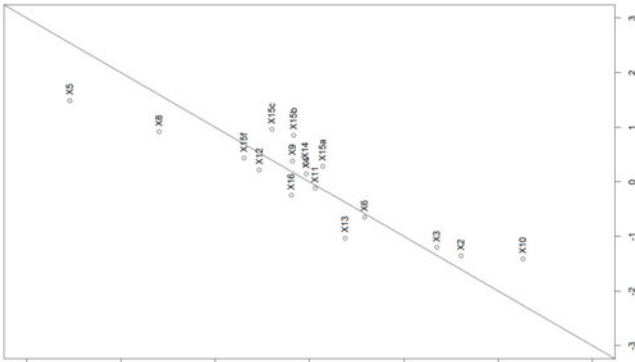
Teilungs- kriterium	Prätest (N = 935)			Posttest (N = 869)			Follow-up Test (N = 922)		
	LR-Wert	df	p	LR-Wert	df	p	LR-Wert	df	p
<i>Median</i>	50.00	16	< .001	122.19	16	< .001	121.99	16	< .001
<i>Mittelwert</i>	43.43	16	< .001	122.19	16	< .001	121.99	16	< .001
<i>zufällige Teilung</i>	18.44	16	= .299	11.95	16	= .746	9.87	16	= .874

Auch im grafischen Modell-Test zeigen sich zu allen drei Messzeitpunkten in den ersten beiden Betrachtungen (Teilungskriterien: Median und Mittelwert) deutliche Abweichungen, so dass nicht von der Personenhomogenität ausgegangen werden kann (siehe Abbildung 43).

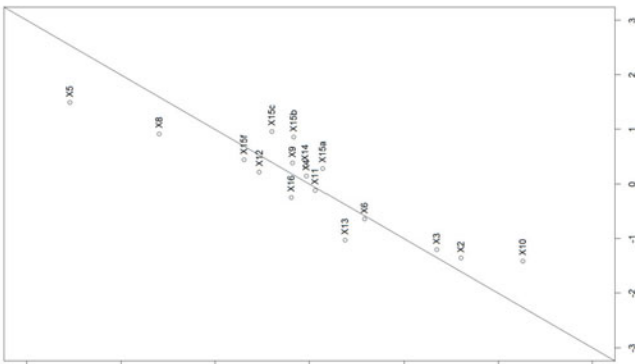
Der grafische Modelltest nutzt die Basisannahme der Personenhomogenität. Sofern das Rasch-Modell zutrifft, misst das zu analysierende Testinstrument in allen beliebigen Teilstichproben dieselbe Fähigkeit. Entsprechend sollten die Schätzungen der Itemparameter für jede Teilstichprobe gleich ausfallen. Grafisch lässt sich diese in einem Streudiagramm darstellen: Fallen die Schätzungen der Itemparameter in den jeweiligen Teilstichproben gleich aus, befinden sich die Itemparameter auf der Diagonalen des Diagramms (vgl. Bühner, 2011, S. 539). Exemplarisch wird hier der grafische Modelltest für den zweiten Messzeitpunkt abgebildet. Für die übrigen beiden Messzeitpunkte ergibt sich jedoch ein ähnliches Bild: Deutliche Abweichungen von der Diagonalen zeigen sich v.a für die Items 5, 10 und 15a-c. Lediglich bei einer zufälligen Teilung kann die Annahme des Raschmodells für den ersten und den zweiten Messzeitpunkt nicht ausgeschlossen werden.



Zufälliges Teilungskriterium



Beta für die Gruppe, Rohwerte \geq Mittelwert



Beta für die Gruppe, Rohwerte \geq Median

Abbildung 43: Grafischer Modelltest: Messzeitpunkt 2 – Posttest

Ein möglicher Grund für die Ablehnung des Modells kann in der Verletzung der Eindimensionalitätsannahme bestehen. Um dies zu prüfen, wurde – entsprechend den Empfehlungen von Bühner (2011) – im nächsten Schritt der Martin-Löf-Test durchgeführt, bei dem es sich um einen „modifizierten Likelihood-Quotienten Test“ handelt (vgl. Rost 2004, zit. n. ebd., S. 538). Anstelle die Personen zwei unterschiedlichen Teilstichproben zuzuweisen, werden im Martin-Löf-Test die Items aufgeteilt und anschließend die beiden Testhälften miteinander verglichen. Unter der Voraussetzung, dass das Rasch-Modell gilt, sollten die Items homogen sein und beide Testhälften die gleiche Fähigkeit erfassen. Auch bei diesem Test wird eine empirisch χ^2 -verteilte Prüfgröße mit einem kritischen χ^2 -Wert verglichen. Fällt die Prüfgröße größer oder gleich als der kritische Wert aus, ist die Eindimensionalitätsannahme verletzt. Als Teilungskriterium wurde für diese Analyse der Median und der Mittelwert der Itemschwierigkeit herangezogen. Wie sich an den Werten ablesen lässt (vgl. Tabelle 17), ist die Annahme der Itemhomogenität zu keinem der drei Messzeitpunkte mit Ausnahme des Konzeptposttests (Teilungskriterium Mittelwert) erfüllt.

Tabelle 17 Ergebnisse des Martin-Löf-Tests

Teilungs- kriterium	Prätest (N = 935)			Posttest (N = 869)			Follow-upTest (N = 922)		
	LR- Wert	df	p	LR-Wert	df	p	LR- Wert	df	p
Median	57.77	71	= .871	61.09	71	= .793	78.32	71	= .258
Mittelwert	48.10	69	= .974	90.10	69	= .021	68.66	69	= .354

Analog zum Vorgehen beim Leistungstest wurde zudem für jeden der drei Messzeitpunkte getrennt ein Generalfaktorenmodell (konfirmatorische Faktorenanalyse) aufgestellt, bei dem alle Variablen auf einen Faktor laden. Das Generalfaktorenmodell basiert auf einem nachträglich dichotomisierten Datensatz, welcher auch für die Rasch-Analyse verwendet wurde.

Da es sich um ordinalskalierte Daten handelt, wurde der konfirmatorischen Faktorenanalyse eine polychorische Korrelationsmatrix zugrunde gelegt, welche ebenfalls in R berechnet wurde (vgl. Eid et al., S. 922).

Die Ergebnisse der konfirmatorischen Faktorenanalyse bestätigen die Verletzung der Eindimensionalitätsannahme. So musste das Generalfaktorenmodell für jeden der drei Messzeitpunkte verworfen werden. Zum Beleg werden exemplarisch wieder die Fit-Indizes des Generalfaktorenmodells für den zweiten

Messzeitpunkt aufgeführt: $\chi^2_{(119)} = 1388.47, p < .001; CFI = 0.60; TLI = 0.54; RMSEA = 0.11; SRMR = 0.10$. Die Fit-Indizes der beiden anderen Messzeitpunkte weichen hierbei geringfügig von den Fit-Indizes für den zweiten Messzeitpunkt ab.

Unter Berücksichtigung der vorgestellten Analysen: Likelihood-Quotienten-Test, grafischer Modelltest, Martin-Löf-Test und des Generalfaktorenmodells der konfirmatorischen Faktorenanalyse, muss die Annahme, dass das Raschmodell gilt, insgesamt verworfen werden. Im Hinblick auf eine künftige Weiterentwicklung des Tests werden an dieser Stelle die Ergebnisse für den zweiten Messzeitpunkt (Posttest) berichtet (vgl. Tabelle 18 und Abbildung 44).

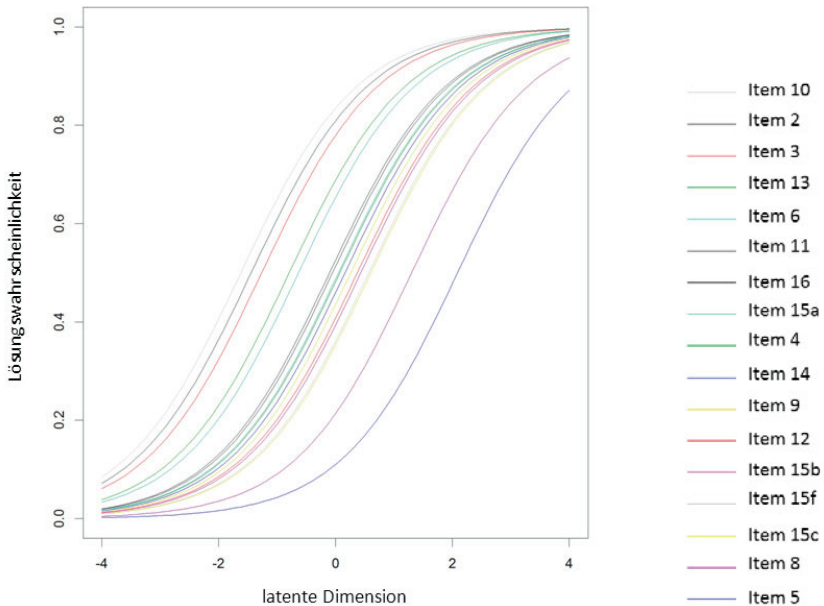


Abbildung 44: Item Characteristic Curve ICC. Die Kurven von links nach rechts entsprechen den Items von oben nach unten (also 10 oben bis 5 unten) in der Legende.

Tabelle 18 Parameterschätzungen der Itemschwierigkeiten für den Konzeptposttest, ($N = 869$)

Item	<i>Eta</i>	<i>SD</i>	Untergrenze KI^a	Obergrenze KI^a
2	-1.45	-	-	-
3	-1.26	0.07	-1.41	-1.12
4	0.07	0.07	-0.07	0.22
5	2.09	0.13	1.83	2.35
6	-0.63	0.07	-0.77	-0.50
8	1.30	0.10	1.11	1.50
9	0.26	0.08	0.11	0.41
10	-1.62	0.08	-1.77	-1.47
11	-0.10	0.07	-0.25	0.04
12	0.37	0.08	0.22	0.53
13	-0.79	0.07	-0.93	-0.66
14	0.16	0.08	0.01	0.31
15a	0.05	0.07	-0.10	0.20
15b	0.43	0.08	0.28	0.59
15c	0.60	0.08	0.44	0.77
15f	0.57	0.08	0.41	0.73
16	-0.05	0.07	-0.20	0.09

^aKonfidenzintervall (KI) 95 %

Die Normierung des Modells wurde so vorgenommen, dass die Summe der *Etas* aller Items 0 ergibt (vgl. Tabelle 18). Der Wert des ersten Items (hier Item2), ist somit durch die Werte der anderen Items festgelegt (vgl. Poinstingl et al., 2007).

Die Ergebnisse der vorangegangenen Analysen weisen darauf hin, dass es sich beim Konzepttest um ein mehrdimensionales Testinstrument handelt, das unterschiedliche Fähigkeitsfacetten oder möglicherweise sogar voneinander unabhängige Fähigkeiten erfasst.

Um die Dimensionalität des Tests zu analysieren, wurde das Instrument im nächsten Schritt mittels exploratorischer und konfirmatorischer Faktorenanalyse kreuz validiert.

2.3.5.5 Kreuzvalidierung des Konzepttests

Im Gegensatz zur Rasch-Analyse wurden beiden Analysen der ursprüngliche ordinalskalierte Datensatz zu Grunde gelegt, welcher die Abstufungen 0, 1 und 2 Punkte enthielt, um möglichst viele Informationen beizubehalten.

Das Modell, welches sich aus der exploratorischen Faktorenanalyse ergab, wurde anschließend mittels der Methode der konfirmatorischen Faktorenanalyse überprüft. Um die Ergebnisse der exploratorischen und der konfirmatorischen Faktorenanalyse miteinander in Beziehung setzen zu können, fiel die Wahl in beiden Fällen auf eine Maximum-Likelihood-Faktorenanalyse.

Das Vorgehen umfasste mehrere Schritte:

1. Die Stichproben beider Studien „SV“ und „Ko“ wurden gepoolt ($N_{prä} = 935$, $N_{post} = 869$, $N_{follow-up} = 922$).
2. Der Datensatz der Gesamtstichprobe wurde zufällig in zwei Hälften geteilt, so dass zu jedem der drei Messzeitpunkte in etwa gleich viele Fälle vorhanden sind. Es wurde geprüft, ob sich die beiden Hälften hinsichtlich der interessierenden Merkmale signifikant unterscheiden. Da dies nicht der Fall war, wurden beide Datensätze für die Kreuzvalidierung verwendet.
3. Es wurde eine exploratorische Faktorenanalyse mit der ersten Hälfte des Datensatzes durchgeführt und zwar zu jedem der drei Messzeitpunkte.
4. Das ermittelte exploratorische Modell wurde an dem zweiten Datensatz je Messzeitpunkt überprüft.
5. Die Schritte 3) und 4) wurden vice versa wiederholt.

Detaillierte Angaben zu den Teilstichproben finden sich in Tabelle 14 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com.

Aus Gründen der Übersichtlichkeit werden an dieser Stelle lediglich die Ergebnisse für den Posttest berichtet. Die Analyseergebnisse konnten für den dritten Messzeitpunkt bestätigt werden (Follow-up Messung). Lediglich zum ersten Messzeitpunkt ergaben sich inhaltliche Abweichungen, welche jedoch angesichts der geringen Vorkenntnisse der Schüler und der daraus resultierenden geringen Reliabilität des Tests nicht überbewertet werden sollten.

Beide Analysen wurden mit der freien Statistiksoftware R unter Verwendung der Pakete psych (Revelle, 2013), lavaan (Rosseel, 2012) und qgraph (Epskamp et al., 2012) durchgeführt. Der Vorteil die freie Statistiksoftware R auch für die Durchführung der exploratorischen Faktorenanalyse zu verwenden, besteht darin, dass R im Gegensatz zu SPSS die Möglichkeit bietet, der Analyse eine polychorische Korrelationsmatrix für ordinalskalierte Daten zugrunde zu legen, welche ebenfalls in R berechnet wurde.

Bei der Durchführung der Analysen zeigte sich, dass (zusätzlich zu den Item 1, 7, 15e-d) weitere Items ausgeschlossen werden mussten, um Doppelladungen zu vermeiden und die Ergebnisse der EFA auch inhaltlich sinnvoll interpretieren zu können. Dies betraf die Items 5, 8, 11, 13 15f, 16 welche bereits in

den Itemstatistiken durch eine vergleichsweise geringe Trennschärfe aufgefallen waren. Von ursprünglich 21 verbleiben nunmehr 11 Items, aus denen sich ein Gesamtscore von maximal 22 Punkten errechnet.

Ausgehend von der Annahme, dass die Faktoren untereinander korrelieren können, wurde in der exploratorischen Faktorenanalyse die Promax-Rotation (Kappa = 4) gewählt. Der Bartlett-Test belegt, dass die Korrelationsmatrix signifikant von der Einheitsmatrix verschieden ist ($\chi^2 = 1072.14$, $df = 55$, $p < .001$) und somit faktorisiert werden kann. Der Verlauf der Eigenwerte und das Ergebnis der Parallelanalyse nach Horn (1965) weisen darauf hin, dass die ersten zwei bis vier Faktoren die Hauptinformationen der Daten repräsentieren (vgl. Abbildung 45).

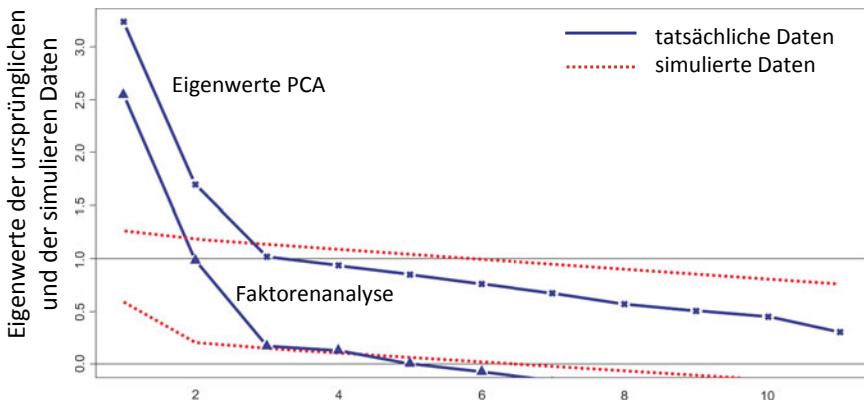


Abbildung 45: Screeplot inklusive Parallelanalyse nach Horn (1965) für den Konzeptposttest

Die Kommunalitäten liegen mit $0.14 \leq 0.75$ in einem für die Stichprobe von $n = 430$ und drei bis fünf Items pro Faktor in einem akzeptablen bis guten Bereich (vgl. Bühner, 2011, S. 345), siehe auch Tabelle 19.

Tabelle 19 Ergebnisse der exploratorischen Faktorenanalyse des Konzeptpostests, (n = 430)

Variable	Itemformulierung	Promax Rotation Mustermatrix			Kommunalität
		F 1	F 2	F3	$\hat{H} (y_i)$
Y ₂	Welche der folgenden Gegenstände / Tiere kann man in einem völlig abgedunkelten Raum sehen?				
	... ein leuchtendes Glühwürmchen				
	... ein weißes Blatt Papier	-0.18	0.42	0.22	0.22
	... einen Fahrrad-Reflektor die Augen einer Katze				
Y ₃	Hat es einen Einfluss auf die Helligkeit in einem Zimmer, ob es helle oder dunkle Tapeten hat?				
	ja, weil die helle Tapete mehr Licht streut, das ins Auge fällt, als eine dunkle Tapete.				
	nein, weil dunkle Tapeten nichts an der Menge des Lichtes im Raum ändern.	0.20	0.51	-0.23	0.31
	ja, weil auf der hellen Tapete mehr Licht liegen bleibt. nein, es kommt auf die Lampe in dem Zimmer an oder das Sonnenlicht, das durch das Fenster fällt und nicht auf die Helligkeit der Tapete.				
	Was ist richtig? Lichtstrahlen sind etwas Wirkliches, so wie dünne Wasserstrahlen aus einer Spritzpistole.				
Y ₄	... etwas Gedachtes, so wie Konstruktionen in der Geometrie, um z.B. Dreiecks-Probleme lösen zu können.	0.10	0.17	0.20	0.14
	... exakt das gleiche wie Lichtbündel. Lichtbündel sind etwas Gedachtes, z.B. um die Bildgröße bestimmen zu können.				
Y ₆	Kreuze an, wo sich das Spiegelbild für Dich als Betrachter befindet:				
	... vor dem Spiegel				
	... hinter dem Spiegel	-0.02	0.50	-0.14	0.22
	... im Spiegel ... auf dem Spiegel				
Y ₉	In einem abgedunkelten Raum ist der helle Fleck einer Taschenlampe an der Wand zu sehen, nicht aber der Lichtstrahl zwischen Taschenlampe und Wand. Warum?				
	... Erst das an Gegenständen gestreute Licht trifft ins Auge und ist sichtbar.				
	... In dem dunklen Raum wird das Licht absorbiert (verschluckt), daher ist es nicht zu sehen.	0.01	0.45	-0.01	0.20
	... Das Licht erhellt die Wand, weil es auf ihr liegen bleibt. Das Licht der Taschenlampe entfernt sich vom Beobachter, erst durch die Wand wird es umgedreht und geht auf den Beobachter zu.				
Y ₁₀	Was passiert, wenn man in dem Lichtstrahl einen Tafellappen aufschüttelt?				
	... Die Staubteilchen wirken wie kleine Linsen, die das Licht auf der Wand bündeln.	-0.06	0.67	0.16	0.52
	... Der feine Kreidestaub sammelt das Licht und dadurch sieht man den hellen Fleck auf der Wand nicht mehr.				
	... Die Staubteilchen werden durch das auftreffende Licht durcheinander gewirbelt.				

	... Die Staubteilchen streuen das Licht in alle Richtungen, dadurch trifft es ins Auge und wird sichtbar.				
Y ₁₁	Wie entsteht durch Verwendung einer Sammellinse ein Bild, das auf einem Schirm aufgefangen werden kann?				
	... Solch ein Bild entsteht durch Spiegelung der Lichtstrahlen an der Linse nach dem Reflexionsgesetz.				
	... Eine Sammellinse hat den Effekt, die Lichtstrahlen aufzuhellen. Lichtstrahlen, die von einem Gegenstandspunkt ausgehen, werden durch die Sammellinse abgelenkt und treffen sich im Bildpunkt.	0.06	0.10	0.29	0.14
	... Das Bild geht als Ganzes durch die Linse zum Schirm, dabei wird es in der Linse unter Einhaltung der Linsengesetze umgedreht.				
Y ₁₄	Welche Aussagen zur Bildkonstruktion und Bildentstehung treffen zu?				
	... Nur die ausgezeichneten Strahlen kann man im Strahlengang zeichnen.				
	... Mit den ausgezeichneten Strahlen kann man den Strahlengang besonders leicht zeichnen.	0.05	-0.03	0.69	0.50
	... Die ausgezeichneten Strahlen erschweren die Zeichnung, machen sie dafür aber besonders genau.				
	... Ohne die ausgezeichneten Strahlen (wenn diese z.B. durch dünne Stifte aufgehalten werden) kann es kein Bild geben.				
Y _{15a}	In einer Versuchsanordnung sind eine Glühlampe, eine Sammellinse und ein Schirm auf einer optischen Bank so montiert, dass ein vergrößertes, umgekehrtes, scharfes Bild des Glühfadens entsteht: Was passiert, wenn die untere Hälfte der Linse abgedeckt.				
	... Die obere Hälfte des Bildes wird abgeschnitten.	0.72	-0.08	0.13	0.58
	... Die untere Hälfte des Bildes wird abgeschnitten.				
	... Das Bild wird dunkler.				
	... Das Bild wird kleiner.				
Y _{15b}	Was passiert, wenn man einen Karton mit großem Loch (ringförmige Blende) vor die Linse hält?				
	... Das Bild wird kleiner.				
	... Das Bild wird dunkler.	0.88	0.01	-0.03	0.75
	... Die Ränder des Bildes werden kreisförmig abgeschnitten.				
	... Das Bild wird heller.				
Y _{15c}	Was passiert, wenn man einen Karton mit einem sehr kleinen Loch 5 mm (ringförmige Blende) vor die Linse hält?				
	... Das Bild wird kleiner.				
	... Das Bild wird dunkler.	0.73	0.09	0.05	0.64
	... Die Ränder des Bildes werden kreisförmig abgeschnitten.				
	... Das Bild wird heller.				

$TLI = 0.91$, $RMSR = 0.03$, $RMSEA = 0.06$

Es zeigt sich, dass die Items 15a-c einen höheren und die Items 4, 11 und 14 einen eher geringeren Varianzanteil aufklären (vgl. Tabelle 20).

Tabelle 20 Eigenwerte der Faktoren, Reliabilität und Anteil aufgeklärter Varianz der resultierenden Skalen für die Faktorenanalyse des Konzeptposttests, ($n = 430$)

F ^a	Interpretation	Variablen	α	Anfänglicher Eigenwert	Anteil erklärter Varianz in %	Anteil kumulierter Varianz in %
1	Verständnisfragen zur Bildentstehung inklusive Abdeckaufgaben	$y_{15a}, y_{15b}, y_{15c}$.81	3.39	18	18
2	Geradlinige Lichtausbreitung und Streuung	$y_2, y_3, y_6, y_9, y_{10}$.65	1.78	13	31
3	Verständnis Bildkonstruktion / Strahlenmodell	y_4, y_{11}, y_{14}	.49	1.37	7	38

^aFaktor

Insgesamt werden 38 % der Varianz durch die drei Faktoren aufgeklärt.

Im zweiten Schritt wurden die Ergebnisse der exploratorischen Faktorenanalyse nun mit der zweiten Hälfte des Datensatzes ($n_{post} = 430$) kreuzvalidiert.

Auf Basis des Modells, das sich an der exploratorischen Faktorenanalyse orientierte, waren die Items, wie folgt, zugeordnet:

- Faktor 1 („Verständnisfragen zur Bildentstehung inklusive Abdeckaufgaben“) = $y_{15a}, y_{15b}, y_{15c}$
- Faktor 2 („geradlinige Lichtausbreitung und Streuung“) = $y_2, y_3, y_6, y_9, y_{10}$
- Faktor 3 („Verständnis Bildkonstruktion / Strahlenmodell“) = y_4, y_{11}, y_{14}

Um zu analysieren, wie die verschiedenen Skalen zusammenspielen, wurde das Modell nun auf unterschiedliche Weise spezifiziert:

- Basismodell mit Korrelationen zwischen Faktoren: Zunächst wurde ein Modell aufgestellt, bei dem die jeweiligen Items auf einen Faktor laden. Das Programm schätzte dabei Korrelationen zwischen den drei Faktoren.
- Bifaktormodell: In einem nächsten Schritt wurde ein bifaktorielles Modell aufgestellt. In diesem Modell wird angenommen, dass ein Generalfaktor existiert, der die Bearbeitung aller Items beeinflusst. Darüber hinaus bestehen jedoch auch latente Merkmale, die über das Merkmal, welches eine übergreifende Kompetenz misst, hinaus auch noch voneinander unabhängige Merkmale erfassen. Korrelationen zwischen den Faktoren wurden daher unterbunden.
- Second-Order-Modell: Im letzten Schritt wurde ein Second-Order-Modell aufgestellt. Dieses Modell geht davon aus, dass ein „generelles“ latentes Merkmal besteht, welches sich wiederum auf alle spezifischen latenten Merkmale auswirkt, die voneinander unabhängig und daher untereinander unkorreliert sind.

Tabelle 21 Konfirmatorische Faktorenanalyse Konzeptposttests – Modellvergleich des empirisch aus der Faktorenanalyse abgeleiteten Modells ($n = 439$)

Kriterium (globale Fitindizes)	Basismodell mit Korrelationen	Bifaktorielles Mo- dell	Second-Order- Modell
$\chi^2_{(df)}$	104.17 ^{(41)*}	104.27 ^{(41)*}	104.34 ^{(42)*}
<i>CFI</i>	0.94	0.94	0.94
<i>TLI</i>	0.92	0.92	0.92
<i>RMSEA</i>	0.06	0.08	0.06
<i>SRMR</i>	0.05	0.07	0.05
<i>AIC</i>	12761.63	12761.63	12759.70
<i>BIC</i>	12863.74	12863.74	12857.72

* $p < .05$

Ein Vergleich der Fit-Indizes zeigt, dass alle Modelle sehr ähnliche globale Fit-Indizes aufweisen (vgl. Tabelle 21). Die angegebenen globalen Fit Indizes der Modelle liegen jeweils unter bzw. über den in der Literatur angegebenen Cut-off-Werten (vgl. Bühner, 2011, S. 428). Nach dem Akaike- und dem Bayes-Informationskriterium schneidet jeweils das Second-Order-Modell am besten ab (geringste Werte). Hinsichtlich der lokalen Fit-Werte erwies sich die Modellspezifikation, bei der die jeweiligen Items auf einen Faktor laden und die Faktoren untereinander korrelieren, am geeignetsten (Basismodell mit Korrelationen), da im Gegensatz zu der bifaktoriellen Variante und dem Second-Order-Modell jeder der betreffenden Pfadkoeffizienten signifikant ist. Die Varianzaufklärung je Item unterscheidet sich zwischen Second-Order-Modell und Basismodell erst ab der dritten Nachkommastelle. Die lokalen Fit-Indizes sprechen insgesamt für das Basismodell mit Korrelationen.

Letzteres kann inhaltlich wie folgt interpretiert werden (vgl. Abbildung 46): Es bestehen drei Dimensionen des konzeptuellen Verständnisses im Bereich der Strahlenoptik:

- AB: Verständnisfragen zur Bildentstehung inklusive Abdeckaufgaben,
- LS: geradlinige Lichtausbreitung und Streuung,
- BS: Verständnis Bildkonstruktion / Strahlenmodell.

Alle drei Dimensionen tragen zum Verständnis der Bildentstehung mittels Sammellinse bei. Sie hängen inhaltlich miteinander zusammen, was sich in den wechselseitigen signifikanten Korrelationen zeigt. Dabei liegen die Korrelationen ($0.42 \leq r \leq 0.71$) im mittleren bis hohen Bereich (siehe ebenfalls Abbildung 46).

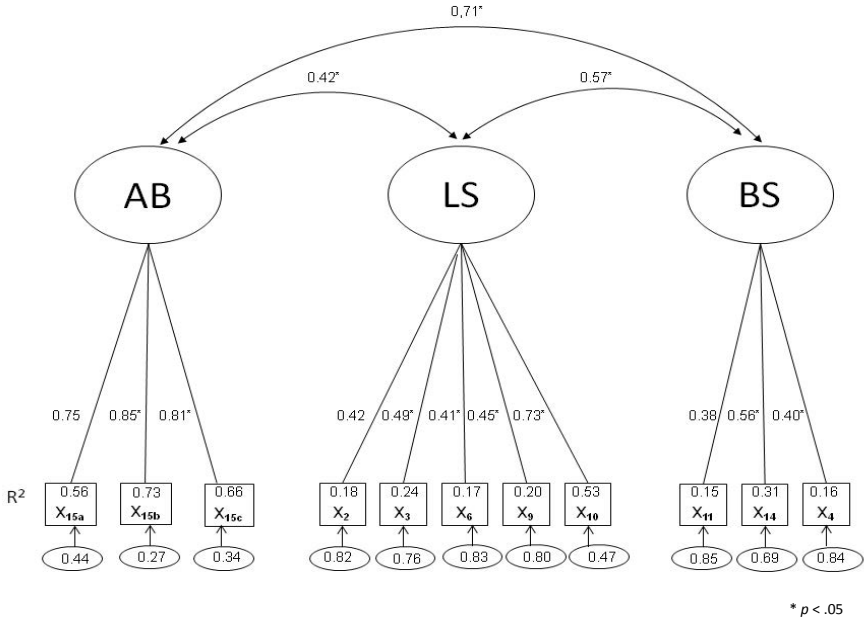


Abbildung 46: Konfirmatorische Faktorenanalyse aus der EFA empirisch abgeleitetes Basismodell mit Korrelationen: standardisierte Lösung, AB = Verständnisfragen zur Bildentstehung inklusive Abdeckaufgaben, LS = geradlinige Lichtausbreitung und Streuung, BS: Verständnis Bildkonstruktion / Strahlenmodell,

Während die globalen Gütekriterien in einem guten bis akzeptablen Bereich liegen, sind die lokalen Gütekriterien der Indikatorreliabilität und der diskriminanten Validität (Formel-Larker-Kriterium) für die beiden Faktoren „LS: geradlinige Lichtausbreitung und Streuung“, und „BS: Verständnis Bildkonstruktion / Strahlenmodell“ nicht zufriedenstellend erfüllt (vgl. Tabelle 22).

Als Anhaltspunkt für die Kriteriumsvalidität wurden des Weiteren Korrelationen zwischen der Gesamtpunktzahl aus den verbleibenden 11 Items und den Fachnoten sowie relevanten Subskalen des I-S-T 2000 R für die letzten beiden Messzeitpunkte berechnet.

Im Ergebnis zeigen sich signifikante mittlere Korrelationen zwischen den Fachnoten in Physik ($r_{\text{post_PH}}(479) = .42, p < .001$, $r_{\text{follow-up-PH}}(476) = .38, p < .001$) und Mathematik ($r_{\text{post_Mathe}}(476) = .38, p < .001$, $r_{\text{follow-up-Mathe}}(483) = .40, p < .001$) und den Gesamtpunktzahlen im Post- und Follow-up Test. Es ergeben sich jedoch nur

geringe Korrelationen ($r < .15$) zwischen der Gesamtpunktzahl post und follow-up und den Subskalen des figural-räumlichen und figural-logischen Schlussfolgern.

Tabelle 22 Lokale Gütekriterien der CFA des Konzeptposttests, ($n = 439$)

F	Interpretation	Item	Indikatorreliabilität	Faktor-reliabilität	DEV	Formel Larcker Kriterium
AB ^a	Verständnisfragen zur Bildentstehung inklusive Abdeckaufgaben	Y _{15a}	0.57	.83	0.62	> 0.50
		Y _{15b}	0.73			
		Y _{15c}	0.66			
LS ^b	Geradlinige Lichtausbreitung und Streuung	Y ₂	0.18	.77	0.42	.64 (nicht erfüllt)
		Y ₃	0.24			
		Y ₆	0.17			
		Y ₉	0.20			
		Y ₁₀	0.53			
BS ^c	Verständnis Bildkonstruktion / Strahlenmodell	Y ₄	0.15	.55	0.30	.50 (nicht erfüllt)
		Y ₁₁	0.31			
		Y ₁₄	0.16			

^aAB = Verständnisfragen zur Bildentstehung inklusive Abdeckaufgaben

^bLS = geradlinige Lichtausbreitung und Streuung

^cBS = Verständnis Bildkonstruktion / Strahlenmodell

2.3.5.6 Itemstatistiken zum Motivationsfragebogen

Der Motivationsfragebogen basiert auf Items, die sich in bisherigen Studien bewährt haben. Somit handelt es sich anders als beim Leistungstest und beim Konzepttest nicht um ein im Rahmen des Projekts entwickeltes Instrument, sondern um eine auf die Fragestellung zugeschnittene Auswahl und Zusammenstellung von Skalen.

Um das Zusammenspiel dieser Skalen zu untersuchen, wird die Itemanalyse des Motivationsfragebogens im Überblick vorgestellt, zumal die Analyse der Wirkung der Intervention auf die Motivation einen Nebenaspekt dieser Arbeit darstellt. Zur Berücksichtigung der Schülermotivation wurden in die Gesamtpunktzahl die zuvor genannten Skalen SK (Selbstkonzept), IE (Intrinsische Motivation / Engagement und GN (Gute Noten - extrinsische Motivation) einbezogen. Die Skala LES: „wahrgenommenes Lehrerengagement aus Schülersicht“ (abgewandelt

nach Seidel et al., 2003) wurde aus der Gesamtwertung der Schülermotivation herausgenommen und ging als gesonderte Kovariate in die Analyse ein.

Die Analyse der Itemschwierigkeit und Mittelwerte zeigt, dass alle Items (inklusive der Items zu LES) zu allen drei Messzeitpunkten in einem guten Bereich von $0.20 \leq P_i \leq 0.80$ bezüglich der Itemschwierigkeit lagen (vgl. Tabelle 15 und 16 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com). In Bezug auf die korrigierte Trennschärfe ergaben sich zufriedenstellende bis gute Werte. So liegt das Minimum bei $r_{it} = .28$ im Follow-up Test und das Maximum jeweils bei $r_{it} = .79$ im Posttest (vgl. Tabelle 16 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com). Cronbachs Alpha als Maß für die interne Konsistenz und Schätzer für die Reliabilität kann für die Schülermotivation und LES als hoch bewertet werden. Für die Test-Retest-Reliabilität ergeben sich für beide Konstrukte zufriedenstellende bis gute Werte (vgl. Tabelle 23).

Tabelle 23 Reliabilitätsstatistiken für den Motivationsfragebogen

Reliabilität- tät- schätzung	Schülermotivation (24 Items)			Lehrerengagement aus Schülersicht (4 Items)		
	prä (N = 387)	post (N = 362)	follow-up (N = 391)	prä (N = 465)	post (N = 451)	follow-up (N = 471)
α	.93	.94	.94	.88	.90	.91
$r_{\text{Test-Retest}}$	prä – post	prä – follow-up	post – follow-up	prä – post	prä – follow-up	post – follow-up
	.76***	.69***	.77***	.67***	.63***	.67***

*** $p < .001$

Abweichende Werte für die Stichprobengrößen zwischen Schülermotivation und Lehrerengagement ergeben sich aus dem listenweisen Fallausschluss, bei dem komplette Bögen ausgeschlossen wurden, sobald mindestens ein Item unbeantwortet blieb. Für die späteren Analysen wurde in diesen Fällen Daten imputiert (siehe Kapitel 2.3.9.1 Methodisches Vorgehen: Mehrebenenanalyse).

2.3.5.7 Diskussion der Ergebnisse

Diskussion der Analysen zum Leistungstest: Die Sicherung der Gütekriterien Objektivität, Reliabilität und Validität des Leistungstests zur Erfassung der Physikleistung bei repräsentationsbezogenen Aufgaben kann zusammenfassend als zufriedenstellend bis gut bewertet werden.

Die Sicherung der Durchführungsobjektivität ergibt sich aus dem standardisierten Verfahren der Instruktion der Schüler. So wurde der Test jeweils von der Autorin durchgeführt und die Durchführung in gleicher Weise in den verschiedenen Schulklassen erklärt. Die Auswertungsobjektivität wurde durch ein zuvor festgelegtes Bewertungsschema gewährleistet, in dem detailliert festgelegt wurde, welche Teilpunkte für welche Antworten oder Bestandteile der Strahlenkonstruktion vergeben waren, an welche sich die beteiligten Auswerter einschließlich der Autorin hielten.

Die Analyse der Itemstatistiken des Leistungstests belegen zufriedenstellende bis gute Werte der Kennwerte (Lösungswahrscheinlichkeit, Trennschärfe und Cronbachs-Alpha, Test-Retest-Reliabilität) für den zweiten und den dritten Messzeitpunkt. Daher kann die Reliabilität als insgesamt zufriedenstellend bis gut eingestuft werden.

Da die Schüler zuvor keinen Unterricht zur Bildentstehung bei der Sammellinse erhalten hatten, waren die niedrigen Werte in den Itemstatistiken (Mittelwert, Standardabweichung, Lösungswahrscheinlichkeit und Trennschärfe) zum ersten Messzeitpunkt zu erwarten. Für die Beurteilung, welche Items beibehalten werden, wurden daher vorwiegend die Daten der Post- und der Follow-up Messung in Betracht gezogen.

Die Items 2b und 6 zeichneten sich durch eine geringere Lösungswahrscheinlichkeit aus. Inhaltlich beziehen sich die beiden Items darauf, aus der schematischen Repräsentation der Strahlenkonstruktion Schlussfolgerungen zu ziehen und diese in Worte zu fassen.³⁰ Spezielle Voraussetzung für die Bearbeitung der Aufgaben sind Kenntnisse in Geometrie. Offenbar wurden die hierzu benötigten Fähigkeiten im Unterricht nicht im ausreichenden Maß vermittelt und eingeübt. Die Lösungswahrscheinlichkeit für Item 2b liegt hierbei mit 18 % in der Postmessung knapp unter der kritischen Grenze. Ergänzend muss jedoch hinzugefügt werden, dass es sich in beiden Fällen um Transferaufgaben handelt, die ein hohes Maß an Repräsentationskompetenz erfordern und durch unterschiedliche Argumentationswege bearbeitet werden können. Vor diesem Hintergrund war eine geringe Lösungswahrscheinlichkeit erwartbar. So ist aus der Forschung zur Problemlösekompetenz bekannt, dass Schülern die Bearbeitung von Transferaufgaben schwerfällt (vgl. Funke & Zumbach, 2006; S. 216 f. im Überblick).

30 Wortlaut des Items 2b: „Begründe mit den Strahlen in der Abbildung oben: Warum verändert sich die Bildgröße so wie oben beschrieben, wenn der Gegenstand weiter von der Linse entfernt wird?“; Wortlaut des Items 6 „Warum kann man die fett hervorgehobenen Dreiecke nicht zur Herleitung der Abbildungsgleichung verwenden?“

Als weiterer bemerkenswerter Befund zeigte sich, dass die Lösungswahrscheinlichkeit für die Wahl der korrekten Gegenstandsweite (5c - Erklärung1) deutlich weniger stark sinkt als die Lösungswahrscheinlichkeit für die Erklärung dieser Wahl (5c - Erklärung2). Gründe könnten zum einen darin liegen, dass sich die Schüler daran erinnern, an welche Stelle sie das Kreuz in der Multiple-Choice Aufgabe gesetzt haben, aber die Begründung nicht mehr herleiten können, da die Inhalte nicht nachhaltig erlernt wurden, oder aber, dass zum dritten Messzeitpunkt die Motivation fehlte, wiederholt eine konsistente Begründung zu entwickeln, zumal der Follow-up Test nicht in die Notengebung einfließt. Gegen die zweite Hypothese spricht, dass die Begründung für Aufgabenteil 5a häufiger korrekt entwickelt wurde. Da die Itemschwierigkeit für Item 5c - Erklärung 2 in der Postmessung jedoch innerhalb des Toleranzbereichs liegt, wurde das Item vorerst beibehalten.

Bei der Analyse der Trennschärfe fielen die Ergebnisse für Item 7 (Anwendung der Abbildungsgleichung) auf. Schüler, die im Vortest Item 7 zumindest teilweise richtig lösten, bearbeiteten die übrigen Aufgaben im Vortest offenbar mit weniger Erfolg. Im Posttest steigt die Trennschärfe für dieses Item deutlich an und sinkt im Follow-up Test wieder. Eine mögliche Erklärung könnte darin bestehen, dass die Bearbeitung dieser klassischen Aufgaben in hohem Maß von erlerntem Wissen beeinflusst wird. Schüler, die im Vortest möglicherweise „raten“, erzielten hierbei ggf. eine Teilpunktzahl durch die Angabe der Gleichung, die ihnen z.B. durch die Lochkamera bekannt sein könnte. Andere Schüler hingegen, die erlerntes Wissen anwenden, ließen die Beantwortung des Items aus, so dass sich Item 7 bei Schülern mit sehr geringen Vorkenntnissen nicht dazu eignet, zwischen leistungsstarken und leistungsschwachen Schülern zu differenzieren. Da zwischen der Post- und der Follow-up Messung ein Zeitabstand von ca. zwei Monaten gegeben war, wurde vermutlich der Lerninhalt der Abbildungsgleichung vergessen, obgleich gegebenenfalls das Grundverständnis für die Bildentstehung erhalten blieb, was den Rückgang der Trennschärfe im Follow-up Test erklären könnte.

Die Analyse der Itemstatistiken und die Ergebnisse der Faktorenanalyse legten nahe, Aufgabe 6 („Warum kann man die fett hervorgehobenen Dreiecke nicht zur Herleitung der Abbildungsgleichung verwenden?“) von der Gesamtwertung auszuschließen, auch wenn sich hierdurch Cronbachs Alpha nicht veränderte.

Die Ergebnisse der Faktorenanalyse weisen darauf hin, dass der Test insgesamt drei Kernkompetenzen erfasst, die sich auf den Umgang mit verschiedenen Repräsentationsformen beziehen: den Umgang mit deskriptiven Repräsentationen, die Aufgabe, auf Basis einer realistisch depiktionalen Repräsentation Bezüge zu einem realen Experiment herzustellen und auf den Umgang mit schematisch depiktionalen Repräsentationen. Die Modellvergleiche belegen, dass die Art der

Repräsentationsform entscheidender ist als die Unterscheidung, ob Schülervorstellung thematisiert werden sollen oder nicht (Vergleich der 2-Faktoren- mit den 3-Faktorenmodellen).

Die guten globalen Fitwerte des Bifaktormodells und die hohen Korrelationen der Faktoren im Basismodell weisen darauf hin, dass die drei Kernkompetenzen vermutlich durch eine generelle „Repräsentations“-Kompetenz oder ein physikalisches Grundverständnis der Bildentstehung beeinflusst werden.

Die Struktur des Tests ist mit den Ausführungen im Theorieteil dieser Arbeit, insbesondere mit den Ausführungen zur Funktion multipler Repräsentationen (siehe Kapitel 1.1.4.1 Funktionen von multiplen Repräsentationen) konform, was als Beleg für die Konstruktvalidität gewertet werden kann. Die Überlegenheit der 3-Faktorenmodelle gegenüber den 2-Faktorenmodellen stützt die Bedeutung der Unterscheidung der repräsentationalen Anforderungen.

Zur Optimierung der lokalen Fit-Werte sollten insbesondere die Aufgaben zur Erfassung des Umgangs mit schematischen Repräsentationen überarbeitet werden, um die interne Konsistenz dieser Skala zu erhöhen. Zudem könnte die Skala um ein bis zwei Aufgaben erweitert werden, welche einen Transfer bei der Bearbeitung der Strahlenkonstruktion erfordern. Weitere Aufgaben, welche in dieser Studie nicht zum Einsatz kamen, finden sich in Scheid (2013).

Die curriculare Validität des Tests wurde in einem Expertenrating von Scheid (2013, S. 139 - 140) untersucht. An dem Rating nahmen 11 Gymnasiallehrer mit einer durchschnittlichen Unterrichtserfahrung von $M = 20.91$ Jahren ($SD = 13.46$ Jahre) teil, deren Urteile nach Kendalls Konkordanzkoeffizient W höchst signifikant übereinstimmten (vgl. Wirtz & Caspar, S. 155). Im Ergebnis zeigte sich, dass die Items bis auf zwei Ausnahmen, als curricular valide eingestuft wurden. Lediglich die Aufgaben 5c) und 6), wurden als für einen Leistungstest weniger geeignet bzw. als weniger curricular valide eingeschätzt und wären entsprechend von den Lehrern eher nicht im eigenen Unterricht verwendet worden (vgl. Scheid, 2013, S. 139 - 141).

2.3.5.8 Fazit zur ersten Zielsetzung: Entwicklung eines tragfähigen Konzepttests

Die Sicherung der Gütekriterien Objektivität, Reliabilität und Validität des Konzepttests zur Erfassung des konzeptuellen Verständnisses in der Strahlenoptik kann zusammenfassend ebenfalls als zufriedenstellend bis gut beurteilt werden.

Die Sicherung der Durchführungsobjektivität ergibt sich analog zum Vorgehen beim Leistungstest aus dem standardisierten Verfahren der Instruktion, welches

jeweils von der Autorin durchgeführt wurde. Die Auswertungsobjektivität wurde durch ein zuvor festgelegtes Bewertungsschema sichergestellt (siehe Kapitel 2.3.4.3 Konzepttest).

Für die Analyse der Itemstatistiken und der internen Konsistenz ergeben sich unter Ausschluss der genannten Items zufriedenstellende bis gute Werte für den zweiten und den dritten Messzeitpunkt (Lösungswahrscheinlichkeit, Trennschärfe und Cronbachs-Alpha, Test-Retest-Reliabilität) – zumal zu berücksichtigen ist, dass das konzeptuelle Verständnis von Schülern, die keine Experten sind, auch nach dem Unterricht nach wie vor von Inkonsistenzen geprägt sein dürfte. Die niedrigeren Werte im Prätest resultieren aus dem geringen Vorwissen der Schüler.

In der Analyse der Struktur belegen die Untersuchungen zur Prüfung auf Rasch-Skalierbarkeit für den nachträglich dichotomisierten Datensatz und die Prüfung der Eindimensionalitätsannahme für den ordinalskalierten Datensatz, dass es sich beim konzeptuellen Verständnis um ein mehrdimensionales Konstrukt handelt. Entsprechend kann der Test nicht Rasch-skaliert werden.

Die Idee der Rasch-Skalierung kann jedoch für die künftige Weiterentwicklung des Tests aufgegriffen werden, unter der Voraussetzung, dass insbesondere die Items 2, 3, 5, 8 und 10, welche sich vorwiegend auf Konzepte zur Lichtausbreitung und Streuung beziehen, ausgeschlossen oder deutlich überarbeitet werden.

In der Kreuzvalidierung (EFA und CFA) zeigte sich, dass der Test insgesamt drei übergeordnete Verständnisbereiche erfasst:

- Bildentstehung inklusive Abdeckaufgaben,
- geradlinige Lichtausbreitung und Streuung,
- Bildkonstruktion und Strahlenmodell.

Die Modellvergleiche belegen, dass die drei Verständnisbereiche unterschiedliche Facetten des konzeptuellen Verständnisses erfassen, welche jedoch miteinander korrelieren. Die Korrelationen sprechen für einen mittleren bis hohen Zusammenhang der drei Facetten untereinander. Die höchste Korrelation besteht zwischen den beiden Faktoren „Abdeckaufgaben“ und „Bildkonstruktion Strahlenmodell“, was als empirischer Anhaltspunkt für Wiesners Annahme (Wiesner, 1992a, 1992b) interpretiert werden kann, dass Schüler, welche Lernschwierigkeiten bei Abdeckaufgaben haben, das Konzept der Punkt-zu-Punkt-Abbildung nicht verstanden haben.

Zur Optimierung der lokalen Fit-Werte in der CFA, sollten insbesondere die Aufgaben zur Erfassung der Bildkonstruktion und des Strahlenmodells überarbeitet werden. Diese Skala könnte um ein bis zwei weitere Aufgaben erweitert werden,

welche speziell das Konzept der Punkt-zu-Punkt-Abbildung thematisieren. Die in sich schlüssigen Ergebnisse und die akzeptablen globalen Fitwerte des Modells sprechen dafür, dass das Gütekriterium der Konstruktvalidität (interne Validität) ausreichend erfüllt ist.

Die signifikanten mittleren Korrelationen mit den Fachnoten in Physik und Mathematik belegen den Zusammenhang der erfassten Konstrukte mit dem relevanten Außenkriterium der Schulleistung in inhaltlich nahen Fächern (Kriteriumsvalidität).

Als Resümee kann festgehalten werden, dass die verbleibenden 11 Items eine tragfähige Testfassung bilden. Neben der Erweiterung der Skala „Bildkonstruktion und Strahlenmodell“ zur Optimierung der internen Konsistenz und internen Konstruktvalidität, könnte die (externe) Konstruktvalidität durch weitere Kriterien – wie offene Interviews mit Teilnehmern, welche Hinweise auf die angewandten Lösungsstrategien bieten – und Expertenratings zur Beurteilung der inhaltlichen Validität des Tests abgesichert werden (vgl. Lindell, Peak & Foster, 2007).

Fazit zu den Analysen des Fragebogens zu Erhebung der Motivation und des Lehrer-Engagements aus Schülersicht: Für die Erhebung der Schülermotivation und des LES ergeben sich gute bis sehr gute Schätzwerte für die Reliabilität. Zur Diskussion der Validität sei auf die genannte Originalliteratur verwiesen aus der die Items stammen (Rheinberg & Wendland 2003, 2004; Seidel et al., 2001).

2.3.6 Ergebnisse zu den Kovariaten

2.3.6.1 Ergebnisse zu kognitiven Fähigkeiten

Zur Erfassung der kognitiven Fähigkeiten wurden drei Einzelwerte erhoben: verbales Schlussfolgern mit der Aufgabengruppe Satzergänzung (SE), figural-räumliches Schlussfolgern mit der Aufgabengruppe Würfelaufgaben (WÜ) und figural-logisches Schlussfolgern mit der Aufgabengruppe Matrizen (MA). Da es sich bei dem Intelligenztest um ein normiertes Instrument handelt, wurde auf detaillierte Angaben zur Itemanalyse verzichtet und pro Skala der Range angegeben. Die Analyse der Itemschwierigkeit zeigt, dass die Items aller drei Skalen in etwa in dem Bereich liegen, welchen die Autoren angeben. Gleiches gilt für die Trennschärfe. Bezüglich beider Kennwerte zeigte sich aber auch, dass die Werte für die Trennschärfe und die Lösungswahrscheinlichkeit v.a. für die jeweils letzte und schwerste Aufgabe der Skala abweichen, da diese Aufgaben nur von wenigen Schülern gelöst wurden (vgl. Tabelle 24).

Tabelle 24 Vergleich der Schätzungen der Reliabilität: gemessene Werte und Werte von Liepmann et al. (2007) im I-S-T 2000 R

Aufgabengruppe	Itemstatistiken	Werte „SV“ (<i>N</i> = 479) ^a	Werte der Autoren (<i>n</i> = 1445) ^b
Satzergänzung (20 Items)	Lösungswahrscheinlichkeiten (Range)	.13 - .82	.19 - .94
	Trennschärfen (Range)	.00 - .43	.09 - .43
	Cronbachs Alpha	.60	.63
	Split Half	.48	.67
Würfelaufgaben (20 Items)	Lösungswahrscheinlichkeiten (Range)	.08 - .88	.10 - .82
	Trennschärfen (Range)	.00 - .49	.23 - .52
	Cronbachs Alpha	.74	.79
	Split Half	.50	.84
Matrizen (20 Items)	Lösungswahrscheinlichkeiten (Range)	.15 - .93	.06 - .96
	Trennschärfen (Range)	.05 - .34	.15 - .41
	Cronbachs Alpha	.54	.66
	Split Half	.44	.70

^a Gesamtstichprobe „SV“: Schülervorstellungen (*N* = 479, *n_{GY}* = 399, *n_{IGS}* = 80)

^b Vergleichswerte für die Teilstichprobe gymnasial (vgl. Liepmann et al, 2007, S. 24 ff.)

Tabelle 25 Mittelwert und Standardabweichung der Subskalen des I-S-T 2000 R nach Bedingung und Schultyp, (*N* = 479)

Skala	TG ^a		KG ^b		Gymnasium		IGS ^c	
	<i>(n</i> = 220)		<i>(n</i> = 229)		<i>(n</i> = 399)		<i>(n</i> = 80)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
SE ^d	99.34	10.65	98.93	10.55	100.80	10.32	90.90	7.70
WÜ ^e	97.70	9.14	98.92	9.06	99.09	8.98	94.28	8.74
MA ^f	100.52	9.44	101.19	9.66	101.74	9.62	96.33	7.77

^aTreatmentgruppe, ^bKontrollgruppe, ^cIntegrierte Gesamtschule, ^dSatzergänzung, ^eWürfelaufgaben, ^fMatrizen

Cronbachs Alpha als Maß für die interne Konsistenz und Schätzer für die Reliabilität kann als ausreichend bewertet werden, liegt jedoch für die drei verwendeten Skalen unter den Werten, welche die Autoren des Tests angeben (vgl. ebenfalls Tabelle 24).

Tabelle 17 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com enthält eine ausführliche Aufstellung deskriptiver Statistiken bezüglich der drei Subskalen. Die Mittelwerte der drei Intelligenzkomponenten der Gymnasiasten hinweg liegen nahe bei 100 und die Standardabweichungen nahe bei 9 (vgl. Tabelle 25). Dies entspricht den Werten der Autoren, welche den Test

auf einen Mittelwert von $M = 100$ und eine Standardabweichung von $SD = 10$ normiert haben (vgl. ebd., S. 40). Die im Rahmen der Studie untersuchte Gruppe entspricht in etwa knapp dem Bevölkerungsschnitt. Abweichungen in den Mittelwerten der Skalen nach unten ergeben sich vor allem für die Gesamtschüler.

Zur grafischen Veranschaulichung der Daten wurden Boxplots der drei Intelligenzkomponenten nach Bedingung (TG versus KG) und Schultyp (Gymnasium versus IGS) erstellt (vgl. Abbildungen 47 und 48). Auf deskriptiver Ebene zeigen sich in den Boxplots im verbalen Bereich höhere Werte in der Treatmentgruppe: So ist der Median nach oben verschoben. Im Bereich des figural-räumlichen Denkens verhält es sich genau umgekehrt. Hier schnitten die Schüler der Kontrollgruppe etwas besser ab. Zudem ist die Streuung der Werte in diesem Bereich in der Treatmentgruppe höher (vgl. Abbildung 47). Im Bereich des logisch-figuralen Schlussfolgerns sind die Verteilungen annähernd identisch.

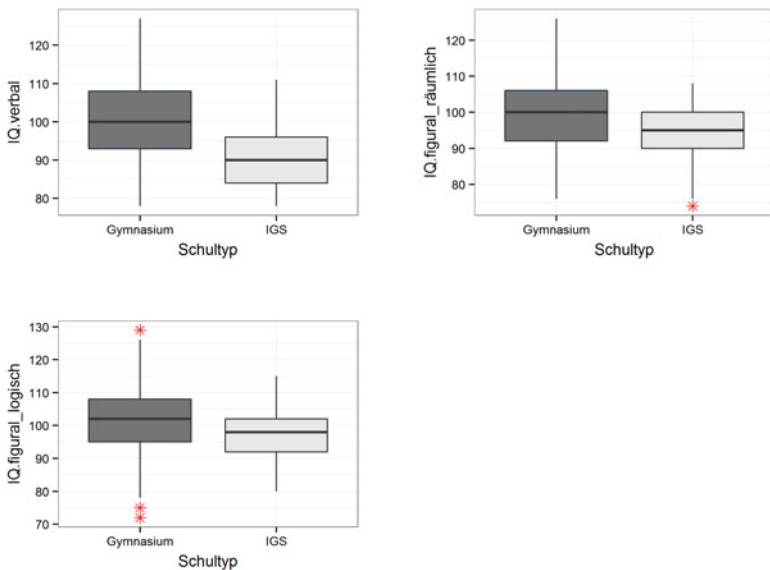


Abbildung 47: Boxplots kognitive Fähigkeiten je Bedingung

Ein Vergleich der Boxplots nach Schultyp und Verteilung zeigt, dass Gymnasiasten im Schnitt höhere Werte in den drei Aufgabenbereichen zur Messung der Intelligenz erzielten. Insgesamt zeigte sich eine höhere Streuung bei den Gymnasiasten als bei den Gesamtschülern. Bei den geringen Werten sind Gymnasiasten ebenso vertreten

wie Gesamtschüler. Die höchsten Werte sind durchweg bei den Gymnasiasten zu finden (vgl. Abbildung 48).

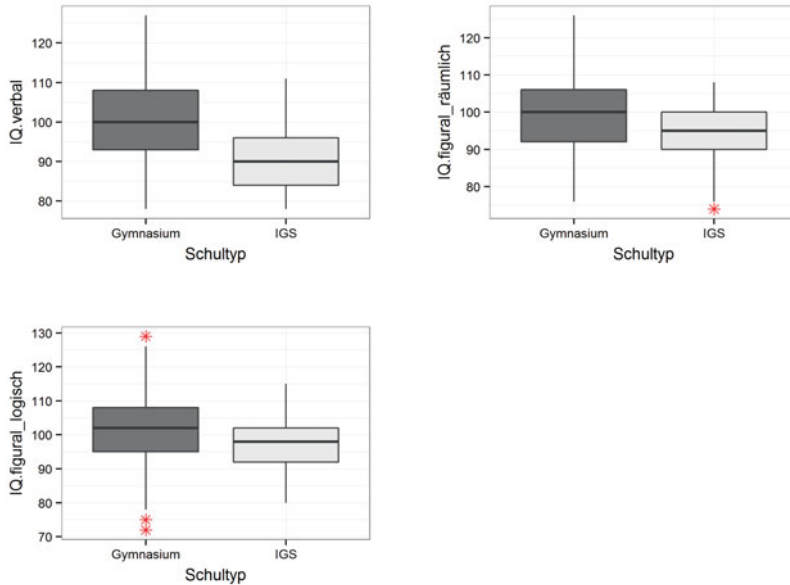


Abbildung 48: Boxplots kognitive Fähigkeiten je Schultyp

Zur Analyse der Zusammenhänge der drei Intelligenzkomponenten untereinander wurden Streudiagramme der drei gemessenen Intelligenzkomponenten (verbales Schlussfolgern, figural-räumliches Schlussfolgern und figural-logisches Schlussfolgern) gegeneinander erstellt. Die Diagramme wurden in Form von sogenannten „Jitter-Plots“ erstellt (siehe Abbildungen 3 bis 5 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com).³¹ Die drei Intelligenz-Komponenten lassen grafisch beim Blick auf die Jitter-Plots keine hohe Korrelation erkennen. Zur Prüfung dieser Schlussfolgerung wurde eine Produkt-Moment-Korrelation nach Bravais-Pearson (vgl. Eid et al., 2011, S. 506) berechnet. Die Ergebnisse be-

³¹ Bei Jitter-Plots werden zu den Punktkoordinaten Zufallszahlen addiert, damit eigentlich übereinanderliegende Punkte getrennt werden. Jede Erzeugung der Diagramme ergibt damit ein geringfügig anderes Bild.

legen eine schwache Korrelation zwischen den Bereichen des verbalen und des figural-räumlichen Schlussfolgerns, eine geringe jedoch signifikante Korrelation zwischen den Bereichen des verbalen und des figural-logischen Schlussfolgerns sowie eine signifikante Korrelation zwischen figural-räumlichem und figural-logischem Schlussfolgern (vgl. Tabelle 26).

Tabelle 26 Pearson-Korrelationen der drei Intelligenzskalen des I-S-T 2000 R ($N = 479$)

	1	2	3
1. Satzergänzung	-		
2. Würfelaufgaben	.08	-	
3. Matrizen	.22**	.26**	-

** Korrelationen $\geq .20$ sind zweiseitig signifikant, $p < .01$

Die drei Korrelationen der Komponenten untereinander wurden insgesamt als eher niedrig eingeschätzt, so dass die Gefahr der Multikollinearität und damit verbundener instabiler Schätzungen der Gewichte des Regressionsmodells als gering zu bewerten ist.

Um für das spätere Modell jedoch die Anzahl an zu berücksichtigenden Variablen zu reduzieren, wurde die Zusammenfassung der Komponenten in Betracht gezogen. Eine Hauptkomponentenanalyse unter Verwendung der Varimax (orthogonale) Rotation erbrachte, dass eine Komponente extrahiert werden konnte, die 46% der Varianz aufklärt (vgl. Tabelle 27).

Tabelle 27 Eigenwerte und Anteil aufgeklärter Varianz Subskalen des I-S-T 2000 R

Komponente	Eigenwert	Anteil aufgeklärter Varianz
1	1.38	46.02 %
2	.92	30.64 %
3	.70	23.34 %

Basierend auf den Ergebnissen wurde davon abgesehen, statt der drei Intelligenzkomponenten eine oder zwei Linearkombinationen derselben zu verwenden, da der Anteil, der durch die Hauptkomponente beschriebenen Gesamtvarianz als nicht ausreichend angesehen wurde. Hinzu kommt, dass eine Zusammenfassung zu einer Hauptkomponente die Interpretierbarkeit des Modells erschwert.

2.3.6.2 Ergebnisse zu vorherigen Schulleistungen

Bezüglich der Noten in den Fächern Deutsch, Physik und Mathematik zeigen sich zwischen Treatment- und Kontrollgruppe kaum deskriptive Unterschiede. Im Schnitt hatten die Schüler die in Tabelle 28 aufgeführten Noten erreicht. Der Vergleich der Notenvergabe zwischen den Schultypen zeigt, dass Schüler der IGS im Schnitt schlechtere Bewertungen erhalten (vgl. Tabelle 28).

Tabelle 28 Mittelwert und Standardabweichung der Fachnoten nach Bedingung (und Schultyp)

Fachnote	TG ^a (n = 269)		KG ^b (n = 243)		Gymnasium (n = 423)		IGS ^c (n = 89)	
	M	SD	M	SD	M	SD	M	SD
Deutsch	2.72	0.05	2.80	0.06	2.68	0.86	3.15	0.75
Mathematik	2.82	0.06	2.82	0.06	2.73	0.98	3.27	0.86
Physik	2.95	0.06	2.83	0.07	2.77	1.01	3.46	1.21

^aTreatmentgruppe, ^bKontrollgruppe, ^cIntegrierte Gesamtschule

Zur Analyse, ob die drei Noten wechselseitig korrelieren, wurden ebenso wie für die Subskalen des I-S-T 2000 R „Jitter-Plots“ (Deutsch, Mathematik, Physik) gegeneinander erstellt. Anhand der Plots (vgl. Abbildungen 6 bis 8, Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com) lässt sich ein statistischer Zusammenhang zwischen den Leistungen im Fach Mathematik und Deutsch sowie zwischen den Fächern Mathematik und Physik vermuten, welcher sich nicht auf Ausreißer zurückführen lässt. Zur Prüfung dieser Schlussfolgerung aus den Jitter-Plots wurde Spearmans Rangkorrelationskoeffizient berechnet, der sich für ordinalskalierte Daten eignet (vgl. Eid et al., 2011, S. 520). Die Ergebnisse weisen auf einen starken Zusammenhang zwischen der Deutsch- und der Mathematiknote sowie der Mathematik- und der Physiknote und einen mittleren Zusammenhang zwischen Deutsch- und Physiknote hin (vgl. Tabelle 29).

Tabelle 29 Rangkorrelationen nach Spearman der drei Fachnoten, (N = 512)

	1	2	3
1. Deutschnote	-		
2. Physiknote	.46**	-	
3. Mathematiknote	.52**	.50**	-

** p < .01, zweiseitig signifikant

Um für das spätere Modell die Anzahl an zu berücksichtigenden Variablen zu reduzieren, wurde ebenfalls die Zusammenfassung der Komponenten in Betracht gezogen. Eine Hauptkomponentenanalyse unter Verwendung der Varimax (orthogonale) Rotation erbrachte, dass eine Komponente extrahiert werden konnte, die 67 % der Varianz aufklärt (vgl. Tabelle 30).

Tabelle 30 Eigenwerte und Anteil aufgeklärter Varianz zu den drei Fachnoten, ($N = 479$)

Komponente	Eigenwert	Anteil aufgeklärter Varianz
1	2.01	67.04 %
2	.53	17.83 %
3	.45	15.14 %

Die Ergebnisse einer Parallelanalyse nach Horn (1965) bestätigen die Extraktion einer Komponente. So lag der zufällige Eigenwert bei 1.07.

Unter Berücksichtigung der deutlichen Korrelationen der Fachnoten und basierend auf den Ergebnissen der PCA wird in Betracht gezogen, statt der drei Fachnoten eine Linearkombination derselben zu verwenden.

2.3.6.3 Zusammenhang zwischen Schulleistungen und kognitiven Fähigkeiten

Abschließend wurden als nichtparametrisches Verfahren die Rangkorrelationen nach Spearman für die Noten und die erhobenen Subskalen des I-S-T 2000 R berechnet. Zur Wahl des Korrelationskoeffizienten bei Variablen mit unterschiedlichen Skalenniveau sei auf Eid et al. (2011, S. 538) hingewiesen. Die Ergebnisse weisen bis auf zwei Ausnahmen auf einen schwachen, jedoch signifikanten Zusammenhang zwischen Noten und den verwendeten Skalen des Intelligenztests auf (vgl. Tabelle 31). Wenig überraschend ist der geringe Zusammenhang zwischen Deutschnote und figural-räumlichem Schlussfolgern (Würfelaufgaben). Unerwartet ist jedoch, dass kein signifikanter Zusammenhang zwischen Physiknote und figural-räumlichem Schlussfolgern (Würfelaufgaben) gefunden wurde. Hier weicht der Wert auch von den Werten ab, welche die Autoren des Tests für die Korrelation zwischen Skalenwerten und Schulnoten berichten (vgl. Liepmann et al., 2007, S. 37).

Tabelle 31 Nichtparametrische Korrelationen der drei Fachnoten mit den verwendeten Subskalen des I-S-T 2000 R, ($N = 479$)

	Deutsch	Physik	Mathematik
Satzergänzung	-.30 **	-.25**	-.26**
Würfelaufgaben	-.07	-.09	-.18**
Matrizen	-.17**	-.17**	-.22**

** $p < .01$, zweiseitig signifikant

Basierend auf den eher niedrig einzustufenden Korrelationen der Intelligenzskalen mit den drei Fachnoten wird die Gefahr der Multikollinearität und damit verbundener instabiler Schätzungen der Gewichte des Regressionsmodells als gering eingeschätzt (vgl. Tabelle 31).

2.3.6.4 Diskussion der Analysen zu den Kovariaten

Möglicher Grund für die Abweichungen in den Itemkennwerten (Trennschärfe und Lösungswahrscheinlichkeit) und den Maßen für die interne Konsistenz könnten darauf zurückgeführt werden, dass es sich bei der Stichprobe um eine vergleichsweise junge Altersgruppe handelt. Auch wenn die Zuordnung der Messwerte zu den Standardwerten auf altersangepassten Normen basiert, wird vermutet, dass insbesondere die letzten und schwierigsten Aufgaben je Skala für die Stichprobe dieser Studie (Schüler der 7. und 8. Klasse) die Schüler im Alter von 12 bis 15 Jahren etwas überfordert haben könnten, da der Test in erster Linie auf die Erhebung der Intelligenz von Erwachsenen zielt.

Da davon auszugehen ist, dass die Schüler im Jugendalter etwas schlechter abschneiden als im späteren Erwachsenenalter, wird insgesamt vermutet, dass die untersuchte Gruppe eine höhere durchschnittliche Intelligenz aufweist als die Gesamtpopulation aller Schüler dieser Altersklasse. Dies ist aufgrund des hohen Anteils an Gymnasiasten (83%, 399 von 479 Schülern) erwartbar.

Die Analysen der Schulnoten ergaben ein ähnliches Bild: Gesamtschüler schnitten hier deutlich schlechter ab als Gymnasiasten. Sowohl in Bezug auf die Noten als auch auf die kognitiven Fähigkeiten zeigte sich eine höhere Streuung bei den Gymnasiasten als bei den Gesamtschülern. In Bezug auf die kognitiven Fähigkeiten waren bei den geringen Werten Gymnasiasten ebenso vertreten wie Gesamtschüler. Die höchsten Werte sind durchweg bei den Gymnasiasten zu finden.

Dieser Befund erscheint zunächst kontraintuitiv, da die Gesamtschule sich insbesondere an Schüler unterschiedlichster Leistungsbereiche richtet. Im Hin-

blick auf die Problematik der relativ frühen Selektion von Schüler im deutschen Schulsystem könnte das Ergebnis vor dem Hintergrund der oft diskutierten Bevorzugung von Kindern mit entsprechendem sozialen Hintergrund interpretiert werden (vgl. auch Ergebnisse der dritten PISA-Studie 2006, PISA-Konsortium Deutschland, 2007).

Für die Vergleichbarkeit von Treatment- und Kontrollgruppe ergeben sich hieraus keine negativen Konsequenzen. In der Folge sollte jedoch der Schultyp in der Auswertung unbedingt berücksichtigt werden und mögliche Interaktionen zwischen Schultyp und Bedingung (TG versus KG) – Interaktionen zwischen kognitiven Fähigkeiten und Bedingung sowie Interaktionen zwischen Schulnoten und Bedingung – analysiert werden.

Bezüglich des Zusammenhangs zwischen Noten und kognitiven Fähigkeiten ergaben sich weitestgehend erwartbare Ergebnisse. Überraschend war der geringe Zusammenhang zwischen Physiknote und figural-räumlichem Schlussfolgern (Würfelaufgaben). Ein möglicher Grund könnte darin bestehen, dass es sich bei der Physiknote in dieser Studie um einen Sonderfall handelt. So basiert diese Note entweder auf einem geringen Zeitabschnitt, weil der Physikunterricht erst in dem Schuljahr startete oder auf der Note im Fach Naturwissenschaften aus Klasse 6. Die Note aus Jahrgangsstufe 6 wurde in den Klassen, in denen noch keine Physiknote aus dem bisherigen Physikunterricht verfügbar war, ersatzweise herangezogen.³²

Umgang mit dem Problem der Multikollinearität: Um instabile und ungenaue Schätzungen der Regressionskoeffizienten zu vermeiden, wird aufgrund der hohen Korrelationen der Fachnoten untereinander für die nachfolgenden Analysen das Ergebnis der PCA als Kovariate verwendet. In Bezug auf die Subskalen des I-S-T 2000 R und die Korrelationen zwischen Fachnoten und kognitiven Fähigkeiten werden die Ausgangswerte beibehalten, da aufgrund der schwachen wechselseitigen Korrelationen keine Multikollinearität besteht.

2.3.7 Einflussfaktoren bei der Anwendung des Lehrmaterials

2.3.7.1 Wahrgenommenes Lehrerengagement aus Schülersicht

Mögliche Einflussfaktoren bei der Anwendung des Lehrmaterials wurden aus zwei Perspektiven erhoben: erstens aus Sicht der Schüler und zweitens aus Sicht

³² Dies war der Fall, wenn der Physikunterricht zu Beginn der Studie erstmals startete und Strahlenoptik von der Lehrkraft als Einführungsthema gewählt wurde.

der Lehrkräfte. Im Rahmen des Motivationsfragebogens schätzten die Schüler in der Skala LES subjektiv das Engagement ihres Lehrers zu den drei Messzeitpunkten auf einer Likert-Skala von 1 (geringe Ausprägung des Engagements) bis 6 (hohe Ausprägung des Engagements) ein. Ein Blick auf die Mittelwerte (vgl. Tabelle 32) und Boxplots (vgl. Abbildung 49) zeigt, dass die Schüler in der Treatmentbedingung das Engagement ihres Lehrers zum ersten und zum dritten Messzeitpunkt, geringfügig höher bewerten als die Schüler der Kontrollbedingung.

Tabelle 32 Lehrerengagement aus Schülersicht je Bedingung zu den drei Messzeitpunkten

	Prätest			Posttest			Follow-up Test		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
TG ^a	243	4.56	1.18	238	4.15	1.39	255	4.36	1.28
KG ^b	222	4.41	1.28	213	4.14	1.33	216	4.19	1.40

Anmerkung. Skala von 1 (geringe Ausprägung) bis 6 (hohe Ausprägung)

^aTreatmentgruppe,

^bKontrollgruppe

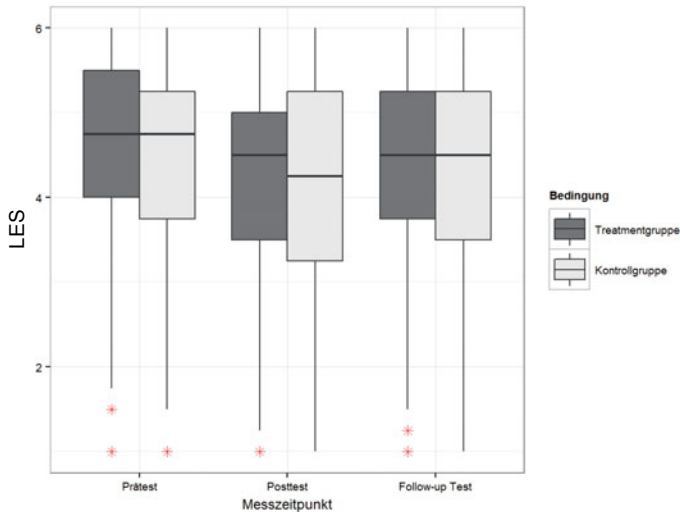


Abbildung 49: Boxplots des Lehrerengagements aus Schülersicht (LES) zu den drei Messzeitpunkten

Für die Durchführung der Studie besonders relevant waren mögliche Unterschiede zwischen Treatment- und Kontrollgruppe im Posttest (unmittelbar nach der Intervention). Die Schüler wurden explizit darum gebeten, ihre Einschätzung auf die letzten sechs Stunden zu beziehen. In den deskriptiven Werten zeigen sich (nahezu) identische Werte für die Mittelwerte und Standardabweichungen von $M = 4.15$ ($SD = 1.39$) bzw. $M = 4.14$ ($SD = 1.33$) in der Einschätzung der Schüler zwischen Treatment- und Kontrollgruppe nach der Intervention (post).

Die Analyse der Daten mittels eines Mehrebenenmodells (ohne Kovariaten) ergab zu keinem der drei Messzeitpunkte signifikante Unterschiede zwischen den Bedingungen (TG versus KG) – weder zu den einzelnen Messzeitpunkten noch in Bezug auf den Vergleich der Entwicklungsverläufe prä – post und prä – follow-up (vgl. Tabelle 33). So bestätigt sich auch in den Daten zum Anteil aufgeklärter Varianz mit Werten nahe 0, dass die Zuordnung zu den Bedingungen zur Erklärung der Varianz bedeutungslos ist. In beiden Bedingungen kam es jedoch zu einer signifikanten Abnahme der Einschätzung des Lehrerengagements im Schuljahresverlauf.

Tabelle 33 Ergebnisse des Mehrebenenmodells zur Analyse der Entwicklung des Lehrerengagements aus Schülersicht

Level (Stichprobengröße)					
Level 1 (N = 1387)	Messzeitpunkte				
Level 2 (N = 470)	Individuen				
Level 3 (N = 21)	Schulklassen				
Fixe Effekte					
<i>Variable</i>	<i>Erläuterung</i>	<i>b</i>	<i>(SE)</i>	<i>F_(numDF, denDF)</i>	<i>p</i>
Interzept	Durchschnittliche Einschätzung, Bedingung = KG	4.44	0.25	312.68 _(1,860)	< .001
Veränderung (prä – post: t ₂ -t ₁)	Durchschnittliche Veränderung, Bedingung = KG	-0.25	0.07	11.81 _(1,860)	< .001
Veränderung (prä – follow-up: t ₃ -t ₁)	Durchschnittliche Veränderung, Bedingung = KG	-0.20	0.07	6.67 _(1,860)	= .01
Bedingung = TG	Unterschied TG und KG zu t ₁	-0.13	0.35	0.14 _(1,19)	= .709
Bedingung * Veränderung prä – post	Zusätzlicher Einfluss Bedingung = TG auf die Veränderung zwischen t ₂ und t ₁	-0.19	0.10	3.63 _(1,860)	= .057
Bedingung * Veränderung prä – follow-up	Zusätzlicher Einfluss Bedingung = TG auf die Veränderung zwischen t ₃ und t ₁	0.01	0.11	0.003 _(1,860)	= .956
Zufällige Effekte					
σ_u (Klasse)	0.77				
σ_p (Individuum)	0.66				
σ_ε (individuen-spezifisch)	0.68				
ρ (Messzeitpunkt)	(ρ_{12}) 0.10	(ρ_{13}) -0.12	(ρ_{23}) 0.16		
g (Messzeitpunkt)	(g_1) 1.00	(g_2) 1.26	(g_3) 1.20		
Modellvergleich					
	<i>Berichtetes Modell</i>	<i>Leermodell^a</i>			
<i>Devianz</i>	13883.94	3840.29			
<i>df</i>	14	11			
<i>AIC</i>	13941.94	3862.29			
<i>BIC</i>	14112.99	3919.88			
„Wirkung“ des Treatments: Anteil aufgeklärter Varianz^a					
R ² Level 1	(t ₁) 0.01	(t ₂) < 0.01	(t ₃) < 0.01		
R ² Level 2	0.01				
R ² Level 3	< 0.01				

^a Das Leermodell unterscheidet sich vom berichteten Modell darin, dass alle fixen Effekte außer dem Interzept und den Dummy-Variablen t₂ und t₃ fehlen. Das Modell sagt für jeden Schüler voraus, dass die Einschätzung des LES zum jeweiligen Messzeitpunkt der mittleren Einschätzung der Stichprobe entspricht. Man beachte, das Leermodell schätzt folgende 11 Parameter (*df* = 11): Interzept, t₂, t₃, g₂, g₃, ρ_{12} , ρ_{13} , ρ_{23} , σ_u , σ_p , σ_ε .

In den Verlaufsdiagrammen (vgl. Abbildung 50) erscheint der Rückgang des LES in der Treatmentgruppe bei den Gesamtschülern in der Postmessung größer als in der entsprechenden Kontrollgruppe, auch wenn diese Tendenz, wie die Mehrebenenanalyse zeigt (vgl. Tabelle 33), nicht signifikant ist.

Insgesamt bestätigen die Verlaufsdiagramme den leichten Abwärtstrend in der Bewertung der Schüler beider Bedingungen und Schularten im Schuljahresverlauf. Darüber weisen die Verlaufsdiagramme darauf hin, dass Gesamtschüler das Engagement ihres Lehrers etwas geringer bewerten als Gymnasiasten (vgl. Abbildung 50). Aufschluss für mögliche Gründe dieser Entwicklung könnte die Analyse des Lehrernotizheftes ergeben.

2.3.7.2 Auswertung des Lehrernotizheftes

Um die Perspektive der Lehrkräfte bei der Umsetzung der Intervention zu erfassen wurden die Lehrkräfte gebeten, jede Unterrichtsstunde der Intervention in einem Lehrernotizheft zu bewerten und besondere Vorkommnisse oder Störungen zu notieren. Die vorliegenden Ergebnisse sind in Tabelle 34 zusammengefasst. Die Bewertung umfasste drei Skalen: erstens die Bewertung der eingesetzten Aufgaben, zweitens die Einschätzung des Unterrichtsablaufs und drittens die Bewertung der Lernbilanz.

Der Mittelwert zur Einschätzung der Aufgabenschwierigkeit und des Aufgabenumfangs liegt bei $M = 1.69$ (0.76) bzw. $M = 2.35$ (0.77) in der Treatmentgruppe und bei $M = 1.53$ (0.73) bzw. $M = 2.02$ (0.86) in der Kontrollgruppe. Damit wurden die Aufgaben im Bereich „nicht sehr schwierig“ bis „schwierig aber lösbar“ bzw. „umfangreich, aber in einer Stunde lösbar“ mit der Tendenz zur Einschätzung eines zu hohen Aufgabenumfangs in der Treatmentgruppe bewertet. Die Beurteilerübereinstimmung je Bedingung (Treatment versus Kontrollgruppe) kann für die Bewertung der Aufgaben und des Unterrichtsablaufs als zufriedenstellend ($ICC = .51 - .57$) bis gut ($ICC = .61 - .90$) bewertet werden (vgl. Cichetti & Sparrow, 1981).³³ Die niedrigeren Übereinstimmungen bezüglich des Ausmaßes an kognitiver Aktivierung (Lernbilanz, Frage 3) spiegeln wieder, dass die Meinungen über das Maß an kognitiver Aktivierung in der Kontrollbedingung unterschiedlich bewertet wird, was auf die gewollte Variation im Ausmaß der kognitiven Bedingungen zurückzuführen ist.

33 Da es sich um ordinalskalierte Daten handelt, wurde als Maß für die Beurteilerübereinstimmung Kendalls Konkordanzkoeffizient W berechnet und der anschaulichere ICC , der jedoch unter dem Vorbehalt zu interpretieren ist, dass die Voraussetzung des Intervallskalenniveaus im strengen Sinn nicht erfüllt ist.

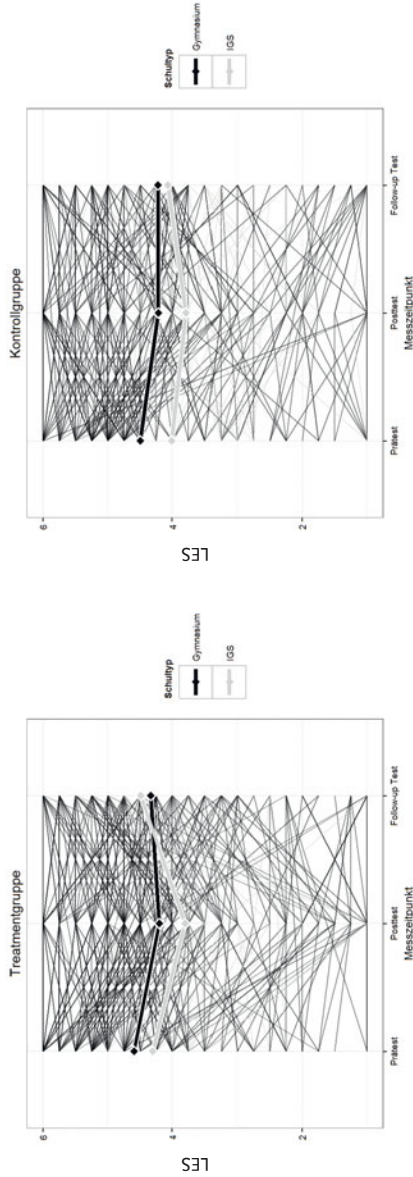


Abbildung 50: Verlaufsdigramme des Lehrerengagements aus Schülersicht (LES) zu den drei Messzeitpunkten

Tabelle 34 Einschätzung der Unterrichtsqualität der $N = 10$ teilnehmenden Lehrkräfte gemittelt über die sechs geplanten Unterrichtsstunden

Item	TG				KG			
	<i>M</i>	<i>SD</i>	<i>Modus</i>	<i>ICC_{unjustiert}</i>	<i>M</i>	<i>SD</i>	<i>Modus</i>	<i>ICC_{unjustiert}</i>
<i>Bewertung der Aufgaben</i>								
Aufgabenschwierigkeit	1.69	0.76	1	.62	1.53	0.73	1	.51
Aufgabenumfang	2.35	0.77	3	.52	2.02	0.86	2	.42
Kendalls <i>W</i>	.50 $\chi^2_{(1)} = 4.50, p = .034$.40 $\chi^2_{(1)} = 3.57, p = .059, n.s.$			
Skala 0 - 3,	0 - zu einfach / zu wenig, 1 - nicht sehr schwierig / gut umsetzbar, 2 - schwierig / umfangreich, aber lösbar, 3 - zu schwierig / zu umfangreich							
<i>Unterrichtsablauf</i>								
Der Unterricht hat pünktlich begonnen	3.10	0.85	3	.65	3.11	0.75	3	.20
Den Schülern war jederzeit klar, was sie tun sollten	2.92	0.85	3	.61	3.13	0.82	3	.44
Die gesamte Unterrichtsstunde wurde für den Lernstoff verwendet	3.38	0.82	4	.56	3.39	0.70	4	.27
Kendalls <i>W</i>	.24 $\chi^2_{(1)} = 4.22, p = 1.21, n.s.$.04 $\chi^2_{(1)} = 0.77, p = 2.06, n.s.$			
<i>Lernbilanz</i>								
Ich schätze: Die Schüler haben in dieser Stunde etwas dazu gelernt	3.16	0.67	3	.61	3.19	0.68	3	.61
Die Lernziele dieser Unterrichtsstunde wurden erreicht	2.91	0.93	3	.90	2.85	0.94	3	.57
Ich schätze: Der Unterricht hat die Schüler zum Nachdenken angeregt	2.93	0.77	3	.67	3.04	0.75	3	.32
Kendalls <i>W</i>	.44 $\chi^2_{(1)} = 4.50, p = .018$.45 $\chi^2_{(1)} = 8.06, p = .018$			
Skala 1-4	1 - stimme nicht, 2 - stimme eher nicht zu, 3 - stimme eher zu, 4 - stimme zu							

Bezüglich des Unterrichtsablaufs divergieren die Bewertungen. Unter Berücksichtigung der Auswertung der Kommentare der Lehrer zu den einzelnen Stunden lässt sich dieses Ergebnis auf Gegebenheiten im Schulalltag zurückzuführen. Die genannten Störungen (z.B. Raumverlegungen, Besprechung von Klassenfahrten u.a.) stellen jedoch keine wesentlichen Einschränkungen für die Vergleichbarkeit der beiden Bedingungen dar, da die Lernzeit in beiden Bedingungen parallel gehalten wurde.

Der Mittelwert der Zustimmung der Fragen zur Unterrichtsqualität rangiert zwischen $M = 2.92$ (0.85) bis $M = 3.38$ (0.82) in der Treatment- und $M = 2.85$ (0.94) bis $M = 3.11$ (0.75) in der Kontrollgruppe und liegt damit im Bereich zwischen „stimme eher zu“ (3) und „stimme zu“ (4).

Ein Mann-Whitney-U-Test für unabhängige Stichproben (hier Treatment- versus Kontrollgruppe) bei ordinalskalierten Daten ergab mit Ausnahme der Einschätzung des Aufgabenumfanges keine signifikanten Unterschiede in den Einschätzungen seitens der Lehrkräfte zwischen Treatment- und Kontrollgruppe.

Während die Aufgabenschwierigkeit zwischen nicht „sehr schwierig“ oder „schwierig aber machbar“ eingestuft wurde, schätzten die Lehrkräfte den Aufgabenumfang in der Treatmentbedingung als signifikant höher ein als in der Kontrollbedingung ($U(10,10) = 1015.00$, $z = 2.05$, $p = .040$). Um diesem Befund nachzugehen, wurde die Einschätzung des Aufgabenumfanges für die einzelnen Unterrichtsstunden betrachtet.

Ein Blick auf die einzelnen Stunden zeigt, dass der Umfang für alle Unterrichtsstunden außer der dritten nicht signifikant unterschiedlich bewertet wurde. So lag die Einschätzung des Aufgabenumfanges der Stunden 1 bis 2 und 4 bis 6 im Schnitt bei 2 „Die Aufgaben sind umfangreich, aber in einer Stunde lösbar“; außer für die dritte Unterrichtsstunde, in welcher die Aufgaben als zu umfangreich eingeschätzt wurden.

Hier weichen die Ergebnisse für die Treatment- und die Kontrollgruppe auch signifikant ab: ($U(10,10) = 10.00$, $z = 2.56$, $p = .021$).

Mit Blick auf den Unterrichtsablauf lässt sich dieser Befund leicht erklären. So wurde in der dritten Unterrichtsstunde die Strahlenkonstruktion behandelt. Zwecks Datenerhebung wurden hier die Arbeitsblätter der Schüler eingesammelt und zwar ausschließlich in der Treatmentgruppe. Die Kontrollgruppe hielt parallel einen Lehrervortrag und übte die Strahlenkonstruktion später. Die Einschätzung der Lehrkräfte deckt sich hierbei mit der Unterrichtsbeobachtung. So war das Ein- und Austeilen der Blätter mit einem deutlich höheren Zeitaufwand verbunden als in der Kontrollgruppe. Zudem musste in der Treatmentgruppe zusätzliche Zeit aufgewandt werden, um den Schülern den Zweck des Einsammelns zu erklären.

Die Analyse der übrigen Items erbrachte keine Unterschiede zwischen Treatment- und Kontrollgruppe. Dies ist ebenso wenig für die dritte Unterrichtsstunde als auch für die übrigen Stunden der Fall. So wurde weder der Lernzuwachs noch das Erreichen der Lernziele für die beiden Gruppen unterschiedlich bewertet.

In einer offenen Frage wurden die teilnehmenden Lehrkräfte zudem gebeten Anmerkungen, besondere Vorkommisse und offene Punkte zu nennen. Die Antworten der Lehrer lassen sich wie folgt zusammenfassen:

- Kritik wurde an der Thematisierung des virtuellen Bildes geübt. Hier wurde vor allem mehr Zeit für die Einführung und Besprechung des virtuellen Bildes gefordert.

- Das Austeilen und Einsammeln der Arbeitsblätter zwecks Datenerhebung wurde als zeitaufwendig eingeschätzt.
- Seitens der teilnehmenden Gesamtschullehrer wurde angemerkt, dass mehr Möglichkeiten der Binnendifferenzierung zwischen schnellen und langsamen Schülern wünschenswert gewesen wären.
- Bezüglich der Arbeitsblätter ergab sich in manchen Klassen das Problem, dass die Schüler die Arbeitsaufträge nicht genau lasen und es hierdurch zu zeitlichen Verzögerungen kam.
- Positiv wurde genannt, dass die Arbeitsblätter den Schülern einen Überblick verschafften, die sechste Unterrichtsstunde Gelegenheit zur Wiederholung und Vertiefung bot und dass die Schüler die Aufgabenblätter zu Teilen mit Interesse und weitgehend selbstständig erarbeiten konnten.

2.3.7.3 Diskussion zu den Einflussfaktoren bei der Anwendung des Lehrmaterials

Das Mehrebenenmodell zur Entwicklung des Lehrerengagements aus Schülersicht belegt, dass keine signifikanten Unterschiede in der Einschätzung des LES bestehen. Insbesondere zum zweiten Messzeitpunkt ergaben sich nahezu gleiche Werte in der Einschätzung der Schüler (siehe arithmetisches Mittel und Standardabweichung des LES), daher kann von einem vergleichbar wahrgenommenen Engagement der Lehrkräfte in den Bedingungen ausgegangen werden. Zur Berücksichtigung von Unterschieden zwischen den Schulen, welche sich insbesondere durch die niedrigeren Werte der Gesamtschulen zeigen, wird das LES vor der Intervention (erster Messzeitpunkt) als Kovariate in den folgenden Analysen einbezogen.

Auch in den Analysen des Lehrernotizhefts zeigen sich auf Schul- und Klassenebene keine Unterschiede, welche die Vergleichbarkeit zwischen Treatment- und Kontrollgruppe einschränken würden.

Die genannten Kritikpunkte ergeben sich vorwiegend aus Erfordernissen der empirischen Erhebung. So zielte das erstellte Unterrichtsmaterial darauf, bis auf den intendierten Unterschied zwischen Treatment- und Kontrollgruppe vergleichbare Bedingungen herzustellen. Entsprechend wurde weniger Wert auf die Möglichkeit der Leistungsdifferenzierung gelegt. Aspekte, welche für die Lerninhalte mit geringerer Priorität bewertet wurden, wie das virtuelle Bild, konnten durch den vorgegebenen Zeitraum weniger ausführlich behandelt werden.

Das Einsammeln der Arbeitsblätter diente dazu, Einblicke in den Lernprozess zu erhalten. Gegebenenfalls hätte hier eine andere Methode der Datenerhebung zu einem geringeren Aufwand für die teilnehmenden Lehrer und Schüler geführt.

Letzterer Punkt könnte eventuell erklären, warum im Verlauf Schüler beider Bedingungen das Engagement ihres Lehrers – bezogen auf den Zeitraum der Intervention geringer bewerten als zuvor oder danach. Möglicherweise lässt sich dieser Effekt auf eine gewisse „Testmüdigkeit“ seitens der teilnehmenden Schüler und Lehrkräfte zurückführen.

2.3.8 Einblick in die Lernprozesse beim Umgang mit Repräsentationen

2.3.8.1 Auswertung der Arbeitsblätter 4 und 7 zur Bildkonstruktion

Um Einblicke in den Lernprozess zu erhalten, wurde eine geschichtete Zufallsauswahl an Arbeitsblättern, die im Rahmen der Intervention eingesammelt wurden, ausgewertet. Sowohl in der Treatment- als auch in der Kontrollbedingung wurde das Aufgabenblatt 7 „Wiederholung zur Bildkonstruktion“ siehe Anhang C1 auf der Produktseite zu diesem Buch unter www.springer.com, eingesammelt, welches die Schüler in der fünften Unterrichtsstunde bearbeiteten. Folgende Fragen wurden im Rahmen der Auswertung von Aufgabenblatt 7 untersucht:

1. Unterscheiden sich Schüler in den beiden Bedingungen (TG versus KG) signifikant hinsichtlich des Erfolgs beim Bearbeiten der Aufgaben?
2. Sind Leistungsunterschiede, die im Leistungspost- und im Konzeptposttest gemessen wurden, bereits im Lernprozess zu erkennen:
 - Unterscheiden sich die Schüler in unterschiedlichen Leistungsbereichen, hohes Leistungsniveau, mittleres Leistungsniveau und unteres Leistungsniveau bereits bei der Bearbeitung des Aufgabenblatts?
 - Korreliert der Erfolg beim Bearbeiten der Aufgaben mit den jeweiligen Posttestergebnissen signifikant? Wenn ja, wären bezüglich der Auswertung des Aufgabenblatts 7 für die A1 (Kohärenzaufgabe, vgl. auch Treatment von Scheid, 2013) höhere Korrelationen mit dem Leistungstest und für die Aufgaben A2 und A3 (Linse < Schirm, Abdeckaufgabe) höhere Korrelationen mit dem Konzeptposttest zu erwarten.

In der Treatmentgruppe wurde zudem das Aufgabenblatt 4 „Die Konstruktion der reellen Bildfälle“ aus der dritten Unterrichtsstunde ausgewertet (siehe Anhang C1 auf der Produktseite zu diesem Buch unter www.springer.com). Folgende Fragen wurden hierzu untersucht:

1. Besteht ein Zusammenhang zwischen Lernprozessen zum Verständnis und Umgang der zentralen schematischen Repräsentation des Strahlengangs mit dem Leistungs- bzw. dem Konzeptposttest?
2. Sofern der Erfolg beim Bearbeiten der Aufgaben mit den jeweiligen Testergebnissen signifikant korreliert, wären bezüglich der Auswertung des Aufgabenblattes 4 höhere Korrelationen mit dem Leistungs- als mit dem Konzeptposttest zu erwarten.

Die geschichtete Zufallsauswahl wurde nach folgenden Kriterien gezogen: pro Bedingung Treatment- versus Kontrollgruppe wurden jeweils fünf Schüler des unteren Quartils des Konzeptpost- und des Leistungsposttests, fünf Schüler des oberen Quartils des Konzeptpost- und Leistungsposttests sowie je fünf Schüler der mittleren Prozenträge des Konzeptpost- und des Leistungsposttest per Zufall (Zufallsstichprobe in SPSS) ausgewählt (vgl. Tabelle 35 und Tabelle 36). Weitere Informationen zur Verteilung von Jungen und Mädchen bzw. den Anteilen an Gymnasiasten und Gesamtschülern finden sich in den Tabellen 18 und 19 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com.

Tabelle 35 Kreuztabelle zur Verteilung der ausgewählten Schüler je Test (abhängige Variable) und Bedingung

		Abhängige Variable		
		LT ^a	KT ^b	Σ
Bedingung	TG ^c	15	15	30
	KG ^d	15	15	30
Σ		30	30	60

^aLeistungstest, ^bKonzepttest, ^cTreatmentgruppe, ^dKontrollgruppe

Die Auswertung des Aufgabenblatts 7 erfolgte nach folgendem Schema: je Aufgabe konnten sowohl in der Treatment- als auch in der Kontrollbedingung maximal 4 Punkte erzielt werden. Daraus ergibt sich für das Aufgabenblatt 7 eine Höchstpunktzahl von 3 x 4 Punkten = 12 Punkten. Die Punktevergabe erfolgte nach einem festgelegten Schema, bei dem sowohl für die Zeichnung als auch für die Begründung je 2 Teilpunkte (TP) erzielt werden konnte. Die Vergabe der Punkte wurde von einem geschulten Auswerter mit fachdidaktischem Hintergrund vor-

genommen und von der Autorin im Hinblick auf die Einhaltung des Auswertungsschemas geprüft.

Tabelle 36 Kreuztabelle zur Verteilung der ausgewählten Schüler je Test (abhängige Variable) und Quartil des Leistungsbereichs Physikleistung und konzeptuelles Verständnis

		Abhängige Variable		
		LT ^a	KT ^b	Σ
Quartil	$Q_{0.25}$	10	10	20
	$Q_{0.25} - Q_{0.75}$	10	10	20
	$Q_{0.75}$	10	10	20
Σ		30	30	60

^aLeistungstest, ^bKonzepttest

Tabelle 37 Deskriptive Statistiken zur Auswertung von A1-A3 des Aufgabenblatt 7 „Wiederholung zur Bildkonstruktion“, ($N = 60$)

	A1	A2	A3	Gesamt
	2 TP Zeichnung, 2 TP Begründung	2 x 1 TP Zeichnung, 2 x 1 TP Begründung	2 x 1 TP Zeichnung, 2 x 1 TP Begründung	3 x 4 Punkte
<i>M</i>	1.61	1.69	1.00	4.05
<i>(SD)</i>	(1.46)	(1.40)	(1.26)	(3.39)
<i>Range</i>	0 - 4	0 - 4	0 - 4	0 - 12
α (A1-A3, 3 Items)	.74			

Anmerkung. A1: 2 x 2TP; A2 und A3 je 4 x 1TP

Wie an den deskriptiven Statistiken ersichtlich ist, ist die Bearbeitung der Aufgaben 2 und 3 durch einen starken Bodeneffekt gekennzeichnet (vgl. Tabelle 37). Insbesondere die Abdeckaufgabe A3 fällt durch einen niedrigen Mittelwert auf. Möglicherweise bereitet die weitverbreitete Schülervorstellung, die Linse schneide das Bild ab, Schülern Probleme im Umgang mit der Strahlenkonstruktion und der verbalen Begründung ihrer Schlussfolgerungen aus der Zeichnung. Die interne Konsistenz der Aufgaben als Schätzer für die Reliabilität kann mit $\alpha = .75$ als zufriedenstellend eingestuft werden.

Zur Untersuchung der ersten Frage wurden die Mittelwerte der Aufgaben mittels eines t-Tests für unabhängige Stichproben verglichen (siehe Tabelle 38).

In der Analyse zeigten sich keine signifikanten Mittelwertunterschiede zwischen Schülern in den beiden Bedingungen (TG versus KG).

Tabelle 38 t-Test zur statistischen Prüfung von Mittelwertunterschieden zwischen TG und KG bei der Bearbeitung des Aufgabenblatts 7 „Wiederholung zur Bildkonstruktion“, ($N = 60$)

	Gesamt	A1	A2	A3
TG - M (SD)	3.40 (2,70)	1.37 (0.96)	1.48 (0.95)	1.03 (0.96)
KG - M (SD)	4.70 (3.90)	1.87 (1.50)	1.90 (1.72)	0.97 (1.52)
t (58)	1.50, $p = 1.89$	1.33, $p = 2.52$	1.16, $p = .84$	0.20, $p = 1.40$
$M_{Diff.}$ ($SD_{Diff.}$)	1.30 (0.86)	0.50 (0.38)	0.42 (0.36)	0.67 (0.33)

Im Hinblick auf den Erfolg in der Aufgabenbearbeitung sind signifikante Unterschiede von Schülern im unteren, mittleren und oberen Leistungsniveau des Leistungspost- bzw. Konzeptposttests ersichtlich (Forschungsfrage 2a).

Mittels einer univariaten Varianzanalyse wurde geprüft, ob signifikante Unterschiede zwischen den Schülern der unterschiedlichen Leistungsbereiche bestanden. Hierbei bildeten die Faktorstufen die Zuteilung zu den drei Leistungsbereichen und die abhängige Variable die Gesamtpunktzahl in den drei Aufgaben. Die Analyse bestätigt, dass signifikante Unterschiede, die in den Posttests festgestellt wurden, sich bereits bei der Aufgabenbearbeitung zeigen ($F(2,57) = 8.61$, $p < .001$, $\eta_p^2 = .23$, $\omega^2 = 0.20$, großer Effekt).

Zudem wurde der Effekt jeder Faktorstufe paarweise mit dem Effekt der vorherigen Faktorstufe verglichen (Post-hoc-Test Tukey). Die Ergebnisse zeigen (vgl. Tabelle 39), dass Schüler im unteren Leistungsposttestniveau die Aufgaben weniger erfolgreich bearbeiteten als Schüler des mittleren und oberen Niveaus, während sich Schüler des mittleren und oberen Leistungsniveaus nicht signifikant unterschieden.

Die Analyse der Korrelationen des Erfolgs beim Bearbeiten der Aufgaben mit den jeweiligen Posttestergebnissen (Forschungsfrage 2b) zeigte signifikante Zusammenhänge zwischen der Bearbeitung der Teilaufgaben A1 bis A3 sowie der Gesamtwertung des Arbeitsblatts und der Gesamtpunktzahl im Konzeptposttest, nicht jedoch zwischen der Bearbeitung des Aufgabenblatts und der Gesamtpunktzahl im Leistungsposttest.

Tabelle 39 Paarweiser Vergleich der Schüler der verschiedenen Leistungsniveaus bezüglich der Gesamtpunktzahl für die Bearbeitung des Aufgabenblatts 7, ($N = 60$)

Quartil	n	Vergleich	M	SD	p
Q_{25}	20	Mittleres gegen unteres Niveau	2.45	0.96	= .034
$Q_{0.25} - Q_{0.75}$	20	Oberes gegen mittleres Niveau	1.47	0.96	= .279
Q_{75}	20	Oberes gegen unteres Niveau	3.93	0.96	< .001

Die höchste Korrelation findet sich zwischen der Gesamtwertung des Arbeitsblatts und der Gesamtpunktzahl im Konzeptposttest, $r(58) = .41$, $p < .001$. Entgegen der Vorannahme besteht zwar ein signifikanter Zusammenhang zwischen der Punktzahl für Aufgabe 1 (Kohärenzaufgabe) und dem Ergebnis des Konzeptposttests, nicht jedoch zwischen der Punktzahl für Aufgabe 1 und dem Leistungsposttest. Die Punktzahlen für A2 und A3 korrelieren, wie erwartet, höher miteinander als die jeweilige Korrelation mit der Punktzahl für Aufgabe 1 (vgl. Tabelle 40).

Tabelle 40 Produkt-Momentkorrelation zwischen den Posttestergebnissen und den Arbeitsaufgaben von Arbeitsblatt 7 ($N = 40$; $n_{Q_{25}} = 10$, $n_{Q_{25}-Q_{75}} = 20$, $n_{Q_{75}} = 10$)

Test / Aufgabe (A)	1	2	3	4	5	6
1 Leistungsposttest	-					
2 Konzeptposttest	.35*	-				
3 A ₁	.15	.33*	-			
4 A ₂	.06	.34*	.39*	-		
5 A ₃	.08	.36*	.36*	.63**	-	
6 A ₁₋₃ Gesamt	.09	.41**	.74**	.83**	.81**	-

* $p < .05$, ** $p < .001$, zweiseitig signifikant

Zwischen der Gesamtwertung des Aufgabenblatts und den Subskalen des I-S-T 2000 R, sowie zwischen der Gesamtwertung und den Fachnoten in Mathematik, Physik und Deutsch konnten in der Teilstichprobe von ($n = 40$) keine signifikanten Korrelationen festgestellt werden. Die relativ stärksten Korrelationen bestanden hier zwischen Physiknote und Gesamtwertung des Aufgabenblatts 7 ($r(38) = -.30$, $n.s.$) sowie zwischen Würfelaufgaben (räumlich-figurales Schlussfolgern) und Gesamtwertung des Aufgabenblatts ($r(38) = .23$, $n.s.$).

Im Zug der Auswertung von Aufgabenblatt 4 wurde untersucht, ob ein Zusammenhang beim Bearbeiten von Aufgabenblatt 4 (bildhaftes, verbales und

mathematisches Verständnis der Strahlenkonstruktion) und dem Lernerfolg im Leistungs- bzw. Konzeptposttest besteht.

Die Auswertung des Aufgabenblatts 4 erfolgte analog zur Auswertung von Aufgabenblatt 7: je Teilaufgabe (4a, b, c) konnten maximal 8 Punkte erzielt werden; daraus ergibt sich für das Aufgabenblatt 4 eine Höchstpunktzahl von 3×8 Punkten = 24 Punkten.

Tabelle 41 Deskriptive Statistiken zur Auswertung von 4a-4c des Aufgabenblatts 4 „Die Konstruktion der realen Bildfälle“ ($N = 30$)

Aufgabe	4a	4b	4c	Gesamt
Teilpunkte (TP)	4 TP Zeichnung 2 TP Begründung 2 TP Messen	4 TP Zeichnung 2 TP Begründung 2 TP Messen	4 TP Zeichnung, 2 TP Begründung 2 TP Messen	3 x 8 Punkte
<i>M</i>	4.78	5.91	5.73	16.43
<i>(SD)</i>	(2.55)	(2.25)	(2.23)	(6.63)
<i>Range</i>	0 - 8	1 - 8	0 - 8	1 - 23.50
α (4a - 4c, 3 Items)		.94		

Wie an den deskriptiven Statistiken ersichtlich ist (vgl. Tabelle 41), findet ein deutlicher Lerneffekt zwischen der Bearbeitung von 4a zu 4b statt (dreigliedrige Strahlenkonstruktion) statt. Die leichte Abwärtstendenz in der Gesamtwertung von Teilaufgabe 4c kann auf die erhöhte Anforderung zurückgeführt werden, die Größen- und Streckenverhältnisse verbal und mathematisch frei zu formulieren. In Teilaufgabe 4b bestanden hier noch mehr Unterstützungsangebote (repräsentationales Fading-out). Die interne Konsistenz der Aufgaben als Schätzer für die Reliabilität kann mit $\alpha \geq .90$ als sehr gut bewertet werden.

Wie erwartet, ergab die Analyse der Korrelationen des Erfolgs beim Bearbeiten der Aufgaben mit den jeweiligen Posttestergebnissen tendenziell stärkere Zusammenhänge zwischen der Bearbeitung der Teilaufgaben 4a-c sowie der Gesamtwertung des Arbeitsblatts und der Gesamtpunktzahl im Leistungsposttest als zwischen der Bearbeitung der Aufgaben und der Gesamtpunktzahl im Konzeptposttest (vgl. Tabelle 42).

Zwischen der Gesamtwertung des Aufgabenblatts und den Subskalen des I-S-T 2000 R, sowie zwischen der Gesamtwertung und den Fachnoten in Mathematik, Physik und Deutsch konnte in der Teilstichprobe von ($n = 30$) lediglich eine signifikante Korrelation zwischen Physiknote und Gesamtwertung des Aufgabenblatts 4 ($r(28) = -.41, p < .05$) festgestellt werden.

Tabelle 42 Produkt-Momentkorrelation zwischen den Posttestergebnissen und den Arbeitsaufgaben von Arbeitsblatt 4, ($n = 30$)

Testwert / Aufgabe	1	2	3	4	5	6
1 Leistungsposttest	-					
2 Konzeptposttest	.35*	-				
3 Aufgabe 4a	.57**	.66**	-			
4 Aufgabe 4b	.62**	.51**	.82**	-		
5 Aufgabe 4c	.70**	.54**	.85**	.83**	-	
6 4 _{a-c} Gesamt	.66**	.61**	.95**	.94**	.95**	-

Anmerkung: Arbeitsblatt 4 wurde in dieser Form nur von der Treatmentgruppe bearbeitet

* $p < .05$, ** $p < .001$, zweiseitig signifikant

2.3.8.2 Diskussion der Ergebnisse zur Analyse der Lernprozesse

Die Auswertung der Arbeitsblätter bestätigt, dass der Umgang mit der zentralen Repräsentation der Strahlenkonstruktion bereits in den Lernprozessen entscheidend für den Lernerfolg ist. Die Frage, ob der erfolgreiche Umgang mit zentralen Repräsentationen ein geeigneter Prädiktor für den Lernerfolg unmittelbar nach dem Unterricht darstellt, kann somit mit „ja“ beantwortet werden. Die guten bis sehr guten Schätzwerte für die Reliabilität sprechen für die Zuverlässigkeit der Datenbasis.

Insgesamt kann die Bearbeitung der Aufgaben als mittelstarker Prädiktor für die Lernleistung im Konzept- und Leistungsposttest gewertet werden. Die signifikanten Korrelationen im mittleren Bereich von Aufgabenbearbeitung und der Physiknote stützen diesen Befund, zumal keine signifikanten Korrelationen mit den anderen Fachnoten bestanden.

Da sich die Analysen wegen des hohen Auswertungsaufwands auf eine kleine Stichprobe von 40 Schülern beziehen, könnte der tatsächliche Zusammenhang zwischen Aufgabenbearbeitung und Lernleistung in der Gesamtstichprobe sogar noch stärker sein. Erwartungsgemäß waren Aufgaben zur Bearbeitung der Strahlenkonstruktion bessere Prädiktoren zur Vorhersage des Leistungsposttest-Ergebnisses und Aufgaben, welche Schülervorstellungen (Abdeckaufgaben) thematisierten, bessere Prädiktoren für die Ergebnisse im Konzeptposttest.

Unterschiede in der Bearbeitung zwischen den Bedingungen (TG versus KG) konnten zu dem Zeitpunkt des Einsammelns der Aufgabenblätter (gegen Mitte bzw. zu Beginn des letzten Drittels der Intervention) nicht festgestellt werden.

2.3.9 Ergebnisse zu den untersuchten Hypothesen

2.3.9.1 Methodisches Vorgehen: Mehrebenenanalyse

Zur Analyse der Fragestellungen wurde eine Mehrebenenanalyse gewählt. Hierbei handelt es sich prinzipiell um ein Regressionsmodell, welches aber im Gegensatz zu klassischen Regressionsmodellen in der Lage ist, die Mehrebenenstruktur der Daten zu berücksichtigen (vgl. Eid et al., 2011, S. 699). Eine solche Mehrebenenstruktur zeichnet sich dadurch aus, dass die Elemente auf den unteren Ebenen in jeweils genau ein Element der nächst höheren Ordnung eingebettet sind (vgl. Ditton, 1998, S. 11; Eid et al., 2011, S. 701). Mehrebenenmodelle werden daher auch als hierarchische lineare Modelle bezeichnet (Bryk & Raudenbush, 1992; Snijders & Bosker, 1999; Hox, 2002): Werden solche hierarchisch strukturierten Datensätze wie einfache Zufallsstichproben behandelt, besteht das Risiko falsche Schlussfolgerungen bei der Interpretation von Zusammenhängen zu ziehen (vgl. Eid et al., 2011, S. 700). Die eigentlich notwendige Unabhängigkeit der Beobachtungen ist in hierarchischen Datenstrukturen nicht erfüllt, da die untersuchten Individuen innerhalb der Gruppen oder Aggregateinheiten gemeinsamen Einflüssen oder Erfahrungen unterliegen, welche für die Einheiten eines Aggregats charakteristisch sind.

Bei der hier vorliegenden Stichprobe handelt es sich um eine solche mehrstufige Stichprobenziehung. So wurden die Elemente nicht unabhängig voneinander ausgewählt, sondern eine Zufallsauswahl von Schulen getroffen, welche wiederum eine Auswahl von Klassen und eine Auswahl von Schülern enthielt.

Die hier vorliegende Struktur stellt sich wie folgt dar: Pro Individuum wurde die fachliche Leistung, das konzeptuelle Verständnis und die Motivation zu drei Messzeitpunkten erhoben. Jedes Individuum war einer der 21 Klassen zugeordnet (mit einer Klassenstärke, welche zwischen 17 bis 28 Schüler variierte). Die Schüler wurden als Klassen der Treatment- und Kontrollbedingung zugeteilt. Je eine Treatment- und eine Kontrollklasse wurden von einer Lehrkraft unterrichtet. Insgesamt erklärten sich zehn Lehrer aus zehn verschiedenen Schulen (darunter sieben Männer und drei Frauen) dazu bereit, die Studie in Ihren Parallelklassen umzusetzen (vgl. Abbildung 51).

In Mehrebenenmodellen können auf jeder Ebene Regressionsgleichungen formuliert werden. Das Hauptziel dieser Arbeit besteht darin zu analysieren, ob Effekte auf das Treatment zurückgeführt werden können.

Im speziellen Fall dieser Studie lassen sich drei Ebenen unterscheiden: die Ebene der Messzeitpunkte, die Ebene der Individuen und die Ebene der Klassen.

Für die Untersuchung der Schulebene ist die Stichprobe mit zehn Aggregateinheiten zu klein.

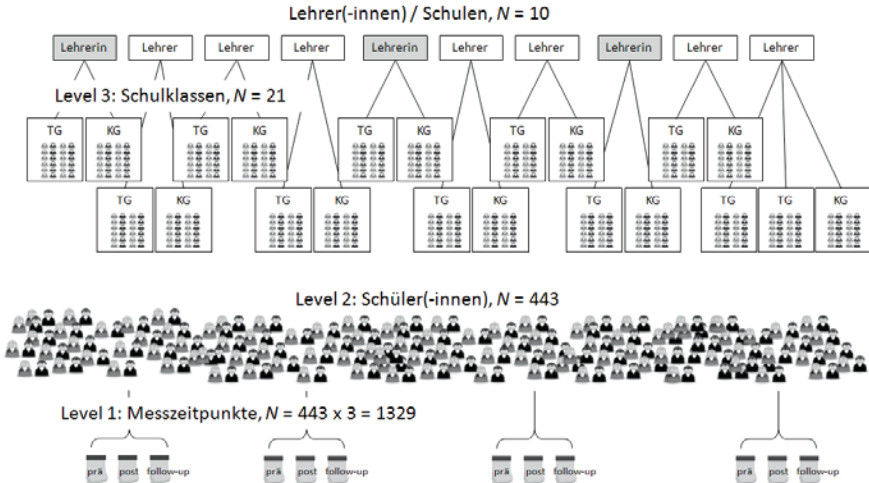


Abbildung 51: Veranschaulichung der Mehrebenenstruktur der Stichprobe, die Anzahl (N) auf Ebene 1 und 2 kann je nach Auswahl an Kovariaten und Messzeitpunkt variieren.

Das Hauptziel der Modellbildung bestand darin, zu analysieren, ob Effekte auf das Treatment zurückgeführt werden können. Hierzu wurde zunächst für jede abhängige Variable ein Mehrebenenmodell aufgestellt: Ein Modell zur Untersuchung des Effekts des Treatments auf die Leistung, ein Modell zur Untersuchung des Effekts des Treatments auf das konzeptuelle Verständnis und ein Modell zur Analyse der Wirkung auf die Motivation. Weitere Modelle ergaben sich aus vertiefenden und weiterführenden Fragestellungen, so der Frage nach der Wirkung des variierten Treatments und der Analyse möglicher ATI-Effekte (Aptitude-Treatment-Interaktionen, zur Erklärung des Begriffs sei auf Kapitel 2.3.9.8 Vertiefende Analysen zu Aptitude-Treatment-Interaktionen hingewiesen). Die Berechnung der Modelle erfolgte mit der freien Statistiksoftware „R“ unter Verwendung des Pakets „nlme“³⁴, das von Pinheiro & Bates (2000) entwickelt wurde.

34 Letztes Update des Pakets Januar 2013 (vgl. Pinheiro & Bates, 2013).

In den folgenden Modellen wird die jeweilige interessierende Variable (Physikleistung, konzeptuelles Verständnis oder Schülermotivation) zum zweiten und dritten Messzeitpunkt in Abhängigkeit von den kognitiven Fähigkeiten, den vorherigen Leistungen (Noten), vom Geschlecht, vom Schultyp, von dem wahrgenommenen Lehrerengagements aus Sicht des jeweiligen Schülers und der Zuordnung zur Untersuchungsbedingung (Treatment- versus Kontrollgruppe) beschrieben.

Ziel des Modells ist es, die Änderung der jeweiligen interessierenden abhängigen Variablen in Relation zum ersten Messzeitpunkt unter Berücksichtigung relevanter Kovariaten zu beschreiben und auf statistische Signifikanz zu testen sowie die Stärke des Effekts zu schätzen.

Im Gegensatz zu klassischen Regressionsanalysen hat sich für Mehrebenenanalysen noch kein Standardverfahren für die Berechnung und den Bericht von Effektstärken etabliert (vgl. auch Hochweber, 2010, S. 153). Im Mittelpunkt dieser Arbeit steht die Analyse der Wirkung des Treatments auf die abhängigen Variablen Physikleistung bei repräsentationsbezogenen Aufgaben und auf das konzeptuelle Verständnis in Strahlenoptik. Zur Berechnung der Effektstärke wurde das Effektstärkemaß berechnet (vgl. Tymms, Merrel & Henderson, 1997; Tymms, 2004). Nach Tymms (2004, S. 56 f.) ergibt sich für dichotome Variablen aus der Differenz der betrachteten Mittelwerte (hier mittlerer Lernzuwachs) geteilt durch die gepoolte Standardabweichung (Wurzel aus der Varianz innerhalb der Gruppen). Angewandt auf die vorliegende Datenstruktur errechnet sich wie folgt:

$$\begin{aligned} \Delta_{\text{prä - post}} &= \frac{\text{Differenz der Klassenmittelwerte des Zuwachses zwischen Treatment und Kontrollgruppe zu } t_2}{\sqrt{\text{Varianz innerhalb einer Klasse zu } t_2}} \\ &= \frac{\beta_4}{\sqrt{\sigma_{\text{innerhalb Klasse zu } t_2}^2}} \\ &= \frac{\beta_4}{\sqrt{\sigma_p^2 + g_2^2 \sigma_\epsilon^2}} \end{aligned}$$

Die Berechnung für die Effekte prä – follow-up erfolgte in entsprechender Weise. Analog zu Cohen's d gilt = 0.20 als kleiner, = 0.50 als mittlerer und = 0.80 als großer Effekt (vgl. Bortz & Döring, 2005, S. 568). Zudem wurde je Level (Messzeitpunkt, Individuum und Schulklasse) der Anteil erklärter Varianz berichtet, welcher sich auf das Treatment zurückführen lässt.

Des Weiteren wurde zum Vergleich der Wirkung von Variablen, bei denen sich die Bandbreite möglicher Werte unterscheidet (z.B. Vergleich der Wirkung

von Noten und Intelligenz auf die jeweilige abhängige Variable), sowohl die Kovariaten als auch die abhängigen Variablen standardisiert. Vollstandardisierte Koeffizienten (standardisiert bezüglich Prädiktor und Kriterium) werden in den folgenden Modellen mit „ β “ angegeben, unstandardisierte Koeffizienten mit „ b “. Liegt das Kriterium als dichotome Variable vor (wie etwa bei den Variablen *Bedingung* oder *Geschlechtszugehörigkeit*), werden ebenfalls bezüglich des Prädiktors standardisierte Koeffizienten mit „ β “ angegeben (vgl. auch Fox, 1997, S. 153).

Da in den folgenden Modellen der Interzept die erwartete Leistung bzw. das konzeptuelle Verständnis eines Schülers darstellt, dessen Kovariaten den Wert 0 aufweisen, kann der Interzept für Variablen, die den Wert 0 nicht annehmen können, nicht sinnvoll interpretiert werden. Aus diesem Grund wurden alle metrischen Kovariaten grand-mean zentriert d.h. von allen verfügbaren Werten wurde der Mittelwert aller verfügbaren Werte abgezogen (vgl. Eid et al., 2011, S. 724; Raudenbush & Bryk, 2002, S. 31 ff.).

Nach der Zentrierung kann der Interzept als erwarteter Wert der abhängigen Variable eines absolut durchschnittlichen Schülers interpretiert werden, der sich in der Gruppe befindet, in der alle Dummy-kodierten kategorialen Variablen den Wert 0 annehmen (hier: Geschlecht = weiblich, Schultyp = IGS, Bedingung = Kontrollgruppe). Ein weiterer Vorteil der Zentrierung besteht darin, dass in Mehrebenenmodellen die Varianz des Interzepts interpretierbar wird (vgl. Hox, 2010, S. 62).

Im Gegensatz zu herkömmlichen Verfahren der linearen Regressionsanalyse ändert eine lineare Transformation der Kovariaten das Modell nicht essentiell (vgl. Eid et al, S. 724): Wird der Wert der Kovariate verdoppelt, halbiert sich der entsprechende Regressionskoeffizient. In Mehrebenenmodellen führt die Aufnahme von Random Slopes in Kombination mit linearen Transformationen (wie Grand-Mean-Centering) zu substantiellen Änderungen des Modells, die sich auf die inferenzstatistische Absicherung auswirken. Eine grafische Veranschaulichung dieses Problems findet sich in Hox (2010, S. 61). Da die im Folgenden aufgestellten Modelle keine Random-Slope-Terme enthalten, stellt sich das Problem in den vorgestellten Analysen nicht.

Mit fehlenden Werten wurde in den berechneten Modellen wie folgt umgegangen:

1. Umgang mit fehlenden Werten auf Ebene der Individuen und Messzeitpunkte:
 - Ein Vorteil des hier verwendeten Modells besteht darin, dass Werte von Schülern, die zu einem Messzeitpunkt fehlen, zu den anderen Messzeitpunkten berücksichtigt werden können, sofern alle relevanten Kovariaten

vorliegen. Wenn in späteren Analysen also Daten von $N = 443$ Schülern eingehen, bedeutet dies, dass pro Schüler alle berücksichtigten Kovariaten vorliegen und zu einem der drei Messzeitpunkte mindestens ein Test- bzw. Fragebogengesamtwert. Pro Messzeitpunkt betrachtet kann die Stichprobengröße von $N = 443$ abweichen.

- Zur Veranschaulichung ein Beispiel: liegen von einem Individuum Daten zu den kognitiven Fähigkeiten und den Schulnoten vor, jedoch keine Daten zu den Prätests, können die Daten der Posttests und Follow-up Tests (Physikleistung und konzeptuelles Verständnis) dennoch als abhängige Variablen in den Modellen für den zweiten und dritten Messzeitpunkt eingehen.
2. Umgang mit fehlenden Werten auf Itemebene:
- Da es sich beim Physikleistungs-, beim Konzepttest und bei den Skalen des I-S-T 2000 R um Wissens- bzw. Fähigkeitstests handelt, wurden unbeantwortete Items mit 0 Punkten (nicht korrekt gelöst) bewertet.
 - Anders verhält es sich bei dem Motivationsfragebogen inklusive der Skala wahrgenommenes Lehrerengagement aus Schülersicht (LES): Sofern mindestens ein Item unbeantwortet blieb, wurden je Skala die jeweilige Antwort mittels einer multiplen linearen Regression vorhergesagt, unter der Voraussetzung, dass 2/3 der Items pro Skala beantwortet wurden: deterministic multiple regression imputation (vgl. Buck, 1960; Little & Rubin, 2002).
 - Im Fall der Skala LES (4 Items) wurde die jeweilige fehlende Antwort imputiert, sobald drei von vier Items beantwortet waren. Für die Skala „SK“ (Selbstkonzept) war diese Bedingung erfüllt, sobald sechs von neun Antworten vorlagen usw.
 - Die Imputation mittels multipler Regression stellt unter der Annahme von „MAR“ (missing at random)³⁵ einen optimalen Schätzer dar und ist der Mittelwertersetzung überlegen. Sind die Bedingungen für MAR nicht erfüllt, führt diese Methode zur Unterschätzung der Varianz des Kriteriums und zur Überschätzung der Kovarianzen zwischen Kriterium und Prädiktoren (vgl. Lüdtke & Robitzsch, 2010, S. 13).
 - Da kein Anlass bestand, der für eine anderweitige Annahme sprach, wurde davon ausgegangen, dass MAR vorliegt – zumal der Aufwand für komplexere Verfahren wie etwa die Multiple Imputation mittels des R-Pakets MICE (vgl.

35 Die Annahme für MAR ist gegeben, wenn nach Kontrolle der beobachteten Variablen, z.B. LES 1-LES 3 das Auftreten der fehlenden Variable, hier z.B. LES 4 nicht mehr von der (Gesamt-)Ausprägung des wahrgenommenen Lehrerengagement aus Schülersicht abhängt – also die Beobachtung zufällig fehlt (vgl. ebd., S. 13).

van Buuren & Groothuis-Oudshoorn, 2011) für Variablen, deren Untersuchung nicht im Mittelpunkt dieser Arbeit stehen, ungerechtfertigt erscheint.

- Unter Anwendung der Imputation mittels multipler Regression konnten statt $N = 387$ nun $N = 489$ Fälle im Prätest, statt $N = 362$ nunmehr $N = 477$ Fälle im Posttest und statt $N = 391$, $N = 495$ Fälle im Follow-up Test berücksichtigt werden.

2.3.9.2 Methodisches Vorgehen: Modellentwicklung

Zum Verständnis des Modells, welches dieses Ziel umsetzt, wird im Folgenden stufenweise von einem „leeren“ Modell zu dem letztendlich verwendeten Modell übergegangen. Die Notation der jeweiligen Modelle erfolgt in Anlehnung an Fahrmeir, Kneib und Lang (2009, S. 254 ff.).

1) Modell 1 LM – Leeres Modell

$$y_{ij} = \beta_0 + \beta_1 \cdot t_2 + \beta_2 \cdot t_3 + \varepsilon_{ij}$$

y_{ij}	Leistung des Individuums i zum Zeitpunkt j
β_0	Mittelwert der Leistung zum Zeitpunkt 1
i	Zeitpunkt ($i = 1, 2, 3$)
β_1	Beschreibt die Änderung der Leistung zwischen Zeitpunkt 1 und Zeitpunkt 2 bzw. Zeitpunkt 1 und Zeitpunkt 3. Dieses Vorgehen entspricht einem einfachen Einzelvergleich (einfacher Kontrast in SPSS) bei einer Varianzanalyse mit Messwiederholung (vgl. Bortz, 2005b, S. 306; Rudolf & Müller, 2004, S. 100)
t_2 / t_3	Dummy-Kodierung der Zeitpunkte
	$t_2 = \begin{cases} 1 & \text{falls } i = 2 \\ 0 & \text{sonst} \end{cases}$ $t_3 = \begin{cases} 1 & \text{falls } i = 3 \\ 0 & \text{sonst} \end{cases}$
j	Individuum ($j = 1, \dots, 525$)
ε_{ij}	Residuum (Abweichung zwischen vorhergesagter Leistung und beobachteter Leistung des Individuums i zum Zeitpunkt j).

Die ε_{ij} werden als identisch und unabhängig normalverteilt angenommen mit Mittelwert 0 und zu schätzender Standardabweichung σ .

Dieses Modell beschreibt nur die Leistung zu den unterschiedlichen Zeitpunkten ohne Berücksichtigung von Gruppenstrukturen, Kovariaten etc.

Um die Vergleichbarkeit mit den nachfolgenden gemischten Modellen zu gewährleisten, wurden die Modellparameter mittels der Maximum Likelihood Methode ermittelt.

Die Nullhypothese entspricht dabei der Annahme: „Es gab im Mittel keine Änderung der Leistung zwischen dem ersten und zweiten Messzeitpunkt“. „ $H_0: \beta_1 = 0$ “. Diese Hypothese kann im nlme-Paket mittels eines Wald-Tests geprüft werden.

Tabelle 43 Modell 1 LM – Leeres Modell

Variable		<i>b</i>	<i>SE</i>	<i>F</i> (numDF,denDF)
β_0	Interzept	2.42	0.29	4436.28 (1, 1469) ***
β_1	Zuwachs prä – post	14.83	0.41	689.15 (1, 1469) ***
β_2	Zuwachs prä – follow-up	11.09	0.40	752.71 (1, 1469) ***

*** $p < .001$, $\sigma_\epsilon = 6.34$

Aufbauend auf dem „leeren Modell“ (vgl. Tabelle 43) wird nun ein Modell spezifiziert, welches berücksichtigt, dass die einzelnen Messwerte je einem Individuum zugeordnet sind. Dieses Modell entspricht einer Varianzanalyse mit Messwiederholung ohne Berücksichtigung von Kovariaten (vgl. Eid et al., 2011, S. 449 ff.).

2) Modell 2PM (Personenparameter) – Varianzanalyse mit Messwiederholung

$$y_{ij} = \beta_0 + p_j + \beta_1 \cdot t_2 + \beta_2 \cdot t_3 + \epsilon_{ij}$$

p_j Individuenparameter (geschätzt pro Individuum).

p_j wird als normalverteilt mit Mittelwert 0, zu schätzender Varianz und unabhängig von ϵ_{ij} angenommen (vgl. Hedeker, 2012, S. 3).

p_j ist der Unterschied in der Leistung der Person j über die Messzeitpunkte hinweg vom Mittelwert aller Leistungen aller Personen über die Messzeitpunkte hinweg (vgl. Eid et al., 2011, S. 450).

In diesem Modell wird die Korrelation der Messzeitpunkte im Individuum durch p_j adressiert. Die Varianz von p_j entspricht der Varianz zwischen den Individuen (vgl. Tabelle 44).

Tabelle 44 Modell 2 PM (Personenparameter) – Varianzanalyse mit Messwiederholung

Variable		<i>b</i>	<i>SE</i>	<i>F</i> _(numDF, denDF)
β_0	Interzept	2.26	0.28	2392.87 (1, 946)***
β_1	Zuwachs prä – post	14.90	0.31	1207.91 (1, 946)***
β_2	Zuwachs prä – follow-up	11.16	0.31	1328.43 (1, 946)***

*** $p < .001$, $\sigma_p = 4.21$, $\sigma_\varepsilon = 4.76$

Die Intraklassen-Korrelation (*ICC*), abgekürzt $\hat{\Omega}$, ergibt sich auf Ebene der Population wie folgt (vgl. Eid et al., 2011, S. 704):

$$\hat{\rho} = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_\varepsilon^2} = \frac{17.73}{17.73 + 22.67} = 0.44$$

p entspricht hier dem Individuenparameter und damit $\hat{\sigma}_p^2$ der Varianz zwischen den Individuen. Damit ist $\hat{\sigma}_{\text{Level-2}}^2 = 17.73$.

ε_{ij} entspricht dem der Abweichung zwischen vorhergesagter Leistung und beobachteter Leistung des Individuums i zum Zeitpunkt j . Damit ist $\hat{\sigma}_{\text{Level-1}}^2 = 22.67$.

In diesem Modell wurde noch nicht berücksichtigt, dass die Individuen je einer Schulklasse zugeordnet sind. Im folgenden Schritt soll geprüft werden, ob ein Modell, welches einbezieht, dass die Leistung von Individuen einer Klasse korreliert sein könnte, dem vorigen Modell überlegen ist.

3) Modell 3KM (Klassenparameter) – Varianzanalyse unter Berücksichtigung der Klassenstruktur

$$y_{ijk} = \beta_0 + u_k + \beta_1 \cdot t_2 + \beta_2 \cdot t_3 + \varepsilon_{ij}$$

u_k

Klassenparameter (geschätzt pro Schulklasse)

u_k wird als normalverteilt mit Mittelwert 0 zu schätzender Varianz und unabhängig von ε_{ij} angenommen.

In Anlehnung an Modell 2 PM (Personenparameter) – Varianzanalyse mit Messwiederholung (siehe Tabelle 44) wird nun die Korrelation der Individuen inner-

halb der Schulklasse untersucht.³⁶ Das Modell berücksichtigt mit dem Random Interzept u_k , dass die in der Stichprobe enthaltenen Schulklassen lediglich eine Auswahl an Schulklassen aus der Grundgesamtheit darstellt (vgl. Tabelle 45). Durch u_k erhalten wir einen Term, der eine Korrelationsstruktur der Individuen (*ICC*), die jeweils einer Schulklasse angehören, schätzt. In dem aufgestellten Modell entspricht die Korrelation der Leistungen der Individuen einer Schulklasse untereinander nun gerade dem *ICC* (vgl. Hochweber, 2010, S. 144). Die Varianz von u_k ist gleich der Varianz zwischen den Schulklassen.

Tabelle 45 Modell 3 KM (Klassenparameter) – Varianzanalyse unter Berücksichtigung der Klassenstruktur

Variable		<i>b</i>	<i>SE</i>	<i>F</i> (<i>numDF</i> , <i>denDF</i>)
β_0	Interzept	2.12	0.75	224.42 (1, 1449)***
β_1	Zuwachs prä – post	15.02	0.35	942.21 (1, 1449)***
β_2	Zuwachs prä – follow-up	11.22	0.35	1026.50 (1, 1449)***

*** $p < .001$, $\sigma_u = 3.25$, $\sigma_\varepsilon = 5.50$

Die Intraklassen-Korrelation ergibt sich wie folgt (vgl. Eid et al., 2011, S. 704)

$$\hat{\rho} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_\varepsilon^2} = \frac{10.55}{10.55 + 30.23} = 0.26$$

u_k entspricht hier dem Klassenparameter. Damit ist $\hat{\sigma}_{\text{Level-2}}^2 = 10.55$.

ε_{ij} entspricht wieder dem der Abweichung zwischen vorhergesagter Leistung und beobachteter Leistung des Individuums *i* zum Zeitpunkt *j*. Damit ist $\hat{\sigma}_{\text{Level-1}}^2 = 30.23$.

Im Ergebnis zeigt sich, dass der *ICC* dabei hinreichend hoch ist, um die Berücksichtigung der Klassenstruktur durch einen Random Interzept in der Schulklasse zu rechtfertigen (vgl. Eid et al., 2011, S. 705). Werden beide Ebenen (Messzeitpunkte und Schulklassen) berücksichtigt, ergibt sich folgendes Modell:

³⁶ Dieses Modell berücksichtigt nicht die Individuenstruktur, es werden lediglich die Messungen zu den verschiedenen Zeitpunkten miteinander verglichen (ohne den Messungen Individuen zuzuordnen).

4) Modell 4 KPM (Klassen- und Personenparameter) – Varianzanalyse unter Berücksichtigung von sowohl der Klassenstruktur als auch der Individuenkomponente

$$y_{ik} = \beta_0 + p_j + u_k + \beta_1 \cdot t_2 + \beta_2 \cdot t_3 + \varepsilon_{ij}$$

u_k / p_j Klassenparameter (geschätzt pro Schulklasse) / Personenparameter (geschätzt pro Individuum)

Beide Parameter werden als normalverteilt mit Mittelwert 0, zu schätzender Varianz und unabhängig voneinander und ε_{ij} angenommen.

Verschiedene Autoren wie Hox und Kreft (1994), Hox (2002) sowie Maas und Hox (2004) fordern überwiegend eine Fallzahl von $N \geq 50$ auf höheren Ebenen. Allerdings bezieht sich diese Forderung insbesondere auf Fragestellungen, in denen explizit die Interaktion zwischen den Ebenen untersucht werden soll. Aufgrund der besseren Schätzung der Messfehler und der Möglichkeit, durch die Mehrebenenstruktur der Verletzung der Unabhängigkeitsannahme gerecht zu werden, wurde die Klassenstruktur in das Modell aufgenommen (vgl. Tabelle 46). Die Ergebnisse sind somit mit dem hier untersuchten Datensatz von $N = 21$ (Schulklassen) nur unter dem Vorbehalt zu interpretieren, dass die Schätzung des Einflusses der Schulklasse mit Unsicherheit behaftet ist.

Folgende Gründe sprachen dafür, diesen Vorbehalt in Kauf zu nehmen:

- Erstens geht es in der Modellbildung vorwiegend um den Einfluss des Treatments (UV = Bedingung). Somit soll der Einfluss der Schulklasse kontrolliert werden und ist nicht selbst Gegenstand der Forschungsfrage. Gemäß Eid et al. (2011, S. 715) genügt bereits ein Stichprobenumfang von $N = 10$ Einheiten der Aggregatebene, sofern „man lediglich an einer Schätzung der Parameter im festen Teil des Modells interessiert (ist)“ (zit.n. ebd.). Daher wird angenommen, dass die fixen Effekte (wie der Einfluss des Treatments) zuverlässig geschätzt werden.
- Zweitens ergibt sich bereits für einen ICC von .10 bei $N = 10$ auf der Aggregatebene ein 50 % Risiko einer statistischen Fehlentscheidung (vgl. Eid et al., 2011, S. 705).

Nach Einschätzung der Autorin wiegt der Vorteil somit die Nachteile auf. Zu den Ergebnissen sei auf Tabelle 46 hingewiesen.

Tabelle 46 Modell 4KPM (Klassen- und Personenparameter) – Varianzanalyse unter Berücksichtigung von sowohl der Klassenstruktur als auch der Individuenkomponente

Variable		<i>b</i>	<i>SE</i>	<i>F</i> _(numDF, denDF)
β_0	Interzept	2.09	0.75	222.33 (1, 946)***
β_1	Zuwachs prä – post	14.98	0.31	1236.87 (1, 946)***
β_2	Zuwachs prä – follow-up	11.21	0.30	1355.58 (1, 946)***

*** $p < .001$, $\sigma_p = 2,81$, $\sigma_u = 3,22$, $\sigma_\varepsilon = 4,74$

Die ICCs auf den entsprechenden Ebenen erscheinen hinreichend groß, um sowohl die Schachtelung der Messzeitpunkte als auch die Schachtelung der Individuen in Klassen zu rechtfertigen:

- Die Varianz innerhalb des Individuums entspricht der Varianz der Residuen $\hat{\sigma}_\varepsilon^2 = 22.56$
- Die Varianz zwischen den Individuen einer Klasse ist gleich der Varianz des Random Interzept auf Ebene des Individuums $\hat{\sigma}_p^2 = 7.87$.
- Die Varianz zwischen den Klassen entspricht der Varianz des Random Interzept auf Ebene der Schulklasse $\hat{\sigma}_u^2 = 10.55$

Die Intraklassen-Korrelation ergibt sich auf Ebene der Population wie folgt (vgl. Eid et al., 2011, S. 704):

$$\begin{aligned} \text{ICC Level 3:} \quad \hat{\rho}_3 &= \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_p^2 + \hat{\sigma}_\varepsilon^2} = \frac{10.55}{10.55 + 7.87 + 22.56} = 0.25 \\ &= \text{„Ähnlichkeit der Individuen einer Klasse“} \end{aligned}$$

$$\begin{aligned} \text{ICC Level 2:} \quad \hat{\rho}_2 &= \frac{\hat{\sigma}_u^2 + \hat{\sigma}_p^2}{\hat{\sigma}_u^2 + \hat{\sigma}_p^2 + \hat{\sigma}_\varepsilon^2} = \frac{10.55 + 7.87}{10.55 + 7.87 + 22.56} = 0.45 \\ &= \text{„Ähnlichkeit der Messzeitpunkte eines Individuums“} \end{aligned}$$

5) Modell 5 KPM&H – Varianzanalyse unter Berücksichtigung von sowohl der Klassenstruktur als auch der Individuenkomponente unter Adressierung der Heteroskedastizität

Die Betrachtung der Boxplots der Ergebnisse des Leistungstests zu den verschiedenen Messzeitpunkten lässt unterschiedliche Varianzen (Heteroskedastizität) erkennen (siehe *Abbildung 53*, S. 289). So zeigt sich in den Ergebnisse des Prätests ein starker „Bodeneffekt“ und damit verbunden eine geringere Varianz (siehe auch die nachfolgenden Histogramme, *Abbildung 52*):

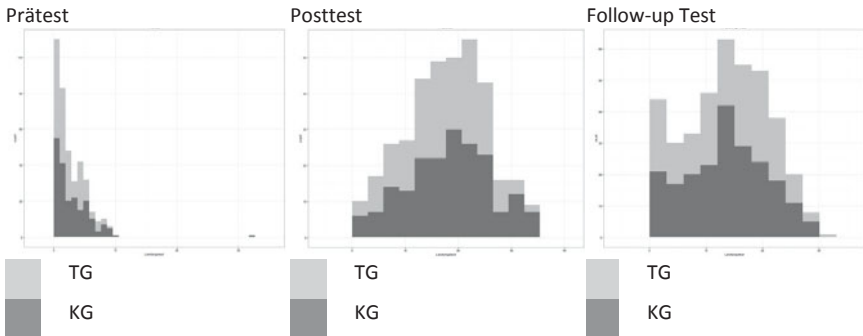


Abbildung 52: Histogramme der Leistungstestergebnisse je Messzeitpunkt, TG = Treatmentgruppe, KG = Kontrollgruppe

Der Grund besteht sehr wahrscheinlich darin, dass die Schüler zum ersten Messzeitpunkt nur geringe Vorkenntnisse mitbringen, da die Bildentstehung an der Sammellinse noch ein neues Thema ist. Diese geringen Vorkenntnisse konnten vermutlich auch nicht durch früheren Physikunterricht aufgefangen werden, weil es sich sowohl für die Siebt- als auch für die Achtklässler um das erste Schuljahr des Physikunterrichts handelt. Nach der Unterrichtsreihe fällt die Varianz entsprechend größer aus, wobei sich die Verteilungen im Post- und im Follow-up Test ähneln.

Aufgrund der Heteroskedastizität wird im nächsten Schritt zusätzlich die Möglichkeit unterschiedlicher Varianzen in den Fehlertermen je Messzeitpunkt aufgenommen, welche durch die Gewichtskomponente g geschätzt wird (siehe auch *Tabelle 47*).

Durch die Hinzunahme der Fehlergewichte können keine erheblichen Änderungen der geschätzten fixen Effekte des Modells beobachtet werden, wohl aber eine Verbesserung des Modells, wie aus *Tabelle 49* ersichtlich wird, welche einen Überblick zu den Informationskriterien bietet.

$$y_{ij} = \beta_0 + p_j + u_k + \beta_1 \cdot t_2 + \beta_2 \cdot t_3 + \varepsilon_{ij}$$

wobei Varianz von ε_{ij} gleich $g_i^2 \cdot \sigma_\varepsilon^2$ ist.

g_i Gewichtungskomponente. Die Gewichtungskomponenten wurden mittels des Befehls „weights“ geschätzt, der sich dafür eignet, Heteroskedastizität innerhalb von Gruppen abzubilden.³⁷

$g_1 = 1$, g_2 und g_3 werden geschätzt.

g_2 beschreibt den (multiplikativen) Unterschied der Standardabweichung der Fehler zum Zeitpunkt 1 (σ_ε) im Vergleich zu Zeitpunkt 2, g_3 analog dazu den (multiplikativen) Unterschied der Standardabweichung der Fehler zum Zeitpunkt 1 im Vergleich zu Zeitpunkt 3.

Man beachte, dass im Programmpaket nlme nicht die Standardabweichungen der Fehlerterme je Messzeitpunkt ausgegeben werden, sondern die Standardabweichung zum Messzeitpunkt 1 und die Gewichte für die anderen Messzeitpunkte. Die Standardabweichung zum Messzeitpunkt 2 bzw. 3 ergibt sich als Produkt aus der Standardabweichung zum Messzeitpunkt 1 und dem Gewicht des Zeitpunktes 2 bzw. 3.

Tabelle 47 Modell 5 KPM&H – Varianzanalyse unter Berücksichtigung von sowohl der Klassenstruktur als auch der Individuenkomponente unter Adressierung der Heteroskedastizität

Variable		<i>b</i>	<i>SE</i>	<i>F</i> (numDF, denDF)
β_0	Interzept	2.30	0.39	91.51 (1, 946)
β_1	Zuwachs prä – post	14.90	0.32	2001.67 (1, 946)
β_2	Zuwachs prä – follow-up	11.15	0.30	1336.99 (1, 946)

*** $p < .001$, $\sigma_p = 1,75$, $\sigma_u = 1,71$, $\sigma_\varepsilon = 1,59$, $g_1 = 1$, $g_2 = 4.26$, $g_3 = 4.12$

6) Modell 6 KPM&H&Korr – Varianzanalyse unter Berücksichtigung von sowohl der Klassenstruktur als auch der Individuenkomponente unter Adressierung der Heteroskedastizität mit unstrukturierter Korrelationsmatrix der Fehlerterme im Individuum

Die durch den Individuenparameter p_j alleine implizierte Korrelation der Leistungen zu den drei Zeitpunkten innerhalb eines Individuums erscheint zu starr. Das R Paket nlme erlaubt im Gegensatz zu dem R Paket lme4 die Schätzung

37 „Weights: an optional varFunc object or one-sided formula describing the within-group heteroscedasticity structure“ (zit. n. Pinheiro & Bates, 2013, S. 145).

von beliebigen Korrelationen der Fehler im Individuum, welches wir in dem Modell mit aufnehmen (vgl. Tabelle 48 zu den Ergebnissen). Die Informationskriterien des *AIC* und *BIC* stützen dieses Vorgehen (siehe Tabelle 49).

$$y_{ij} = \beta_0 + p_j + u_k + \beta_1 \cdot t_2 + \beta_2 \cdot t_3 + \varepsilon_{ij}$$

$$= \beta_0 + p_j + u_k + \varepsilon_{1j} \cdot t_1 + (\beta_1 + \varepsilon_{2j}) \cdot t_2 + (\beta_2 + \varepsilon_{3j}) \cdot t_3 \quad (\text{identische Umformulierung})$$

ε_{ij} Fehlerterm: Abweichung der tatsächlich gemessenen Leistung des Schülers j der Klasse k zum Zeitpunkt i von der durch das Modell vorhergesagten Leistung.
(Random Variable)

ε_{ij} ist für jede Kombination aus i/j Realisation einer (univariat) normalverteilten Zufallsvariablen mit Mittelwert $\mu = 0$ und Standardabweichung $g_i \cdot \sigma_\varepsilon$. Die Standardabweichung ist damit für jeden Zeitpunkt $i = 1, 2, 3$ unterschiedlich.

(man beachte, dass $g_1 = 1$ ist / siehe oben)

Damit adressiert der Fehler über die Gewichte g_i die unterschiedlichen Varianzen der Leistungstestergebnisse zu den verschiedenen Messzeitpunkten (Heteroskedastizität).

Ferner wird die Korrelation der Messzeitpunkte untereinander berücksichtigt, welche sich aus dem repeated Measurement Design ergeben. Dies ist zwar bereits durch den Individuenparameter p adressiert. Durch diesen ist jedoch die Korrelation der Messzeitpunkte immer gleich, wohingegen die beliebige Korrelation der Fehlerterme höhere Flexibilität bietet.

Man beachte auch die Darstellung:

$$\varepsilon_{ij} = t_1 \varepsilon_{1j} + t_2 \varepsilon_{2j} + t_3 \varepsilon_{3j}$$

Für die Korrelation der Fehlerterme ergibt sich:

$$\begin{pmatrix} 1 & 0.15 & 0.18 \\ 0.15 & 1 & 0.61 \\ 0.18 & 0.61 & 1 \end{pmatrix}, \text{ d.h. } \begin{matrix} \rho_{12} = 0.15 \\ \rho_{13} = 0.18 \\ \rho_{23} = 0.61 \end{matrix}$$

Die Korrelation der Leistungen eines Individuums zu den Messzeitpunkten ergibt sich im Modell nun als Kombination der Effekte durch die Individuenkomponente p_j und der Korrelation der Fehlerterme.

Als Element zur Erzeugung einer Korrelationsstruktur innerhalb des Individuums benötigt man die Individuenkomponente p_j nun im Grunde nicht

mehr. Die Komponente wurde dennoch beibehalten, um einerseits die Gruppenstruktur der Messungen weiterhin deutlich zu machen und andererseits die Varianz zwischen den Individuen weiterhin abschätzen zu können. Eine Herausnahme des Terms verbessert jeweils *AIC* und *BIC* bei gleichbleibender *Devianz* und Loglikelihood, wobei sich die Korrelationsschätzungen der Fehlerterme entsprechend anpassen. Die durch p_j „erklärte“ Varianz wird dann durch höhere Fehlervarianzen aufgefangen.

Tabelle 48 Modell 6 KPM&H&Korr – Varianzanalyse unter Berücksichtigung von sowohl der Klassenstruktur als auch der Individuenkomponente unter Adressierung der Heteroskedastizität mit unstrukturierter Korrelationsmatrix der Fehlerterme im Individuum

Variable	<i>b</i>	<i>SD</i>	$F_{(numDF, denDF)}$
β_0 Interzept	2.33	0.37	65.29 _(1, 946) ***
β_1 Zuwachs prä – post	14.75	0.32	955.74 _(1, 946) ***
β_2 Zuwachs prä – follow-up	11.01	0.31	1274.79 _(1, 946) ***

*** $p < .001$, $\sigma_p = 1,76$, $\sigma_u = 1,61$, $\sigma_\epsilon = 1,97$, $g_1 = 1$, $g_2 = 3,62$, $g_3 = 3,53$

Zur Beurteilung der Modellgüte wurden die Informationskriterien *AIC* und *BIC* sowie die *Devianz* des Modells herangezogen. Beim *AIC* und *BIC* handelt es sich um Maße der Anpassungsgüte des Modells an die vorliegenden Daten unter Berücksichtigung der Modellkomplexität. *AIC* und *BIC* sind Maßzahlen für den Informationsverlust. Beide errechnen sich aus der logarithmierten Likelihood-Funktion und einem Strafterm, dessen Größe von der Anzahl der geschätzten Parameter abhängt. Im Gegensatz zum *AIC* fallen beim *BIC* komplexere Annahmen stärker ins Gewicht. Weniger sparsame Modelle werden beim *BIC* also stärker bestraft als beim *AIC* (vgl. Moosbrugger & Kelava, 2007, S. 390).

AIC und *BIC* lassen sich nicht direkt interpretieren, sondern können nur als relative Werte innerhalb eines Modellvergleichs auf Basis derselben Stichprobe zueinander in Beziehung gesetzt werden. Die *Devianz* stellt ein weiteres Kriterium dar mit dessen Hilfe die Passung des Modells auf die jeweiligen Daten beurteilt werden kann. Sie wird aus der maximierten doppelten negativen Log-Likelihood berechnet (vgl. Hox, 2002, S. 42; Eid et al., 2011, S. 715):

$$DEV = -2\ln(L)$$

Auch für die *Devianz* gilt, je größer der Wert, desto schlechter ist die Passung des Modells. Wie aus der Tabelle 49 ersichtlich ist, zeichnet sich das letzte Modell (KPM&H&Korr) durch die besten Modelleigenschaften aus: So weist es bezüglich des *AIC* (Akaike's Informationskriterium), des *BIC* (Bayesian Informationskriterium) und der *Devianz* die geringsten Werte auf. Daher wurde das Modell, welches sowohl die Klassenstruktur, die Messwiederholung durch die Individuenkomponente als auch die Heteroskedastizität berücksichtigt und eine unstrukturierte Korrelationsmatrix der Fehlerterme im Individuum zugrunde legt, als Ausgangsmodell gewählt.

Tabelle 49 Überblick über die Modelleigenschaften bei steigender Komplexität

Nr.	Modell- kürzel	Berücksichtigte Para- meter	AIC	BIC	Devianz	σ_u^2	σ_p^2	σ_ε^2 zu t_1	σ_ε^2 zu t_2	σ_ε^2 zu t_3
1	LM	Leeres Modell	9623.66	9644.83	9615.66	-	-	-	40.20	-
2	PM	LM + Personen-para- meter (PM)	9386.67	9413.14	9376.67	-	17.73	-	22.67	-
3	KM	LM + Klassen-para-me- ter (KM)	9273.04	9299.52	9263.04	10.55	-	-	30.23	-
4	KPM	LM + KM + PM	9192.13	9223.90	9180.13	10.35	7.87	-	22.56	-
5	KPM&H	KPM + Heteros-keda- stizität (H)	8850.14	8892.49	8834.13	2.92	3.05	2.54	46.22	43.20
6	KPM&H &Korr	KPM&H + unstruktu- rierte Korrelation der Fehler zu den Mess- zeitpunkten	8669.04	8727.28	8647.04	2.60	1.38	3.88	51.01	48.45

Tabelle 50 Überblick zum Einfluss der Kovariaten auf Wissen und Problemlösen bei repräsentationsbezogenen Aufgaben (Physikleistung)

	Prätest			Posttest			Follow-up Test		
	<i>b</i>	β	$F_{(numDF, denDF)}$	<i>b</i>	β	$F_{(numDF, denDF)}$	<i>b</i>	β	$F_{(numDF, denDF)}$
Notenfaktor PCA (D, M, Ph)	-0.39	-0.06	30.35 (1, 496)***	-2.30	-0.36	122.45 (1, 936)***	-2.21	-0.34	119.18 (1, 936)***
IQ-verbal (Satzergänzung)	0.05	0.06	30.01 (1, 458)***	0.17	0.20	30.39 (1, 873)***	0.13	0.16	20.47 (1, 873)***
IQ-räumlich (Würfelaufgaben)	0.00	0.00	0.05 (1, 458)	0.18	0.18	25.81 (1, 873)***	0.13	0.13	13.10 (1, 873)**
IQ-figural-logisch (Matrizen)	0.02	0.02	3.03 (1, 456)	0.15	0.16	18.80 (1, 869)***	0.13	0.14	15.16 (1, 869)**
LES ^a	0.00	0.00	0.02 (1, 465)	0.17	0.09	6.38 (1, 900)*	0.02	0.01	0.11 (1, 900)
Geschlecht	0.04	0.00	0.06 (1, 502)	0.49	0.06	0.60 (1, 942)	-0.91	-0.10	2.22 (1, 942)
Klassengröße	0.02	0.01	0.05 (1, 18)	0.34	0.11	1.61 (1, 942)	0.38	0.13	14.80 (1, 942)**
Schultyp	1.88	0.21	7.24 (1, 18)*	7.86	0.88	104.27 (1, 942)***	6.27	0.70	66.04 (1, 942)***

* $p < .05$, ** $p < .01$, *** $p < .001$

^aLehrerengagement aus Schülersicht

Nachdem das Ausgangsmodell festgelegt ist, steht die Überlegung an, welche Kovariaten in das Modell eingehen sollen. Basierend auf dem Ausgangsmodell wurde jeweils eine Kovariante hinzugenommen und analysiert, ob die jeweilige Kovariante je Messzeitpunkt einen signifikanten Erklärungswert für die abhängige Variable aufweist. Die betrachteten Modelle sehen damit jeweils wie folgt aus:

$$y_{ij} = \beta_0 + p_j + u_k + \beta_1 \cdot t_2 + \beta_2 \cdot t_3 + \beta_3 \cdot x + \beta_4 \cdot t_2 \cdot x + \beta_5 \cdot t_3 \cdot x + \varepsilon_{ij}$$

x jeweilige Kovariante

Die Analyse zeigt, dass sich insbesondere die kognitiven Fähigkeiten (Subskalen des I-S-T 2000 R), die vorherigen Schulleistungen (Noten) und der Schultyp signifikant auf das Abschneiden im Leistungstest zu allen drei Messzeitpunkten auswirken (vgl. Tabelle 50). Das Lehrerengagement aus Sicht der Schüler hat einen signifikanten Einfluss auf den Lernzuwachs prä – post ($b = 0.17$, $\beta = 0.09$, $p < .01$). Die Klassengröße wirkt sich lediglich signifikant auf den Lernzuwachs prä – follow-up aus ($b = 0.38$, $\beta = 0.13$, $p < .01$). Da somit jede der infrage kommenden Variablen außer der Geschlechtszugehörigkeit zu einem der drei Zeitpunkte einen signifikanten Erklärungswert aufweist, wurden alle Variablen bis auf das Geschlecht in das Modell aufgenommen.

Unter Berücksichtigung der genannten Variablen ergibt sich nun folgendes Gesamtmodell:

$$\begin{array}{lll}
 y_{ijk} = & \beta_0 + p_j + u_k & + \beta_1 t_2 & + \beta_2 t_3 \\
 & + \beta_3 \text{ Bedingung}_k & + \beta_4 t_2 \text{ Bedingung}_k & + \beta_5 t_3 \text{ Bedingung}_k \\
 & + \beta_6 \text{ Notenfaktor}_j & + \beta_7 t_2 \text{ Notenfaktor}_j & + \beta_8 t_3 \text{ Notenfaktor}_j \\
 & + \beta_9 \text{ IQ}_{\text{verbal } j} & + \beta_{10} t_2 \text{ IQ}_{\text{verbal } j} & + \beta_{11} t_3 \text{ IQ}_{\text{verbal } j} \\
 & + \beta_{12} \text{ IQ}_{\text{figural-räumlich } j} & + \beta_{13} t_2 \text{ IQ}_{\text{figural-räumlich } j} & + \beta_{14} t_3 \text{ IQ}_{\text{figural-räumlich } j} \\
 & + \beta_{15} \text{ IQ}_{\text{figural-logisch } j} & + \beta_{16} t_2 \text{ IQ}_{\text{figural-logisch } j} & + \beta_{17} t_3 \text{ IQ}_{\text{figural-logisch } j} \\
 & + \beta_{18} \text{ Klassengröße}_k & + \beta_{19} t_2 \text{ Klassengröße}_k & + \beta_{20} t_3 \text{ Klassengröße}_k \\
 & + \beta_{21} \text{ Schultyp}_k & + \beta_{22} t_2 \text{ Schultyp}_k & + \beta_{23} t_3 \text{ Schultyp}_k \\
 & + \beta_{24} \text{ LES}_j & + \beta_{25} t_2 \text{ LES}_j & + \beta_{26} t_3 \text{ LES}_j \\
 & + \varepsilon_{ij} & &
 \end{array}$$

Erklärungen zu den Variablen:

i	Zeitpunkt (i = 1,2,3).
j	Individuum (j = 1,...,525).

k	Klasse (k = 1,...,21).
y _{ijk} (abhängige Variable)	Gemessene Leistung des Individuums j aus der Klasse k zum Zeitpunkt i.
β ₀ (fixed variable)	Geschätzte Leistung einer durchschnittlichen weiblichen Schülerin, aus der Kontrollbedingung, welche eine durchschnittlich große Klasse einer Gesamtschule besucht und auch bezüglich kognitiver Fähigkeiten und Noten durchschnittlich abschneidet: der Mittelwert der Kovariaten Lehrereengagement aus Schülersicht (LES), IQ _{verbal} , IQ _{figural-räumlich} , IQ _{figural-logisch} , Notenfaktor beträgt jeweils 0, da die genannten Variablen grand-mean-zentriert wurden.
	Mittlerer Unterschied der Leistungen des Schülers j zum Schnitt seiner Klasse. Dieser mittlere Unterschied wird über alle Messzeitpunkte hinweg geschätzt. D.h. wenn für j = 1 die geschätzte Größe p ₁ = 2 ist, so gehen wir davon aus, dass der Schüler zu allen Messzeitpunkten um 2 Punkte besser ist, als seine Klassenkameraden (bei gleichen Werten der Kovariaten).

p_j ist für jeden Schüler j eine Realisation einer normalverteilten Zufallsvariablen mit Mittelwert μ = 0 und Standardabweichung σ = σ_p. p_j ist unabhängig von allen anderen Random Termen des Modells. Die Komponente p_j ermöglicht zusammen mit der unstrukturierten Korrelationsmatrix der Fehlerterme ε (siehe unten) die Korrelation der Messzeitpunkte i₁ und i₂ innerhalb eines Individuums j. Man kann errechnen:

p_j
(Random Variable)

$$\rho(y_{i_1 j k}, y_{i_2 j k}) = \frac{\sigma_p^2 + \sigma_u^2 + \rho_{i_1 i_2} \cdot g_{i_1} \cdot g_{i_2} \cdot \sigma_\epsilon^2}{\sqrt{\sigma_u^2 + \sigma_p^2 + g_{i_1}^2 \cdot \sigma_\epsilon^2} \cdot \sqrt{\sigma_u^2 + \sigma_p^2 + g_{i_2}^2 \cdot \sigma_\epsilon^2}}$$

- σ_u² Varianz der Zufallsvariablen u
- σ_p² Varianz der Zufallsvariablen v
- σ_ε² Varianz der Zufallsvariablen ε zum Zeitpunkt 1
- g_{i1}/g_{i2} Gewicht des Zeitpunkts i₁/i₂ (siehe unten)

p_j adressiert damit Unterschiede zwischen Schülern, welche auf nicht erhobene Kovariaten (z.B. sozio-ökonomischer Status der Eltern, soziale Kompetenz, sonstige unbekannte Einflüsse) zurückzuführen sind.

Mittlerer Unterschied der Leistungen von Schülern in Klasse k zum Schnitt über alle Klassen. Dieser mittlere Unterschied wird über alle Messzeitpunkte hinweg geschätzt. D.h. wenn für $k = 1$ die geschätzte Größe $u_1 = 2$ ist, so gehen wir davon aus, dass alle Schüler der Klasse 1 zu allen Messzeitpunkten um 2 Punkte besser sind, als das Modell es andernfalls vorhersagen würde.

u_k ist für jede Klasse k eine Realisation einer normalverteilten Zufallsvariablen mit Mittelwert $\mu = 0$ und Standardabweichung $\sigma = \sigma_u$; u_k ist unabhängig von allen anderen Random Termen des Modells; u_k sorgt für die Korrelation ρ der Leistungen y aller Individuen j_1, j_2, \dots einer Klasse k zum selben Messzeitpunkt i :

u_k
(Random Variable)

$$\rho(y_{i j_1 k}, y_{i j_2 k}) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_p^2 + \sigma_{\epsilon_i}^2}$$

σ_u^2 Varianz der Zufallsvariablen u
 σ_p^2 Varianz der Zufallsvariablen v
 $\sigma_{\epsilon_i}^2$ Varianz der Zufallsvariablen ϵ zum Zeitpunkt i
 (= $g_i^2 \sigma_\epsilon^2$)

u_k adressiert damit Unterschiede zwischen Klassen, welche auf nicht erhobene Kovariaten (z.B. Klassenklima, regionale Einflüsse wie z.B. wohnortsgebundener sozio-ökonomischer Status, Stadt vs. Land etc.) zurückzuführen sind.

$$t_1 \text{ (Dummy-Kodierung)} = \begin{cases} 1 & \text{falls } i = 1 \\ 0 & \text{sonst} \end{cases}$$

$$t_2 \text{ (Dummy-Kodierung)} = \begin{cases} 1 & \text{falls } i = 2 \\ 0 & \text{sonst} \end{cases}$$

$$t_3 \text{ (Dummy-Kodierung)} = \begin{cases} 1 & \text{falls } i = 3 \\ 0 & \text{sonst} \end{cases}$$

β_1
(fixed variable)

Mittlere Änderung der Leistung (= Leistungszuwachs) zwischen Zeitpunkt $i = 1$ und Zeitpunkt $i = 2$.
 Mittlere Änderung bedeutet dabei: die Änderung der Leistung, wenn von einer „durchschnittlichen“ Schülerin (siehe β_0) ausgegangen wird.

β_2 (fixed variable)	Mittlere Änderung der Leistung (= Leistungszuwachs) zwischen Zeitpunkt $i = 1$ und Zeitpunkt $i = 3$. Mittlere Änderung bedeutet dabei: die Änderung der Leistung, wenn von einer „durchschnittlichen“ Schülerin (siehe β_0) ausgegangen wird.
Bedingung _k (Dummy Kodierung / unabhängige Variable)	$= \begin{cases} 1 & \text{falls Klasse } k \text{ in der Treatmentgruppe ist} \\ 0 & \text{sonst} \end{cases}$
β_3 Bedingung _k (fixed variable)	„Einfluss“ der Variable Bedingung zum Zeitpunkt $i = 1$. Da zu diesem Zeitpunkt noch kein Einfluss des Treatments vorliegt, misst β_{13} den Leistungsunterschied zwischen Treatment- und Kontrollgruppe zum Zeitpunkt $i = 1$.
β_4 Bedingung _k (fixed variable)	Einfluss der Variable Bedingung auf die Änderung der Leistung zum Zeitpunkt $i = 2$ im Vergleich zum Zeitpunkt $i = 1$. Bsp.: Es sei $\beta_4 = 2$. Dann sagt das Modell für Schüler aus der Treatmentgruppe zum Zeitpunkt $i = 2$ einen um 2 Punkte <i>höheren Leistungszuwachs</i> zwischen $i = 1$ und $i = 2$ voraus, als in der Kontrollgruppe.
β_5 Bedingung _k (fixed variable)	Einfluss der Variable Bedingung auf die Änderung der Leistung zum Zeitpunkt $i = 3$ im Vergleich zum Zeitpunkt $i = 1$. Bsp.: Es sei $\beta_5 = 2$. Dann sagt das Modell für Schüler aus der Treatmentgruppe zum Zeitpunkt $i = 3$ einen um 2 Punkte <i>höheren Leistungszuwachs</i> zwischen $i = 1$ und $i = 3$ voraus, als in der Kontrollgruppe.
Notenfaktor _j (Kovariate)	Gewichtete Summe aus der Deutsch-, Physik- und Mathematiknote des Schülers j . Entspricht der ersten Komponente einer Principal Component Analysis der drei Noten. Der aus den Einzelnoten berechnete Notenfaktor wurde grand-mean-zentriert.

β_6 Notenfaktor _j (fixed variable)	Einfluss der Noten auf die Leistung.
β_7 Notenfaktor _j (fixed variable)	Zusätzlicher Einfluss der Noten auf die Änderung der Leistung zum Zeitpunkt $i = 2$ im Vergleich zu Zeitpunkt $i = 1$.
β_8 Notenfaktor _j (fixed variable)	Zusätzlicher Einfluss der Noten auf die Änderung der Leistung zum Zeitpunkt $i = 3$ im Vergleich zu Zeitpunkt $i = 1$.
$IQ_{\text{verbal } j}$ (Kovariate)	(Grand-mean-)zentrierte verbale Intelligenz des Schülers j : Bsp.: Es sei 105 die gemessene verbale Intelligenz des Schülers j und der Durchschnitt der verbalen Intelligenz sei 99. Dann ist $IQ_{\text{verbal } j} = 6$
β_9 $IQ_{\text{verbal } j}$ (fixed variable)	Einfluss der verbalen Intelligenz auf die Leistung. Bsp.: Es sei $\beta_3 = 0.5$. Dann erhöht sich die Vorhersage der Leistung des Schülers zum Zeitpunkt $i = 1$ um 0.50 Punkte je zusätzlichem Punkt IQ_{verbal}
β_{10} $IQ_{\text{verbal } j}$ (fixed variable)	Zusätzlicher Einfluss der verbalen Intelligenz auf die Änderung der Leistung zum Zeitpunkt $i = 2$ im Vergleich zu Zeitpunkt $i = 1$. Bsp.: Es sei $\beta_4 = 0.50$. Dann erhöht sich die Vorhersage der Leistungszunahme des Schülers zum Zeitpunkt $i = 2$ um Vergleich zu Zeitpunkt $i = 1$ um 0.50 Punkte je zusätzlichem Punkt IQ_{verbal}
β_{11} $IQ_{\text{verbal } j}$ (fixed variable)	Zusätzlicher Einfluss der verbalen Intelligenz auf die Änderung der Leistung zum Zeitpunkt $i=3$ im Vergleich zu Zeitpunkt $i = 1$. Bsp.: Es sei $\beta_5 = 0.5$. Dann erhöht sich die Vorhersage der Leistungszunahme des Schülers zum Zeitpunkt $i = 3$ um Vergleich zu Zeitpunkt $i = 1$ um 0.50 Punkte je zusätzlichem Punkt IQ_{verbal}
$IQ_{\text{figural-räumlich } j}$ (Kovariate)	(Grand-mean-)zentrierte figural-räumliche Intelligenz des Schülers j : Bsp.: Es sei 105 die gemessene figural-räumliche Intelligenz des Schülers j und der Durchschnitt der figural-räumlichen Intelligenz sei 99. Dann ist $IQ_{\text{figural-räumlich } j} = 6$

β_{12} IQ _{figural-räumlich j} (fixed variable)	Einfluss der figural-räumlichen Intelligenz auf die Leistung. (Grand-mean-) zentrierte Intelligenz des Schülers j
β_{13} IQ _{figural-räumlich j} (fixed variable)	Zusätzlicher Einfluss der figural-räumlichen Intelligenz auf die Änderung der Leistung zum Zeitpunkt $i = 2$ im Vergleich zu Zeitpunkt $i = 1$.
β_{14} IQ _{figural-räumlich j} (fixed variable)	Zusätzlicher Einfluss der figural-räumlichen Intelligenz auf die Änderung der Leistung zum Zeitpunkt $i = 3$ im Vergleich zu Zeitpunkt $i = 1$.
IQ _{figural-logisch j} (Kovariate)	(Grand-mean-)zentrierte figural-logische Intelligenz des Schülers j: Bsp.: Es sei 105 die gemessene figural-logische Intelligenz des Schülers j und der Durchschnitt der figural-logischen Intelligenz sei 99. Dann ist $IQ_{\text{figural-logisch } j} = 6$
β_{15} IQ _{figural-logisch j} (fixed variable)	Einfluss der figural-logischen Intelligenz auf die Leistung. (Grand-mean-)zentrierte figural-logische Intelligenz des Schülers j
β_{16} IQ _{figural-logisch j} (fixed variable)	Zusätzlicher Einfluss der figural-logischen Intelligenz auf die Änderung der Leistung zum Zeitpunkt $i = 2$ im Vergleich zu Zeitpunkt $i = 1$.
β_{17} IQ _{figural-logisch j} (fixed variable)	Zusätzlicher Einfluss der figural-logischen Intelligenz auf die Änderung der Leistung zum Zeitpunkt $i = 3$ im Vergleich zu Zeitpunkt $i = 1$.
Klassengröße _k (Kovariate)	(Grand-mean-)zentrierte Klassengröße der Klasse k.
β_{18} Klassengröße _k (fixed variable)	Einfluss der Klassengröße auf die Leistung.
β_{19} Klassengröße _k (fixed variable)	Zusätzlicher Einfluss der Klassengröße auf die Änderung der Leistung zum Zeitpunkt $i = 2$ im Vergleich zu Zeitpunkt $i = 1$.
β_{20} Klassengröße _k (fixed variable)	Zusätzlicher Einfluss der Klassengröße auf die Änderung der Leistung zum Zeitpunkt $i = 3$ im Vergleich zu Zeitpunkt $i = 1$.

Schultyp _k (Dummy-Kodierung / Kovariate)	Schultyp der Klasse k. $\text{Schultyp}_k = \begin{cases} 1 & \text{falls Klasse } k \text{ an Gymnasium} \\ 0 & \text{falls Klasse } k \text{ an IGS} \end{cases}$
β_{21} Schultyp _k (fixed variable)	Einfluss des Schultyps auf die Leistung.
β_{22} Schultyp _k (fixed variable)	Zusätzlicher Einfluss des Schultyps auf die Änderung der Leistung zum Zeitpunkt $i = 2$ im Vergleich zu Zeitpunkt $i = 1$.
β_{23} Schultyp _k (fixed variable)	Zusätzlicher Einfluss des Schultyps auf die Änderung der Leistung zum Zeitpunkt $i = 3$ im Vergleich zu Zeitpunkt $i = 1$.
LES _j (Kovariate)	(Grand-mean-)zentrierte Einschätzung des Lehrerengagement aus Sicht des Schülers (LES) j zum Zeitpunkt $i = 1$, d.h. vor Beginn der Unterrichtsreihe.
β_{24} LES _j (fixed variable)	Einfluss von LES auf die Leistung des Schülers j .
β_{25} LES _j (fixed variable)	Zusätzlicher Einfluss von LES auf die Änderung der Leistung zum Zeitpunkt $i = 2$ im Vergleich zu Zeitpunkt $i = 1$.
β_{26} LES _j (fixed variable)	Zusätzlicher Einfluss von LES auf die Änderung der Leistung zum Zeitpunkt $i = 3$ im Vergleich zu Zeitpunkt $i = 1$.

ε_{ij}

(Random Variable)

Fehlerterm. Abweichung der tatsächlich gemessenen Leistung des Schülers j der Klasse k zum Zeitpunkt i von der durch das Modell vorhergesagten Leistung.

ε_{ij} ist für jede Kombination aus i/j Realisation einer (univariat) normalverteilten Zufallsvariablen mit Mittelwert $\mu = 0$ und Standardabweichung $g_i \cdot \sigma_\varepsilon$. Die Standardabweichung ist damit für jeden Zeitpunkt $i = 1, 2, 3$ unterschiedlich.

Die ε_{ij} sind unabhängig von den Random Interzepten u und p aber untereinander in jedem Individuum korreliert.

D.h.: $(\varepsilon_{1j}, \varepsilon_{2j}, \varepsilon_{3j})$ sind gemeinsam multivariat normalverteilt mit Varianz-Kovarianz-Matrix Σ :

$$\Sigma = \sigma_\varepsilon^2 \cdot \begin{pmatrix} 1 & \rho_{12}g_2 & \rho_{13}g_3 \\ \rho_{12}g_2 & g_2^2 & \rho_{23}g_2g_3 \\ \rho_{13}g_3 & \rho_{23}g_2g_3 & g_3^2 \end{pmatrix}$$

ρ_{12} Korrelation zwischen ε_{1j} und ε_{2j} .

ρ_{13} Korrelation zwischen ε_{1j} und ε_{3j} .

ρ_{23} Korrelation zwischen ε_{2j} und ε_{3j} .

σ_ε^2 Varianz von ε_{1j}

$g_2^2 \cdot \sigma_\varepsilon^2$ Varianz von ε_{2j}

$g_3^2 \cdot \sigma_\varepsilon^2$ Varianz von ε_{3j}

(man beachte dass $g_1 = 1$ ist / siehe oben)

Damit adressiert der Fehler über die Gewichte g_i die unterschiedlichen Varianzen der Leistungstestergebnisse zu den verschiedenen Messzeitpunkten (Heteroskedastizität). Ferner wird die Korrelation der Messzeitpunkte untereinander berücksichtigt, welche sich aus dem Repeated-Measurement-Design ergeben. Dies ist zwar bereits durch den Individuenparameter p adressiert. Durch diesen ist jedoch die Korrelation der Messzeitpunkte immer gleich, wohingegen die beliebige Korrelation der Fehlerterme höhere Flexibilität bietet.

Man beachte auch die Darstellung:

$$\varepsilon_{ij} = t_1 \varepsilon_{1j} + t_2 \varepsilon_{2j} + t_3 \varepsilon_{3j}$$

Darstellung in der Schreibweise als Hierarchisches Modell:**Messzeitpunkt**

$$\begin{aligned} \gamma_{ijk} &= \gamma_{0jk} + \varepsilon_{ij} = \\ &= \gamma_{0jk} + t_1 \varepsilon_{1j} + t_2 \varepsilon_{2j} + t_3 \varepsilon_{3j} \end{aligned}$$

Individuum

$$\begin{aligned} \gamma_{0jk} = & \pi_{0k} + \pi_{11} t_2 + \pi_{12} t_3 + \\ & \pi_{16} \text{Notenfaktor}_j + \pi_{17} t_2 \text{Notenfaktor}_j + \pi_{18} t_3 \text{Notenfaktor}_j + \\ & \pi_{19} \text{IQ}_{\text{verbal } j} + \pi_{10} t_2 \text{IQ}_{\text{verbal } j} + \pi_{11} t_3 \text{IQ}_{\text{verbal } j} + \\ & \pi_{12} \text{IQ}_{\text{figural-räumlich } j} + \pi_{13} t_2 \text{IQ}_{\text{figural-räumlich } j} + \pi_{14} t_3 \text{IQ}_{\text{figural-räumlich } j} + \\ & \pi_{15} \text{IQ}_{\text{figural-logisch } j} + \pi_{16} t_2 \text{IQ}_{\text{figural-logisch } j} + \pi_{17} t_3 \text{IQ}_{\text{figural-logisch } j} + \\ & \pi_{24} \text{LES}_j + \pi_{25} t_2 \text{LES}_j + \pi_{26} t_3 \text{LES}_j + \\ & \rho_j \end{aligned}$$

Klasse

$$\begin{aligned} \pi_{0k} = & \beta_0 + \beta_3 \text{Bedingung}_k + \beta_4 \text{Bedingung}_k t_2 + \beta_5 \text{Bedingung}_k t_3 + \\ & \beta_{18} \text{Klassengröße}_k + \beta_{19} \text{Klassengröße}_k t_2 + \beta_{20} \text{Klassengröße}_k t_3 + \\ & \beta_{24} \text{Schultyp}_k + \beta_{25} \text{Schultyp}_k t_2 + \beta_{26} \text{Schultyp}_k t_3 + \\ & u_k \\ \pi_m = & \beta_m \text{ für } m = 6-17, 21-23, 27-29 \end{aligned}$$

Das Modell kann in Anlehnung an Fahrmeir et al. (2009, S. 259) auch wie folgt dargestellt werden (für Individuen $j = 1 \dots 525$, und Klassen $k = 1 \dots 21$):

$$y_{jk} = \begin{pmatrix} y_{1jk} \\ y_{2jk} \\ y_{3jk} \end{pmatrix} = \mathbf{X}_{jk} \cdot \boldsymbol{\beta} + \mathbf{U} \cdot \boldsymbol{\gamma}_{jk} + \boldsymbol{\varepsilon}_j$$

Wobei

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_{26})^T \in \mathbb{R}^{27 \times 1}$$

(Vektor der festen Effekte)

und

$$\mathbf{X}_{jk} = \begin{pmatrix} 1 \cdot \mathbf{A}, & \text{Bedingung}_j \cdot \mathbf{A}, & \text{Notenfaktor}_j \cdot \mathbf{A}, \\ \text{IQ}_{\text{verb. } j} \cdot \mathbf{A}, & \text{IQ}_{\text{fig.räuml. } j} \cdot \mathbf{A}, & \text{IQ}_{\text{fig.log. } j} \cdot \mathbf{A}, \\ \text{Klassengrösse}_k \cdot \mathbf{A}, & \text{Schultyp}_k \cdot \mathbf{A}, & \text{LES}_j \cdot \mathbf{A} \end{pmatrix} \in \mathbb{R}^{3 \times 27}$$

(Designmatrix der festen Effekte)

mit

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{3 \times 3}$$

und

$$\mathbf{U} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \in \mathbb{R}^{3 \times 2} \quad (\text{Designmatrix für Random Effekte})$$

$$\boldsymbol{\gamma}_{jk} = \begin{pmatrix} p_j \\ u_k \end{pmatrix} \in \mathbb{R}^{2 \times 1} \quad (\text{Random Interzept je Individuum/Klasse})$$

$$\boldsymbol{\varepsilon}_j = \begin{pmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \varepsilon_{3j} \end{pmatrix} \in \mathbb{R}^{3 \times 1} \quad (\text{Fehlervektor})$$

Hierbei gilt für die zufälligen Effekte, dass sie gemeinsam normalverteilt sind:

$$\begin{pmatrix} p_j \\ u_k \\ \varepsilon_{1j} \\ \varepsilon_{2j} \\ \varepsilon_{3j} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_p^2 & 0 & & & \\ 0 & \sigma_u^2 & & & \\ & & \sigma_\varepsilon^2 & & \\ \mathbf{0} & & \sigma_\varepsilon^2 \cdot \rho_{12} \cdot g_2 & \sigma_\varepsilon^2 \cdot \rho_{12} \cdot g_2 & \sigma_\varepsilon^2 \cdot \rho_{13} \cdot g_3 \\ & & \sigma_\varepsilon^2 \cdot \rho_{13} \cdot g_3 & \sigma_\varepsilon^2 \cdot \rho_{13} \cdot g_2 \cdot g_3 & \sigma_\varepsilon^2 \cdot g_3^2 \end{pmatrix} \right)$$

Ferner sind die zufälligen Einflussgrößen unabhängig voneinander:

$$p_1, \dots, p_{525}, u_1, \dots, u_{21}, \varepsilon_1, \dots, \varepsilon_{525} \text{ unabhängig}$$

Allgemeine Anmerkung zum Verständnis der übrigen Modelle zur Untersuchung der Wirkung des Treatments auf das konzeptuelle Verständnis und die Motivation:

1. Zur Modellbildung wurde jeweils in gleicher Weise vorgegangen: Basierend auf dem Ausgangsmodell wurde jeweils eine Kovariate hinzugenommen und analysiert, ob die jeweilige Kovariate je Messzeitpunkt einen signifikanten Erklärungswert für die abhängige Variable aufweist. Die Ergebnisse dieser Analysen sind in den Tabellen 20 und 21 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com zu finden.
2. In Übereinstimmung mit dem Modell zur Analyse der Treatmentwirkung auf die Physikleistung wurde jeweils ein Mehrebenenmodell gewählt, das sowohl die Klassenstruktur, die Messwiederholung durch die Individuenkomponente als auch die Heteroskedastizität einbezieht und eine unstrukturierte Korrelationsmatrix der Fehlerterme im Individuum zugrunde legt.
3. Des Weiteren wurde für jedes Modell die Intra-Klassenkorrelation (*ICC*) je Untersuchungsebene errechnet und geprüft, ob der *ICC* die Verwendung eines Mehrebenenmodells rechtfertigt oder eine klassische Regressionsanalyse das sparsamere und angemessenere Modell darstellen würde. Zur Erinnerung: gemäß Eid et al. (2011, S. 705) besteht bereits bei zehn Gruppen, einem Signifikanzniveau von $\alpha = 0.05$ und einer Intra-Klassenkorrelation von $\rho = 0.10$ das Risiko einer statistischen Fehlentscheidung von 50 %.
4. Die Notation der Modellgleichung der übrigen Modelle kann in analoger Weise dargestellt werden wie die Notation für das Modell zur Analyse der Treatmentwirkung auf die Physikleistung.

2.3.9.3 Erste Hypothese: Wirkung des Treatments auf die Physikleistung

Zur Untersuchung der ersten Hypothese wurde geprüft, ob die Treatmentgruppe im Vergleich zur Kontrollgruppe im Wissen und Problemlösen beim Umgang mit fachspezifischen Repräsentationen im Bereich der Strahlenoptik zum zweiten und dritten Messzeitpunkt (nach der Unterrichtsreihe sowie zwei Monate später) einen höheren Lernzuwachs verzeichnen kann als die Kontrollgruppe. Erfasst wurden Wissen, Problemlösen beim Umgang mit fachspezifischen Repräsentationen durch den Physikleistungstest, der aus repräsentationsbezogenen Aufgaben bestand.

Ein Blick auf die deskriptiven Statistiken (vgl. Tabelle 51) zeigt, dass beide Gruppen annähernd mit dem gleichen Ausgangsniveau starten, in der Postmessung die Treatmentgruppe im Vorteil zu sein scheint, dieser Vorteil jedoch zum Messzeitpunkt des Follow-up Tests nicht erhalten bleibt, so dass beide Gruppen in der letzten Messung ein nahezu identisches Niveau erzielen.

Tabelle 51 Deskriptive Statistiken Leistungstest je Bedingung

	Prätest ($N = 492$)		Posttest ($N = 484$)		Follow-up Test ($N = 496$)	
	TG ^a ($n = 259$)	KG ^b ($n = 233$)	TG ^a ($n = 252$)	KG ^b ($n = 232$)	TG ^a ($n = 264$)	KG ^b ($n = 232$)
<i>M</i>	2.61	2.14	18.21	16.79	13.61	13.36
<i>(SD)</i>	(2.62)	(2.14)	(7.80)	(7.48)	(7.14)	(7.57)
<i>Range</i>	0 - 10.00	0 - 9.00	0 - 35.00	0 - 32.50	0 - 28.50	0 - 30.50
<i>Im folgenden Mehrebenenmodell berücksichtigte Daten ($N = 443$)^c</i>						
	($n = 230$)	($n = 210$)	($n = 252$)	($n = 232$)	($n = 264$)	($n = 232$)
<i>M</i>	2.63	2.18	18.30	16.70	14.02	13.60
<i>(SD)</i>	(2.66)	(2.17)	(7.80)	(7.46)	(7.00)	(7.54)

Anmerkungen. Maximal erreichbare Punktzahl im Leistungstest: 38 Punkte.

^aTreatmentgruppe, ^bKontrollgruppe.

^cDie Verminderung der Stichprobengröße ergibt sich aus fehlenden Kovariaten.

In den Boxplots (vgl. Abbildung 53) sticht besonders der starke Bodeneffekt zum ersten Messzeitpunkt hervor. Dieser Effekt kann darauf zurückgeführt werden, dass die Schüler zuvor keinerlei Unterricht zu dem Lerninhalt erhalten hatten und entsprechend wenige Aufgaben lösen konnten. Vom Prä- zum Posttest ist ein deutlicher Lernzuwachs in beiden Bedingungen zu verzeichnen. Bis zum dritten Messzeitpunkt haben die Schüler zwar offenbar einige der Lerninhalte vergessen, das

Niveau liegt jedoch deutlich über dem Niveau der Ausgangswerte, so dass auch ein mittelfristiger Lernzuwachs zu erkennen ist.

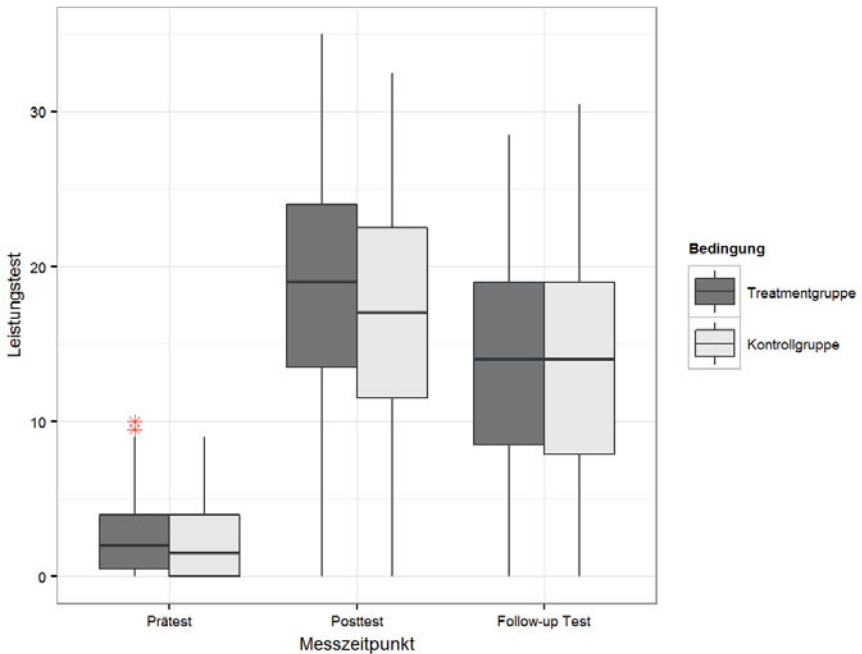


Abbildung 53: Boxplots der Leistungstestergebnisse je Bedingung zu den drei Messzeitpunkten: prä, post und follow-up

In den Verlaufplots (vgl. Abbildung 54) lässt sich erkennen, dass der Lernzuwachs in beiden Bedingungen nahezu parallel verläuft. Das Leistungsniveau der Gymnasiasten liegt hierbei zu allen drei Messzeitpunkten über dem Niveau der Gesamtschüler.

Basierend auf den zuvor dargestellten Überlegungen wurde nun ein Mehrebenenmodell aufgestellt, das sowohl die Klassenstruktur, die Messwiederholung durch die Individuenkomponente als auch die Heteroskedastizität berücksichtigt und eine unstrukturierte Korrelationsmatrix der Fehlerterme im Individuum zugrunde legt (vgl. Tabelle 52). Entsprechend den vorigen Ergebnissen (vgl. Tabelle 50) wurden folgende Kovariaten berücksichtigt: die erhobenen Schulnoten (Mathematik, Deutsch, Physik), alle drei Subskalen des I-S-T 2000 R und das Lehrerengagement aus Schülersicht vor der Intervention. Des Weiteren wurde der Einfluss des Schultyps und der Klassengröße geschätzt.

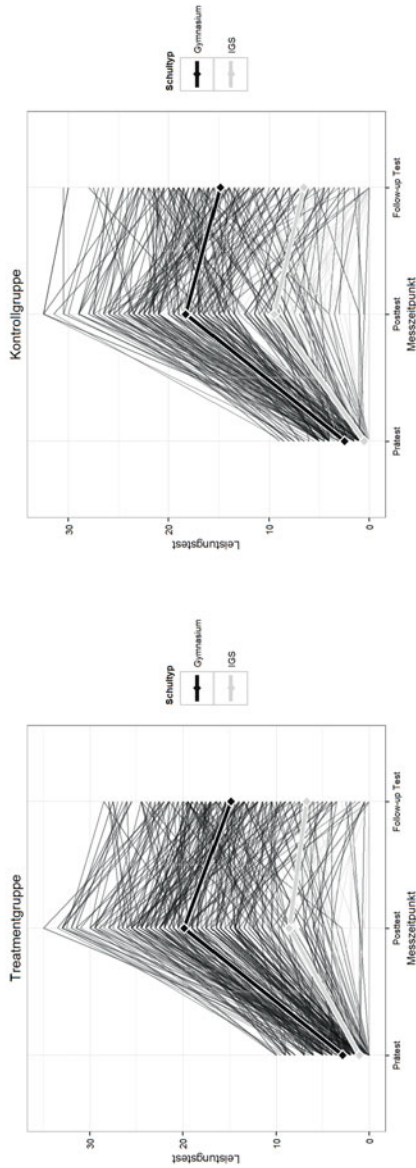


Abbildung 54: Verlaufplots der Leistungstestergebnisse je Bedingung und Schultyp

Tabelle 52 Ergebnisse des Mehrebenenmodells zu Wissen und Problemlösen bei repräsentationsbezogenen Aufgaben (Leistungstest)

Level (Stichprobengröße)	Erläuterung								
Variable	<i>b</i>	(SE)	β	(SE)	<i>F</i> (<i>numDf</i> , <i>denDf</i>)	<i>p</i>			
Level 1 (N = 1263)	Messzeitpunkte								
Level 2 (N = 443)	Individuen: Noten, Intelligenz, LES (= Lehrerengagement aus Schülersicht)								
Level 3 (N = 21)	Schulklassen: Bedingung Treatment (TG) vs. Kontrollgruppe (KG), Schultyp, Klassengröße								
Fixe Effekte									
<i>Erläuterung</i>	<i>b</i>	(SE)	β	(SE)	<i>F</i> (<i>numDf</i> , <i>denDf</i>)	<i>p</i>			
Durchschnittliche Leistung: Kontrollgruppe, Geschlechtszugehörigkeit = weiblich, IGS, mit durchschnittlichen Werten bezüglich IQ und Noten	1.02	0.61	0.11	0.07	2.76 _(1, 814)	.097			
Durchschnittlicher Leistungszuwachs: Kontrollgruppe, Geschlechtszugehörigkeit = weiblich, IGS, mit durchschnittlichen Werten bezüglich IQ und Noten	9.97	0.87	1.11	0.10	131.02 _(1, 814)	< .001			
Durchschnittlicher Leistungszuwachs: Kontrollgruppe, Geschlechtszugehörigkeit = weiblich, IGS, mit durchschnittlichen Werten bezüglich IQ und Noten	7.29	0.87	0.81	0.10	70.80 _(1, 814)	< .001			
Unterschied zwischen TG und KG zu t ₁	0.38	0.46	0.04	0.05	0.69 _(1, 17)	.417			
Zusätzlicher Einfluss Bedingung = TG auf den Zuwachs zwischen t ₂ und t ₁	0.92	0.59	0.10	0.07	2.43 _(1, 814)	.119			
Zusätzlicher Einfluss der Bedingung = TG auf den Zuwachs zwischen t ₃ und t ₁	0.19	0.58	0.02	0.06	0.11 _(1, 814)	.745			
Genereller Einfluss auf Basis von t ₁	-0.31	0.08	-0.05	0.01	15.88 _(1, 417)	< .001			
Zusätzlicher Einfluss auf den Zuwachs zwischen t ₂ und t ₁	-1.72	0.23	-0.26	0.04	53.88 _(1, 814)	< .001			
Zusätzlicher Einfluss auf den Zuwachs zwischen t ₃ und t ₁	-1.97	0.24	-0.30	0.04	70.05 _(1, 814)	< .001			
Genereller Einfluss auf Basis von t ₁	0.04	0.01	0.05	0.01	15.30 _(1, 417)	< .001			
Zusätzlicher Einfluss auf den Zuwachs zwischen t ₂ und t ₁	0.03	0.03	0.04	0.04	1.18 _(1, 814)	.277			
Zusätzlicher Einfluss auf den Zuwachs zwischen t ₃ und t ₁	-0.01	0.03	-0.01	0.04	0.13 _(1, 814)	.724			
Genereller Einfluss auf Basis von t ₁	-0.01	0.01	-0.01	0.01	0.79 _(1, 417)	.375			

IQ figurál-räumlich * Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_1	0.09	0.03	0.09	0.03	7.05 _(1,814)	.008
IQ figurál-räumlich * Prä – Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_1	0.03	0.03	0.03	0.03	0.70 _(1,814)	.402
IQ figurál-logisch * Ausgangswert	Genereller Einfluss auf Basis von t_1	0.01	0.01	0.01	0.01	0.57 _(1,417)	.450
IQ figurál-logisch * Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_1	0.01	0.03	0.01	0.03	0.18 _(1,814)	.673
IQ figurál-logisch * Prä – Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_1	0.03	0.03	0.03	0.03	0.91 _(1,814)	.340
Klassengröße	Genereller Einfluss auf Basis von t_1	-0.08	0.08	-0.03	0.03	0.92 _(1,17)	.352
Klassengröße * Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_1	-0.08	0.11	-0.03	0.04	0.58 _(1,814)	.446
Klassengröße * Prä – Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_1	0.07	0.10	0.02	0.04	0.39 _(1,814)	.535
Schultyp * Ausgangswert	Genereller Einfluss Schultyp = Gymnasium auf Basis von t_1	1.39	0.64	0.16	0.07	4.76 _(1,17)	.043
Schultyp * Posttest	Zusätzlicher Einfluss Schultyp = Gymnasium auf den Zuwachs zwischen t_2 und t_1	5.16	0.93	0.57	0.10	31.05 _(1,814)	< .001
Schultyp * Prä – Follow-up Test	Zusätzlicher Einfluss Schultyp = Gymnasium auf den Zuwachs zwischen t_3 und t_1	4.22	0.92	0.47	0.10	20.96 _(1,814)	< .001
LES (vor der Intervention)	Genereller Einfluss auf die Basis von t_1	-0.02	0.10	0.00	0.01	0.03 _(1,417)	.873
LES (vor der Intervention) * Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_1	0.29	0.25	0.04	0.03	1.33 _(1,814)	.249
LES (vor der Intervention) * Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_1	-0.27	0.25	-0.04	0.03	1.22 _(1,814)	.271
Zufällige Effekte							
σ_{η_1} (Klasse)		0.95					
σ_{η_2} (Individuum)		0.83					
σ_{ϵ} (Individuenerfisch)		1.76					
ρ (Messetpunkt)		(ρ_{12}) 0.12				(ρ_{23}) 0.44	
g (Messetpunkte)		(g_1) 1.00				(g_2) 3.35	
Modellvergleich							
Berichtetes Modell		Leermodell ^g					
Devianz		7104.25					7358.20
df		35					11

AIC	7380.20	7436.85
BIC	7174.25	7354.52
LR-Test ^a (berichtetes Modell versus Leermodell)	253.95, $p < .001$	
Un erklärte Varianz Level 1	(t ₁) 4.67	(t ₂) 36.20
Un erklärte Varianz Level 2	25.71	(b) 55.51
Un erklärte Varianz Level 3	1.92	38.07
Erklärte Varianz des berichteten Modells im Vergleich zum Leermodell		3.53
R ² Level 1	(t ₁) 0.22	(t ₂) 0.31
R ² Level 2	0.32	
R ² Level 3	0.43	
„Wirkung“ des Treatments		
R ² Level 1	(t ₁) 0.01	(t ₂) 0.01
R ² Level 2	< 0.01	(t ₂) < 0.01
R ² Level 3	0.02	
Anteil auf erklärter Varianz^b		
R ² Level 1	(t ₁) 0.01	(t ₂) 0.01
R ² Level 2	< 0.01	(t ₂) < 0.01
R ² Level 3	0.02	
Effektstärke^c		
R ² Level 1	(t ₁) 0.01	(t ₂) 0.01
R ² Level 2	< 0.01	(t ₂) < 0.01
R ² Level 3	0.02	
Δ_{ppra} – follow-up) = 0.03		
Δ_{ppra} – post) = 0.15		

^a Das Leermodell unterscheidet sich vom berichteten Modell darin, dass alle fixen Effekte außer dem Interzept und den Dummy-Variablen t_{and} t₃ fehlen, d.h. das Modell sagt für jeden Schüler voraus, dass seine Leistung zum jeweiligen Messzeitpunkt der mittleren Leistung der Stichprobe entspricht unter Berücksichtigung der individuellen- und klassenspezifischen Unterschiede. Man beachte, das Leermodell schätzt folgende 11 Parameter ($df = 11$): Interzept, t₁, t₂, t₃, B₁, B₂, B₃, ρ₁₃, ρ₂₃, σ₁, σ₂, σ₃.

^b Der Anteil auf erklärter Varianz errechnet sich Snijder und Bosker (1999, S. 39 ff.) folgend, aus:

$$\begin{aligned} \text{Level 1 (Messzeitpunkt):} & 1 - \frac{\sigma_{\epsilon}^2 + \sigma_{\eta}^2 + \sigma_{\tau}^2}{\sigma_{\epsilon}^2 + \sigma_{\eta}^2 + \sigma_{\tau}^2 + \sigma_{\mu}^2} \\ \text{Level 2 (Individuum):} & 1 - \frac{\sigma_{\epsilon}^2 + \sigma_{\eta}^2}{\sigma_{\epsilon}^2 + \sigma_{\eta}^2 + \sigma_{\tau}^2 + \sigma_{\mu}^2} \\ \text{Level 3 (Klasse):} & 1 - \frac{\sigma_{\epsilon}^2 + \sigma_{\eta}^2 + \sigma_{\tau}^2}{\sigma_{\epsilon}^2 + \sigma_{\eta}^2 + \sigma_{\tau}^2 + \sigma_{\mu}^2} \end{aligned}$$

h ist hierbei das harmonische Mittel der Klassengrößen $h = 24.60$. Ferner stehen die mit einer Tilde gekennzeichneten Variablen für Variablen aus dem Leermodell und die ungekennzeichneten Variablen für Variablen aus dem berichteten Modell.

^c Die Effektstärke ergibt sich nach Tymms (2004, S. 56 f.) für dichotome Variablen aus der Differenz der gefragten Mittelwerte (hier mittlerer Lernzuwachs) geteilt durch die gepoolte Standardabweichung (Wurzel der Varianz innerhalb der Gruppen)

$$\Delta_{ppra} - post = \frac{\text{Differenz der Klassennittelwerte des Zwischen- und Kontrollgruppen zu 2}}{\sqrt{\text{Varianz innerhalb einer Klasse zu 2}}}$$

$$= \frac{\beta_1}{\sqrt{\sigma_{\text{innerhalb Klasse zu 2}}^2}}$$

$$= \frac{\beta_1}{\sqrt{\sigma_{\epsilon}^2 + \sigma_{\eta}^2 + \sigma_{\tau}^2}}$$

Δ_{ppra} – follow-up analog

Unter Berücksichtigung der genannten Kovariaten ergaben sich folgende Ergebnisse bezüglich des Lernzuwachses im Post- und im Follow-up Test (Untersuchung der ersten Hypothese): Die Physikleistung bei repräsentationsbezogenen Aufgaben der Schüler war im Prätest zwischen Treatment- und Kontrollgruppe nicht signifikant verschieden ($F_{(1,17)} = 0.69$, n.s.). Betrachtet man den Lernzuwachs vom Prä zum Posttest, erzielten Schüler in der Treatmentbedingung einen leicht höheren Lernzuwachs als die Kontrollgruppe. So erreichten die Schüler in der Treatmentgruppe etwa einen Leistungszuwachs, der um einen Leistungspunkt höher liegt ($b = 0.92$) als der Zuwachs der Kontrollgruppe bei einer Maximalpunktzahl von 38 Punkten im Leistungstest, dies entspricht einem Zehntel der Standardabweichung ($\beta = 0.10$). Allerdings handelt es sich um keinen signifikanten Effekt ($F_{(1,814)} = 2.43$, n.s.). Der geringe und auch nicht signifikante Vorteil der Treatmentgruppe bleibt zum Zeitpunkt des Follow-up Tests *nicht* erhalten.

Die Analyse des Anteils aufgeklärter Varianz mit Werten nahe 0 und die berechnete Effektstärke $\Delta < 0.20$ belegen, dass die Bedingung (TG versus KG) keine bzw. kaum eine Relevanz für die Erklärung von Unterschieden im Lernzuwachs zeigt. Von den genannten Kovariaten erwiesen sich die folgenden beiden Kovariaten zu allen drei Messzeitpunkten als hoch signifikant: die Schulnoten, berücksichtigt durch den Notenfaktor und der Schultyp.

Schüler, deren Schulnote um eine Notenstufe besser ist als der Mittelwert der Stichprobe, erzielten über alle Messzeitpunkte hinweg ein Leistungstestergebnis, das knapp einen Drittel Punkt höher lag ($b = -0.31$, $\beta = -0.05$, $p < .001$). Schüler, deren Schulnote um eine Standardabweichung besser ist, erzielten im Schnitt einen Lernzuwachs prä – post, der um ein Viertel einer Standardabweichung höher liegt ($\beta = -0.26$, $p < .001$) und einen Lernzuwachs prä – follow-up, der knapp um ein Drittel einer Standardabweichung ($\beta = -0.30$, $p < .001$) besser ist, als das Ergebnis ihrer Altersgenossen. Dies entspricht einem zusätzlichen Lernzuwachs von knapp 2 Punkten von maximal 38 zu erreichenden Punkten des Leistungstests. Die negativen Vorzeichen ergeben sich aus der Notenskala: höhere Beträge entsprechen schlechteren Bewertungen der Fachleistung.

Gymnasiasten erreichten im Schnitt einen Lernzuwachs im prä – post, der um mehr als eine halbe Standardabweichung ($\beta = 0.57$) höher liegt als der Lernzuwachs der Gesamtschüler und einen Lernzuwachs prä - follow-up, der knapp um eine halbe Standardabweichung ($\beta = 0.47$) von den Ergebnissen der Gesamtschüler abweicht. Dies entspricht einem Vorteil von rund 4 bis 5 Punkten von maximal 38 zu erreichenden Punkten des Leistungstests.

Von den erhobenen Intelligenzkomponenten erweist sich im Gesamtmodell die verbale Intelligenz als über alle Messzeitpunkte hinweg als relevant ($F_{(1,417)} = 15.30$,

$p < .001$). Sie wirkt sich aber nicht auf den Lernzuwachs aus. Die figural-räumliche Intelligenz beeinflusst signifikant nur den Lernzuwachs prä – post. Der Erklärungsanteil der figural-logischen Intelligenz, die sich in der Einzeltestung als wichtig erwies, wird vermutlich durch die Schulleistung (beispielsweise durch die Noten in Mathematik und Physik) erklärt.

Ein Einfluss der Klassengröße und des Lehrer-Engagements aus Schülersicht ist gemäß des Ergebnisses des Gesamtmodells nicht nachweisbar. Von allen Variablen wirkt sich also der Einfluss des Schultyps am deutlichsten auf den Lernzuwachs prä – post bzw. prä – follow-up aus. Gymnasiasten profitieren mehr von der Intervention als Gesamtschüler. Dies wurde zum Anlass genommen, die Wirkung des Treatments gesondert für Gymnasiasten auszuwerten (siehe Tabelle 53).

Für die gesonderte Auswertung der Stichprobe der Gymnasiasten bestätigt sich, dass auch Schüler an Gymnasien in beiden Bedingungen mit dem gleichen Ausgangsniveau in der Physikleistung starten ($F_{(1, 14)} = 0.71, n.s.$). Betrachtet man den Lernzuwachs vom Prä- zum Posttest, erzielten Schüler in der Treatmentbedingung einen höheren Lernzuwachs als die Kontrollgruppe. So erreichten die Schüler in der Treatmentgruppe etwa einen Leistungszuwachs, der knapp um 1.5 Leistungspunkte höher liegt ($b = 1.45$) als die Kontrollgruppe von 38 maximal zu erreichenden Punkten im Leistungstest; dies entspricht $\beta = 0.16$ Standardabweichungen. Im Gegensatz zur Gesamtstichprobe handelt es sich, um einen signifikanten Effekt ($F_{(1, 692)} = 5.04, p < .05$). Der Anteil aufgeklärter Varianz muss mit Werten nahe 0, ebenso wie die Effektstärke ($\Delta_{\text{prä-post}} = .24$) als gering bewertet werden. Der signifikante Vorteil der Treatmentgruppe bleibt zum Zeitpunkt des Follow-up Tests jedoch auch für die Gymnasiasten alleine *nicht* erhalten. Ebenso wie für die Gesamtstichprobe sind die Schulnoten zu allen drei Messzeitpunkten relevante Einflussfaktoren. Schüler, deren Bewertung der Schulleistung um Notenstufe besser war, erzielten über alle Messzeitpunkte hinweg ein Leistungstestergebnis, das knapp einen halben Punkte höher lag ($b = -0.42, \beta = -0.06, p < .001$). Zudem kam es im Schnitt zu einem Lernzuwachs prä – post, der knapp um ein Viertel einer Standardabweichung höher lag ($\beta = -0.24, p < .001$) und einen Lernzuwachs im Leistungs-Follow-up-Test-Ergebnis, der ebenfalls um in etwas mehr als ein Viertel einer Standardabweichung ($\beta = -0.27, p < .001$) höher lag, als der Mittelwert der Stichprobe. Dies entspricht einem zusätzlichen Lernzuwachs von ca. 1.5 Punkten von maximal 38 zu erreichenden Punkten des Leistungstests.

Tabelle 53 Ergebnisse des Mehrebenenmodells zu Wissen und Problemlösen bei repräsentationsbezogenen Aufgaben (Physikleistung), Stichprobe nur Gymnasiasten

Level (Stichprobengröße)	Erläuterung	b	(SE)	β	(SE)	F_(numDF, denomDF)	p
Level 1 (N = 1086)	Messzeitpunkte						
Level 2 (N = 376)	Individuen: Noten, Intelligenz, LES (= Lehrerengagement aus Schülersicht)						
Level 3 (N = 17)	Schulklassen: Bedingung Treatment (TG) vs. Kontrollgruppe (KG), Schultyp, Klassengröße						
Fixe Effekte							
Variable	Erläuterung	b	(SE)	β	(SE)	F_(numDF, denomDF)	p
Interzept	Durchschnittliche Leistung: Kontrollgruppe, weiblich, mit durchschnittlichen Werten bezüglich IQ und Noten	2.18	0.42	0.24	0.05	26.59 _(1,692)	< .001
Zuwachs (prä – post: t ₂ -t ₁)	Durchschnittlicher Leistungszuwachs: Kontrollgruppe, weiblich, mit durchschnittlichen Werten bezüglich IQ und Noten	14.66	0.57	1.62	0.06	658.00 _(1,692)	< .001
Zuwachs (prä – follow-up: t ₃ -t ₁)	Durchschnittlicher Leistungszuwachs: Kontrollgruppe, weiblich, mit durchschnittlichen Werten bezüglich IQ und Noten	11.73	0.55	1.30	0.06	457.66 _(1,692)	< .001
Bedingung = Treatmentgruppe (TG)	Unterschied zwischen TG und KG zu t ₁	0.47	0.56	0.05	0.06	0.71 _(1,14)	.413
Bedingung * Zuwachs prä – post	Zusätzlicher Einfluss Bedingung = TG auf den Zuwachs zwischen t ₂ und t ₁	1.45	0.65	0.16	0.07	5.04 _(1,692)	.025
Bedingung * Zuwachs prä – follow-up	Zusätzlicher Einfluss der Bedingung = TG auf den Zuwachs zwischen t ₃ und t ₁	-0.02	0.62	0.00	0.07	0.00 _(1,692)	.972
Noten (D, M, PH)	Genereller Einfluss auf Basis von t ₁	-0.42	0.09	-0.06	0.01	20.80 _(1,353)	< .001
Noten * Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t ₂ und t ₁	-1.62	0.26	-0.24	0.04	37.65 _(1,692)	< .001
Noten * Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t ₃ und t ₁	-1.84	0.25	-0.27	0.04	52.84 _(1,692)	.001

IQ verbal	Genereller Einfluss auf Basis von t_1	0.05	0.01	0.05	0.01	16.81 _(1,353)	= .001
IQ verbal* Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_1	0.03	0.03	0.03	0.04	0.72 _(1,692)	.396
IQ verbal* Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_1	-0.02	0.03	-0.03	0.04	0.51 _(1,692)	.475
IQ figural-räumlich* Ausgangswert	Genereller Einfluss auf Basis von t_1	-0.02	0.01	-0.02	0.01	2.69 _(1,353)	.102
IQ figural-räumlich* Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_1	0.12	0.04	0.12	0.04	9.81 _(1,692)	.002
IQ figural-räumlich* Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_1	0.05	0.04	0.05	0.04	2.16 _(1,692)	.142
IQ figural-logisch* Ausgangswert	Genereller Einfluss auf Basis von t_1	0.05	0.04	0.01	0.01	1.12 _(1,692)	.291
IQ figural-logisch* Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_1	0.01	0.01	0.00	0.04	0.01 _(1,692)	.905
IQ figural-logisch* Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_1	0.00	0.03	0.03	0.04	0.92 _(1,692)	.338
Klassengröße	Genereller Einfluss auf Basis von t_1	0.03	0.03	-0.04	0.03	1.39 _(1,4)	.258
Klassengröße* Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_1	-0.13	0.11	0.00	0.04	0.02 _(1,692)	.898
Klassengröße* Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_1	-0.02	0.13	0.14	0.04	15.70 _(1,692)	< .001
LES (vor der Intervention)	Genereller Einfluss auf Basis von t_1	-0.87	0.62	0.00	0.02	0.02 _(1,353)	.877
LES (vor der Intervention)* Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_1	-0.02	0.11	0.02	0.04	0.31 _(1,692)	.577
LES (vor der Intervention)* Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_1	0.15	0.27	-0.05	0.04	2.30 _(1,692)	.130
Zufällige Effekte							
σ_1 (Klasse)		1.05					
σ_2 (Individuum)		0.93					
σ_3 (individuumsspezifisch)		1.78					

ρ (Messzeitpunkt)	(ρ_{13}) 0,14	(ρ_{13}) 0,19	(ρ_{23}) 0,44
g (Messzeitpunkt)	(g_1) 1,00	(g_2) 3,36	(g_3) 3,24
Modellvergleich	Berichtetes Modell ^a		
Devianz	6072,56		Leermodell ^b 6261,87
df	35		11
AIC	6142,556		6283,872
BIC	6317,215		6338,765
LR-Test (berichtetes Modell versus Leermodell)	189,32, $p < .001$		
Unerklärt Varianz Level 1	(t_1) 5,15	(t_2) 37,39	(t_3) 35,31
Unerklärt Varianz Level 2	25,95		(t_1) 6,25
Unerklärt Varianz Level 3	2,11		(t_2) 48,80 (t_3) 48,59
Erklärte Varianz des berichteten Modells im Vergleich zum Leermodell			34,54
R^2 Level 1	(t_1) 0,18	(t_2) 0,23	2,97
R^2 Level 2	0,25		
R^2 Level 3	0,30		
„Wirkung“ des Treatments	Anteil aufgeklärter Varianz ^b		
R^2 Level 1	(t_1) 0,01	(t_2) 0,02	$(t_3) < 0,01$
R^2 Level 2	0,01		$\Delta(\rho^2 - \text{follow-up}) = 0,24$
R^2 Level 3	0,03		$\Delta(\rho^2 - \text{follow-up}) < 0,01$

^a Das Leermodell unterscheidet sich vom berichteten Modell darin, dass alle fixen Effekte außer dem Intercept und den Dummy-Variablen bund t_3 fehlen, d.h. das Modell sagt für jeden Schüler voraus, dass seine Leistung zum jeweiligen Messzeitpunkt der mittleren Leistung der Stichprobe entspricht unter Berücksichtigung der individuellen- und klassenspezifischen Unterschiede. Man beachte, das Leermodell schätzt folgende 11 Parameter ($df = 11$): Interzept t_2 , t_3 , β_2 , β_3 , ρ_{12} , ρ_{13} , ρ_{23} , σ_{η} , σ_{θ_1} , σ_{θ_2} .

^b Der Anteil aufgeklärter Varianz errechnet sich wie in Tabelle 52 angegeben nach Snijder und Bosker (1999, S. 99 ff.)
Für das harmonische Mittel der Klassengrößen ergibt sich $n_{\text{geymassim}} = 25,20$.

Von den erhobenen Intelligenzkomponenten erweist sich für die Stichprobe der Gymnasiasten ebenfalls die verbale Intelligenz als relevant ($F_{(1, 353)} = 16.81, p < .001$). Die figural-räumliche Intelligenz beeinflusst wie zu erwarten ebenfalls den Lernzuwachs prä – post ($F_{(1, 692)} = 9.81, p < .01$). Ebenfalls relevant war die Klassengröße, wobei Schüler in größeren Klassen einen höheren Lernzuwachs prä – follow-up erzielten ($\beta = 0.14, F_{(1, 692)} = 15.70, p < .001$).

2.3.9.4 Zweite Hypothese: Wirkung des Treatments auf das konzeptuelle Verständnis

Zur Untersuchung der zweiten Hypothese wurde geprüft, ob die Treatmentgruppe im Vergleich zur Kontrollgruppe bezüglich des konzeptuellen Verständnisses zum zweiten und dritten Messzeitpunkt (nach der Intervention sowie zwei Monate später) einen höheren Lernzuwachs verzeichnen kann als die Kontrollgruppe. Das konzeptuelle Verständnis wurde durch den Konzepttest Strahlenoptik erhoben. Ein Blick auf die deskriptiven Statistiken zeigt, dass beide Gruppen mit dem annähernd gleichen Ausgangsniveau starten, nach der Unterrichtsreihe ihr konzeptuelles Verständnis verbessern und diese Verbesserung auch erhalten bleibt (vgl. Tabelle 54).

Tabelle 54 Deskriptive Statistiken Konzepttest je Bedingung

	Prätest (N = 491)		Posttest (N = 480)		Follow-up Test (N = 486)	
	TG ^a (n = 260)	KG ^b (n = 231)	TG ^a (n = 248)	KG ^b (n = 232)	TG ^a (n = 258)	KG ^b (n = 228)
M	7.50	7.41	10.73	11.20	10.91	11.18
(SD)	(3.24)	(3.23)	(4.70)	(5.29)	(4.64)	(4.86)
Range	0 - 18.00	0 - 16.00	0 - 21.00	0 - 22.00	0 - 22.00	0 - 22.00
<i>Im folgenden Mehrebenenmodell berücksichtigte Daten (N = 443)^c</i>						
	(n = 231)	(n = 210)	(n = 211)	(n = 200)	(n = 217)	(n = 194)
M	7.49	7.49	10.79	11.28	10.91	11.39
(SD)	(3.20)	(3.10)	(4.74)	(5.27)	(4.61)	(4.83)

Anmerkung. Maximal erreichbare Punktzahl im Konzepttest: 22 Punkte.

^aTreatmentgruppe, ^bKontrollgruppe.

^cDie verminderte Stichprobengröße ergibt sich aus fehlenden Kovariaten.

An den Boxplots (vgl. Abbildung 55) zeigt sich, dass Schüler in der Kontrollbedingung zu allen drei Messzeitpunkten leicht im Vorteil zu sein scheinen.

Während zum ersten Messzeitpunkt von Schülern in beiden Bedingungen im Schnitt etwa ein Drittel der Maximalpunktzahl erreicht wird, steigt dieser Wert zum zweiten und dritten Messzeitpunkt auf knapp die Hälfte der maximal erreichbaren Punktzahl. Dieses Niveau bleibt zum dritten Messzeitpunkt in beiden Bedingungen in etwa erhalten.

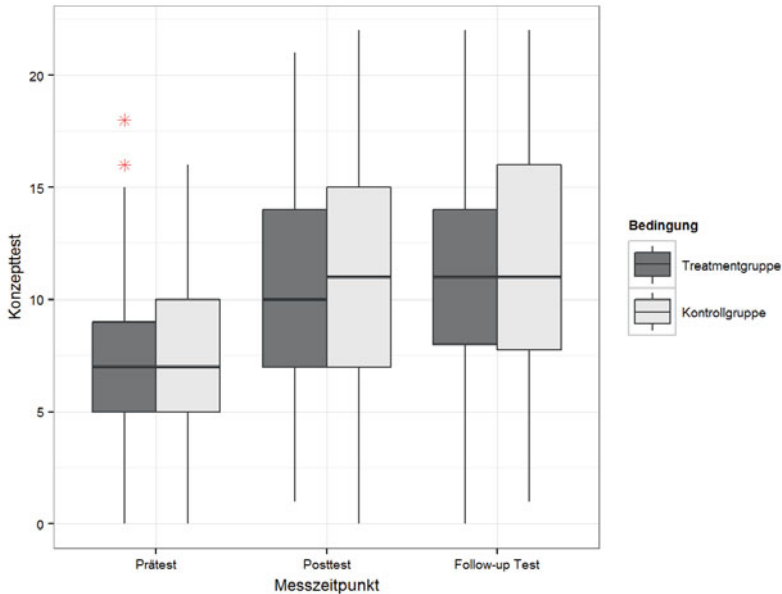


Abbildung 55: Boxplots der Konzepttestergebnisse je Bedingung zu den drei Messzeitpunkten: prä, post und follow-up

Die Verlaufplots legen nahe, dass sich das konzeptuelle Verständnis in beiden Bedingungen ähnlich entwickelt (vgl. Abbildung 56). Auch im Konzepttest schneiden die Gymnasiasten im Mittel zu allen drei Messzeitpunkten besser ab als die Gesamtschüler.

Ebenso wie im Modell zum Leistungstest, erweist sich auch für den Konzepttest jede der infrage kommenden Variablen außer der Geschlechtszugehörigkeit zu mindestens einem der drei Zeitpunkte als relevant. Entsprechend wurden alle infrage kommenden Variablen außer der Geschlechtszugehörigkeit im Gesamtmodell berücksichtigt. Die ICCs des Modells zur Untersuchung des konzeptuellen

Verständnisses bestätigten, dass die Verwendung eines hierarchischen linearen Modells ($ICC_{(Level-Schulklasse)} = 0.25$, $ICC_{(Level-Individuen)} = 0.51$) die angemessene Analyseform darstellte. Die Berücksichtigung der Heteroskedastizität verbesserte im Hinblick auf die Informationskriterien AIC und BIC das Modell, daher wurde auch hier mathematisch das gleiche Ausgangsmodell gewählt wie im Leistungstest. Die Ergebnisse des Modells zur Analyse der Wirkung der Intervention auf das konzeptuelle Verständnis finden sich in Tabelle 55.

Unter Berücksichtigung der genannten Kovariaten ergaben sich folgende Ergebnisse bezüglich des Lernzuwachses im Post- und im Follow-up Test (Untersuchung der zweiten Hypothese): Schüler in beiden Bedingungen starteten mit dem gleichen Ausgangsniveau bezüglich des konzeptuellen Verständnisses ($F_{(1, 17)} = 0.03$, *n.s.*). Betrachtet man den Lernzuwachs, erzielten Schüler in der Treatmentbedingung weder prä – post ($F_{(1, 802)} = 1.85$), noch prä – follow-up ($F_{(1, 802)} = 1.47$) einen signifikant höheren Lernzuwachs als die Kontrollgruppe (*n.s.*). Werte nahe 0 für den Anteil aufgeklärter Varianz (der sich auf das Treatment zurückführen lässt) und die geringe Effektstärke Δ unterstützen den Befund, dass keine Unterschiede feststellbar sind.

Von den genannten Kovariaten erwiesen sich lediglich die Schulnoten (berücksichtigt durch den Notenfaktor durch das Ergebnis der PCA, siehe Kapitel 2.3.6.2 Ergebnisse zu vorherigen Schulleistungen) im Gesamtmodell zu allen drei Messzeitpunkten als hoch signifikant. Schüler, deren Bewertung der Schulleistung um eine Notenstufe besser war, erzielten über alle Messzeitpunkte hinweg ein Testergebnis, das um einen Drittel Punkt besser war ($b = -0.33$). Ausgedrückt in Standardabweichungen bedeutet dies: Weichen die Noten um eine Standardabweichung ab, ergibt sich ein Unterschied im Leistungstest von $\beta = -0.10$. Schüler, deren Bewertung der Schulleistung um eine Standardabweichung besser war, erzielten darüber hinaus im Schnitt noch zusätzlich einen Lernzuwachs prä – post, der ebenfalls um knapp ein Drittel einer Standardabweichung höher lag ($\beta = -0.27$) sowie einen Lernzuwachs vom Prä- zum Follow-up Test, der etwa um ein Viertel einer Standardabweichung ($\beta = -0.26$) höher lag als der Mittelwert. Dies entspricht bezüglich des Lernzuwachses in beiden Fällen einem Vorteil von knapp einem Punkt von maximal 22 zu erreichenden Punkten des Konzepttests.

Von den erhobenen Intelligenzkomponenten erweisen sich im Gesamtmodell die verbale Intelligenz und das figural-räumliche Schlussfolgern prinzipiell als relevant ($p < .01$). Weicht der gemessene Wert im I-S-T 2000 R um eine Standardabweichung ab, so erzielen die Schüler ein Ergebnis im Konzepttest, das um (etwa) eine Zehntel Standardabweichungen höher liegt als der Mittelwert: $\beta_{IQ\text{-figural-räumlich}} = 0.08$, $\beta_{IQ\text{-verbal}} = 0.10$.

* Ausgangswert

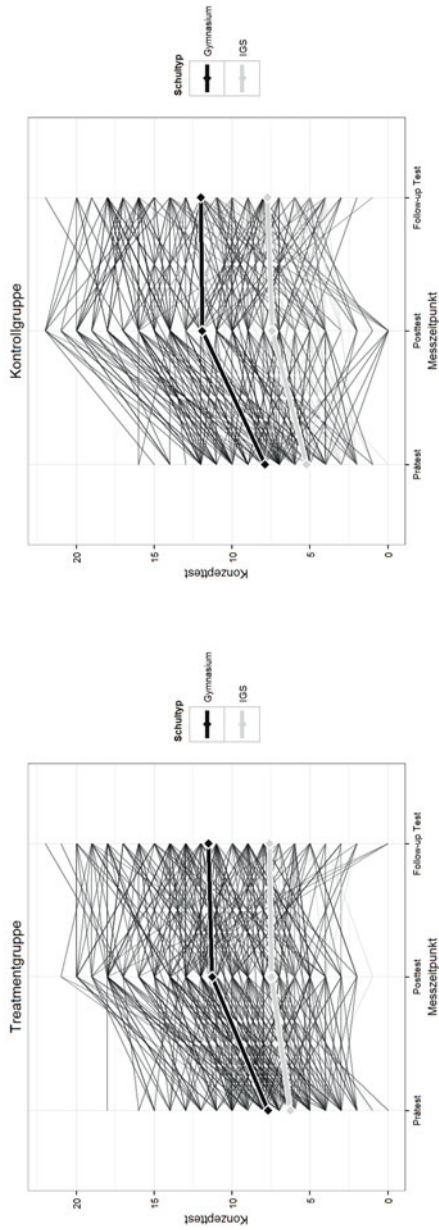


Abbildung 56: Verlaufsplots der Konzepttestergebnisse je Bedingung und Schultyp

Tabelle 55 Ergebnisse des Mehrebenenmodells zum konzeptuellen Verständnis in der Strahlenoptik

Level (Stichprobengröße)	Erläuterung	b	(SE)	β	(SE)	F _(numDF, denDF)	p
Level 1 (N = 1263)	Messzeitpunkte Individuen: Noten, Intelligenz, LES (= Lehrerengagement aus Schülersicht)	6.76	0.73	1.45	0.16	86.71 _(1,802)	<.001
Level 2 (N = 443)	Schulklassen: Bedingung Treatment- (TG) vs. Kontrollgruppe (KG), Schultyp, Klassengröße	3.08	0.62	0.66	0.13	25.04 _(1,802)	<.001
Level 3 (N = 21)		2.72	0.59	0.58	0.13	21.10 _(1,802)	<.001
Fixe Effekte		0.09	0.54	0.02	0.12	0.03 _(1,17)	.873
<i>Variable</i>	<i>Erläuterung</i>	-0.57	0.42	-0.12	0.09	1.85 _(1,802)	.174
Interzept	Durchschnittliches Verständnis: Kontrollgruppe, weiblich, IGS, mit durchschnittlichen Werten bezüglich IQ und Noten	-0.49	0.40	-0.10	0.09	1.47 _(1,802)	.226
Zuwachs (prä – post: t ₂ -t ₁)	Durchschnittlicher Verständniszuwachs: Kontrollgruppe, weiblich, IGS, mit durchschnittlichen Werten bezüglich IQ und Noten	-0.33	0.11	-0.10	0.03	9.46 _(1,417)	.002
Zuwachs (prä – follow-up: t ₃ -t ₁)	Durchschnittlicher Verständniszuwachs: Kontrollgruppe, weiblich, IGS, mit durchschnittlichen Werten bezüglich IQ und Noten	-0.91	0.16	-0.27	0.05	31.28 _(1,802)	<.0001
Bedingung = Treatmentgruppe (TG)	Unterschied zwischen TG und KG zu t ₁	-0.88	0.16	-0.26	0.05	29.01 _(1,802)	<.0001
Bedingung * Zuwachs prä – post	Zusätzlicher Einfluss der Bedingung = TG auf den Zuwachs zwischen t ₂ und t ₁	0.04	0.01	0.10	0.03	9.45 _(1,417)	.002
Bedingung * Zuwachs prä – follow-up	Zusätzlicher Einfluss der Bedingung = TG auf den Zuwachs zwischen t ₃ und t ₁	0.01	0.02	0.03	0.05	0.27 _(1,802)	.604
Noten (D, M, PH)	Genereller Einfluss auf Basis von t ₁	0.01	0.02	0.02	0.05	0.17 _(1,802)	.683
Noten * Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t ₂ und t ₁	0.04	0.02	0.08	0.03	6.59 _(1,417)	.011
Noten * Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t ₃ und t ₁	-0.03	0.02	-0.06	0.05	1.72 _(1,802)	.190
IQ, verbal	Genereller Einfluss auf Basis von t ₁	-0.06	0.02	-0.11	0.05	5.68 _(1,802)	.017
IQ, verbal* Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t ₂ und t ₁						
IQ, verbal* Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t ₃ und t ₁						
IQ, figural-räumlich* Ausgangswert	Genereller Einfluss auf Basis von t ₁						
IQ, figural-räumlich* Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t ₂ und t ₁						
IQ, figural-räumlich* Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t ₃ und t ₁						

IQ figural-logical* Ausgangswert	Genereller Einfluss auf Basis von t_1	0.02	0.01	0.05	0.03	2.59 _(1,417)	.108
IQ figural-logical* Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_3	-0.01	0.02	-0.02	0.05	0.22 _(1,802)	.637
IQ figural-logical* Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_4	-0.01	0.02	-0.03	0.05	0.37 _(1,802)	.543
Klassengröße	Genereller Einfluss auf Basis von t_1	0.07	0.09	0.04	0.06	0.55 _(1,17)	.467
Klassengröße * Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_3	0.11	0.07	0.07	0.05	2.34 _(1,802)	.127
Klassengröße * Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_4	0.10	0.07	0.07	0.05	2.19 _(1,802)	.140
Schultyp * Ausgangswert	Genereller Einfluss des Schultyps = Gymnasium auf Basis von t_1	0.68	0.76	0.15	0.16	0.80 _(1,17)	.384
Schultyp * Posttest	Zusätzlicher Einfluss des Schultyps = Gymnasium auf den Zuwachs zwischen t_2 und t_3	0.68	0.65	0.15	0.14	1.10 _(1,802)	.296
Schultyp * Follow-up Test	Zusätzlicher Einfluss des Schultyps = Gymnasium auf den Zuwachs zwischen t_3 und t_4	1.26	0.63	0.27	0.14	4.02 _(1,802)	.045
LES (vor der Intervention) * Posttest	Genereller Einfluss auf Basis von t_1	0.18	0.13	0.05	0.03	2.03 _(1,417)	.155
LES (vor der Intervention) * Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_3	0.31	0.18	0.08	0.05	3.08 _(1,802)	.080
	Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_4	-0.06	0.17	-0.02	0.04	0.11 _(1,802)	.735
Zufällige Effekte							
σ_u (Klasse)	1.08						
σ_B (Individuum)	1.55						
σ_ϵ (Individuengruppe)	2.15						
ρ (Messzeitpunkt)	$(\rho_{12}) - 0.01$	(ρ_{13})	0.04	(ρ_{23})	0.25		
g (Messzeitpunkt)	(g_1) 1.00	(g_2) 1.66		(g_3) 1.61			
Modellvergleich							
Berichtetes Modell							
Devianz	6597.94	Leermode ^l a					
df	35	6764.41					
AIC	6667.94	11					
BIC	6847.88	6786.411					
LR-Test (berichtetes Modell versus Leermode ^l)	166.47***	6842.964					
Unerklärt Varianz Level 1	(t_1) 8.20	(t_2) 16.31	(t_3) 15.59	(t_4) 10.83	(t_5) 21.49	(t_6) 21.01	

Unerklärte Varianz Level 2	13.37	17.78
Unerklärte Varianz Level 3	1.66	3.76
Erklärte Varianz des berichteten Modells im Vergleich zum Leermodell		
R ² Level 1	(t ₁) 0.24	(t ₃) 0.26
R ² Level 2	0.25	
R ² Level 3	0.57	
„Wirkung“ des Treatments		
Anteil aufgeklärter Varianz ^b		
R ² Level 1	(t ₁) 0.01	Effektstärke
R ² Level 2	< 0.01	$\Delta_{\text{pre} - \text{post}} = -0.15$
R ² Level 3	< 0.01	$\Delta_{\text{pre} - \text{follow-up}} = -0.13$

* $p < .05$, ** $p < .01$, *** $p < .001$

^a Das Leermodell unterscheidet sich vom berichteten Modell darin, dass alle fixen Effekte außer dem Intercept und den Dummy-Variablen t₁ und t₃ fehlen. Das Modell sagt also für jeden Schüler voraus, dass seine Leistung zum jeweiligen Messzeitpunkt dem mittleren Konzeptuellen Verständnis der Stichprobe entspricht - unter Berücksichtigung der individuellen- und klassenspezifischen Unterschiede. Man beachte, das Leermodell schätzt folgende 11 Parameter (df = 11): Intercept, t₂, t₃, β_2 , β_3 , β_{22} , β_{33} , β_{23} , σ_w , σ_{t1} , σ_{t3} .

^b Der Anteil aufgeklärter Varianz errechnet sich wie in Tabelle 52 angegeben nach Snijder und Bosker (1999, S. 99 ff.).

Für das harmonische Mittel der Klassengrößen ergibt sich $n = 24.60$

Der Schultyp zeigt im Gesamtmodell lediglich auf, dass Gymnasiasten mittelfristig (prä – follow-up) das Gelernte besser behalten als Gesamtschüler ($F_{(1, 802)} = 4.02, p < .05$). Sie erzielten durchschnittlich einen Lernzuwachs, der um knapp ein Drittel einer Standardabweichung ($\beta = 0.27$) prä – follow-up höher liegt als der Zuwachs der Gesamtschüler. Dies entspricht einem Vorteil von etwas mehr als einem Punkt von maximal 22 zu erreichenden Punkten des Konzepttests. Zu den anderen Messzeitpunkten konnten keine signifikanten Unterschiede im Erwerb des konzeptuellen Verständnisses zwischen Gymnasiasten und Gesamtschülern festgestellt werden. Ein Einfluss der Klassengröße und des Lehrer-Engagements aus Schülersicht (LES) ist im Gesamtmodell nicht mehr nachweisbar.

2.3.9.5 Dritte Hypothese: Wirkung der Treatmentvariation auf das konzeptuelle Verständnis

Zur Untersuchung der dritten Hypothese wurde geprüft, ob Schüler, welche ein variiertes Treatment erhalten hatten, zum zweiten und dritten Messzeitpunkt (nach der Intervention sowie zwei Monate später) in Bezug auf das konzeptuelle Verständnis einen höheren Lernzuwachs verzeichnen können, als Schüler, welche das reguläre Treatment erhalten hatten bzw. als Schüler in der Kontrollbedingung.

Die deskriptiven Statistiken zeigen, dass alle Gruppen in etwa mit dem gleichen Ausgangsniveau starten, wobei die Schüler in der variierten Treatmentbedingung die höchsten Werte im Prätest erzielten.

Nach der Unterrichtsreihe verbessern Schüler in allen Bedingungen ihr konzeptuelles Verständnis und diese Verbesserung bleibt auch in allen Bedingungen erhalten (vgl. Tabelle 56).

Bei dem Blick auf die Boxplots deutet sich an, dass Schüler in der variierten Treatmentbedingung im Prä- und Posttest leicht im Vorteil zu sein scheinen (vgl. Abbildung 57). Im Posttest erreichen Schüler in der variierten Treatmentbedingung im Schnitt die höchsten Punktzahlen und auch den höchsten Zuwachs im Vergleich zum Prätest. Im Follow-up Test scheinen Schüler in der Kontrollbedingung am besten abzuschneiden und im Vergleich zu den vorherigen Messzeitpunkten den höchsten Lernzuwachs zu erzielen. Zu prüfen ist, ob diese Unterschiede signifikant sind.

Tabelle 56 Deskriptive Statistiken Konzepttest je Bedingung

	Prätest (N = 491)			Posttest (N = 480)			Follow-up Test (N = 486)		
	TG var. ^a	TG reg. ^b	KG ^c	TG var. ^a	TG reg. ^b	KG ^b	TG var. ^a	TG reg. ^b	KG ^b
	(n = 91)	(n = 169)	(n = 231)	(n = 91)	(n = 157)	(n = 232)	(n = 92)	(n = 166)	(n = 228)
M	7.83	7.28	7.41	12.01	10.04	11.20	10.68	11.05	11.18
(SD)	3.32	3.11	(3.23)	4.96	4.45	(5.29)	5.09	4.33	(4.86)
Range	0 - 15.00	1 - 18	0 - 16.00	1 - 20.00	1 - 21.00	0 - 22.00	0 - 20.00	2 - 22.00	0 - 22.00

Im folgenden Mehrebenenmodell berücksichtigte Daten (N = 443)^c

	(n = 87)	(n = 144)	(n = 210)	(n = 80)	(n = 131)	(n = 200)	(n = 80)	(n = 137)	(n = 194)
M	7.83	7.28	7.49	12.01	10.04	11.28	10.68	11.05	11.39
(SD)	3.32	3.11	(3.10)	4.96	4.45	(5.27)	5.09	4.33	(4.83)

Anmerkung: Maximal erreichbare Punktzahl im Konzepttest: 22 Punkte,

^avariiertes Treatment, ^breguläres Treatment, ^cKontrollgruppe.

^cDie Verminderung der Stichprobengröße ergibt sich aus fehlenden Kovariaten.

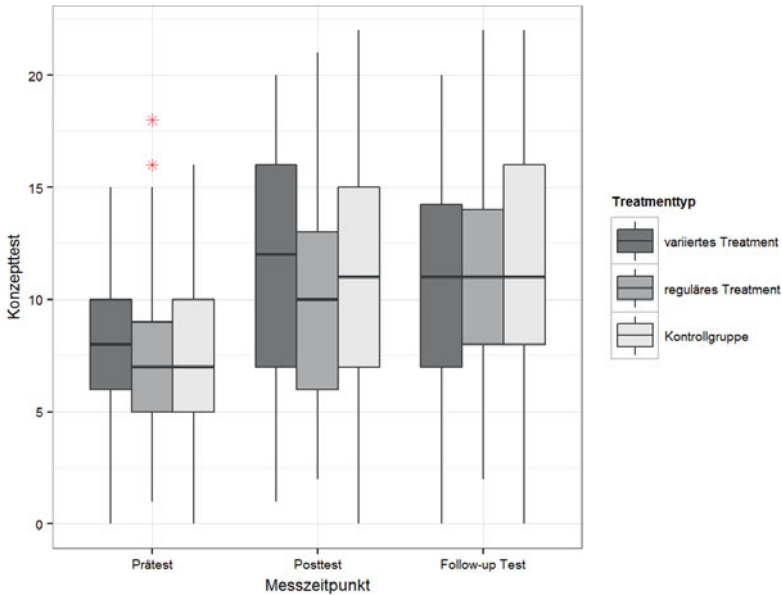


Abbildung 57: Boxplots der Konzepttestergebnisse je Bedingung zu den drei Messzeitpunkten: prä, post und follow-up

Um die Ergebnisse direkt mit den Ergebnissen des vorigen Mehrebenenmodells zum konzeptuellen Verständnis vergleichen zu können (vgl. Tabelle 55), wurden die gleichen Kovariaten wie in dem vorherigen Modell berücksichtigt. Die Ergebnisse sind in Tabelle 57 wieder gegeben.

Die Ergebnisse des Modells (vgl. Tabelle 57) zeigen, dass Schüler in den verschiedenen Bedingungen mit dem gleichen Ausgangsniveau bezüglich des konzeptuellen Verständnisses starten: Dies gilt sowohl für den Vergleich zwischen regulärem Treatment und Kontrollgruppe ($F_{(1,16)} = 0.00, n.s.$), für den Vergleich zwischen variiertem Treatment und Kontrollgruppe ($F_{(1,16)} = 0.12, n.s.$) als auch für den Vergleich zwischen regulärem Treatment und variiertem Treatment ($F_{(1,16)} = 0.10, n.s.$).

Betrachtet man den Lernzuwachs erzielten Schüler in der Kontrollgruppe prä – post ($F_{(1,800)} = 4.72, p = .030$) einen signifikant höheren Lernzuwachs als in der regulären Treatmentbedingung, dies gilt jedoch nicht für den Lernzuwachs prä – follow-up ($F_{(1,800)} = .013, n.s.$). Das variierte Treatment erwies sich als signifikant besser als das reguläre Treatment prä – post ($F_{(1,800)} = 4.05, p = .045$), dieser Unterschied blieb prä – follow-up ($F_{(1,800)} = 2.27, n.s.$) jedoch nicht erhalten. Zwischen variiertem Treatment- und Kontrollgruppe ergaben sich weder ein Unterschied prä – post ($F_{(1,800)} = 0.10, n.s.$) noch prä – follow-up ($F_{(1,800)} = 3.68, n.s.$).

Die in den Boxplots ersichtlichen deskriptiven Vorteile zugunsten des variierten Treatments (vgl. Abbildung 57) werden also bestätigt. Werte nahe 0 für den Anteil aufgeklärter Varianz und die geringe Effektstärke Δ belegen jedoch, dass die genannten Unterschiede praktisch wenig bedeutsam sind (vgl. Tabelle 57).

Von den genannten Kovariaten erwiesen sich lediglich die Schulnoten (berücksichtigt durch den Notenfaktor als Ergebnis der PCA, siehe Kapitel 2.3.6.2 Ergebnisse zu vorherigen Schulleistungen) im Gesamtmodell zu allen drei Messzeitpunkten als hoch signifikant. Je besser die Schulnoten waren, desto höher fiel die Testleistung aus. Schüler, deren Bewertung der Schulleistung um eine Notenstufe besser war, erzielten über alle Messzeitpunkte hinweg im Schnitt ein Testergebnis, das um einen Drittel Punkt höher lag ($b = 0.33, \beta = -0.10$). Schüler, deren Bewertung der Schulleistung um Standardabweichung besser war, erzielten darüber hinaus im Schnitt noch zusätzlich einen Lernzuwachs prä – post, der ebenfalls um knapp ein Drittel einer Standardabweichung höher lag ($\beta = -0.27$) sowie einen Lernzuwachs vom Prä- zum Follow-up Test, der etwa um ein Viertel einer Standardabweichung ($\beta = -0.26$) höher ausfiel, als der Mittelwert der Gesamtstichprobe. Dies entspricht bezüglich des Lernzuwachses in beiden Fällen einem Vorteil von knapp einem Punkt von maximal 22 zu erreichenden Punkten des Konzepttests.

Tabelle 57 Ergebnisse des Mehrebenenmodells zum Vergleich der Wirkung des regulären Treatments mit der Treatmentvariation auf das konzeptuelle Verständnis in der Strahlenoptik

Level (Stichprobengröße)	Variable	b	(SE)	β	(SE)	F (numDF, denDF)	p
Level 1 (N = 1263)	Erläuterung Messzeitpunkte Individuen: Noten, Intelligenz, LES (= Lehrerengagement aus Schülersicht)	6.74	0.73	1.44	0.16	86.00(1,800)	< .001
Level 2 (N = 443)	Schulklassen: Bedingung Treatment-(TG) vs. Kontrollgruppe (KG), Schultyp, Klassengröße						
Level 3 (N= 21)							
Fixe Effekte							
	<i>Variable</i>						
Interzept	<i>Erläuterung</i> Durchschnittliches Verständnis: Kontrollgruppe, Geschlechtszugehörigkeit = weiblich, IGS, mit durchschnittlichen Werten bezüglich IQ und Noten	2.94	0.62	0.63	0.13	22.72(1,800)	< .001
Zuwachs (prä – post: t ₂ -t ₁)	Durchschnittlicher Verständniszuwachs: Kontrollgruppe, Geschlechtszugehörigkeit = weiblich, IGS, mit durchschnittlichen Werten bezüglich IQ und Noten	2.84	0.60	0.61	0.13	22.68(1,800)	< .001
Zuwachs (prä – follow-up: t ₃ -t ₁)	Durchschnittlicher Verständniszuwachs: Kontrollgruppe, Geschlechtszugehörigkeit = weiblich, IGS, mit durchschnittlichen Werten bezüglich IQ und Noten	0.01	0.61	-0.01	0.13	0.00(1,116)	.997
Bedingung = Treatmentgruppe regulär (TG)	Unterschied zwischen regulärem Treatment und Kontrollgruppe zum Zeitpunkt 1	-1.03	0.48	-0.22	0.10	4.72(1,800)	.030
Bedingung * Zuwachs prä – post	Unterschied im Lernzuwachs zwischen t ₂ und t ₁ zwischen regulärem Treatment und Kontrollgruppe	-0.16	0.46	-0.04	0.10	0.13(1,800)	.719
Bedingung * Zuwachs prä – follow-up	Unterschied im Lernzuwachs zwischen t ₃ und t ₁ zwischen regulärem Treatment und Kontrollgruppe	0.25	0.78	0.05	0.17	0.10(1,16)	.751
Bedingung = Treatmentgruppe variiert (TG)	Unterschied zwischen regulärem Treatment und Treatmentvariation zum Zeitpunkt 1	1.21	0.60	0.26	0.13	4.02(1,800)	.045
Bedingung * Zuwachs prä – post	Unterschied im Lernzuwachs zwischen t ₂ und t ₁ zwischen regulärem Treatment und Treatmentvariation	-0.86	0.58	0.12	0.12	2.27(1,800)	.132
Bedingung * Zuwachs prä – follow-up	Unterschied im Lernzuwachs zwischen t ₃ und t ₁ zwischen regulärem Treatment und Treatmentvariation	0.25	0.73	0.05	0.16	0.12(1,116)	.737
Bedingung = Treatment regulär + Treatment variiert	Unterschied zwischen Treatmentvariation und Kontrollgruppe zum Zeitpunkt 1 (Summe der obigen Effekte)						

Unterschied im Lernzuwachs zwischen t_2 und t_1 zwischen Treatmentvariation und Kontrollgruppe zum (Summe der obigen Effekte)	0.18	0.56	0.04	0.12	0.10 _(1,800)	.747
Unterschied im Lernzuwachs zwischen t_3 und t_1 zwischen Treatmentvariation und Kontrollgruppe (Summe der obigen Effekte)	-1.03	0.54	-0.22	0.12	3.68 _(1,800)	.056
Genereller Einfluss auf Basis von t_1	-0.33	0.11	-0.10	0.03	9.44 _(1,417)	.002
Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_1	-0.92	0.16	-0.27	0.05	31.82 _(1,800)	.722
Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_1	-0.88	0.16	-0.26	0.05	29.03 _(1,800)	.601
Genereller Einfluss auf Basis von t_1	0.04	0.01	0.10	0.03	9.36 _(1,417)	.011
Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_1	0.01	0.02	0.02	0.05	0.13 _(1,800)	.126
Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_1	0.01	0.02	0.02	0.05	0.27 _(1,800)	.026
Genereller Einfluss auf Basis von t_1	0.04	0.02	0.08	0.03	6.52 _(1,417)	.109
Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_1	-0.04	0.02	-0.07	0.05	2.35 _(1,800)	.580
Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_1	-0.05	0.02	-0.10	0.04	4.99 _(1,800)	.592
Genereller Einfluss auf Basis von t_1	0.02	0.01	0.05	0.05	2.58 _(1,417)	.472
Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_1	-0.01	0.02	-0.03	0.04	0.31 _(1,800)	.167
Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_1	-0.01	0.02	-0.02	0.06	0.29 _(1,800)	.121
Genereller Einfluss auf Basis von t_1	0.07	0.07	-0.04	0.05	10.54 _(1,16)	.369
Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_1	0.10	0.76	0.07	0.05	1.92 _(1,800)	.193
Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_1	0.11	0.66	0.07	0.05	2.41 _(1,800)	.078
Genereller Einfluss des Schultyps = Gymnasium auf Basis von t_1	0.70	0.63	0.15	0.16	0.86 _(1,16)	.157
Bedingung * Zuwachs prä – post						
Bedingung * Zuwachs prä – follow-up						
Noten (D, M, PH)						
Noten * Posttest						
Noten * Follow-up Test						
IQ, verbal						
IQ, verbal* Posttest						
IQ, verbal* Follow-up Test						
IQ, figurativ-räumlich* Ausgangswert						
IQ, figurativ-räumlich* Posttest						
IQ, figurativ-räumlich* Follow-up Test						
IQ, figurativ-logisch* Ausgangswert						
IQ, figurativ-logisch* Posttest						
IQ, figurativ-logisch* Follow-up Test						
Klassengröße						
Klassengröße * Posttest						
Klassengröße * Follow-up Test						
Schultyp * Ausgangswert						

Schultyp * Posttest	Zusätzlicher Einfluss des Schultyps = Gymnasium auf den Zuwachs zwischen t_2 und t_1	0.86	0.03	0.18	0.14	1.70 _(1,800)	.142
Schultyp * Follow-up Test	Zusätzlicher Einfluss des Schultyps = Gymnasium auf den Zuwachs zwischen t_3 und t_1	1.12	0.04	0.24	0.14	3.12 _(1,800)	.881
LES (vor der Intervention)	Genereller Einfluss auf Basis von t_1	0.18	0.13	0.05	0.03	2.01 _(1,417)	< .001
LES (vor der Intervention) * Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t_2 und t_1	0.26	0.18	0.07	0.05	2.16 _(1,800)	< .001
LES (vor der Intervention)* Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t_3 und t_1	-0.03	0.17	-0.01	0.04	0.02 _(1,800)	< .001
Zufällige Effekte							
σ_u (Klasse)	1.07						
σ_p (Individuum)	1.57						
σ_e (Individuenspezifisch)	2.14						
ρ (Messzeitpunkt)	(ρ_{12}) -0.01						
g (Messzeitpunkt)	(g_1) 1.00						
Modellvergleich							
Berichtetes Modell		Leermodell ^a					
Devianz	6597.94	6261.87					
df	38	11					
AIC	6662.72	6786.41					
BIC	6858.08	6842.96					
LR-Test (berichtetes Modell versus Leermodell)	177.69***						
Un erklärte Varianz Level 1	(t_1) 8.20	(t_2) 16.12	(t_3) 15.52	(t_3) 21.49			
Un erklärte Varianz Level 2	13.28	17.78					
Un erklärte Varianz Level 3	1.65	3.76					
Erklärte Varianz des berichteten Modells im Vergleich zum Leermodell							
R^2 Level 1	(t_1) 0.24	(t_2) 0.25	(t_3) 0.26				
R^2 Level 2	0.25						
R^2 Level 3	0.56						

„Wirkung“ des Treatments	Anteil aufgeklärter Varianz ^b	(t ₃)	Effektstärke
R ² Level 1	(t ₁) < 0.01	(t ₃) 0.01	reguläres Treatment vs. Kontrollgruppe $\Delta_{(pr3 - post)} = - 0.27$ $\Delta_{(pr3 - follow-up)} = - 0.04$
R ² Level 2	0.01		variirtes Treatment vs. Kontrollgruppe $\Delta_{(pr3 - post)} = 0.05$ $\Delta_{(pr3 - follow-up)} = - 0.27$
R ² Level 3	0.01		

* $p < .05$, ** $p < .01$, *** $p < .001$

^a Das Leermmodell unterscheidet sich vom berichteten Modell darin, dass alle fixen Effekte außer dem Interzept und den Dummy-Variablen t₂ und t₃ fehlen. Das Modell sagt für jeden Schüler voraus, dass seine Leistung zum jeweiligen Messzeitpunkt dem mittleren Konzeptuellen Verständnis der Stichprobe entspricht - unter Berücksichtigung der individuellen- und Klassenspezifischen Unterschiede. Man beachte, das Leermmodell schätzt folgende 11 Parameter (df = 11): Interzept, t₂, t₃, β₂, β₃, p₁₂, p₁₃, p₂₃, σ_u, σ_v, σ_ε.

^b Der Anteil aufgeklärter Varianz errechnet sich wie in Tabelle 52 angegeben nach Snijder und Bosker (1999, S. 99 ff.). Für das harmonische Mittel der Klassengrößen ergibt sich $h = 24.60$.

Wie zu erwarten war, unterschieden sich die Ergebnisse bezüglich des Einflusses der Kovariaten kaum im Vergleich zu dem Mehrebenenmodell zum konzeptuellen Verständnis, welches nicht zwischen regulärem und variiertem Treatment differenziert.

Von den erhobenen Intelligenzkomponenten erwies sich ebenfalls das verbale Schlussfolgern als relevant: Wich der gemessene Wert im I-S-T 2000 R um eine Standardabweichung ab, so erzielten die Schüler ein Ergebnis im Konzepttest, das um knapp ein Zehntel einer Standardabweichung höher lag ($\beta = 0.10, p < .05$).

Der Schultyp erwies sich in diesem Modell nunmehr nicht mehr als signifikant. Die Klassengröße und das Lehrer-Engagement aus Schülersicht spielten ebenfalls keine signifikante Rolle.

2.3.9.6 Vierte Hypothese: Wirkung des Treatments auf die Schülermotivation

Zur Untersuchung der Wirkung des Unterrichts in der Treatment- und in der Kontrollbedingung auf die Schülermotivation wurde untersucht, ob sich die Motivation im Verlauf in den beiden Bedingungen unterscheidet. Die deskriptiven Statistiken ergeben, dass beide Gruppen mit dem gleichen Ausgangsniveau starten, im Verlauf des Schuljahrs die Motivation jedoch in beiden Bedingungen sinkt und sich zum zweiten und dritten Messzeitpunkt auf einem etwas niedrigeren Niveau stabilisiert (vgl. Tabelle 58).

Tabelle 58 Deskriptive Statistiken Motivation je Bedingung

	Prätest (N = 489)		Posttest (N = 477)		Follow-up Test (N = 495)	
	TG ^a (n = 259)	KG ^b (n = 230)	TG ^a (n = 253)	KG ^b (n = 224)	TG ^a (n = 264)	KG ^b (n = 231)
<i>M</i>	3.59	3.60	3.40	3.43	3.43	3.42
<i>(SD)</i>	(0.80)	(0.78)	(0.88)	(0.85)	(0.86)	(0.90)
<i>Range</i>	1.08 - 5.63	1.33 - 5.36	1.00 - 6.00	1.21 - 5.58	1.00 - 6.00	1.00 - 5.71
<i>Im folgenden Mehrebenenmodell berücksichtigte Daten (N = 443)^f</i>						
	(n = 231)	(n = 212)	(n = 219)	(n = 193)	(n = 222)	(n = 222)
	3.57	3.60	3.43	3.43	3.40	3.38
	(0.82)	(0.81)	(0.91)	(0.88)	(0.87)	(0.91)

Anmerkung. Skala von 1 (niedrige Ausprägung) - 6 (hohe Ausprägung)

^aTreatmentgruppe, ^bKontrollgruppe

^fDie verminderte Stichprobenzahl ergibt sich aus fehlenden Kovariaten.

Augenscheinlich sind in den Boxplots (vgl. Abbildung 58) kaum Unterschiede zwischen den Bedingungen zu erkennen. Wie sich bereits an den Mittelwerten erkennen ließ, sinkt das Motivationsniveau nach der Unterrichtsreihe leicht ab und stabilisiert sich. D.h. das etwas verminderte Motivationsniveau nach der Intervention bleibt zum dritten Messzeitpunkt in beiden Bedingungen in etwa erhalten.

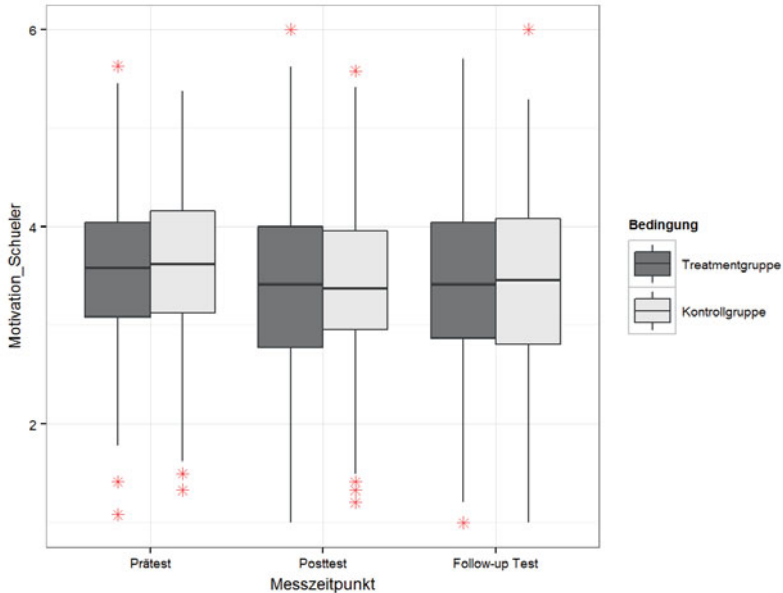


Abbildung 58: Boxplots zur Schülermotivation je Bedingung

Ein interessantes Detailergebnis zeigt sich im Hinblick auf den Motivationsverlauf je Schultyp in der Treatmentbedingung (vgl. Abbildung 59). Im Gegensatz zum Motivationsverlauf der Gymnasiasten, deutet sich für die Gesamtschüler tendentiell ein u-förmiger Verlauf an: D.h. die Motivation der Schüler erreicht ihren Tiefpunkt unmittelbar nach der Unterrichtsreihe und steigt zwei Monate später nach dem regulären Physikunterricht etwas an, wobei zum letzten Messzeitpunkt in etwa das gleiche Motivationsniveau wie an den Gymnasien erreicht wird.

Dabei gilt es zu prüfen, ob der Unterschied im Motivationsverlauf prä – post zwischen den Schultypen Gymnasium und Gesamtschule signifikant ist und ob in Bezug auf den Schultyp ATI-Effekt vorliegt.

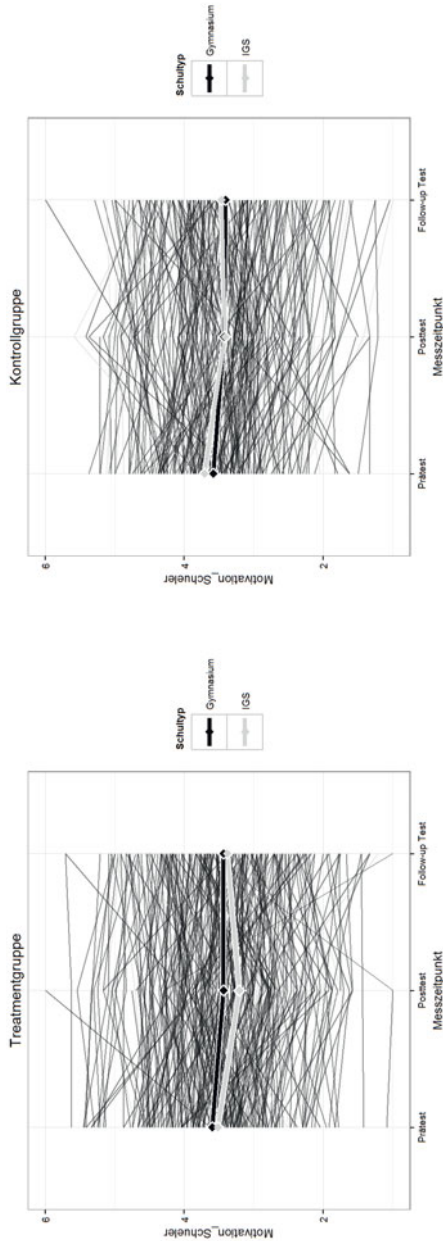


Abbildung 59: Verlaufplots zur Schülermotivation je Bedingung und Schultyp

Auf Basis der Ergebnisse der Einzeltestung relevanter Kovariaten wurden die Variablen: Schulleistungen (Noten), das figural-logische Schlussfolgern, das Lehrerengagement aus Sicht der Schüler (LES) und die Geschlechtszugehörigkeit der Schultyp in dem Mehrebenenmodell zur Analyse des Motivationsverlaufs berücksichtigt.

Der in den Verlaufplots deskriptiv ersichtliche Unterschied im Motivationsverlauf prä – post zwischen Gymnasiasten und Gesamtschülern ist in der Einzeluntersuchung der Kovariaten knapp nicht signifikant ($F_{(1, 931)} = 3.01, p = .083$). Um dennoch den deskriptiv ersichtlichen Unterschieden nachzugehen, wurde die Variable „Schultyp“ in das Gesamtmodell aufgenommen.

Die Werte des ICC's des Modells zur Untersuchung Schülermotivation sprechen für die Verwendung eines hierarchischen linearen Modells anstelle einer klassischen Regressionsanalyse ($ICC_{(Klasse)} = 0.14, ICC_{(Individuen)} = 0.73$). Gemessen an den Informationskriterien *AIC* und *BIC* verbesserte die Berücksichtigung der Heteroskedastizität das Modell.

Unter Berücksichtigung der genannten Kovariaten ergaben sich folgende Ergebnisse zur Entwicklung der Schülermotivation (vgl. Tabelle 59): Schüler in beiden Bedingungen starten mit dem gleichen Ausgangsniveau ($F_{(1, 18)} = 0.72, n.s.$). Betrachtet man die Veränderung im weiteren Verlauf des Schuljahrs nach der Unterrichtsreihe, ist in der Entwicklung der Schülermotivation weder prä – post ($F_{(1, 816)} = 0.06, n.s.$), noch prä – follow-up ($F_{(1, 816)} = 2.19, n.s.$) eine signifikant unterschiedliche Entwicklung zwischen Treatment- und Kontrollgruppe zu beobachten. Die Werte für den Anteil der aufgeklärten Varianz durch das Treatment und die Effektstärke unterstreichen diesen Befund.

Von den berücksichtigten Kovariaten erwiesen sich folgende Variablen im Gesamtmodell als relevant: die Fachnoten, die Geschlechtszugehörigkeit, der Schultyp und das Lehrerengagement aus Schülersicht (LES).

Schüler, deren Bewertung der Schulleistung um eine Notenstufe höher liegt, stuften ihre Motivation im Schnitt über alle Messzeitpunkte hinweg um 4 Punkte höher ein ($b = -0.19, \beta = -0.30, p < .001$), allerdings sank die Schülermotivation „besserer“ Schüler auch signifikant stärker von der Prä- zur Postmessung. Dabei handelt es sich jedoch nur um einen kleinen Effekt: Weichen die Fachnoten um eine Standardabweichung nach oben ab, sinkt die Schülermotivation prä – post um $\beta = 0.03$ Standardabweichungen. Dies entspricht einem Rückgang von etwa 0.09 Punkten auf einer Skala von 1-6 Punkten.

Tabelle 59 Ergebnisse des Mehrebenenmodells zur Schülermotivation

Level (Stichprobengröße)	Erörterung	b	(SE)	β	(SE)	$F_{(numDF, denDF)}$	p
Level 1 (N = 1273)	Messzeitpunkte Individuen: Noten, Intelligenz, LES (= Lehrerengagement aus Schüler Sicht)	3.72	0.13	4.29	0.14	875.21 _(1,816)	< .001
Level 2 (N = 443)	Schulklassen: Bedingung Treatment- (TG) vs. Kontrollgruppe (KG), Schultyp, Klassengröße						
Level 3 (N = 21)							
Fixe Effekte							
<i>Variable</i>	<i>Erörterung</i>						
Interzept	Durchschnittlicher Wert: Kontrollgruppe, Geschlechtszugehörigkeit = weiblich, IGS, mit durchschnittlichen Werten bezüglich IQ und Noten	-0.28	0.09	-0.32	0.10	10.67 _(1,816)	=.001
Zuwachs (prä – post: t ₂ -t ₁)	Durchschnittliche Veränderung: Kontrollgruppe, Geschlechtszugehörigkeit = weiblich, IGS, mit durchschnittlichen Werten bezüglich IQ und Noten	-0.29	0.09	-0.33	0.11	9.64 _(1,816)	.002
Zuwachs (prä – follow-up: t ₃ -t ₁)	Durchschnittlicher Veränderung: Kontrollgruppe, Geschlechtszugehörigkeit = weiblich, IGS, mit durchschnittlichen Werten bezüglich IQ und Noten	-0.08	0.09	-0.09	0.11	0.72 _(1,18)	.407
Bedingung = Treatmentgruppe (TG)	Zusätzlicher Einfluss der Bedingung = TG auf die Veränderung zwischen t ₂ und t ₁	0.01	0.06	0.02	0.07	0.06 _(1,816)	.815
Bedingung * Zuwachs prä – post	Zusätzlicher Einfluss der Bedingung = TG auf die Veränderung zwischen t ₂ und t ₁	0.09	0.06	0.11	0.07	2.19 _(1,816)	.139
Bedingung * Zuwachs prä – follow-up	Zusätzlicher Einfluss der Bedingung = TG auf die Veränderung zwischen t ₃ und t ₁	-0.19	0.02	-0.30	0.04	60.62 _(1,418)	< .001
Noten (D, M, PH)	Genereller Einfluss auf Basis der Schätzung zu t ₁	0.06	0.02	0.09	0.03	7.72 _(1,816)	.006
Noten * Posttest	Zusätzlicher Einfluss auf die Veränderung zwischen t ₂ und t ₁	0.01	0.03	0.01	0.04	0.09 _(1,816)	.766
Noten * Follow-up Test	Zusätzlicher Einfluss auf die Veränderung zwischen t ₃ und t ₁	-0.00	0.03	-0.02	0.04	0.21 _(1,418)	.646
IQ figural-logisch* Ausgangswert	Genereller Einfluss auf Basis der Schätzung zu t ₁	0.00	0.03	0.04	0.03	1.26 _(1,816)	.262
IQ figural-logisch* Posttest	Zusätzlicher Einfluss auf die Veränderung zwischen t ₂ und t ₁	0.01	0.03	0.07	0.04	3.43 _(1,816)	.064
IQ figural-logisch* Follow-up Test	Zusätzlicher Einfluss auf die Veränderung zwischen t ₃ und t ₁	0.32	0.06	0.37	0.07	26.04 _(1,418)	< .001
Geschlecht * Ausgangswert	Einfluss Geschlechtszugehörigkeit = männlich zu t ₁						

Geschlecht * Posttest	Zusätzlicher Einfluss der Geschlechtszugehörigkeit = männlich auf die Veränderung zwischen t_2 und t_1	0.02	0.06	0.02	0.07	0.09 _(t, .816)	.764
Geschlecht * Follow-up Test	Zusätzlicher Einfluss der Geschlechtszugehörigkeit = männlich auf die Veränderung zwischen t_3 und t_1	-0.08	0.06	-0.09	0.07	1.41 _(t, .816)	.236
Schultyp * Ausgangswert	Genereller Einfluss des Schultyps = Gymnasium auf Basis der Schätzung zu t_1	-0.30	0.13	-0.35	0.15	5.74 _(t, 1.18) *	.028
Schultyp * Posttest	Zusätzlicher Einfluss des Schultyps = Gymnasium auf die Veränderung zwischen t_2 und t_1	0.13	0.08	0.15	0.10	2.43 _(t, .816)	.120
Schultyp * Follow-up Test	Zusätzlicher Einfluss des Schultyps = Gymnasium auf die Veränderung zwischen t_3 und t_1	0.09	0.09	0.10	0.11	0.96 _(t, .816)	.327
LES (vor der intervention)	Genereller Einfluss auf Basis der Schätzung zu t_1	0.29	0.03	0.41	0.04	101.20 _(t, .418)	<.001
LES (vor der intervention) * Posttest	Zusätzlicher Einfluss auf die Veränderung zwischen t_2 und t_1	-0.01	0.02	-0.02	0.03	0.31 _(t, .816)	.576
LES (vor der intervention) * Follow-up Test	Zusätzlicher Einfluss auf die Veränderung zwischen t_3 und t_1	-0.09	0.03	-0.13	0.04	12.26 _(t, .816)	<.001
Zufällige Effekte							
σ_k (Klasse)		0.16					
σ_P (Individuum)		0.53					
σ_ε (Individuenpezifisch)		0.32					
ρ (Messzeitpunkt)		(ρ_{12}) 0.13			(ρ_{23}) 0.43		
g (Messzeitpunkt)		(g_1) 1.00			(g_2) 1.63		
Modellvergleich							
Berichtetes Modell							
Devianz		2237.29				Leermodell ^b	
df		26				2414.37	
AIC		2295.29				11	
BIC		2444.61				2436.36	
LR-Test (berichtetes Modell versus Leermodell)		170.17***				2493.00	
Unerklärte Varianz Level 1		(t_1) 0.41		(t_2) 0.58		(t_3) 0.65	(t_3) 0.83
Unerklärte Varianz Level 2		0.55				0.75	
Unerklärte Varianz Level 3		0.05				0.14	
Erklärte Varianz des berichteten Modells im Vergleich zum Leermodell							

R^2 Level 1	(t ₁) 0.37	(t ₂) 0.23	(t ₃) 0.20	
R^2 Level 2	0.26			
R^2 Level 3	0.67			
„Wirkung“ des Treatments	Anteil aufklärter Varianz ^b			Effektstärke
R^2 Level 1	(t ₁) < 0.01	(t ₂) < 0.01	(t ₃) < 0.01	$\Delta_{\text{pre} - \text{follow-up}} = 0.11$
R^2 Level 2	< 0.01			
R^2 Level 3	0.02			

* $p < .05$, ** $p < .01$, *** $p < .001$

^a Das Leermodell unterscheidet sich vom berichteten Modell darin, dass alle fixen Effekte außer dem Interzept und den Dummy-Variablen t₂ und t₃ fehlen. Das Modell sagt für jeden Schüler voraus, dass die Motivation zum jeweiligen Messzeitpunkt den Mittelwert der Motivation in der Stichprobe entspricht - unter Berücksichtigung der individuellen- und klassenspezifischen Unterschiede. Man beachte, das Leermodell schätzt folgende 11 Parameter (df = 11): Interzept, t₂, t₃, β_2 , β_3 , ρ_2 , ρ_3 , σ_u , σ_v , σ_e .

^b Der Anteil aufklärter Varianz errechnet sich wie in Tabelle 52 angegeben nach Snijder und Bosker (1999, S. 99 ff).

Für das harmonische Mittel der Klassengrößen ergibt sich $h = 24.60$.

Jungen stufen ihre Motivation über alle Messzeitpunkte hinweg um 0.32 Punkte signifikant höher ein als Mädchen ($F_{(1,418)} = 26.04, p < .001$). Dies entspricht etwas mehr als einem Drittel einer Standardabweichung ($\beta = 0.37$). Bezüglich der weiteren Motivationsentwicklung konnten jedoch keine signifikanten Unterschiede zwischen Mädchen und Jungen festgestellt werden.

Gymnasiasten schätzten ihre Motivation über alle Messzeitpunkte hinweg um etwa $b = -0.30$ Punkte niedriger ein als Gesamtschüler, was in etwa einem Drittel einer Standardabweichung entspricht ($\beta = -0.35, F_{(1,18)} = 5.74, p < .05$). Bezüglich der weiteren Motivationsentwicklung konnten jedoch keine signifikanten Unterschiede zwischen Gymnasiasten und Gesamtschülern festgestellt werden.

Schultyp und Geschlechtszugehörigkeit wirken sich also in etwa gleichem Maß auf die Schülermotivation über alle Messzeitpunkte hinweg aus, tragen jedoch nicht dazu bei, Unterschiede in Entwicklungsverläufen der Motivation aufzuklären.

Das Lehrer-Engagement aus Schülersicht wirkt sich über alle Messzeitpunkte hinweg höchst signifikant auf die Schülermotivation aus ($F_{(1,418)} = 101.20, p < .001$). Weicht das subjektiv individuell bewertete Engagement des Lehrers um eine Standardabweichung nach oben ab, liegt die selbsteingeschätzte Motivation um $\beta = 0.41$ Standardabweichungen höher. Dies entspricht $b = 0.29$ Punkten auf einer Skala von 1 bis 6 Punkten. Allerdings sank die Motivation von Schülern, die das Lehrerengagement höher einstufen, auch signifikant stärker von der Prä- zur Follow-up-Messung (nicht jedoch von der Prä- zur Postmessung). Dabei handelt es sich jedoch nur um einen kleineren Effekt: weicht die Bewertung des Lehrerengagements um eine Standardabweichung nach oben ab, sinkt die Schülermotivation prä – follow-up um $\beta = -0.13$ Standardabweichungen. Dies entspricht einem Rückgang von $b = -0.09$ Punkten auf einer Skala von 1 bis 6 Punkten.

Die figural-logische Intelligenz spielt gemäß der Ergebnisse des Gesamtmodells nunmehr keine signifikante Rolle für die Motivationsentwicklung weder generell über alle Messzeitpunkte hinweg noch im weiteren Schuljahresverlauf.

2.3.9.7 Fünfte Hypothese: Vergleich inhaltlich aufeinander bezogener Studien

Abschließend wurden die Daten des Konzepttests mit den Daten der Studie von Scheid (2013) verglichen. Ziel war es zu analysieren, ob sich der in der Pilotstudie gefundene starke Effekt für die Überlegenheit der Thematisierung von Schülervorstellungen mit einer größeren Stichprobe, wenn auch unter variierten Randbedingungen, bestätigen lässt. Zudem wurde überprüft, ob die Thematisierung von Schülervorstellungen im Unterricht nicht nur unmittelbar nach der Intervention,

sondern auch mittelfristig (ca. zwei Monate später) das konzeptuelle Verständnis fördern kann.

Gemeinsamkeiten beider Studien: In beiden Stichproben wurde exakt das gleiche Testinstrument verwendet. Für die im Folgenden berichteten deskriptiven Statistiken und für das Mehrebenenmodell zur Prüfung signifikanter Unterschiede im Lernzuwachs wurde der Gesamtscore aus den gleichen Items ermittelt.

Je eine Treatment- und eine Kontrollgruppe wurden in beiden Studien je Schule vom gleichen Lehrer unterrichtet. Schüler beider Studien besuchten die 7. oder 8. Klassenstufe einer weiterführenden Schule im Bundesland Rheinland-Pfalz ($M = 13$ [Jahre], $SD = 8$ [Monate]). Lerninhalt in beiden Studien war die Bildentstehung bei der Sammellinse gemäß Lehrplan des Bundeslandes Rheinland-Pfalz im Fach Physik der Klassenstufen 7 bzw. 8.

In beiden Studien fanden drei Datenerhebungen statt: vor der Unterrichtsreihe, nach der Unterrichtsreihe, die in beiden Studien sechs bis sieben Schulstunden umfasste, sowie einige Wochen nach der Unterrichtsreihe (follow-up). Die zeitlichen Abstände unterscheiden sich lediglich zum dritten Messzeitpunkt, während die Schüler der Stichprobe „Ko“ sechs Wochen später getestet wurden, fand die Follow-up Erhebung der Stichprobe „SV“ acht bis neun Wochen später statt.

Unterschiede beider Studien: Während in dieser Studie („SV“) sowohl in der Treatment- als auch in der Kontrollgruppe Schülervorstellungen adressiert worden sind, lag der Schwerpunkt in der Studie („Ko“) von Scheid (2013) auf der Förderung der Kohärenz im Umgang mit Repräsentationen. Entsprechend spielte die Thematisierung von Schülervorstellungen keine explizite Rolle im Unterricht der Treatment- und der Kontrollbedingung. Eine ausführliche Darstellung des Unterrichts in der Treatment- und Kontrollbedingung kann in der Dissertationsschrift von Scheid (2013) nachgelesen werden. Die unterschiedlichen Werte des Medians des konzeptuellen Verständnisses vor der Intervention (siehe Boxplots, Abbildung 60) resultieren daraus, dass an der Studie „SV“ (dunkelgrau) auch Gesamtschüler teilnahmen, an der Vergleichsstichprobe „Ko“ (hellgrau) jedoch nur Gymnasiasten. Weitere Angaben zur Stichprobe der Studie „Ko“ finden sich in der Dissertationsschrift von Scheid (2013).

Vergleichbarkeit der Daten von Treatment- und Kontrollgruppe innerhalb der beiden Studien: Da innerhalb der Stichproben in keiner der beiden Studien signifikante Unterschiede zwischen Treatment- und Kontrollbedingung festzustellen waren, wurde keine weitere Differenzierung nach Bedingung innerhalb der Stichproben der Studien vorgenommen. So bestanden weder Unterschiede im Ausgangsniveau vor der Unterrichtsreihe noch im späteren Lernzuwachs zwischen Treatment- und Kontrollgruppe in der Stichprobe „SV“ (siehe Kapitel 2.3.9.4).

Zweite Hypothese: Wirkung des Treatments auf das konzeptuelle Verständnis) und auch keine Unterschiede im Ausgangsniveau vor der Unterrichtsreihe und im Lernzuwachs zwischen Treatment- und Kontrollgruppe in der Stichprobe „Ko“ (siehe Scheid, 2013, S. 173, 174). In den folgenden Analysen wurde also Unterricht unter Berücksichtigung von Schülervorstellungen („SV“) mit Unterricht verglichen, in dem das nicht der Fall war („Ko“).

Im Hinblick auf die deskriptiven Statistiken zeigt sich (vgl. Tabelle 60), dass Schüler in beiden Stichproben mit annähernd gleichen Kenntnissen starten, wobei die Stichprobe „SV“ im Schnitt einen Mittelwert, der um einen halben bzw. um einen ganzen Punkt auf einer Skala von 0 - 22 nach unten abweicht. Diese leicht unterschiedlichen Ausgangswerte zu Beginn der Studien resultieren vermutlich daraus, dass an der Studie „SV“ auch Gesamtschüler teilnahmen, an der Vergleichsstichprobe „Ko“ (Scheid, 2013) jedoch nur Gymnasiasten. Zu prüfen ist, ob dieser Anfangsunterschied signifikant ist, zumal bezüglich der Variablen Schultyp keine signifikanten Unterschiede für die Stichprobe „SV“ festgestellt werden konnten (siehe Kapitel 2.3.9.4 Zweite Hypothese: Wirkung des Treatments auf das konzeptuelle Verständnis).

Tabelle 60 Deskriptive Statistiken – Konzepttest Stichprobenvergleich ($N = 988$)

	Prätest ($N = 938$)		Posttest ($N = 889$)		Follow-up Test ($N = 926$)	
	„SV“ ^a ($n = 480$)	„Ko“ ^b ($n = 444$)	„SV“ ^a ($n = 472$)	„Ko“ ^b ($n = 384$)	„SV“ ^a ($n = 476$)	„Ko“ ^b ($n = 436$)
<i>M</i>	7.44	8.08	10.96	9.41	11.04	9.57
<i>(SD)</i>	3.23	3.01	5.01	3.68	4.75	3.54
<i>Range</i>	0 - 18.00	0 - 18.00	0 - 22.00	0 - 20.00	0 - 22.00	0 - 22.00

Anmerkung. Maximal erreichbare Punktzahl im Konzepttest: 22 Punkte.

^aSchülervorstellungen, ^bKohärenz

^cIm Stichprobenvergleich liegen von allen teilnehmenden Schüler vollständige Kovariaten vor, daher entspricht die berichtete Stichprobengröße der Datenbasis des folgenden Mehrebenenmodells. Die Schülerleistungen zählten zur Datenbasis, wenn das jeweilige Individuum zu mindestens einem der Messzeitpunkte den Test bearbeitete und alle Fachnoten bekannt waren.

Kovariaten vor, daher entspricht die berichtete Stichprobengröße der Datenbasis des folgenden Mehrebenenmodells. Die Schülerleistungen zählten zur Datenbasis, wenn das jeweilige Individuum zu mindestens einem der Messzeitpunkte den Test bearbeitete und alle Fachnoten bekannt waren.

An den Boxplots wird ersichtlich (vgl. Abbildung 60), dass Schüler beider Stichproben ihr konzeptuelles Verständnis prä – post und prä – follow-up ver-

bessern, wobei der Zuwachs der Stichprobe „SV“ den Zuwachs der Stichprobe „Ko“ übertrifft. Zum Zeitpunkt der Follow-up Messung bleibt das Wissensniveau in beiden Stichproben in etwa erhalten. Zudem zeigt sich besonders deutlich, dass die Berücksichtigung von Schülervorstellungen (Stichprobe „SV“) zu einem noch deutlicheren Lernzuwachs im konzeptuellen Verständnis führt.

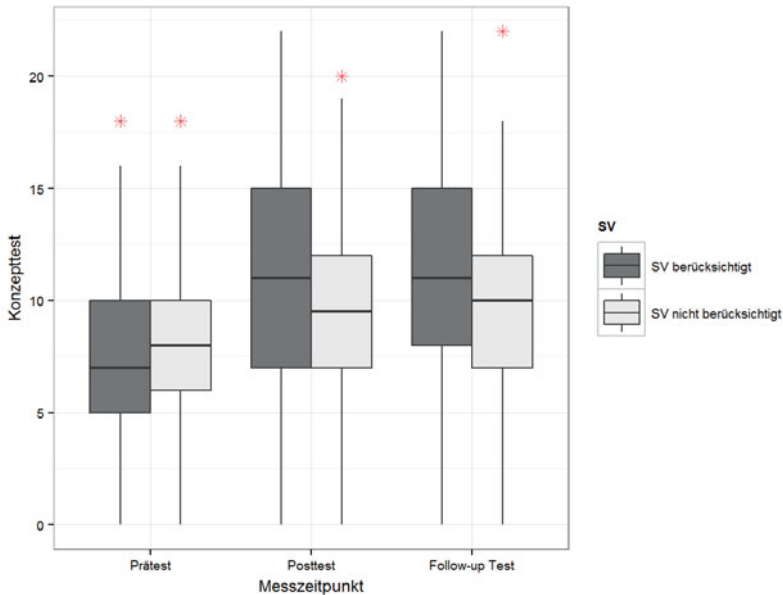


Abbildung 60: Boxplots zum grafischen Vergleich des konzeptuellen Verständnisses der beiden Stichproben je Messzeitpunkt

Auch der Verlaufplot (vgl. Abbildung 61) veranschaulicht, dass in beiden Bedingungen das konzeptuelle Verständnis steigt und sich auf einem höheren Niveau stabilisiert, die Mittelwerte der Stichprobe „SV“ liegen dabei zu den späteren Messzeitpunkten über den Werten der Stichprobe „Ko“ (SV nicht berücksichtigt). Während im unteren Bereich (unterhalb des Mittelwertes, hier angezeigt durch die Linien) Schüler beider Stichproben vertreten sind, scheinen in den oberen Bereichen Schüler der Stichprobe „SV“ zu überwiegen. Daher empfiehlt es sich ATI-Effekte zu untersuchen.

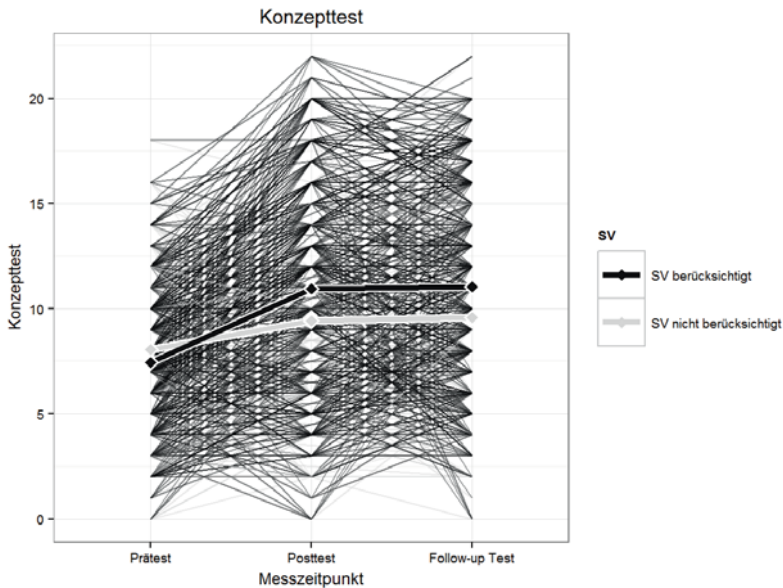


Abbildung 61: Verlaufsplot zum grafischen Vergleich des konzeptuellen Verständnisses der beiden Stichproben je Messzeitpunkt

Zu prüfen ist, ob diese Unterschiede im Lernzuwachs prä – post bzw. prä – follow-up signifikant sind: Analog zu den Mehrebenenmodellen zur Physikleistung und zum konzeptuellen Verständnis der Gesamtstichprobe „SV“ wurde ein Modell aufgestellt, das sowohl die Klassenstruktur, die Messwiederholung durch die Individuenkomponente als auch die Heteroskedastizität berücksichtigt und eine unstrukturierte Korrelationsmatrix der Fehlerterme im Individuum zugrunde legt.

Folgende Kovariaten fanden Eingang in das Modell, da hierzu in beiden Studien die gleichen Werte zur Verfügung standen: Schulnoten (letzte Zeugnisnote) in den Fächern Mathematik, Deutsch und Physik, Geschlecht und Klassengröße. Die Kovariaten kognitive Fähigkeiten (Intelligenzsubskalen) und Lehrerengagement aus Schülersicht (LES) entfielen wegen mangelnder Vergleichbarkeit in der Erhebung. Da an der Studie von Scheid (2013) keine Gesamtschüler teilnahmen, war es nicht sinnvoll die Variable Schultyp einzubeziehen.

Tabelle 61 Ergebnisse der Mehrebenenanalyse zum konzeptuellen Verständnis: Stichprobenvergleich „SV“ und „Ko“

Level (Stichprobengröße)	Erklärung	b	(SE)	β	(SE)	$F_{(numDF, denDF)}$	p
Level 1 (N = 2692)	Erläuterung						
Level 2 (N = 988)	Messzeitpunkte						
Level 3 (N = 37)	Individuen: Noten, Intelligenz, LES (= Lehrerengagement aus Schülersicht) Schulklassen: Bedingung Treatment (TG) vs. Kontrollgruppe (KG), Schultyp, Klassengröße						
Fixe Effekte							
<i>Variable</i>	<i>Erläuterung</i>						
Interzept	Durchschnittliches Verständnis: Stichprobe „Ko“, Geschlechtszugehörigkeit = weiblich, mit durchschnittlichen Noten	8.05	0.44	1.92	0.11	320.82 _(1,1.690)	< .001
Zuwachs (prä – post: t ₂ -t ₁)	Durchschnittlicher Verständniszuwachs: Stichprobe „Ko“, Geschlechtszugehörigkeit = weiblich, mit durchschnittlichen Noten	1.11	0.26	0.26	0.06	18.69 _(1,1.690)	< .001
Zuwachs (prä – follow-up: t ₃ -t ₁)	Durchschnittlicher Verständniszuwachs: Stichprobe „Ko“, Geschlechtszugehörigkeit = weiblich, mit durchschnittlichen Noten	1.79	0.24	0.42	0.06	53.68 _(1,1.690)	< .001
Bedingung = SV	Unterschied zwischen „SV“ und „Ko“ zu t ₁	-0.90	0.61	-0.21	0.15	2.19 _(1,34)	.148
Bedingung * Zuwachs prä – post	Zusätzlicher Einfluss Bedingung = „SV“ auf den Zuwachs zwischen t ₂ und t ₁	2.27	0.31	0.54	0.07	51.94 _(1,1.690)	< .001
Bedingung * Zuwachs prä – follow-up	Zusätzlicher Einfluss der Bedingung = „SV“ auf den Zuwachs zwischen t ₃ und t ₁	1.99	0.30	0.47	0.07	45.10 _(1,1.690)	< .001
Klassengröße	Genereller Einfluss auf Basis der Schätzung zu t ₁	0.01	0.05	0.00	0.01	0.15 _(1,34)	.907
Klassengröße * Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t ₂ und t ₁	-0.01	0.02	-0.01	0.01	0.30 _(1,1.690)	.699
Klassengröße * Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t ₃ und t ₁	-0.01	0.02	-0.01	0.01	0.66 _(1,1.690)	.585
Geschlecht * Ausgangswert	Genereller Einfluss Geschlechtszugehörigkeit = männlich auf Basis der Schätzung zu t ₁	0.19	0.23	0.04	0.06	0.36 _(1,947)	.417
Geschlecht * Posttest	Zusätzlicher Einfluss Geschlechtszugehörigkeit = männlich auf den Zuwachs zwischen t ₂ und t ₁	0.17	0.27	0.04	0.07	0.30 _(1,1.690)	.551
Geschlecht * Follow-up Test	Zusätzlicher Einfluss Geschlechtszugehörigkeit = männlich auf den Zuwachs zwischen t ₃ und t ₁	-0.35	0.26	-0.08	0.06	1.85 _(1,1.690)	.174
Mathematiknote	Genereller Einfluss auf Basis der Schätzung zu t ₁	-0.31	0.13	-0.07	0.03	6.19 _(1,947)	.013

Mathe-Note * Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t ₂ und t ₁	-0.10	0.18	-0.02	0.04	0.31 _(1,1690)	.575
Mathe-Note * Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t ₃ und t ₁	-0.39	0.17	-0.09	0.04	5.42 _(1,1690)	.020
Deutschnote (D, M, PH)	Genereller Einfluss auf Basis der Schätzung zu t ₁	-0.13	0.13	-0.03	0.03	0.87 _(1,947)	.351
Deutschnote * Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t ₂ und t ₁	-0.17	0.19	-0.04	0.04	0.78 _(1,1690)	.376
Deutschnote* Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t ₃ und t ₁	-0.12	0.18	-0.24	0.04	0.45 _(1,1690)	.501
Physiknote (D, M, PH)	Genereller Einfluss auf Basis der Schätzung zu t ₁	-0.11	0.12	-0.03	0.03	1.00 _(1,947)	.319
Physiknote * Posttest	Zusätzlicher Einfluss auf den Zuwachs zwischen t ₂ und t ₁	-0.83	0.16	-0.21	0.04	26.23 _(1,1690)	< .001
Physiknote * Follow-up Test	Zusätzlicher Einfluss auf den Zuwachs zwischen t ₃ und t ₁	-0.51	0.14	-0.13	0.04	11.95 _(1,1690)	< .001

Zufällige Effekte

σ_k (Klasse)	1.50
σ_B (Individuum)	1.70
σ_E (individuumsspezifisch)	2.27
ρ (Messzeitpunkt)	(ρ_{12}) -0.05
g (Messzeitpunkt)	(g_1) 1.00

Modellvergleich

	Berichtetes Modell	Leermode ^{ll} ^a
Devianz	13883.94	14136.72
df	29	11
AIC	13941.94	14158.72
BIC	14112.99	14223.60
LR-Test (berichtetes Modell versus Leermode ^{ll})	45.89, p < .001	
Un erklärte Varianz Level 1	(t ₁) 10.30	(t ₂) 14.70
Un erklärte Varianz Level 2	13.41	(t ₁) 11.27
Un erklärte Varianz Level 3	2.69	15.29
		3.24
		(t ₂) 17.63
		(t ₃) 16.95

(ρ_{13}) -0.03
(g_2) 1.40
(ρ_{23}) 0.27
(g_3) 1.36

Erklärte Varianz des berichteten Modells im Vergleich zum Leermodell

R^2 Level 1	(t ₁) 0.09	(t ₂) 0.14	(t ₃) 0.13
R^2 Level 2	0.12		
R^2 Level 3	0.17		

„Wirkung“ des Treatments

Anteil aufgeklärter Varianz^b

R^2 Level 1	(t ₁) 0.01	(t ₂) 0.02	(t ₃) 0.01
R^2 Level 2	0.01		
R^2 Level 3	< 0.01		

Effektstärke

$\Delta_{(pre-post)}$	= 0.63
$\Delta_{(pre-follow-up)}$	= 0.56

^a Das Leermodell unterscheidet sich vom berichteten Modell darin, dass alle fixen Effekte außer dem Interzept und den Dummy-Variablen t₂ und t₃ fehlen. Das Modell sagt für jeden Schüler voraus, dass seine Leistung zum jeweiligen Messzeitpunkt dem mittleren Konzeptuellen Verständnis der Stichprobe entspricht - unter Berücksichtigung der individuellen- und klassenspezifischen Unterschiede. Man beachte, das Leermodell schätzt folgende 11 Parameter (df = 11): Interzept, t₂, β_2 , β_3 , ρ_{23} , σ_{ϵ} , σ_{ϵ_2} , σ_{ϵ_3} .
^b Der Anteil aufgeklärter Varianz errechnet sich wie in Tabelle 52 angegeben nach Snijder und Bosker (1999, S. 99 ff).
 Für das harmonische Mittel der Klassengrößen ergibt sich ($h = 26.47$).

Die Ergebnisse der Mehrebenenanalyse zeigen (vgl. Tabelle 61), dass die auf Basis der deskriptiven Statistiken zu vermutenden Mittelwertsunterschiede bei den Startbedingungen zwischen den Stichproben „SV“ und „Ko“ nicht signifikant sind ($F_{(1,34)} = 2.19$, n.s.). Die Null-Hypothese, dass Schüler in beiden Bedingungen mit dem gleichen Ausgangsniveau starten, wird beibehalten.

Der Vergleich des Lernzuwachses ergibt, dass es im weiteren Entwicklungsverlauf zu höchst signifikanten Unterschieden kommt. So erreichten Schüler, welche Unterricht besucht hatten, der gezielt auf Schülervorstellungen einging (Stichprobe „SV“), einen Lernzuwachs prä – post, der durchschnittlich um $b = 2.06$ Punkte höher lag und einen Lernzuwachs prä – follow-up, der um $b = 2.27$ Punkte den Lernzuwachs der Schüler in der Stichprobe „Ko“ übertraf. Dies entspricht einem Lernzuwachs prä – post ($\beta_{\text{post}} = 0.54$), der in etwa eine halbe Standardabweichung über dem Lernzuwachs der Stichprobe „Ko“ ($F_{(1,1690)} = 51.94$, $p < .001$) lag und einem Zuwachs prä – follow-up, der ebenfalls knapp um eine halbe Standardabweichung ($\beta_{\text{follow-up}} = 0.47$) den Zuwachs der Stichprobe „Ko“ ($F_{(1,1690)} = 45.10$, $p < .001$) übertraf. Es ergibt sich nach Bortz und Döring (2005) ein relevanter mittlerer Effekt von $\Delta_{(\text{prä} - \text{post})} = 0.63$ und $\Delta_{(\text{prä} - \text{follow-up})} = 0.56$.

Von den analysierten Kovariaten erwiesen sich die Mathematik- und die Physiknote als signifikante Variablen zur Erklärung des Abschneidens im Konzepttest. Schüler, deren Mathematiknote um eine volle Notenstufe besser liegt, erzielten über alle Messzeitpunkte hinweg ein Testergebnis, das um einen Drittel Punkt höher lag ($b = -0.31$, $\beta = -0.07$, $F_{(1,947)} = 6.19$, $p < .05$) als das Ergebnis ihrer Altersgenossen. Schüler, deren Mathematiknote um eine Notenstufe besser ist, erzielten darüber hinaus im Schnitt noch zusätzlich einen Lernzuwachs prä – follow-up, der ebenfalls um $b = 0.39$ Punkte höher lag, was in etwa einem Zehntel einer Standardabweichung entspricht ($\beta = -0.09$) als das Ergebnis ihrer Altersgenossen ($F_{(1,1690)} = 5.42$, $p < .05$). Schüler, deren Physiknote um eine volle Notenstufe besser liegt, erzielten im Schnitt zusätzlich einen Lernzuwachs prä – post, der um knapp einen Punkt höher liegt ($b = -0.83$, $\beta = -0.21$, $F_{(1,1690)} = 26.23$, $p < .001$) und einen Lernzuwachs prä – follow-up, der einen halben Punkt höher liegt als das Ergebnis ihrer Altersgenossen ($b = -0.51$, $\beta = -0.13$, $F_{(1,1690)} = 11.95$, $p < .01$). Die Leistungen im Fach Deutsch, die Klassengröße und die Geschlechtszugehörigkeit spielen gemäß der Ergebnisse des Modells (vgl. Tabelle 61) keine signifikante Rolle.

2.3.9.8 Vertiefende Analysen zu Aptitude-Treatment-Interaktionen

Abschließend wurde untersucht, ob die gewählte Lehr-Lernmethode (Treatment-versus Kontrollbedingung) mit ausgewählten Schülermerkmalen interagiere. Solche Wechselwirkungs-Effekte zwischen Begabungen bzw. Fähigkeiten und Lehrmethoden werden in der pädagogischen Psychologie unter dem Fachbegriff der Aptitude-Treatment-Interaktion diskutiert (vgl. Cronbach & Snow, 1977; Hasebrook, 2006).

Insbesondere wurde der Frage nachgegangen, ob Schüler mit höheren kognitiven Fähigkeiten oder besseren Noten vom Unterricht der Treatmentbedingung in stärkerem Ausmaß profitierten. Neben möglichen ATI-Effekten wurde zusätzlich auch die Interaktion zwischen Schultyp und Bedingung sowie Geschlechtszugehörigkeit und Bedingung untersucht. Hierzu wurde ein Mehrebenenmodell aufgestellt, welches Interaktionsterme zwischen Bedingung und dem jeweiligen Schülermerkmal zu den verschiedenen Messzeitpunkten enthielt.

Zur Analyse möglicher ATI-Effekt zwischen Schülermerkmalen und Treatment bezüglich der Physikleistung bei repräsentationalen Aufgaben und des konzeptuellen Verständnisses wurde folgenden Fragen nachgegangen:

- Profitieren Schüler mit besseren Schulleistungen in höherem Maß vom Treatment als Schüler mit schlechteren Schulleistungen?
- Profitieren Schüler mit besseren Testergebnissen im verbalen, im figural-räumlichen oder im figural-logischen Schlussfolgern deutlicher vom Treatment als Schüler mit geringeren Testergebnissen in diesen Bereichen?

Weitere Wechselwirkungen:

- Profitieren Gymnasiasten deutlicher von der Treatmentbedingung als Gesamtschüler?
- Profitieren Jungen in höherem Maß vom Treatment als Mädchen oder umgekehrt?

Auf Basis der Ergebnisse zu den Kovariaten des Mehrebenenmodells zur Physikleistung könnten Interaktionseffekte zwischen Schulleistung und Treatment sowie Schultyp und Treatment bestehen. So erwiesen sich Schulleistungen und Schultyp als höchst signifikante Merkmale zur Vorhersage der Physikleistung und des Lernzuwachses.

Für die verwendeten Subskalen des I-S-T 2000 R ist ein Interaktionseffekt am wahrscheinlichsten zwischen Treatment und figural-räumlicher Intelligenz zu erwarten, da sich das figural-räumliche Schlussfolgern signifikant auf den Lernzuwachs prä – post auswirkte. Eine Interaktion zwischen verbalem respektive figural-logischen Schlussfolgern und Treatment ist weniger wahrscheinlich, da das verbale Schlussfolgern keinen signifikanten Erklärungswert für den Lernzuwachs prä – post bzw. prä – follow-up aufwies und sich für das figural-logische Schlussfolgern über alle Messzeitpunkte hinweg kein signifikanter Erklärungswert im Gesamtmodell zeigte.

Die Geschlechtszugehörigkeit erwies sich in den bisherigen Ergebnissen (siehe Kapitel 2.3.9.3 Erste Hypothese: Wirkung des Treatments auf die Physikleistung) nicht als signifikante Variable zur Vorhersage der Physikleistung. So schnitten Jungen zwar etwas besser ab als Mädchen und erzielten auch leicht höhere Lernzuwächse, diese Effekte waren aber jeweils nicht signifikant. Um der Vollständigkeit willen wurde jedoch auch die Interaktion zwischen Geschlechtszugehörigkeit und Treatment geprüft.

Im Ergebnis (siehe Tabelle 22 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com) konnte lediglich zwischen vorherigen Schulleistungen (Noten) und Bedingung ein signifikanter Interaktionseffekt bezüglich des Lernzuwachses festgestellt werden. „Gute Schüler“ (d.h. Schüler mit besseren Noten) in der Treatmentbedingung erzielten einen höheren Lernzuwachs prä – follow-up als „gute Schüler“ in der Kontrollbedingung ($b = -1.45, \beta = -0.22, F_{(1,800)} = 10.03, p < .01$). Für den Lernzuwachs prä – follow-up bestand jedoch kein ATI-Effekt. Zwischen den übrigen Merkmalen: Intelligenz, Schultyp, Geschlechtszugehörigkeit und Treatment waren zu keinem Zeitpunkt ATI-Effekte nachweisbar.

Im Hinblick auf die Entwicklung des konzeptuellen Verständnisses wurden ebenso wie für die Entwicklung der Physikleistung bei repräsentationsbezogenen Aufgaben die oben genannten Fragen untersucht.

Im Ergebnis (siehe Tabelle 23 Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com) wurde lediglich zwischen Geschlechtszugehörigkeit und Bedingung ein signifikanter Interaktionseffekt bezüglich des Lernzuwachses prä – follow-up festgestellt. So erzielten Mädchen in der Treatmentbedingung einen signifikant höheren Lernzuwachs im konzeptuellen Verständnis prä – follow-up als Jungen in der Kontrollbedingung ($b = -1.94, \beta = -0.42, F_{(1,788)} = 5.77, p < .05$). Da Jungen jedoch über alle Messzeitpunkte hinweg einen etwas höheren (wenn auch nicht signifikant höheren) Lernzuwachs prä – follow-up erreichten ($b = 0.76, \beta = 0.16, F_{(1,788)} = 1.63, n.s.$), kann nicht die Schlussfolgerung gezogen werden, dass Mädchen in der Treatmentbedingung in der Summe von einem signifikant

höheren Lernzuwachs profitierten. Die Interaktion zwischen Geschlechtszugehörigkeit und Bedingung zeigte sich auch nicht für den Lernzuwachs prä – post. Zwischen den übrigen Merkmalen: Intelligenz, Schulnoten und Schultyp waren zu keinem Zeitpunkt ATI-Effekte nachweisbar.

Auch für die abhängige Variable Schülermotivation wurden mögliche Interaktions-Effekte untersucht. Im Gesamtmodell hatten sich die Fachnoten, die Geschlechtszugehörigkeit, der Schultyp und das Lehrerengagement aus Schülersicht (LES) als relevante Kovariaten erwiesen. Auf Basis dieser Ergebnisse wurde folgenden Fragen nachgegangen:

- Unterscheidet sich die motivationale Entwicklung im Schuljahresverlauf von Gymnasiasten in der Treatmentbedingung von der Entwicklung der Gesamtschüler in der Treatmentbedingung?
- Steigern Schüler mit besseren Schulleistungen (guten Noten) in der Treatmentbedingung ihre Motivation deutlicher als Schüler mit besseren Schulleistungen (guten Noten) in der Kontrollbedingung?
- Unterscheidet sich die motivationale Entwicklung im Schuljahresverlauf von Schülern mit besseren Testergebnissen im figural-logischen Schlussfolgern in der Treatmentgruppe von Schüler mit geringeren Testergebnissen in diesem Bereich?
- Unterscheidet sich die motivationale Entwicklung im Schuljahresverlauf von Jungen in der Treatmentbedingung von der Entwicklung der Mädchen in der Treatmentbedingung?
- Unterscheidet sich die motivationale Entwicklung im Schuljahresverlauf von Schülern in der Treatmentbedingung, die ihren Lehrer als engagierter wahrnehmen, von Schülern in der Kontrollbedingung?

Im Ergebnis (siehe Tabelle 24 Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com) konnte lediglich zwischen LES und Bedingung ein signifikanter Interaktionseffekt bezüglich des Motivationsverlaufs prä – follow-up festgestellt werden. Die Motivation von Schülern in der Treatmentbedingung, die ihren Lehrer als engagierter wahrnehmen, sank vom Prä- zum Follow-up Test stärker als die Motivation von Schülern in der Kontrollbedingung ($b = -0.14$, $\beta = -0.21$, $F_{(1,806)} = 7.96$, $p < .01$).

Interessanterweise ergab sich kein genereller Interaktionseffekt (über alle Messzeitpunkte hinweg) zwischen Bedingung und LES. Auch für die Entwicklung prä – post ist kein Interaktionseffekt nachweisbar. Eine Interaktion zwischen Treatment und LES ist also im Anschluss an die Unterrichtsreihe nicht zu beobachten.

Offenbar werden jedoch Schüler in der Treatmentbedingung, die ihren Lehrer als motivierter wahrnehmen, mittelfristig (prä – follow-up) leicht demotiviert: Liegt die anfänglich wahrgenommene Einschätzung des Lehrerengagements eine Standardabweichung über dem Mittelwert der Gesamtstichprobe, kommt es zu einem mittelfristig nachweisbaren Motivationsverlust der einem Fünftel einer Standardabweichung entspricht. Dies drückt sich in einem Verlust von $b = -0.14$ Punkten auf einer Skala von 1 bis 6 aus.

Zwischen den übrigen Merkmalen: Schulleistungen (Fachnoten), Schultyp, figural-logische Intelligenz, Geschlechtszugehörigkeit und Treatment waren zu keinem Zeitpunkt Interaktions-Effekte in Bezug auf die Motivation im Schuljahresverlauf nachweisbar.

Abschließend wurde auch für den Stichprobenvergleich der Frage nachgegangen, ob zwischen Fachleistungen bzw. Begabungen eine Aptitude-Treatment-Interaktion vorliegen könnte. Da in den beiden Studien der Intelligenztest nicht identisch war, wurden die jeweiligen Fachnoten in Mathematik, Deutsch und Physik herangezogen und für jede einzelne Fachnote der jeweilige Interaktionsterm überprüft. Im Ergebnis (siehe Tabelle 25 in Anhang C9 auf der Produktseite zu diesem Buch unter www.springer.com) war jedoch zu keinem der Messzeitpunkt ein ATI-Effekt nachweisbar.

2.3.9.9 Zusammenfassung der Ergebnisse zur Untersuchung der Hypothesen 1-5

Abbildung 62 veranschaulicht die Ergebnisse zu den Hypothesen 1-4. Sie fasst alle untersuchten signifikanten Einflussfaktoren auf die drei abhängigen Variablen zusammen. Die Darstellung ist nicht mit einem Strukturgleichungsmodell zu verwechseln. Die verwendeten Pfeile beziehen sich auf diejenigen standardisierten Regressionsgewichte (β), welche signifikante Prädiktoren für das jeweilige Kriterium darstellen: (1) Wissen und Problemlösen beim Umgang mit fachspezifischen Repräsentationen, erfasst im Leistungstest (LT), (2) das konzeptuelle Verständnis, erfasst im Konzepttest (KT) und (3) die Schülermotivation, erhoben im Motivationsfragebogen (Mo).

Wie sich an Abbildung 62 ablesen lässt, ist die Bedingung (Treatment- versus Kontrollgruppe) lediglich für die Teilstichprobe der Gymnasiasten (hier dargestellt durch einen grauen Pfeil) ein signifikanter jedoch eher schwacher Prädiktor mit einer im Vergleich zu den übrigen Prädiktoren eher niedrigen Ausprägung von $\beta = 0.16, p < .05$.

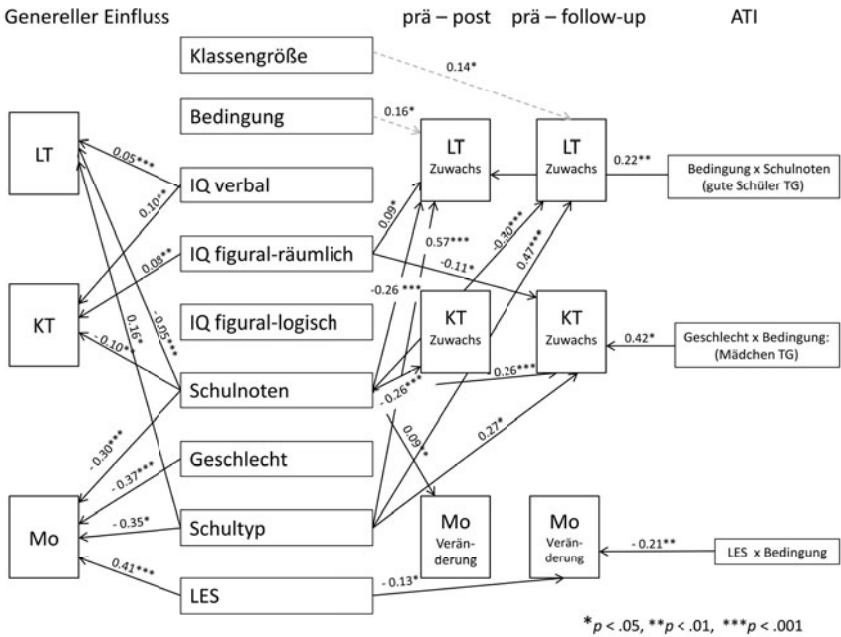


Abbildung 62: Zusammenschau relevanter Prädiktoren und Aptitude-Treatment-Interaktionen (ATI) in Bezug auf die Test- bzw. Fragebogenergebnisse im Physikleistungstest (LT), Konzepttest (KT) und die Schülermotivation (Mo). Gestrichelte Pfeile beziehen sich auf Prädiktoren, die lediglich in der Teilstichprobe der Gymnasiasten relevant waren.

Von den übrigen Variablen sind für die Vorhersage der leistungsbezogenen abhängigen Variablen Physikleistung (LT) und konzeptuelles Verständnis (KT) vor allem die Fachnoten aussagekräftig. Verbales und figural-räumliches Schlussfolgern sowie der Schultyp spielen ebenfalls eine Rolle. In Bezug auf die Motivation sind die Geschlechtszugehörigkeit, die Fachnoten und das Lehrerengagement aus Schülersicht relevant. Klassengröße und figural-logisches Schlussfolgern spielen eine untergeordnete oder überhaupt keine Rolle.

Die Ergebnisse zum Vergleich der inhaltlich aufeinander bezogenen Studien „SV“ versus „Ko“ (Scheid, 2013) – Hypothese 5 – sind in Abbildung 63 zusammenfassend dargestellt.

Es zeigt sich, dass die Bedingung (wurden Schülervorstellungen thematisiert oder nicht) in vergleichsweise hohem Maß entscheidend für die Erklärung des Zuwachses des konzeptuellen Verständnisses ist.

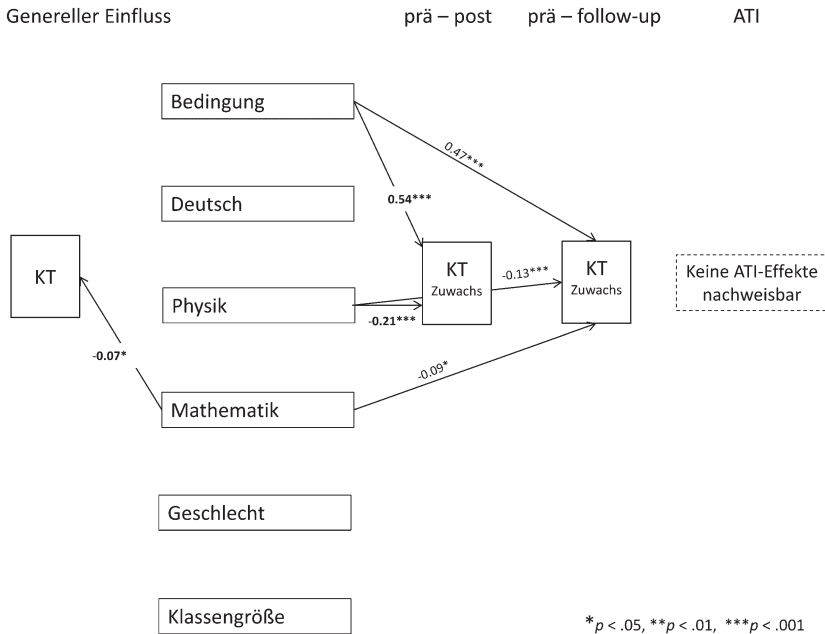


Abbildung 63: Zusammenschau relevanter Prädiktoren für den Vergleich inhaltlich aufeinander bezogener Studien „SV“ versus „Ko“ in Bezug auf die Test- bzw. Fragebogenergebnisse im Konzepttest (KT)

Von den Kovariaten ist (erwartungsgemäß) die Physiknote ein aussagekräftiger Prädiktor, die Mathematikvorleistung (ebenfalls erfasst durch die letzte Zeugnisnote) spielt ebenfalls eine signifikante jedoch weniger ausgeprägte Rolle. Die Deutschvorleistung, die Klassengröße und Geschlechtszugehörigkeit sind zu keinem Zeitpunkt signifikante Einflussfaktoren für das Abschneiden im Konzepttest.

2.3.9.10 Diskussion der Ergebnisse zur Untersuchung der Hypothesen 1-5

Im Rahmen der Hauptstudie wurden fünf Hypothesen untersucht, welche sich auf die drei abhängigen Variablen beziehen: Wissen und Problemlösen bei repräsentationsbezogenen Aufgaben, das konzeptuelle Verständnis in der Strahlenoptik und die Motivation der Schüler im Physikunterricht.

Hypothese 1: Die Treatmentgruppe erreicht im Wissen und Problemlösen beim Umgang mit fachspezifischen Repräsentationen im Bereich der Strahlenoptik unmittelbar nach der Intervention und ca. zwei Monate später einen höheren Lernzuwachs als die Kontrollgruppe.

In dem Mehrebenenmodell zur Physikleistung bei repräsentationsbezogenen Aufgaben zeigte sich kein signifikanter Unterschied zwischen Treatment- und Kontrollgruppe für die Gesamtstichprobe. Die Verlaufsdigramme veranschaulichen, dass sich Wissen und Problemlösen in beiden Gruppen parallel entwickelten. Die erste Hypothese kann daher nicht bestätigt werden.

In Bezug auf die Kovariaten fielen die heterogenen Ergebnisse bezüglich der kognitiven Fähigkeiten auf, so erwies sich die verbale Intelligenz über alle Messzeitpunkte hinweg als relevant, wirkt sich aber nicht auf den Lernzuwachs aus. Die figural-räumliche Intelligenz beeinflusst signifikant nur den Lernzuwachs prä – post. Beides lässt sich plausibel erklären, so mag ein gutes Verständnis des Aufgabentextes Schüler prinzipiell bei der Bearbeitung unterstützen. Für die Erklärung des Lernzuwachses zu den beiden späteren Messzeitpunkten ist diese Fähigkeit wenig relevant. Hier entscheidet vermutlich ein gutes räumliches Vorstellungsvermögen darüber, ob fachlich adäquate mentale Modelle gebildet werden.

Der Einfluss der Klassengröße bei Gymnasiasten auf den Lernzuwachs prä – follow-up wird als Artefakt gewertet (Schüler in größeren Klassen schneiden besser ab). Dafür spricht, dass weder ein Einfluss der Klassengröße über alle Messzeitpunkte hinweg noch ein signifikanter Einfluss der Klassengröße auf den Lernzuwachs prä – post festgestellt werden konnte. Jedoch erzielten Schüler in größeren Klassen an Gymnasien einen höheren Lernzuwachs prä – follow-up als Schüler in kleineren Klassen. Vermutlich kovariieren hier Einflussfaktoren auf Klassen- und Lehrerebene systematisch mit Variablen, die dafür sorgen, dass Schüler aus größeren Klassen weniger Lerninhalte vergessen als Schüler aus kleineren Klassen, z.B. könnte es in „besseren“ Gymnasien mehr Schüler-Anmeldungen gegeben haben. In allen übrigen Modellen spielt die Klassengröße keine Rolle. Hier bestätigt sich das aus der Bildungsforschung bekannte Ergebnis, dass kleine Klassen für den Lernerfolg nicht von signifikantem Vorteil sind (vgl. auch Ergebnisse der dritten PISA-Studie 2006, PISA-Konsortium Deutschland, 2007).

Unterschiedliche Ergebnisse zeigten sich für die Teilstichprobe der Gymnasiasten, die zu allen Messzeitpunkten im Schnitt bessere Ergebnisse erzielten als die Gesamtschüler. Hier ergaben sich unmittelbar nach der Intervention signifikante Vorteile zu Gunsten der Treatmentgruppe, welche jedoch mittelfristig nicht bestehen blieben. Die resultierende Effektstärke war jedoch gering, so dass

dieser Befund für sich genommen für die Unterrichtspraxis zunächst wenig bedeutsam erscheint.

Insgesamt wird die erste Hypothese – mit Ausnahme des Befundes der signifikanten Vorteile der Treatmentgruppe an Gymnasien – nicht unterstützt. Entweder spielt das Ausmaß an kognitiver Aktivierung keine Rolle für die untersuchten Lernprozesse oder die Unterschiede im Ausmaß der kognitiven Aktivierung waren nicht klar genug operationalisiert. Die zuletzt genannte These wird von den Einschätzungen der Lehrkräfte (Zustimmung zur Frage: „Ich schätze, der Unterricht hat die Schüler zum Nachdenken angeregt“) gestützt. So zeigten sich im Mittel für beide Gruppen gleich hohe Werte in der Zustimmung zu dem Item, obgleich sich geringere Maße der Übereinstimmung für die Bewertung des Unterrichts in der Kontrollgruppe ergaben. Wäre der Unterschied deutlicher zum Tragen gekommen, hätte die Zustimmung zu dem Item in der Treatmentgruppe signifikant höher ausfallen müssen als in der Kontrollgruppe.

Hypothese 2: Die Treatmentgruppe erreicht im konzeptuellen Verständnis für den Bereich der Strahlenoptik unmittelbar nach der Intervention und ca. zwei Monate später einen höheren Lernzuwachs als die Kontrollgruppe

In den Mehrebenenmodellen zur Analyse des konzeptuellen Verständnisses zeigten sich ebenfalls keine signifikanten Unterschiede zwischen den Bedingungen TG und KG.

Im Gegensatz zum Physikleistungstest ergaben sich zwischen dem Schultyp Gymnasium versus IGS kein bzw. ein geringer Vorteil der Gymnasiasten für den Vergleich prä – follow-up. Dieser Befund kann als Hinweis gewertet werden, dass das Abschneiden im Konzepttest weniger von der generellen Leistungsfähigkeit beeinflusst wird, welche bei den Gymnasien durchschnittlich höher ausgeprägt sein dürfte: Mehrere Indikatoren (höhere Werte in den drei Subskalen des I-S-T 2000 R und generell bessere Fachnoten) weisen auf eine höhere Leistungsfähigkeit der Gymnasiasten hin.

Im Resümee konnte die zweite Hypothese nicht bestätigt werden. Da in beiden Bedingungen Schülervorstellungen thematisiert wurden, war vermutlich hier ebenso wie für den Leistungstest das Ausmaß an kognitiver Aktivierung entweder wenig entscheidend oder nicht deutlich genug operationalisiert.

Hypothese 3: Ein Teil der Treatmentgruppe, die ein variiertes Treatment erhielt, welches die Predict-Observe-Explain-Strategie (POE-Sequenz) unter der Perspektive des Lernens mit multiplen Repräsentationen beinhaltet, erreicht im Vergleich zu dem regulären Treatment und im Vergleich zur Kontrollgruppe einen höheren Lernzuwachs im konzeptuellen Verständnis unmittelbar nach der Intervention und ca. zwei Monate später.

Ein positiver Effekt des Einsatzes der POE-Sequenz auf den Lernerfolg der Schüler wurde nur sehr eingeschränkt nachgewiesen (Analyse der dritten Hypothese: Vergleich des regulären mit dem variierten Treatment), obgleich etliche andere Studien deutliche signifikante Effekte berichten (vgl. White & Gunstone, 1992; Kearney et al., 2001; Crouch et al., 2004 u.a.).

Möglicherweise war jedoch die relative Lernzeit innerhalb der Unterrichtssequenz zu gering, da die jeweiligen POE-Sequenzen nur in zwei von sechs Stunden und auch nur innerhalb eines Teils der Schulstunden zum Einsatz kamen.

Hypothese 4: Die Entwicklung der Motivation der Schüler in der Treatmentgruppe unterscheidet sich zu keinem der drei Messzeitpunkte von der Entwicklung in der Kontrollgruppe.

In Bezug auf die Entwicklung der Motivation waren keine Unterschiede feststellbar (Ablehnung der Unterschiedshypothese). Die intendierte geringere kognitive Aktivierung in der Kontrollgruppe wirkt sich somit nicht negativ auf die Motivation der Schüler aus. Für die Treatmentgruppe ergibt sich umgekehrt auch keine Verbesserung der Motivation (prä – post). Die generelle Abnahme der Motivation im Schuljahresverlauf wird zwar als bedauerlich, jedoch auch als wenig problematisch bewertet, da sie etlichen Ergebnissen zum Motivationsverlust im Schuljahresverlauf und in der weiteren Schulkarriere entspricht, welche aus der Bildungsforschung bekannt sind (vgl. Hidi, 2002; Wild & Hofer, 2000; Krapp, 2002; zitiert nach Schiefele, 2009, S. 171).

Hypothese 5: Schüler, die ein Treatment zur Förderung des konzeptuellen Verständnisses erhielten (TG und KG dieser Studie), erzielten einen höheren Lernzuwachs in Bezug auf das konzeptuelle Verständnis als Schüler, die ein Treatment zur Förderung der Kohärenz (Scheid, 2013) erhielten bzw. der Kontrollgruppe der entsprechenden Stichprobe angehörten.

Der Vergleich der beiden Studien innerhalb des Projekts zeigt, dass repräsentationsbezogene Aufgaben, welche Schülervorstellungen adressieren, zu einer signifikanten Verbesserung des konzeptuellen Verständnisses führen, die mittelfristig (zwei Monate später) erhalten bleibt (Bekräftigung der Ergebnisse der Pilotstudie). Die fünfte Hypothese wird bestätigt. Der resultierende Effekt wird als relevanter, mittlerer Effekt bewertet. Im Vergleich zu den Ergebnissen der Pilotstudie ergibt sich eine Relativierung des Befundes eines starken Effekts, der darauf zurückgeführt werden könnte, dass auch die Schüler in der Vergleichsstichprobe deutlicher von der Intervention profitieren, was sich in den Lernzuwächsen der Stichprobe „Ko“ zeigt.

Vertiefende Analysen zu ATI-Effekten: Obgleich die Verlaufsdigramme das Bestehen von ATI-Effekten zwischen Treatment und Schultyp nahe legen, konnte

für keine der abhängigen Variablen ATI-Effekte gefunden werden. Offenbar wirkt das Treatment in beiden Schularten in vergleichbarer Weise, wobei sich das Leistungsniveau der Gesamtschüler im Mittel auf einem geringeren Niveau befinden. Bemerkenswerte ATI-Effekte bestanden lediglich zwischen Physikleistung und Schulnoten. Offenbar profitieren leistungsstärkere Schüler in höherem Maß von kognitiv aktivierenden Aufgaben im Umgang mit Repräsentationen als leistungsschwächere.

Eine mögliche Erklärung könnte darin bestehen, dass die aktive Auseinandersetzung mit fachspezifischen Repräsentationen ein bestimmtes Maß an Anstrengungsbereitschaft (und fachlichen Fähigkeiten) erfordert. Ist hier ein möglicher Schwellenwert überschritten, kommt es zur Bildung adäquater mentaler Modelle, welche zu einem höheren Niveau beim Wissen und Problemlösen führt.

Etwaige Effekte zwischen Schultyp und Schulnoten wurden vermutlich durch das höhere Niveau der Fachnoten an Gymnasien erklärt.

Die gefundene Interaktion zwischen LES und Treatment prä – post in Bezug auf die Motivation, sollte nicht überbewertet werden, da es sich um einen minimalen Unterschied handelt. Der positive Effekt für den Lernzuwachs des konzeptuellen Verständnisses der Mädchen, hebt sich, wie zuvor dargestellt, in der Summe auf.

3 Gesamtdiskussion

3.1 Integration der Befunde und Kritik am Vorgehen

Die hier vorgestellte Studie weist darauf hin, dass der Einsatz anspruchsvoller Lernstrategien beim Umgang mit Repräsentationen (kognitive Aktivierung) zu einer signifikanten, wenn auch geringen Verbesserung der Physikleistung von Gymnasiasten in einem unmittelbar nachfolgenden Test führt, nicht jedoch zu einem höheren Lernzuwachs bezüglich des konzeptuellen Verständnisses oder zu einer Veränderung der Motivation.

Der Vergleich der beiden Studien innerhalb des Projekts zeigt darüber hinaus, dass Unterricht mit repräsentationsbezogene Aufgaben, welche Schülervorstellungen adressieren, zu einem signifikant höheren Lernerfolg bezüglich des konzeptuellen Verständnisses führt als Unterricht ohne Thematisierung von Schülervorstellungen. Dieser höhere Lernerfolg bleibt nachweislich mittelfristig erhalten.

Auch wenn die Analyse der Lernprozesse auf Basis einer Teilstichprobe wertvolle Hinweise liefert, dass der Umgang mit multiplen Repräsentationen entscheidend für den Lernerfolg ist, bleibt unklar, ob die geringen Effekte des Treatments darauf zurückzuführen sind, dass der Ansatz des Lernens mit multiplen Repräsentationen durch kognitiv aktivierende Aufgaben nur einen geringen zusätzlichen Lernerfolg bringt, oder darauf, dass sich die Interventionsbedingungen zu wenig unterscheiden haben, also auf die Operationalisierung der Bedingungen.

Ein mögliches Problem bei der Operationalisierung des Treatments könnte einerseits darin bestanden haben, dass die Schüler sowohl in der Treatment- als auch in der Kontrollbedingung gleichermaßen kognitiv aktiviert wurden: Zwar wurden die Schüler in der Kontrollbedingung in geringerem Maß angehalten, sich aktiv mit Repräsentationen auseinanderzusetzen, sie wurden jedoch ebenfalls mit unterschiedlichen Darstellungsformen konfrontiert, welche den Umgang mit Schülervorstellungen thematisieren. Ein Beispiel hierzu findet sich in Arbeitsblatt 7 des Unterrichtsmaterials: den Abdeckaufgaben. So brauchte die Kontrollgruppe die entsprechende Strahlenkonstruktion nicht selbst zu erarbeiten, sie wurde jedoch durch gezielte Hilfen aktiviert, diese zu verstehen.

Da die Auseinandersetzung mit Schülervorstellungen u.a. von Lipowsky (2009, S. 94) sowie von Baumert und Kunter (2011, S. 13) selbst als Maßnahme zur kognitiven Aktivierung genannt wird, könnte die Thematisierung von Schülervorstellungen alleine schon zu einem entsprechenden Lernerfolg in der Kontrollbedingung geführt haben, sodass keine Unterschiede beim Vergleich der beiden Bedingungen nachweisbar waren.

Andererseits kann als Verdienst dieser Studie herausgestellt werden, dass der Kontrollgruppe ebenfalls die gleichen Lerninhalte vermittelt wurden. So stellt die Bildung einer Kontrollgruppe im Bereich der Forschung zu Conceptual Change eine große Herausforderung dar (vgl. dazu Limón, 2001, im Überblick). Forschungsansätze, welche die Thematisierung eines Lerninhalts mit einer Kontrollgruppe vergleichen, in denen die Lerninhalte nicht parallelisiert sind, sehen sich der Kritik eines „unfairen“ Vergleichs ausgesetzt. Kritiker können Folgendes anmerken: Wenn ein bestimmter Inhalt thematisiert wird, ist auch zu erwarten, dass dieser Inhalt zumindest von einem Teil der Schüler auch verstanden und wiedergegeben wird, solange der Unterricht ein Mindestmaß an Qualität erfüllt.

Kritikern, die dieser Argumentationsweise folgen, kann jedoch entgegen gehalten werden, dass der Lernerfolg gerade im Bereich der Forschung zum Konzeptwechsel nicht auf den einfachen Zusammenhang, *die Thematisierung von Schülervorstellungen führe zur Überwindung derselben und resultiere in einem höheren konzeptuellen Verständnis*, heruntergebrochen werden kann. So belegen zahlreiche Studien, dass das Thematisieren von Schülervorstellungen alleine, z.B. in Form von Demonstrationsexperimenten, in etlichen Fällen nicht zu einem messbaren Lernerfolg führt und schon gar nicht zu einem Lernerfolg, der zeitlich stabil bleibt (vgl. auch Kapitel 1.2 Die Rolle von Konzeptwechseln in Lehr-Lernprozessen).

Insofern können die Ergebnisse der Pilotstudie und die Ergebnisse zum Stichprobenvergleich (in welchem Unterricht mit und ohne Thematisierung von Schülervorstellungen verglichen wurde) durchaus als Belege für die Wirksamkeit der hier dargestellten Lehr-Lernstrategie gewertet werden. Zumal die jeweilige Vergleichsgruppe ebenfalls in demselben Zeitumfang Unterricht zum gleichen Lernthema erhielt und sich der Lernerfolg als mittelfristig stabil erwies. Dieser Unterricht führte – wie ebenfalls, durch die Ergebnisse der Leistungstests in beiden Studien (Pilot- und Hauptstudie) belegt wurde – zu einem signifikanten Lernerfolg prä – post (Pilot- und Hauptstudie) bzw. auch prä – follow-up (erfasst in der Hauptstudie).

Aus fachlich wie forschungsmethodischer Sicht kritisch zu bewerten sind die Befunde zum Einsatz der Predict-Observe-Explain-Sequenz (POE-Sequenz). So

ergab der Vergleich des variierten Treatments, in dem die Sequenz zum Einsatz kam, mit dem regulären Treatment keinen signifikant höheren Lernerfolg, obgleich etliche Studien die Effektivität des Ansatzes belegen (White & Gunstone, 1992; Palmer, 1995; Kearney et al., 2001; Crouch et al., 2004). Es steht zu vermuten, dass die „Dosis“ zu gering war: So kam die Sequenz nur in zwei Unterrichtsstunden der Intervention zum Einsatz, wobei jeweils in einer Stunde nur 10 Minuten und in einer späteren Stunde nur 25 bis 30 Minuten aufgewendet wurden, also insgesamt 35 bis 40 Minuten von insgesamt 315 Minuten der Gesamtzeit der Intervention.

Gegebenenfalls hätte dem Einsatz der Sequenz, insbesondere innerhalb der Stunden, mehr Zeit eingeräumt werden müssen, damit die Vorannahmen der Schüler intensiver diskutiert werden können und die Wirkung der POE-Sequenz zum Tragen kommt. Diese zeitliche und inhaltliche Intensivierung der Sequenzen wiederum wäre aus Gründen der Vergleichbarkeit mit der Kontrollgruppe mit anderweitigen Problemen verbunden gewesen (ungleiche Lernzeit für Treatment- und Kontrollgruppe oder mangelnde Thematisierung des virtuellen Bildes in der Treatmentgruppe). Künftige Forschung könnte an diesem Aspekt unter der Perspektive des Lernens mit multiplen Repräsentationen ansetzen, um den Einsatz der POE-Sequenz unter der Perspektive des Lernens mit multiplen Repräsentationen unter optimierten Bedingungen zu erforschen.

Schließlich ergaben sich bezüglich der Wirkung auf die Lernmotivation, welche nicht im Mittelpunkt des Forschungsinteresses dieser Studie stand, keine Unterschiede in den Bedingungen (auch keine zu Ungunsten der Treatmentbedingung).

3.2 Ausblick

Im Mittelpunkt der vorliegenden Studie stand der Vergleich der Lernergebnisse (summative Evaluation), künftige Forschung könnte vertieft die kognitiven Prozesse in den Mittelpunkt rücken, welche den kompetenten Umgang mit Repräsentationen beeinflussen. Ein erster Schritt in diese Richtung wurde mit der Umsetzung des Forschungsziels der Analyse der Lernprozesse auf Basis der stichprobenartigen Auswertung der Arbeitsblätter vorgenommen. Alternativ könnten die Interaktionen der Schüler bei der Auseinandersetzung mit den Lerninhalten in der Peergruppe auf Video aufgenommen werden. Mittels eines Kategoriensystems zum Umgang mit unterschiedlichen Repräsentationsformen könnte analysiert werden, welche Repräsentationsformen und welche Prozesse beim Übersetzen von Repräsentationen mit komplementärem oder äquivalentem Informationsgehalt (vgl. hierzu Kapitel 1.1.4.1 Funktionen von multiplen Repräsentationen)

für die Förderung der Physikleistung und das konzeptuelle Verständnis besonders lernwirksam sind.

Eine beispielhafte Studie, in der ein ähnliches Vorgehen zur Analyse der verwendeten Repräsentationen angewandt wurde, findet sich bei Botzer und Reiner (2005). Die hierzu vorliegende Datenbasis bezieht sich jedoch auf eine sehr kleinen Stichprobe ($N = 16$) und eine Lerneinheit zu Magnetfeldern für 15- bis 16-jährige Schüler. Die Generalisierbarkeit der Ergebnisse für den Physikanfangsunterricht in der Sekundarstufe I im Bereich der Strahlenoptik ist somit eingeschränkt. Hier könnte eine Anschlussstudie ansetzen, um die Ergebnisse auf einer größeren Datenbasis zu prüfen.

Um die Lösungsprozesse der Schüler beim Bearbeiten der Tests besser kennenzulernen, könnten des Weiteren in kleineren Stichproben außerhalb des Unterrichts Interviews mit der Methode des lauten Denkens durchgeführt werden.

Video- und Interviewaufzeichnungen könnten auch dabei hilfreich sein zu klären, ob sich Konzeptwechselprozesse im Mittelstufenunterricht besser durch die Restrukturierung inadäquater Rahmentheorien (Vosniadou & Brewer, 1992) oder durch die Reorganisation von Wissensfragmenten (diSessa, 1983, 1988, 1993) beschreiben lassen, was die Perspektive auf psychologische Grundlagenforschung zu Lernprozessen im Bereich Konzeptwechsel lenkt.

Unmittelbar im Anschluss an die positiven Ergebnisse des Stichprobenvergleichs der Studien „Schülvorstellungen („SV“) und „Kohärenz („Ko“) ergibt sich die Frage, ob der mittelfristig nachgewiesene signifikant höheren Lernzuwachs für die Intervention „SV“ auch langfristig, z.B. nach einem oder sogar nach mehreren Jahren, nachweisbar wäre.

Ebenfalls könnte untersucht werden, ob Aufgaben, die Schüler durch kooperatives Bearbeiten von Repräsentationen kognitiv aktivieren, physikalische Phänomene im Kontext von (Schüler-)Experimenten zu verstehen, den Effekt auf die Lernleistung verstärken und sich positiv auf die Motivation der Schüler auswirken. Es sei an dieser Stelle auf diesbezügliche Forschungsergebnisse von Shayer und Adhami (2007) für den Bereich Mathematik hingewiesen, welche in Kapitel 1.3.3 Kognitive Aktivierung vorgestellt werden.

Neben dem kognitionspsychologischen Fokus auf Konzeptwechsel, der in dieser Arbeit im Mittelpunkt stand, wurden Prozesse des Konzeptwechsels u.a. aus soziokonstruktivistischer Perspektive untersucht. Die soziokonstruktivistische Perspektive geht davon aus, dass Konzepte immer in kognitive, und linguistische und spezifische situationale Kontexte geteilter sozialer Praktiken eingebunden sind. So argumentieren Halldén (1999) sowie Caravita und Halldén (1994), dass Probleme der Lernenden beim Anwenden wissenschaftlicher Konzepte

vorwiegend Probleme der Konzeptualisierung seien. Sogenannte „Fehlvorstellungen“³⁸ resultierten aus Schwierigkeiten der Lernenden, passende Kontexte für spezifische Konzepte zu erkennen. Wenn keine Hinweise zur adäquaten Kontextualisierung zur Verfügung stünden, neigten die Lernenden dazu, auf die ihnen vertrautesten Kontextualisierungen zurückzugreifen, auch wenn diese manchmal wenig geeignet seien. Das Vermitteln von Konzepten sollte daher auf eine passende Kontextualisierung bei der Anwendung der Konzepte zielen. Nicht das Wissen solle verändert werden, sondern die Situierung des Wissens. Lernende sollten die Kompetenz erwerben, zwischen verschiedenen Kontexten zu unterscheiden und zu erkennen, welche Konzepte in welcher Situation nützlich seien (vgl. auch Spada, 1994).

In Bezug auf das Lernen mit multiplen Repräsentationen könnte im Anschluss an diese Perspektive auf Prozesse des Konzeptwechsels ein Ansatz darin bestehen, die Verwendung von Repräsentationen als fachspezifische Werkzeuge zu interpretieren. Schülern sollte vor diesem Hintergrund nicht nur der Werkzeuggebrauch vermittelt, sondern auch ein adäquater Kontext geboten werden, um sie von der Nützlichkeit Ihres Werkzeugs zu überzeugen und Anreize zum diesbezüglichen Expertiseerwerb zu setzen.

Offen bleibt beim genannten Punkt, wie die Wahl eines geeigneten Kontextes mit dem expliziten Thematisieren des Umgangs mit multiplen Repräsentationen in Bezug auf die Förderung des konzeptuellen Verständnisses der Schüler interagiert.

Vor dem Hintergrund, dass sowohl Forschungsergebnisse zum Umgang mit Repräsentationen (vgl. Prain & Waldrip, 2006; Botzer & Reiner, 2005; Waldrip et al., 2010; Hubber et al., 2010; Mortimer & Buty, 2009) als auch Ergebnisse zum Thema kognitiver Aktivierung auf die Lernwirksamkeit der Diskussion von Lösungsstrategien in der Peergruppe im Unterricht hinweisen, könnte die Analyse dieses Aspekts ein künftiges Forschungsfeld bilden.

Obgleich es in der dargestellten Intervention in der Unterrichtspraxis zu Diskussionen der Schüler untereinander und Diskussion der Klasse mit den Fachlehrern kam, wurde diese Lehrstrategie nicht gezielt als Maßnahme zur kognitiven Aktivierung eingesetzt.

Der Neuigkeitswert eines solchen Ansatzes bestünde in der Analyse der Lernprozesse unter der Perspektive „Umgang mit multiplen Repräsentationen zur Förderung des konzeptuellen Verständnisses“.

38 Englisch „misconception“

Forschungsergebnisse, welche in die gleiche Richtung weisen, sich jedoch nicht auf das Lernen mit multiplen Repräsentationen beziehen, stammen aus Studien, welche Konzeptwechsel unter dem Fokus epistemologischer Überzeugungen beleuchteten.

So belegten Pintrich, Marx und Boyle (1993), dass Lernende, welche der Überzeugung sind, dass Fähigkeiten variabel und Wissen vorläufig ist und welche bereit waren, das vermittelte Wissen kritisch zu hinterfragen, stärker involviert waren und häufiger zu einem Konzeptwechsel gelangten als Lernende, die eine gegenteiligen Sichtweise einnahmen.

3.3 Handlungsempfehlungen für die Gestaltung von Lernumgebungen

Abschließend werden einige Handlungsempfehlungen unter der Perspektive des Lernens mit multiplen Repräsentationen zur Förderung des Verstehens physikalischer Konzepte zusammengetragen oder gefolgert. Etliche Handlungsempfehlungen für die Gestaltung von Lernumgebungen und das Begleiten von Lernprozessen ergeben sich aus den Theorien und empirischen Befunden, welche im ersten Teil dieser Arbeit dargestellt wurden. Weitere Handlungsempfehlungen können auf Basis der empirischen Ergebnisse dieser Studie(n) (zweiter Teil dieser Arbeit) abgeleitet werden.

Folgende allgemeine Handlungsempfehlungen ergeben sich aus den im ersten Teil dieser Arbeit vorgestellten empirischen Befunden:

- Im Anschluss an die Cognitive Load Theory sollte bei Gestaltung der Lernumgebung darauf geachtet werden, den Lernenden eine geeignete Menge an Material und Information auf einmal zu präsentieren (maximal ca. 7 neue Informationseinheiten), um eine Überlastung des Arbeitsgedächtnisses zu vermeiden.
- Im Hinblick auf den Umgang mit Repräsentationen könnten insbesondere schwächere Lerner gezielte Unterstützung benötigen, die sie bei der komplexen (Doppel-) Anforderung entlasten, Fach- und Repräsentationskompetenz zu erwerben. Beide Kompetenzen stehen sich nicht gegenüber, sondern bedingen sich wechselseitig, da Fachwissen in Form von internen Repräsentationen im Langzeitgedächtnis gespeichert wird und als Baustein für mentale Modelle bei der jeweiligen konkreten Problemstellung dient.
- Die verwendeten Repräsentationen sollten auf das Vorwissen der Lernenden zugeschnitten werden, um auf diese Weise einer Überlastung des Arbeitsgedächtnisses durch das Darbieten weniger zentraler Informationen vorzubeugen. In

Anlehnung an das Structure-Mapping Prinzip³⁹ nach Schnotz (2005), genügt fortgeschrittenen Lernern vermutlich eine sehr aussagekräftige Repräsentation für die Bearbeitung, welche durch einen hohen Abstraktionsgrad eine hohe Informationsdichte aufweist (in diesem Kontext z.B. das Linsengesetz)⁴⁰, während Novizen und schwächere Lerner strukturierte Lernmaterialien benötigen, die durch Lernhilfen ergänzt werden (z.B. Hilfen zum Umgang mit Gleichungen in Bezug auf die Anwendung des Abbildungsgesetzes).

- Unter Berücksichtigung der Ergebnisse dieser Studie entspricht diese Empfehlung damit auch der Beobachtung der Gesamtschullehrer, die mehr Möglichkeiten zur Binnendifferenzierung bei heterogenen Lernvoraussetzungen gefordert hatten.
- Für erfolgreiches Schlussfolgern und Problemlösen hatten sich auf Basis der Expertiseforschung zwei Faktoren als entscheidend herausgestellt: Erstens das Finden einer zutreffenden und zweckmäßigen Repräsentation der Ausgangslage und zweitens die Anwendung von Lösungsstrategien und Regeln auf das Lernmaterial, sodass schrittweise über Zwischenstufen der Zielzustand erreicht werden kann.
- Ein Ansatzpunkt, der diesen Aspekt aufgreift, könnte in der Entwicklung von Aufgaben bestehen, die Schüler ermutigen, an (offenen) physikalischen Problemstellungen zu arbeiten, die durch unterschiedliche Repräsentationsformen dargestellt werden können. Im Rahmen des Unterrichts sollten die Schüler verschiedene Lösungswege diskutieren und es sollte thematisiert werden, welche Lösungswege sich als elegante und zeitsparende Strategien herausstellen. Beispiele zu dieser Thematik in Bezug auf das Finden einer geeigneten Repräsentation der Ausgangslage aus den Bereichen Mathematik und Physik finden sich bei Schnotz et al. (2011).
- Die Diskussion unterschiedlicher Repräsentationsformen und Lösungswege schließt dabei auch an Befunde von Prain und Waldrup (2006), Botzer und Reiner (2005), Waldrup et al. (2010), Hubber et al. (2010) sowie von Mortimer und Buty (2009) an, welche in Studien mit kleinen Stichproben (meist auf Basis qualitativer Forschungsmethoden) zu dem Schluss kamen, das Generieren eigener Repräsentationen und ihre Diskussion im Unterricht als effektive Lernstrategie zu empfehlen.

³⁹ Stehen mehrere Bilder mit äquivalenten Inhalten zur Verfügung, sollte dasjenige ausgewählt werden, welches die Wissenseinheit am angemessensten repräsentiert.

⁴⁰ $\frac{1}{b} + \frac{1}{g} = \frac{1}{f}$ aus dem Linsengesetz kann man leicht die Abbildungsgleichung folgern und noch weitere Berechnungen ausführen, welche die Abbildungsgleichung nicht ermöglicht.

- Forschungsergebnisse zum Thema kognitive Aktivierung weisen in die gleiche Richtung: So belegten Studien von Stein und Lane (1996), Hiebert und Wearne, (1993), Klieme et al. (2006), Shayer und Adhami (2007) sowie Baumert und Kunter (2011), dass Lernende höhere Lernerfolge erzielten, wenn sie gefördert wurden, verschiedene Lösungswege zu erforschen, Zusammenhänge zwischen diesen Wegen herzustellen und wenn sie einen prozessorientierten Umgang mit Problemen pflegten, bei dem sie angeregt wurden, inhaltliche Konzepte und Ideen intensiv zu diskutieren (Hiebert & Wearne, 1993).
- Im Anschluss an Prain und Waldrip (2006), Waldrip et al. (2010) sowie Mortimer und Buty (2009) sollte zudem das Herstellen von Verbindungen zwischen unterschiedlichen Repräsentationen, die in unterschiedlichen Formaten gegeben werden oder auf unterschiedlichen Abstraktionsebenen liegen, gefördert werden. Forschungsergebnisse von Scheid (2013) aus dem Projekt, welchem auch diese Studie zugeordnet war, bekräftigen die Effektivität dieser Strategie.

Speziell zur Förderung des konzeptuellen Verständnisses wird empfohlen:

- Lerninhalte auf die vorhandenen Lernvoraussetzungen abzustimmen;
- den Lernenden bewusst zu machen, dass sie ihre bisherigen Annahmen überprüfen müssen, ohne sie zu stark überfordern, so dass die neuen Erklärungen eine echte verständliche Alternative bieten (eine Konsequenz, die mehr oder weniger aus allen vorgestellten Konzeptwechsel-Ansätzen gezogen werden kann);
- Die bisherigen Sichtweisen der Lernenden zu reflektieren und im Licht der neuen Annahmen zu reinterpreten, so dass die Lernenden ihr Denken über einen Sachverhalt von inadäquaten (oft explizit nicht zugänglichen) Grundannahmen befreien können, siehe Rahmentheorieansatz nach Vosniadou und Brewer (1992);
- Lernende zu aktivieren, ihre bisherigen Annahmen kritisch zu durchdenken und ihre Wissensbestände auf Kohärenz hin zu überprüfen, siehe Konzeptwechsel als Reorganisation von Wissensfragmenten nach diSessa (1983, 1988, 1993); hierbei könnte das Arbeiten mit externen Repräsentationen (vgl. Cox, 1999) zur Veranschaulichung des eigenen Wissensstandes hilfreich sein;
- und last but not least, explizit das Arbeiten mit multiplen Repräsentationen zu fördern (vgl. Wilhelm, 2005; Prain & Waldrip, 2006; Botzer & Reiner, 2007; Waldrip et al., 2010; Mortimer & Buty, 2009). Ein Befund, der ebenfalls durch Ergebnisse des Stichprobenvergleichs der Studien „Schülervorstellungen“ und „Kohärenz“ gestützt wird, in dem sich die Kombination aus Lernen mit

multiplen Repräsentationen und Thematisieren von Schülervorstellungen als wirksame Methode für einen nachweislich mittelfristig stabilen Lernzuwachs herausgestellt hatte.

Folgende Handlungsempfehlungen ergeben sich aus den empirischen Befunden dieser Arbeit:

1. Auf Basis der Ergebnisse der Hauptstudie wird die Empfehlung bekräftigt, anspruchsvolle Lernstrategien beim Lernen aus Experimenten im Physikunterricht mit multiplen Repräsentationen zu verwenden, welche gezielt auf Unterschiede im Leistungsniveau der Schüler abgestimmt sind.
 - So zeigte sich in den dargestellten Ergebnissen ein schwacher, jedoch signifikanter Effekt des Treatments für den Umgang mit repräsentationsbezogenen Aufgaben bei Gymnasiasten. Aufgrund der geringen Effektstärke erschien dieser Befund zunächst für sich genommen für die Unterrichtspraxis wenig bedeutsam. Unter Berücksichtigung des ATI-Effekts zwischen Fachnoten und Lernzuwachs beim Wissen und Problemlösen, werden beide Befunde zusammen als Hinweis gewertet, das Leistungsniveau der Schüler gezielt zu berücksichtigen.
 - Während Schüler mit guten Lernvoraussetzungen signifikant von kognitiv aktivierenden Aufgaben beim Umgang mit Repräsentationen profitieren, benötigen Schüler mit ungünstigeren Lernvoraussetzungen (geringere kognitive Fähigkeiten und schwächere Schulleistungen) vermutlich verstärkte Hilfestellungen und gezielte Hinführungen zum Umgang mit Repräsentationen. Allgemeine Forschungsergebnisse zum Thema der Individualisierung als ein Merkmal für Unterrichtsqualität finden sich bei Helmke (2009).
2. Des Weiteren empfiehlt es sich, weitverbreitete Schülervorstellungen unter der Perspektive des Lernens mit multiplen Repräsentationen zu thematisieren. So profitierten Schüler aller Leistungsbereiche und berücksichtigten Schultypen mittelfristig von dieser Strategie.

Fazit: Dieses ermutigende Ergebnis weist darauf hin, dass es durch gezielte Aufgaben möglich ist, ein physikalisches Grundverständnis zu vermitteln, das jenseits von spezifischen Aufgaben im Physikunterricht als ein Baustein naturwissenschaftlicher Grundbildung (vgl. hierzu Rönnebeck et al., 2010) verstanden werden kann, welche Schülern die Teilhabe an einer von Naturwissenschaften und Technik geprägten Welt ermöglichen soll.

Abkürzungen

A	Aufgabe
A	Abbildungsmaßstab
α	Cronbachs Alpha (Maß für die interne Konsistenz als Schätzer für die Reliabilität)
AB	Faktor: Verständnisfragen zur Bildentstehung inklusive Abdeckaufgaben
AIC	Akaike Informationskriterium
ANOVA	analysis of variance (Varianzanalyse)
ANCOVA	analysis of covariance (Kovarianzanalyse)
AV	Abhängige Variable
B	Bildgröße
b	Bildweite
BIC	Bayes Informationskriterium
BFM	Bifaktormodell
BMK	Basismodell mit Korrelationen
BRE	Faktor: Bezüge zu einem realen Experiment herstellen
BS	Faktor: Verständnis Bildkonstruktion / Strahlenmodell
CFA	Konfirmatorische Faktorenanalyse
CFI	Comparative Fit Index
d''	Effektgröße für Mittelwertunterschiede zwischen abhängigen Gruppen nach Cohen (1988): Schätzer für den Effekt in der Population
δ''	Effektgröße für Mittelwertunterschiede zwischen abhängigen Gruppen auf Ebene der Population
Δ	Effektstärkemaß, welches sich in Mehrebenenanalysen eignet.
df	degrees of freedom (Freiheitsgrade)

<i>DEV</i>	Devianz
DR	Faktor: Umgang mit deskriptiven Repräsentationen
EFA	Exploratorische Faktorenanalyse
ε	Epsilon, Fehler bzw. Residualwert
η_p^2	partielles Eta-Quadrat: Angabe des Anteils der aufgeklärten Varianz auf der Stichprobenebene
F	Faktor
G	Generalfaktor
<i>G</i>	Gegenstandsgröße
<i>g</i>	Gegenstandsweite
GY	Gymnasium
<i>h</i>	Harmonisches Mittel
\hat{H}	Kommunalität
H_0	Nullhypothese
H_1	Alternativhypothese
IGS	Integrierte Gesamtschule
ICC	Intra-Klassenkorrelation (Beurteilerübereinstimmung / Mehrebenenanalyse)
	Item Characteristic Curve (Rasch-Analyse)
KG	Kontrollgruppe
<i>KI</i>	Konfidenzintervall
KM	Klassenparameter
KPM	Klassen- und Personenparameter
KPM&H	Klassen- und Personenparameter unter Adressierung der Heteroskedastizität
KPM&H&Korr	Klassen- und Personenparameter unter Adressierung der Heteroskedastizität mit unstrukturierter Korrelationsmatrix der Fehlerterme im Individuum
„Ko“	Studie, in der die <u>Kohärenz</u> im Umgang mit Repräsentationen gefördert wurde (Scheid, 2013)
KT	Konzeptuelles Verständnis
LES	(wahrgenommenes) Engagement aus Schülersicht

LM	Leeres Modell
LT	Leistung bei repräsentationsbezogenen Aufgaben
LS	Faktor: geradlinige Lichtausbreitung und Streuung
$M_{Diff.}$	Mittlere Differenz
μ	Mittelwert in der Population
N	Stichprobenumfang
n	(Teil-)Stichprobe
ω^2	Effektstärkemaß in der Varianzanalyse: Schätzer für den Effekt in der Population
Ω^2	Effektstärkemaß in der Varianzanalyse auf Ebene der Population
P_i	Lösungswahrscheinlichkeit (bzw. Itemschwierigkeit)
PM	Personenparameter
p-Prim	phenomenological primitive (phänomenologische Primitive)
PCA	Principal component analysis (Hauptkomponentenanalyse)
Q	Quartil
QQ-Plot	Quantile-Quantile-Plot: Diagramm, in dem die Quantile zweier Variablen gegeneinander abgetragen werden, um ihre Verteilungen zu vergleichen.
r	Korrelationskoeffizient nach Pearson
r_{it}	korrigierte Trennschärfe
RMSEA	Root Mean Square Error of Approximation
r_{tet}	tetrachorischer Korrelationskoeffizient
q	Korrelationskoeffizient für den ICC
SD	Standardabweichung
SE	Standardfehler
$SD_{Diff.}$	Standardabweichung der mittleren Differenz
SOM	Second-Order-Modell
SR	Umgang mit schematisch depiktionalen Repräsentationen
$SRMR$	Standardized Root Mean Square Residual
„SV“	In dieser Arbeit vorgestellte Studie(n), in der (denen) Schüler- vorstellungen thematisiert wurden

<i>TLI</i>	Tucker-Lewis Index
TG	Treatmentgruppe
<i>U</i>	<i>U</i> (Mann-Whitney-U-Test, nicht-parametrisches Verfahren)
UV	Unabhängige Variable
W_k	Kendalls Konkordanzkoeffizient (Maß für die Beurteilerübereinstimmung)
<i>z</i>	<i>z</i> -Wert (<i>z</i> -Verteilung)

Tabellenverzeichnis

Tabelle 1	Deskriptive Ergebnisse Mittelwerte und Standardabweichungen für Treatment- (TG) und Kontrollgruppe (KG).....	135
Tabelle 2	Ergebnisse Vergleich des Mittelwertevergleichs für abhängige Stichproben prä und post.....	137
Tabelle 3 und Tabelle 4	Informationen zur Stichprobe nach Bedingung (Treatment-versus Kontrollgruppe) Klassenstufe, Geschlecht und Schultyp	148
Tabelle 5	Informationen zur Stichprobe je Bedingung und Treatmentform	149
Tabelle 6	Überblick über Formen der kognitiven Aktivierung in der Treatmentbedingung.....	151
Tabelle 7	Überblick über thematisierte Konzepte und damit verbundene Schülervorstellungen in der Treatment- (TG) und der Kontrollbedingung (KG)	153
Tabelle 8	Überblick über Variablen und die Erhebungsinstrumente	176
Tabelle 9	Übersicht über die repräsentationalen Anforderungen der Aufgabentypen im Leistungstest.....	178
Tabelle 10	Erfasste Grundkonzepte und zugeordnete Items des Konzepttests	180
Tabelle 11	Skalen des Motivationsfragebogens	182
Tabelle 12	Auswahl an Subskalen zur Erfassung relevanter kognitiver Fähigkeiten	184
Tabelle 13	Fit-Indizes: Konfirmatorische Faktorenanalysen Leistungstest, ($N = 484$).....	191
Tabelle 14	Zuordnung der Items des Leistungstests zu den Skalen „Umgang mit verschiedenen Repräsentationsformen, Reliabilität der Skalen und lokale Gütemaße, ($N = 484$).....	194
Tabelle 15	Reliabilitätsstatistiken für den Konzepttest	202
Tabelle 16	Ergebnisse des Likelihood-Quotienten-Test.....	203

Tabelle 17	Ergebnisse des Martin-Löf-Tests.....	205
Tabelle 18	Parameterschätzungen der Itemschwierigkeiten für den Konzeptposttest, ($N = 869$)	207
Tabelle 19	Ergebnisse der exploratorischen Faktorenanalyse des Konzeptposttests, ($n = 430$).....	210
Tabelle 20	Eigenwerte der Faktoren, Reliabilität und Anteil aufgeklärter Varianz der resultierenden Skalen für die Faktorenanalyse des Konzeptposttests, ($n = 430$).....	212
Tabelle 21	Konfirmatorische Faktorenanalyse Konzeptposttests – Modellvergleich des empirisch aus der Faktorenanalyse abgeleiteten Modells ($n = 439$)	213
Tabelle 22	Lokale Gütekriterien der CFA des Konzeptposttests, ($n = 439$) .	215
Tabelle 23	Reliabilitätsstatistiken für den Motivationsfragebogen.....	216
Tabelle 24	Vergleich der Schätzungen der Reliabilität: gemessene Werte und Werte von Liepmann et al. (2007) im I-S-T 2000 R.....	222
Tabelle 25	Mittelwert und Standardabweichung der Subskalen des I-S-T 2000 R nach Bedingung und Schultyp, ($N = 479$).....	222
Tabelle 26	Pearson-Korrelationen der drei Intelligenzskalen des I-S-T 2000 R ($N = 479$)	225
Tabelle 27	Eigenwerte und Anteil aufgeklärter Varianz Subskalen des I-S-T 2000 R	225
Tabelle 28	Mittelwert und Standardabweichung der Fachnoten nach Bedingung (und Schultyp)	226
Tabelle 29	Rangkorrelationen nach Spearman der drei Fachnoten, ($N = 512$)	226
Tabelle 30	Eigenwerte und Anteil aufgeklärter Varianz zu den drei Fachnoten, ($N = 479$)	227
Tabelle 31	Nichtparametrische Korrelationen der drei Fachnoten mit den verwendeten Subskalen des I-S-T 2000 R, ($N = 479$).....	228
Tabelle 32	Lehrerengagement aus Schülersicht je Bedingung zu den drei Messzeitpunkten.....	230
Tabelle 33	Ergebnisse des Mehrebenenmodells zur Analyse der Entwicklung des Lehrerengagements aus Schülersicht.....	232
Tabelle 34	Einschätzung der Unterrichtsqualität der $N = 10$ teilnehmenden Lehrkräfte gemittelt über die sechs geplanten Unterrichtsstunden	235

Tabelle 35	Kreuztabelle zur Verteilung der ausgewählten Schüler je Test (abhängige Variable) und Bedingung.....	239
Tabelle 36	Kreuztabelle zur Verteilung der ausgewählten Schüler je Test (abhängige Variable) und Quartil des Leistungsbereichs Physikleistung und konzeptuelles Verständnis	240
Tabelle 37	Deskriptive Statistiken zur Auswertung von A1-A3 des Aufgabenblatt 7 „Wiederholung zur Bildkonstruktion“, ($N = 60$)	240
Tabelle 38	t-Test zur statistischen Prüfung von Mittelwertunterschieden zwischen TG und KG bei der Bearbeitung des Aufgabenblatts 7 „Wiederholung zur Bildkonstruktion“, ($N = 60$)	241
Tabelle 39	Paarweiser Vergleich der Schüler der verschiedenen Leistungsniveaus bezüglich der Gesamtpunktzahl für die Bearbeitung des Aufgabenblatts 7, ($N = 60$).....	242
Tabelle 40	Produkt-Momentkorrelation zwischen den Posttestergebnissen und den Arbeitsaufgaben von Arbeitsblatt 7 ($N = 40$; $nQ25 = 10$, $nQ25-Q75 = 20$, $nQ75 = 10$).....	242
Tabelle 41	Deskriptive Statistiken zur Auswertung von 4a-4c des Aufgabenblatts 4, „Die Konstruktion der realen Bildfälle“ ($N = 30$).....	243
Tabelle 42	Produkt-Momentkorrelation zwischen den Posttestergebnissen und den Arbeitsaufgaben von Arbeitsblatt 4, ($n = 30$)	244
Tabelle 43	Modell 1 LM – Leeres Modell	251
Tabelle 44	Modell 2 PM (Personenparameter) – Varianzanalyse mit Messwiederholung.....	252
Tabelle 45	Modell 3 KM (Klassenparameter) – Varianzanalyse unter Berücksichtigung der Klassenstruktur	253
Tabelle 46	Modell 4KPM (Klassen- und Personenparameter) – Varianzanalyse unter Berücksichtigung von sowohl der Klassenstruktur als auch der Individuenkomponente	255
Tabelle 47	Modell 5 KPM&H – Varianzanalyse unter Berücksichtigung von sowohl der Klassenstruktur als auch der Individuenkomponente unter Adressierung der Heteroskedastizität	257

Tabelle 48	Modell 6 KPM&H&Korr – Varianzanalyse unter Berücksichtigung von sowohl der Klassenstruktur als auch der Individuenkomponente unter Adressierung der Heteroskedastizität mit unstrukturierter Korrelationsmatrix der Fehlerterme im Individuum.....	259
Tabelle 49	Überblick über die Modelleigenschaften bei steigender Komplexität	261
Tabelle 50	Überblick zum Einfluss der Kovariaten auf Wissen und Problemlösen bei repräsentationsbezogenen Aufgaben (Physikleistung).....	262
Tabelle 51	Deskriptive Statistiken Leistungstest je Bedingung	274
Tabelle 52	Ergebnisse des Mehrebenenmodells zu Wissen und Problemlösen bei repräsentationsbezogenen Aufgaben (Leistungstest)	277
Tabelle 53	Ergebnisse des Mehrebenenmodells zu Wissen und Problemlösen bei repräsentationsbezogenen Aufgaben (Physikleistung), Stichprobe nur Gymnasiasten.....	282
Tabelle 54	Deskriptive Statistiken Konzepttest je Bedingung.....	285
Tabelle 55	Ergebnisse des Mehrebenenmodells zum konzeptuellen Verständnis in der Strahlenoptik.....	289
Tabelle 56	Deskriptive Statistiken Konzepttest je Bedingung.....	293
Tabelle 57	Ergebnisse des Mehrebenenmodells zum Vergleich der Wirkung des regulären Treatments mit der Treatmentvariation auf das konzeptuelle Verständnis in der Strahlenoptik	295
Tabelle 58	Deskriptive Statistiken Motivation je Bedingung	299
Tabelle 59	Ergebnisse des Mehrebenenmodells zur Schülermotivation.....	303
Tabelle 60	Deskriptive Statistiken – Konzepttest Stichprobenvergleich	308
Tabelle 61	Ergebnisse der Mehrebenenanalyse zum konzeptuellen Verständnis: Stichprobenvergleich „SV“ und „Ko“	311

Abbildungsverzeichnis

Abbildung 1	Beschrifteter Versuchsaufbau zur Entstehung reeller Bilder bei der Sammellinse und schriftliche Erläuterung des Versuchsaufbaus	34
Abbildung 2	Das Arbeitsgedächtnis nach Baddeley (1988), entnommen aus Zumbach (2010, S. 74)	40
Abbildung 3	Gedächtnissysteme nach Paivio, entnommen aus Zumbach (2010, S. 75)	43
Abbildung 4	Theorie des multimedialen Lernens nach Mayer (2005) S. 43, übersetzt von der Autorin	45
Abbildung 5	Integratives theoretisches Rahmenmodell des Text- und Bildverstehens nach Schnotz (2005), S. 57, übersetzt von der Autorin	48
Abbildung 6	Funktionale Taxonomie multipler Repräsentationen nach Ainsworth (2006, S.6), übersetzt von der Autorin	52
Abbildung 7	Experiment zur Bildentstehung bei der Sammellinse	110
Abbildung 8	Im Unterricht verwendete Messwertetabelle zur Bildentstehung bei der Sammellinse	110
Abbildung 9	Strahlenkonstruktion zur Bildentstehung bei der Sammellinse.....	111
Abbildung 10	Generalisierende Beschreibung der Verhältnisse von Gegenstandsweite, Bildweite und Abbildungsmaßstab	111
Abbildung 11	Abbildungsgleichung	112
Abbildung 12	Geometrische Veranschaulichung der Abbildungsgleichung ..	112
Abbildung 13	Aufgabenblatt 4: Übungen zur Strahlenkonstruktion und zur Abbildungsgleichung am Hohlspiegel, Aufgabe 1 [Material 7-TG] bzw. [Material 7-KG], links Aufgaben der Treatmentgruppe (TG), rechts Aufgabe der Kontrollgruppe (KG).....	126
Abbildung 14	Aufgabenbeispiel aus dem Leistungsposttest – Anwenden der Strahlenkonstruktion	128

Abbildung 15	Aufgabenbeispiel aus dem Leistungsposttest: Lochblende vor Hohlspiegel	129
Abbildung 16	Aufgabenbeispiel (Item 9) aus dem Konzepttest	130
Abbildung 17 und Abbildung 18	Gemeinsame Dichte zweier multivariat normalverteilter Zufallsvariablen aus unterschiedlichen Perspektiven	135
Abbildung 19	Interaktion zwischen Bedingung Treatment- (TG) versus Kontrollgruppe (KG) und Zeit zur Veranschaulichung des höheren Lernzuwachses der TG in Bezug auf das konzeptuelle Verständnis	141
Abbildung 20	Aufgabenblatt 1 – Die Sammellinse[Mat 2_K1] TG bzw. [Mat 2_ K2] KG links Aufgabe der Treatmentgruppe (TG), rechts Aufgabe der Kontrollgruppe (KG)	155
Abbildung 21	[Experimentieranleitung Mat. 5_ K1] bzw. [Mat5 1]	157
Abbildung 22	Aufgabenblatt 3: Bildkonstruktion für die Sammellinse [Mat.6 _K1] bzw. [Mat. 6 _ K2] links Aufgaben der TG, rechts Aufgaben der KG 1	159
Abbildung 23	Aufgabenblatt 6: Übungen zur Bildkonstruktion [Mat. 9 _ K1] bzw. [Mat. 9_K2] links Aufgaben der TG, rechts Aufgabe der KG	163
Abbildung 24	Demonstrationsversuch „Abgedeckte Sammellinse“ [Zusatz-1-K1]	165
Abbildung 25	Ausgang des Demonstrationsversuchs „Abgedeckte Sammellinse“	165
Abbildung 26	Demonstrationsversuch „Abgedeckte Sammellinse“ [Zusatz-2-K1]	166
Abbildung 27	Konstruktion des reellen Bildes einer Kerze, welches durch eine Sammellinse entsteht, wobei die obere Hälfte der Linse abgedeckt wurde. Die hervorgehobene Fläche zeigt den Bereich des Lichts, welcher nach Abdeckung der Linse zur Entstehung des reellen Bildes beiträgt.....	166
Abbildung 28	Aufgabenblatt 7 – Wiederholung zur Bildkonstruktion [Mat. 10_K1], A2 bzw. [Mat 10_K2], A2 links Aufgabe der TG, rechts Aufgabe der KG	169
Abbildung 29	Aufgabenblatt 7: Wiederholung zur Bildkonstruktion [Mat_10K1], A3 bzw. [Mat. 10_K2] A3, links Aufgabe TG, rechts Aufgabe der KG	170

Abbildung 30	Aufgabenblatt 8: Die Abbildungsgleichung bei der Sammellinse [Mat.12_K1] bzw. [Mat. 12_K2] geometrische Veranschaulichung zur Herleitung der Abbildungsgleichung	170
Abbildung 31	Aufgabenblatt 9 – Übungen zur Abbildungsgleichung und zur Bildkonstruktion [Mat. 13_K1], A1 bzw. [Mat. 13_K2], A1 links Aufgabe der TG, rechts Aufgabe der KG	174
Abbildung 32	Aufgabenblatt 9 – Übungen zur Abbildungsgleichung und zur Bildkonstruktion [Mat. 13_K1] bzw. [Mat. 12_K2], A2 links Aufgabe der TG, rechts Aufgabe KG	175
Abbildung 33	Die mit dem I-S-T 2000 R erfasste Fähigkeitsstruktur, (Lipmann et al., 2007, S. 13) In der Studie eingesetzte Subskalen wurden von der Autorin grau hervorgehoben, gf = schlussfolgerndes Denken, gc = Wissen (ebd., S. 12).....	183
Abbildung 34	Itemschwierigkeiten des Leistungstests prä, post und follow-up, 2Z: Aufgabenteil Zeichnung, 4b-E: Aufgabenteil Erklärung, 5c-E1: Aufgabenteil Erklärung 1, 5c-E2: Aufgabenteil-Erklärung 2.....	186
Abbildung 35	Korrigierte Trennschärfen der Items des Leistungstests, prä, post, und follow-up, 2Z: Aufgabenteil Zeichnung, 4b-E: Aufgabenteil Erklärung, 5c-E1: Aufgabenteil Erklärung 1, 5c-E2: Aufgabenteil-Erklärung 2.....	188
Abbildung 36	Bifaktormodell „Umgang mit unterschiedlichen Repräsentationsformen“, DR = Umgang mit deskriptiven Repräsentationen, BRE = Bezüge zu einem realen Experiment herstellen, SR = Umgang mit schematisch depiktionalen Repräsentationen, G = Generalfaktor	192
Abbildung 37	BMK: Basismodell mit Korrelationen zwischen Faktoren „Umgang mit unterschiedlichen Repräsentationsformen“, DR = Umgang mit deskriptiven Repräsentationen, BRE = Bezüge zu einem realen Experiment herstellen, SR = Umgang mit Umgang mit schematisch depiktionalen Repräsentationen	193
Abbildung 38	Mittelwerte und Standardabweichungen des Konzeptprätest der Stichproben „SV“ (= Schülervorstellungen) und „Ko“ (= Kohärenz) im Vergleich (nur Gymnasiasten)	196

Abbildung 39	Abbildung: Itemschwierigkeiten des Konzepttests für die Stichprobe „SV“ (= Schülervorstellungen) je Messzeitpunkt	197
Abbildung 40	Itemschwierigkeit für den Konzeptprätest: Prätest: Gesamt, Post- und Follow-up Test je Stichprobe („SV“ = Schülervorstellungen versus „Ko“ = Kohärenz)	198
Abbildung 41	Korrigierte Trennschärfe für den Konzeptprätest: Prätest: Gesamt, Post- und Follow-up Test je Stichprobe „SV“ (Schülervorstellungen) versus „Ko“ (Kohärenz)	199
Abbildung 42	korrigierte Trennschärfen des Konzepttests für die Stichprobe „SV“ (Schülervorstellungen) je Messzeitpunkt	199
Abbildung 43	Grafischer Modelltest: Messzeitpunkt 2 – Posttest	204
Abbildung 44	Item Characteristic Curve ICC	206
Abbildung 45	Screepplot inklusive Parallelanalyse nach Horn (1965) für den Konzeptposttest	209
Abbildung 46	Konfirmatorische Faktorenanalyse – aus der EFA empirisch abgeleitetes Basismodell mit Korrelationen: standardisierte Lösung, AB = Verständnisfragen zur Bildentstehung inklusive Abdeckaufgaben, LS = geradlinige Lichtausbreitung und Streuung, BS: Verständnis Bildkonstruktion / Strahlenmodell,	214
Abbildung 47	Boxplots kognitive Fähigkeiten je Bedingung	223
Abbildung 48	Boxplots kognitive Fähigkeiten je Schultyp	224
Abbildung 49	Boxplots des Lehrerengagements aus Schülersicht (LES) zu den drei Messzeitpunkten	230
Abbildung 50	Verlaufsdiagramme des Lehrerengagements aus Schülersicht (LES) zu den drei Messzeitpunkten	234
Abbildung 51	Veranschaulichung der Mehrebenenstruktur der Stichprobe, die Anzahl (N) auf Ebene 1 und 2 kann je nach Auswahl an Kovariaten und Messzeitpunkt variieren.	246
Abbildung 52	Histogramme der Leistungstestergebnisse je Messzeitpunkt, TG = Treatmentgruppe, KG = Kontrollgruppe	256
Abbildung 53	Boxplots der Leistungstestergebnisse je Bedingung zu den drei Messzeitpunkten: prä, post und follow-up	275
Abbildung 54	Verlaufplots der Leistungstestergebnisse je Bedingung und Schultyp	276
Abbildung 55	Boxplots der Konzepttestergebnisse je Bedingung zu den drei Messzeitpunkten: prä, post und follow-up	286

Abbildung 56	Verlaufplots der Konzepttestergebnisse je Bedingung und Schultyp	288
Abbildung 57	Boxplots der Konzepttestergebnisse je Bedingung zu den drei Messzeitpunkten: prä, post und follow-up	293
Abbildung 58	Boxplots zur Schülermotivation je Bedingung	300
Abbildung 59	Verlaufplots zur Schülermotivation je Bedingung und Schultyp	301
Abbildung 60	Boxplots zum grafischen Vergleich des konzeptuellen Verständnisses der beiden Stichproben je Messzeitpunkt	309
Abbildung 61	Verlaufplot zum grafischen Vergleich des konzeptuellen Verständnisses der beiden Stichproben je Messzeitpunkt	310
Abbildung 62	Zusammenschau relevanter Prädiktoren und Aptitude-Treatment-Interaktionen (ATI) in Bezug auf die Test- bzw. Fragebogenergebnisse im Physikleistungstest (LT), Konzepttest (KT) und die Schülermotivation (Mo)	319
Abbildung 63	Zusammenschau relevanter Prädiktoren für den Vergleich inhaltlich aufeinander bezogener Studien „SV“ versus „Ko“ in Bezug auf die Test- bzw. Fragebogenergebnisse im, Konzepttest (KT)	320

Literaturverzeichnis

- Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education*, 33 (2-3), 131–152.
- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, 16 (3), 183–198.
- Anderson, J. R. (1996). *Kognitive Psychologie* (2. Aufl.). Heidelberg: Spektrum, Akademischer Verlag.
- Andersson, B. & Kärrqvist, C. (1983). How Swedish pupils, aged 12–15 years, understand light and its properties. *European Journal of Science Education*, 5 (4), 387–402.
- Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York: Holt, Rinehart and Winston.
- Backhaus, U., Boysen, G., Fösel, A., Heise, H., Hilscher, H., Lichtenberger, J., Liebers, K., Schepers, H., Schlichting, H.-J., Schön, L.-H., Schweitzer, S., Thanner, A., Wilke, H.-J. & Wörlen, F. (2008). *Fokus Physik. Gymnasium Gesamtband Rheinland-Pfalz*. (1. Aufl.). Berlin: Cornelsen.
- Baddeley, A. D. (1999). *Essentials of human memory*. Hove: Psychology Press.
- Baumert, J., Bos, W., Brockmann, J., Gruehn, S., Klieme, E., Köller, O. et al. (2000). *TIMSS/III–Deutschland Der Abschlussbericht: Zusammenfassung ausgewählter Ergebnisse der Dritten Internationalen Mathematik- und Naturwissenschaftsstudie zur mathematischen und naturwissenschaftlichen Bildung am Ende der Schullaufbahn*. Berlin. Zugriff am 1. September 2012 unter http://www.timss.mpg.de/TIMSS_im_Ueberblick/TIMSSIII-Broschuere.pdf
- Baumert, J. & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss & M. Neubrand (Hrsg.). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 29–53). Münster: Waxmann.
- Carey, S. (Ed.). (1985). *Conceptual change in childhood* (1st ed.). Cambridge, Mass: MIT Press.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler. Mit 242 Tabellen* (6. Aufl.). Berlin: Springer.
- Bortz, J. & Döring, N. (2005). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (Springer-Lehrbuch, 3. Aufl.). Heidelberg: Springer.
- Botzer, G. & Reiner, M. (2005). Imagery in Physics Learning - From Physicists' Practice to Naive Students' Understanding. In J. K. Gilbert (Ed.), *Visualization in Science Education*. (pp. 147–168). Springer.
- Bransford, J. (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). Washington, D.C: National Academy Press.
- Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models. Applications and data analysis methods*. Newbury Park, Calif: SAGE Publications.
- R. Hettmannsperger, *Lernen mit multiplen Repräsentationen aus Experimenten*, DOI 10.1007/978-3-658-07436-4, © Springer Fachmedien Wiesbaden 2015

- Buck, S. F. (1960). A method of estimation of missing values in multivariate data suitable for use with an electronic Computer. *Journal of the Royal Statistical Society, Series B* (22), 302–306.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München: Pearson Studium.
- Caravita, S. & Halldén, O. (1994). Re-framing the problem of Conceptual Change. *Learning and Instruction*, 4 (1), 89–111.
- Carey, S. (Hrsg.). (1985). *Conceptual change in childhood* (1st ed.). Cambridge, Mass: MIT Press.
- Cicchetti, D. V. & Sparrow, S. S. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, 86, 127–137.
- Chan, C, Burtis, J. & Bereiter, C. (1997). Knowledge Building as a Mediator of Conflict in Conceptual Change. *Cognition and Instruction*, 15 (1), 1–40.
- Chi, M. T. H. (1992). Conceptual Change within and across Ontological Categories: Examples from Learning and Discovery Science. In R. N. Giere (Ed.), *Cognitive models of science: Minnesota studies in the Philosophy of Science* (pp. 129–186). Minneapolis: University of Minnesota Press.
- Chi, M. T. H. & Slotta, J. D. (1993). The Ontological Coherence of Intuitive Physics. *Cognition and Instruction*, 10 (2-3), 249–260.
- Chi, M. T. H., Feltovich, P. J. & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science*, 5 (2), 121–152.
- Chi, M. T. H., Slotta, J. D. & Leeuw, N. de. (1994). From things to processes: A Theory of Conceptual Change for Learning Science Concepts. *Learning and Instruction*, 4 (1), 27–43. Zugriff am 25. Januar 12. unter <http://www.mendeley.com/research/from-things-to-processes-a-theory-of-conceptual-change-for-learning-science-concepts/>
- Clark, D. B. (2006). Longitudinal conceptual change in students' understanding of thermal equilibrium: An examination of the process of conceptual restructuring. *Cognition and Instruction*, 24 (4), 467–563.
- Clark, J. M. & Paivio, A. (1991). Dual coding theory and education. *Educ Psychol Rev*, 3 (3), 149–210.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J: Lawrence Erlbaum.
- Cox, R. (1999). Representation construction, externalised cognition and individual differences. *Learning and Instruction*, 9 (4), 343–363.
- Craik, F. I. M. & Lockhart, R. (1972). Levels of processing:: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684. Zugriff am 31. Mai.2012 unter http://www.numyspace.co.uk/~unn_tsmc4/prac/labs/depth/craiklock.pdf
- Crouch, C. H., Fagen, A. P., Callan, J. P. & Mazur, E. (2004). Classroom demonstrations: Learning tools or entertainment? *American Journal of Physics*, 72 (6), 835–838.
- Cronbach, L. J. & Snow, R. E. (1977). *Aptitudes and instructional methods. A handbook for research on interactions*. New York: Irvington.
- diSessa, A. (1983). Phenomenology and the Evolution of Intuition. In D. Gentner & A. L. Stevens (Hrsg.), *Mental Models* (Cognitive Science, pp. 15–34). Hillsdale, New Jersey, London: Lawrence Erlbaum Associates.

- diSessa, A. (1988). Knowledge in Pieces. In G. E. Forman & P. B. Pufall (Eds.), *Constructivism in the computer age* (pp. 49–70). Hillsdale, N.J: L. Erlbaum.
- diSessa, A. (1993). Towards an Epistemology of Physics. *Learning and Instruction*, 10 (2/3), 105–225.
- diSessa, A. (2008). Bird's Eye View of the "Pieces" vs. "Coherence" Controversy (From the "Pieces" Side of the Fence). In S. Vosniadou (Eds.), *International handbook of research on conceptual change* (pp. 35–60). New York: Routledge.
- Ditton, H. (1998). *Mehrebenenanalyse. Grundlagen und Anwendungen des Hierarchisch Linearen Modells*. Weinheim, München: Juventa-Verlag.
- Dolin, J. (2007). Science education standards and science assessment in Denmark. In D. J. Waddington (Eds.), *Standards in science education. Making it comparable* (pp. 71–82). Münster: Waxmann.
- Donald Peterson. (1996). Introduction. In Donald Peterson (Eds.), *Forms of Representation*, (pp. 7–27). Exeter: Intellect Books
- Dreyfus, A., Jungwirth, E. & Eliovitch, R. (1990). Applying the "Cognitive Conflict" Strategy for Conceptual Change--Some Implications, Difficulties, and Problems. *Science Education*, 74 (5), 555–569.
- Driver, R. & Easley, J. (1978). Pupils and Paradigms: a Review of Literature Related to Concept Development in Adolescent Science Students. *Studies in Science Education*, 5 (1), 61–84.
- Duit, R. (1993). Alltagsvorstellungen berücksichtigen! *Praxis der Naturwissenschaften Physik*, 42 (6), 7–11.
- Duit, R. (1999). Conceptual Change Approaches in Science Education. In W. Schnotz & Vosniadou S. & Carretero M. (Eds.), *New Perspectives on Conceptual Change* (1st, pp. 263–282). Oxford: Pergamon.
- Duit, R., Roth, W. M., Komorek, M. & Wilbers, J. (2001). Fostering conceptual change by analogies— between Scylla and Charybdis. *Learning and Instruction*, 11 (4-5), 283–303. Zugriff am 25. Januar 2012 unter <http://www.sciencedirect.com/science/article/pii/S0959475200000347>
- Dutke, S. (1994). *Mentale Modelle: Konstrukte des Wissens und Verstehens : kognitionspsychologische Grundlagen für die Software-Ergonomie* (Bd. 4). Göttingen: Verlag für Angewandte Psychologie.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2011). *Statistik und Forschungsmethoden. Lehrbuch ; mit Online-Materialien* (2. Aufl.). Weinheim: Beltz.
- Epskamp, S., Cramer, A.O.J., Waldorp, L.J., Schmittmann, V.D. and Borsboom, D. (2012) qgraph: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software*, 48 (4), 118. Zugriff am 15. September 2012 unter <http://www.jstatsoft.org/v48/i04/>
- Fahrmeir, L., Kneib, T. & Lang, S. (2009). *Regression. Modelle, Methoden und Anwendungen* (2. Aufl.). Berlin: Springer.
- Field, A. P. (2009). *Discovering statistics using SPSS. (and sex and drugs and rock 'n' roll)* (3rd ed.). Los Angeles [i.e. Thousand Oaks, Calif.], London: SAGE Publications.
- Fischer, U. (2001). OMEGA 3.1: Programm zur Bestimmung der Effektstärke. Landau: Zentrum für empirische pädagogische Forschung der Universität Koblenz-Landau, Campus Landau.
- Flammer, A. (2003). *Entwicklungstheorien: Psychologische Theorien der menschlichen Entwicklung* (3. Aufl.). Bern: Huber.

- Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Thousand Oaks, Calif: SAGE Publications.
- Fox, J. (2010). Package 'polycor': Polychoric and Polyserial Correlations. Version 0.7-8. Freie Statistiksoftware R. Zugriff am 02. Februar 2012 unter <http://cran.r-project.org/web/packages/polycor/>.
- Funke, J. & Zumbach, J. (2006). Problemlösen. In H. Mandl & Friedrich H. F. (Hrsg.), *Handbuch Lernstrategien* (S. 206–220). Göttingen: Hogrefe.
- Gagné, R. M., Briggs, L. J. & Wager, W. W. (1988). *Principles of instructional design* (3rd ed.). Fort Worth: Holt, Rinehart, and Winston.
- Gentner, D. & Gentner, D. (1983). Flowing Waters or Teeming Crowds: Mental Models of Electricity. In D. Gentner & A. L. Stevens (Eds.), *Mental Models* (Cognitive Science, pp. 99–129). Hillsdale: Lawrence Erlbaum Associates.
- Gilbert, J. K. (2008). Science Communication: Towards a Proper Emphasis on the Social Aspects of Science and Technology. *Alexandria Revista de Educação em Ciência e Tecnologia*, 1 (1), 3–25.
- Goldberg, F. M. & McDermott, L. C. (1987). An investigation of student understanding of the real image formed by a converging lens or concave mirror. *Am. J. Phys.*, 55 (2), 108–119. Zugriff am 30. Januar 2012 unter http://ajp.aapt.org/resource/1/ajpias/v55/i2/p108_sl?isAuthorized=no
- Guesne, E. (1985). Light. In R. Driver, E. Guesne & A. Tiberghien (Ed.), *Children's ideas in science* (pp. 10–32). Milton Keynes [Buckinghamshire], Philadelphia: Open University Press.
- Gunstone, R. F. & Champagne, A. (1990). Promoting Conceptual Change in the Laboratory. In E. Hegarty-Hazel (Eds.), *The Student laboratory and the science curriculum*. London: Routledge.
- Guzzetti, B. J., Snyder, T. E., Glass, G. V. & Gamas, W. S. (1993). Promoting Conceptual Change in Science: A Comparative Meta-Analysis of Instructional Interventions from Reading Education and Science Education. *Reading Research Quarterly*, 28 (2), 116.
- Hasebrook, J. (2006). Aptitude-Treatment-Interaktion. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (3. Aufl., S. 20–33). Weinheim: Beltz, PVU.
- Halldén, O. (1999). Conceptual Change and Contextualization. In W. Schnotz, S. Vosniadou & M. Carretero (Eds.), *New Perspectives on Conceptual Change* (1st ed, S. 53–65). Oxford: Pergamon.
- Hardy, I., Schneider, M., Jonen, A., Stern, E. & Möller, K. (2005). Fostering Diagrammatic Reasoning in Science Education. *Swiss Journal of Psychology*, 64 (3), 207–217.
- Harlen, W. (1999). *Effective teaching of science: A review of research*. Edinburgh: Scottish Council for Research in Education.
- Herrmann, T. (1993). Mentale Repräsentationen - ein erläuterungswürdiger Begriff. In J. Engelkamp (Hrsg.), *Mentale Repräsentation* (1. Aufl., S. 17–30). Bern: Hans Huber.
- Hedeker, D. (2012). *BSTT537 Longitudinal Data Analysis - Fall 2012*, University of Illinois at Chicago. Zugriff am 19. Dezember 2012 unter <http://www.uic.edu/classes/bstt/bstt513/anova.pdf>.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts* (1. Aufl.). Seelze-Velber: Kallmeyer.
- Helmke, A., Helmke, T., Lenske, G., Pham, G., Praetorius, A.-K., Schrader, F.-W. & Ade-Thurow, M. (2010). *EMU Unterrichtsdiagnostik*. Universität Koblenz-Landau, Campus Landau.

- Hestenes, D., Wells, M. & Swackhamer, G. (1992). Force concept inventory. *Phys. Teach*, 30 (3), 141.
- Hettmannsperger, R., Schnotz, W., Müller, A. & Scheid, J. (in Druck). Understanding Basic Concepts in Ray Optics: A Representational Approach. A. Kauertz, H. Ludwig, A. Müller, J. Pretsch, W. Schnotz (Eds.). *Multidisciplinary Research on Teaching and Learning*. Palgrave Macmillan Humanities and Social Sciences.
- Hiebert, J. & Wearne, D. (1993). Instructional Tasks, Classroom Discourse, and Students' Learning in Second-Grade Arithmetic. *American Educational Research Journal*, 30 (2), 393–425.
- Hochweber, J. (2010). *Was erfassen Mathematiknoten? Korrelate von Mathematik-Zeugniszensuren auf Schüler- und Schulklassenebene in Primar- und Sekundarstufe*. Münster: Waxmann.
- Homberger, D. (2003). *Sachwörterbuch zur Sprachwissenschaft*. Stuttgart: Reclam.
- Hopf, M. (2007). *Problemorientierte Schülerexperimente*. Dissertation, Ludwig-Maximilians-Universität. München
- Horn, J. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* (30), 179–185.
- Hox, J. J. (2002). *Multilevel analysis. Techniques and applications* (Quantitative methodology series). Mahwah: Lawrence Erlbaum Associates.
- Hox, J. J. (2010). *Multilevel analysis* (2. Aufl.). London: Routledge Academic.
- Hox, J. J. & Kreft, I. G. (1994). *Multilevel analysis methods* (Sociological methods and research, Vol. 22, nr. 3 (feb. 1994)). Thousand Oaks: Sage Periodicals Press.
- Hubber, P., Tytler, R. & Haslam, F. (2010). Teaching and Learning about Force - Pedagogy and Teacher Change. *Research in Science Education*, 40 (1), 5–28.
- Hunt, E. & Minstrell, J. (1994). A cognitive approach to the teaching of physics. In K. McGilly (Hrsg.), *Classroom lessons: Integrating cognitive theory and classroom practice*. (S. 51–74). Cambridge, MA.
- Kearney, M. & Treagust, D. F. (2001). Constructivism as a referent in the design and development of a computer program using interactive digital video to enhance learning in physics. *Australian Journal of Educational Technology*, 17 (1), 64–79.
- Kearney, M., Treagust, D., Yeo, S. & Zadnik, M. G. (2001). Student and Teacher Perceptions of the Use of Multimedia Supported Predict–Observe–Explain Tasks to Probe Understanding. *Research in Science Education*, 31 (4), 589–615. Zugriff am 18. Januar 2012 unter <http://www.springerlink.com/content/x4265174781k4580/>
- Kircher, E., Girwidz, R. & Häußler, P. (2000). *Physikdidaktik: Eine Einführung in Theorie und Praxis*. Braunschweig: Vieweg.
- Klauer, K. J. & Leutner, D. (2007). *Lehren und Lernen: Einführung in die Instruktionspsychologie* (1. Aufl.). Weinheim: Beltz, PVU.
- Klieme, E., Lipowsky, F., Rakoczy, K. & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts „Pythagoras“. In M. Prenzel (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (S. 127–160). Münster: Waxmann.

- Kozma, R. & Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. *Journal of Research in Science Teaching*, 34 (9), 949–968.
- Kozma, R. & Russell, J. (2005). Students becoming chemists: Developing representational competence. In J. K. Gilbert (Hrsg.), *Visualization in Science Education*. (S. 121–146). Berlin: Springer.
- Krüger, D. (2007). Die Conceptual Change-Theorie. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biogiedidaktischen Forschung. Ein Handbuch für Lehramtsstudenten und Doktoranden* (1. Aufl., S. 81–92). Berlin: Springer.
- Kuhn, D. (1991). *The skills of argument*. New York: Cambridge University Press.
- Kuhn, D., Amsel, E. & O'Loughlin, M. (1988). *The development of scientific thinking skills*. San Diego, CA: Academic Press.
- Kuhn, J. (2008). *Authentische Aufgaben in theoretischen Rahmen von Instruktions- und Lehr-Lern-Forschung: Effektivität und Optimierung von Ankermedien für eine neue Aufgabenkultur im Unterricht*. Habilitationsschrift, Universität Koblenz - Landau. Campus Landau.
- Langley, D., Ronen, M. & Eylon, B.-S. (1997). Light propagation and visual patterns: Preinstruction learners' conceptions. *J. Res. Sci. Teach*, 34 (4), 399–424.
- Larkin, J. (1983). The Role of Problem Representation in Physics. In D. Gentner & A. L. Stevens (Eds.), *Mental Models* (Cognitive Science, pp. 75–98). Hillsdale: Lawrence Erlbaum Associates.
- Larkin, J. & Simon, H. (1987). Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science*, 11 (1), 65–99.
- Leisen, J. (1998). Förderung des Sprachlernens durch den Wechsel von Symbolisierungsformen im Physikunterricht. *Praxis der Naturwissenschaften Physik*, 47 (2), 9–13.
- Leeuw, E. D. d., Hox, J. J. & Snijders, G. (2002). The Effect of Computer-Assisted Interviewing on Data Quality: A Review. *Social surveys*.
- Levin, A. & Arnold, K. (2006). Selbstgesteuertes und selbstreguliertes Lernen. In K. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (S. 206–214). Bad Heilbrunn: Klinkhardt.
- Liepmann, D., Beauducel, A., Brocke, B. & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R*. (2. Aufl.). Göttingen: Hogrefe.
- Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: a critical appraisal. *Learning and Instruction*, 11 (pp. 357–380).
- Lindell, R. S., Peak, E. & Foster, Th. M. (2006): Are They All Created Equal? A Comparison of Different Concept Inventory Development Methodologies. *Physics Education Research Conference 2006, Part of the PER Conference Invited Paper Series* 883 (July 26-27). S. 14–17.
- Lipowsky, F. (2009): Unterricht. In: Elke Wild (Hrsg.): *Pädagogische Psychologie*. (1. Aufl.) Berlin: Springer, S. 73–102.
- Little R. J. A. & Rubin D.B. (2002). *Statistical analysis with missing data*. New York: Wiley.
- Lüdtke, O., Robitzsch, A. (2010). Umgang mit fehlenden Daten in der empirischen Bildungsforschung. Quantitative Forschungsmethoden (Maschke, S. & Stecher, L., Hrsg.): Juventa Verlag Weinheim und München.
- Maas, C. & Hox, J. (2004). Sufficient sample sizes for multilevel modeling. *Methodology*, 1, 86–92.

- Mandl, H. & Friedrich, H. F. (1992). *Lern- und Denkstrategien: Analyse und Intervention*. Göttingen: Hogrefe.
- Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge: Cambridge University Press.
- Metzinger, T. (2010). Generelle Einführung. In T. Metzinger (Hrsg.), *Grundkurs Philosophie des Geistes. (Band 3): Intentionalität und mentale Repräsentation* (S. 11–32). Paderborn: mentis.
- Ministerium für Bildung, Wissenschaft, Jugend und Kultur. Rheinland-Pfalz. (2007). Rahmenlehrplan Mathematik. Klassenstufen 5-9/10. Zugriff am 5. September 2010 unter Bildungsserver Rheinland-Pfalz. [http://lehrplaene.bildung-rp.de/schulart.html?tx_abdownloads_pi1\[action\]=getviewcatalog&tx_abdownloads_pi1\[category_uid\]=118&tx_abdownloads_pi1\[cid\]=5785&cHash=f3307f600efed5f07cf048f6532f7d0b](http://lehrplaene.bildung-rp.de/schulart.html?tx_abdownloads_pi1[action]=getviewcatalog&tx_abdownloads_pi1[category_uid]=118&tx_abdownloads_pi1[cid]=5785&cHash=f3307f600efed5f07cf048f6532f7d0b)
- Ministerium für Bildung, Ministerium für Bildung, Wissenschaft, Jugend und Kultur: Rheinland-Pfalz.(Hrsg.). (1997). Lehrplan-Entwürfe Lernbereich Naturwissenschaften Biologie Physik Chemie. Orientierungsstufe (Klassen 5-6) Hauptschule Realschule Gymnasium (Klassen 7-9/10). Zugriff am 1. September 2010 unter http://lehrplaene.bildung-rp.de/nocache/lehrplaene-nach-faechern.html?tx_abdownloads_pi1%5Baction%5D=getviewclickeddownload&tx_abdownloads_pi1%5Buid%5D=192.
- Ministerium für Bildung, Wissenschaft, Jugend und Kultur. Rheinland-Pfalz. Lehrplan-Entwürfe Lernbereich Naturwissenschaften Biologie Physik Chemie. Orientierungsstufe (Klassen 5-6) Hauptschule Realschule Gymnasium (Klassen 7-9/10) Beispiele für fachübergreifenden Unterricht. Verfügbar unter Bildungsserver Rheinland-Pfalz. Zugriff am 1. September 2010 unter [http://lehrplaene.bildung-rp.de/lehrplaene-nach-faechern.html?tx_abdownloads_pi1\[action\]=getviewcatalog&tx_abdownloads_pi1\[category_uid\]=105&tx_abdownloads_pi1\[cid\]=5786&cHash=ee8ea91787597029336f19799076da00](http://lehrplaene.bildung-rp.de/lehrplaene-nach-faechern.html?tx_abdownloads_pi1[action]=getviewcatalog&tx_abdownloads_pi1[category_uid]=105&tx_abdownloads_pi1[cid]=5786&cHash=ee8ea91787597029336f19799076da00).
- Moosbrugger, H. (2007). *Testtheorie und Fragebogenkonstruktion. Mit 43 Tabellen*. Heidelberg: Springer.
- Mortimer, E. & Buty, C. (2009). What Does “In the Infinite” Mean?: The Difficulties with Dealing with the Representation of the “Infinite” in a Teaching Sequence on Optics. In C. Andersen, N. Scheuer, M. d. P. Pérez Echeverría & E. Teubal (Eds.), *Representational Systems and Practices as Learning Tools* (pp. 225–242). Rotterdam: Sense Publishers.
- Nachtigall, C. & Wirtz, M. (2004). *Statistische Methoden für Psychologen* (Bd. 2, 3. Aufl.). Weinheim: Juventa.
- Niegemann, H. M., Domagk, S., Hessel, S., Hein, A., Hupfer, M. & Zobel, A. (2008). *Kompendium multimediales Lernen*. Berlin: Springer.
- Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M. T. Keane & K. J. Gilhooly (Eds.), *Advances in the psychology of thinking* (Vol 1, pp. 1–44). New York: Harvester Wheatsheaf.
- Özdemir, G. & Clark, D. B. (2007). An Overview of Conceptual Change Theories. *Eurasia Journal of Mathematics, Science & Technology Education*, 3 (4), 351–361.
- Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, (44), 443–460.
- Özdemir, G. & Clark, D. B. (2007). An Overview of Conceptual Change Theories. *Eurasia Journal of Mathematics, Science & Technology Education*, 3 (4), 351–361.

- Paas, F., Tuovinen, J. E., Merriënboer, J. J. & Darabi, A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Research and Development*, 53 (3), 25–34.
- Paas, F., van Merriënboer, J. J. & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, 79, 419–430.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press; Clarendon Press.
- Palmer, D. (1995). The POE in the Primary School: An Evaluation. *Research in Science Education*, 25 (3), 323–332.
- Piaget Jean. (1996). *Einführung in die genetische Erkenntnistheorie* (6. Auflage). Frankfurt am Main: Suhrkamp.
- Piaget, J. (1979). *Sprechen und Denken des Kindes* (4. Aufl.). Düsseldorf: Schwann.
- Pinheiro, J. & Bates, D. M. (2013). nlme [Computer Software]: freie Statistiksoftware R. Zugriff am 17. September 2011 unter <http://cran.r-project.org/web/packages/nlme/nlme.pdf>.
- Pinheiro, J. C. & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.
- Pintrich, P. & Garcia, T. (1994). Self-regulated learning in college students. In P. Pintrich & D. W. C. Brown (Eds.), *Student motivation, cognition and learning* (pp. 113–133). Hillsdale (N.J.): Erlbaum.
- Pintrich, P., Marx, R. & Boyle, R. (1993). Beyond Cold Conceptual Change: The Role of Motivational Beliefs and Classroom Contextual Factors in the Process of Conceptual Change. *Review of educational research*, 63 (2), 167–199.
- PISA-Konsortium Deutschland (Hrsg.). (2007). *PISA 2006: Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.
- Plötzner, R. & Spada, H. (1998). Inhalt, Struktur und Anwendung von Physikwissen: Eine psychologische Perspektive. *Zeitschrift für Didaktik der Naturwissenschaften*, 4 (2), 81–100.
- Posner, G. J., Strike, K. A., Hewson, P. W. & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Sci. Ed*, 66 (2), 211–227.
- Poinstingl, H., Mair, P. & Hatzinger, R. (2007). *Manual zum Softwarepackage eRM (extended Rasch modeling). Anwendung des Rasch-Modells (1-PL-Modell)* (Dt. Version.). Lengerich: Pabst Science Publ.
- Prain, V. & Waldrip, B. (2006). An exploratory study of teachers' and students' use of multimodal representations of concepts in primary science. *International Journal of Science Education*, 28, 1843–1866.
- Prenzel, M., Carstensen, C., Frey, A., Drechsel, B. & Rönnebeck, S. (2007). PISA 2006 - Eine Einführung in die Studie. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 31–58). Münster: Waxmann.
- Rasch, B., Frieese, M., Hofmann, W. & Naumann, E. (2010). *Quantitative Methoden. Einführung in die Statistik für Psychologen und Sozialwissenschaftler*, Bd. 2., 3. Auflage. Heidelberg: Springer.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods* (Advanced quantitative techniques in the social sciences, Vol. 1, 2nd ed.). Thousand Oaks, Calif.: Sage Publ.

- Reiner, M., Slotta, J. D., Chi, M. T. H. & Resnick, L. B. (2000). Naive Physics Reasoning: A Commitment to Substance-Based Conceptions. *Cognition and Instruction*, 18 (1), 1–34.
- Rey, G. D. (2009). *E-Learning: Theorien, Gestaltungsempfehlungen und Forschung* (1. Aufl.). Bern: Huber.
- Reinmann-Rothmeier, G., Mandl, H., Prenzel, M., Arzberger, H. & Brehm, K.-H. (op. 1994). *Computerunterstützte Lernumgebungen. Planung, Gestaltung und Bewertung*. Erlangen: Publicis MCD Verlag.
- Revelle, W. (2013) psych [Computer Software]: freie Statistiksoftware R. Zugriff am 22. September 2011 unter <http://cran.r-project.org/web/packages/psych/index.html>.
- Rheinberg, F. & Wendland, M. (2003). DFG Projekt – Veränderung der Lernmotivation in Mathematik und Physik: Eine Komponentenanalyse und der Einfluss elterlicher und schulischer Kontextfaktoren. Itemübersicht zum Fragebogen.
- Rheinberg, F. & Wendland, M. (2004). Welche Motivationsfaktoren beeinflussen die Mathematikleistung? – Eine Längsschnittanalyse. In J. Doll (Hrsg.), Bildungsqualität von Schule. Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsentwicklung (S. 309–328). Münster: Waxmann.
- Rönnebeck, S., Schöps, K., Prenzel, M. & Hammann, M. Naturwissenschaftliche Kompetenzen im Ländervergleich. In *PISA-Konsortium Deutschland (Hg.) 2006 – PISA 2006 in Deutschland* (S. 67–92).
- Rönnebeck, S., Schöps, K., Prenzel, M., Mildner, D. & Hochweber, J. (2010). Naturwissenschaftliche Kompetenzen von PISA 2006 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel et al. (Hrsg.), *PISA 2009 Bilanz nach einem Jahrzehnt* (S. 177–198). Münster: Waxmann.
- Rudolf, M. & Müller, J. (2004). *Multivariate Verfahren. Eine praxisorientierte Einführung mit Anwendungsbeispielen in SPSS*. Göttingen: Hogrefe.
- Savelsbergh, E. R., Ferguson-Hessler, M. G. M. & Jong, T. de. (1997). *The Importance of an Enhanced Problem Representation: On the Role of Elaborations in Physics Problem Solving*. (University of Twente, F. o. E. S. a. T. D. o. I. T. P. B. 2. 7. A. E. T. N., Eds.).
- Scheid, Jochen (2013). Multiple Repräsentationen, Verständnis physikalischer Experimente und kognitive Aktivierung: Ein Beitrag zur Entwicklung der Aufgabenkultur. In H. Niedderer, H. Fischler, E. Sumfleth (Hrsg.), *Studien zum Physik- und Chemielernen*, Band 151. Berlin: Logos Verlag.
- Schiefele, U. (2009). Motivation. In E. Wild und J. Möller (Hrsg.). *Pädagogische Psychologie* (S. 152–177) Heidelberg: Springer.
- Schnotz, W. (1994). *Aufbau von Wissensstrukturen: Untersuchungen zur Kohärenzbildung bei Wissenserwerb mit Texten*. Weinheim: Beltz, PVU. Zugriff am 14. September 2011 unter <http://katalog.ub.uni-heidelberg.de/cgi-bin/titel.cgi?katkey=9417963>.
- Schnotz, W. (2005). An Integrated Model of Text and Picture Comprehension. In R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (S. 49–69). Cambridge: Cambridge University Press.
- Schnotz, W. (2006a). *Pädagogische Psychologie: Workbook*. Weinheim: Beltz, PVU.
- Schnotz, W. (2006b). Conceptual Change. In D. H. Rost (Hrsg.), *Handwörterbuch pädagogische Psychologie* (3. Aufl., S. 75–81). Weinheim: Beltz, PVU.
- Schnotz, W. & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13 (2), (pp. 141–156).

- Schnotz, W. & Rasch, T. (2005). Enabling, Facilitating, and Inhibiting Effects of Animations in Multimedia Learning: Why Reduction of Cognitive Load Can Have Negative Results on Learning. *Educational Technology Research and Development*, 53 (3), 47–58.
- Schnotz, W. & Rasch, T. (2007). Functions of animation in comprehension and learning. In R. Lowe & W. Schnotz (Hrsg.), *Learning with animation. Research and implications for design* (pp. 92–113). Cambridge: Cambridge University Press.
- Schnotz, W., Baadte, C., Müller, A., & Rasch, R. (2011). Kreatives Denken und Problemlösen mit bildlichen und beschreibenden Repräsentationen. In R. Sachs-Hombach, & R. Totzke (Hrsg.), *Bilder – Sehen – Denken. Zum Verhältnis von begrifflich-philosophischen und empirisch-psychologischen Ansätzen in der bildwissenschaftlichen Forschung* (S. 204 – 252). Köln: Halem Verlag.
- Schnotz, W., Vosniadou, S. & Carretero, M. (Hrsg.). (1999). *New Perspectives on Conceptual Change* (1st ed.). Oxford: Pergamon.
- Schrader, F.-W. & Helmke, A. (2006). Lernstrategien. In *Brockhaus Enzyklopädie* (S. 638–639). Leipzig: Brockhaus.
- Seidel, T., Prenzel, M., Duit, R. & Lehrke, M. (Hrsg.). (2003). *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“*. Kiel: Universität, Leibniz-Institut für die Pädagogik der Naturwissenschaften.
- Séré, M. G. (2002). Towards Renewed Research Questions from the Outcomes of the European Project Labwork in Science Education. *Science Education*, 86 (5), 624–644.
- Shayer, M. & Adhami, M. (2007). Fostering Cognitive Development Through the Context of Mathematics: Results of the Came Project. *Educational Studies in Mathematics*, 64 (3), 265–291.
- Snijders, T. A. B. & Bosker, R. J. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. London: SAGE Publications.
- Southerland, S. A., Abrams, E., Cummins, K. & Anzelmo, J. (2001). Understanding Students' Explanations of Biological Phenomena: Conceptual Frameworks or P-Prims? *Science Education*, 85 (4), 328–348.
- Spada, H. (1994). Commentary - Conceptual Change or Multiple Representations? *Learning and Instruction*, 4 (1), 113–116. Zugriff am 03.03. 2012 unter <http://www.sciencedirect.com/science/article/pii/0959475294900213>.
- Stark, R. (2002). *Conceptual Change: kognitivistisch oder kontextualistisch?* (Ludwig-Maximilians-Universität, L. f. E. P. u. P. P., Hrsg.) (Forschungsbericht Nr. 149). München: Institut für Pädagogische Psychologie und Empirische Pädagogik.
- Strike, K. A. & Posner, G. J. (1982). Conceptual change and science teaching. *European Journal of Science Education*, 4 (3), 231–240.
- Stein, M. K. & Lane, S. (1996). Instructional Tasks and the Development of Student Capacity to Think and Reason: An Analysis of the Relationship between Teaching and Learning in a Reform Mathematics Project. *Educational Research and Evaluation*, 2 (1), 50–80.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4 (4), 295–312.
- Sweller, J. (2005). Implications of Cognitive Load Theory for Multimedia Learning. In R. E. Mayer (Eds.), *The Cambridge handbook of multimedia learning* (pp. 19–30). Cambridge: Cambridge University Press.

- Sweller, J. & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12 (3), 185–233.
- Tabachneck, H. J. M., Leonardo, A. M. & Simon, H. A. (1994). How does an expert use a graph? A model of visual & verbal inferecing in economics. In A. Ram & K. Eiselt (Eds.), *Proceedings of the sixteenth annual conference of the cognitive science society, August 13 to 16, 1994, Georgia Institute of Technology* (pp. 842–847). Hillsdale (N.J.): Lawrence Erlbaum.
- Tabachneck-Schijf, H. J. M. & SIMON, H. (1996). Alternative representations of instructional material. In D. Peterson (Eds.), *Forms of representation. An interdisciplinary theme for Cognitive Science* (pp. 28–46). Wiltshire: Intellect.
- Taylor, B. M., Pearson, D. P. & Rodriguez, M. C. (2003). Reading Growth in High-Poverty Classrooms: The Influence of Teacher Practices That Encourage Cognitive Engagement in Literacy Learning. *The Elementary School Journal*, 104 (1), 3–28.
- Teichert, D. (2006). *Einführung in die Philosophie des Geistes*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Tymms, P. (2004). But what does it mean? The use of effect sizes in educational research. *National Foundation for Educational Research*, 55–66.
- Tymms, P., Merrell, C. & Henderson, B. (1997). The first year at school. A quantitative investigation of the attainment and progress of pupils. *Educational Research and Evaluation*.
- Tyson, L. M., Venville, G. J., Harrison, A. G. & Treagust, D. F. (1997). A Multidimensional Framework for Interpreting Conceptual Change Events in the Classroom. *Science Education*, 81 (4), 387–404.
- van Buuren, S. & Groothuis-Oudshoorn, K. (2011). MICE: Flexible imputation by Chained Equations in R. *Journal of Statistical Software*.
- Vosniadou, S. & Brewer, W. (1992). Mental Models of the Earth: A Study of Conceptual Change in Childhood. *Cognitive Psychology*, 24 (5), 535–585.
- Vosniadou, S. (1994). Capturing and Modeling the Process of Conceptual Change. *Learning and Instruction*, 4 (1), 45–69. Zugriff am 19. September 2011 unter <http://www.sciencedirect.com/science/article/pii/S0959475294900183>.
- Vosniadou, S. & Brewer, W. F. (1994). Mental Models of the Day/Night Cycle. *Cognitive Science*, 18 (1), 123–183.
- Vosniadou, S., Ioannides, C., Dimitrakopoulou, A. & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction*, 11 (4-5), 381–419. Zugriff am 28. September 2011 unter <http://www.sciencedirect.com/science/article/pii/S0959475200000384>.
- Waldrip, B., Prain, V. & Carolan, J. (2006). Learning Junior Secondary Science through Multi-Modal Representations. *Electronic Journal of Science Education*, 11 (1), 87–107. Zugriff am 14. September 2011.
- Waldrip, B., Prain, V., & Carolan, J. (2010). Using multi-modal representations to improve learning in junior secondary science. *Research in Science Education*, 40 (1), 65–80
- Weidenmann, B. (1997). Multicodierung und Multimodalität im Lernprozess. In L. J. Issing & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia* (2. Aufl., S. 65–84). Weinheim: Beltz.
- White, R. & Gunstone, R. (1992). *Probing understanding* (1st ed.). London: The Falmer Press.

- Wiesner, H. (1986). Schülervorstellungen und Lernschwierigkeiten im Bereich der Optik. *Naturwissenschaften im Unterricht - Physik*, 34 (13), 25–29.
- Wiesner, H. (1992a). Schülervorstellungen und Lernschwierigkeiten mit dem Spiegelbild. *Naturwissenschaften im Unterricht - Physik*, 3 (14), 16–18.
- Wiesner, H. (1992b). Verbesserung des Lernerfolgs im Unterricht über Optik (1). *Physik in der Schule*, 30 (9), 286–290.
- Wiesner, H. (1994). Ein neuer Optikkurs für die Sekundarstufe I, der sich an Lernschwierigkeiten und Schülervorstellungen orientiert. *Naturwissenschaften im Unterricht – Physik*, 5 (22), 7–15.
- Wilhelm, T. (2005). *Konzeption und Evaluation eines Kinematik/Dynamik-Lehrgangs zur Veränderung von Schülervorstellungen mit Hilfe dynamisch ikonischer Repräsentationen und graphischer Modellbildung*. Dissertation, Julius-Maximilians-Universität. Würzburg.
- Wodzinski R. (2004). Schülervorstellungen und Lernschwierigkeiten. In R. Müller, Wodzinski R. & M. Hopf (Hrsg.), *Schülervorstellungen in der Physik. Festschrift für Hartmut Wiesner* (1. Aufl., S. 23–36). Köln: Aulis Verlag Deubner.
- Wygotski, L. S. (1986). *Denken und Sprechen* (Ungekürzte Ausg.). Frankfurt am Main: Fischer-Taschenbuch-Verlag.
- Yves Rosseel. (2012) Package lavaan [Computer software]: freie Statistiksoftware R. Zugriff am 7. September 2012 unter <http://lavaan.org>.
- Zeyer, G. (1981). Vorstellungen von Erwachsenen zum freien Fall. In A. Scharmann, A. Hofstaetter & W. Kuhn (Hrsg.), *Didaktik der Physik. Vorträge Physikertagung 1981 Gießen (Deutsche Physikalische Gesellschaft (DPG), Fachausschuß Didaktik der Physik)*, 87–92.
- Zhang, J. (1997). The Nature of External Representations in Problem Solving. *Cognitive Science*, 21 (2), 179–217. Zugriff am 03. November 2010 unter <http://csjarchive.cogsci.rpi.edu/1997v21/i02/p0179p0217/MAIN.PDF>.
- Zumbach, J. (2010). *Lernen mit neuen Medien: Instruktionspsychologische Grundlagen* (1. Aufl.). Stuttgart: Kohlhammer.

Verzeichnisse für den Anhang

zu finden auf der Produktseite zu diesem Buch unter www.springer.com

Anhang	A	Aufgabenanalyse zur Lehrplaneinheit Hohlspiegel	3
Anhang	B	Pilotstudie.....	13
	B1	Gegenüberstellung Unterrichtsmateriel Hohlspiegel	13
	B2	Unterrichtsentwurf Pilotstudie Hohlspiegel Treatment- und Kontrollgruppe	58
	B3	Pilotstudie Leistungstests Physik	67
	B4	Konzepttest.....	76
	B5	Ergänzende Statistiken zur Pilotstudie	94
Anhang	C	Hauptstudie	100
	C1	Gegenüberstellung des Unterrichtsmaterials Treatment- und Kontrollgruppe	100
	C2	Unterrichtsentwurf Bilder der Sammellinse Treatment- und Kontrollgruppe	180
	C3	Treatmentvariation Gegenüberstellung Treatment- und Kontrollgruppe	198
	C4	Treatmentvariation in der 6. Unterrichtsstunde	202
	C5	Leistungstest Treatmentgruppe und Kontrollgruppe	206
	C6	Konzepttest Strahlenoptik Treatmentgruppe und Kontrollgruppe	213
	C7	Fragebogen zur Motivation und zum Lehrerengagement	225
	C8	Notizheft Hauptstudie Linse	231
	C9	Ergänzende Statistiken zur Hauptstudie	233

Tabellenverzeichnis

Tabelle 1	Itemschwierigkeiten des Leistungstests prä und post ($N = 52$).....	94
Tabelle 2	Trennschärfen des Leistungstests prä und post ($N = 52$)	94
Tabelle 3	Itemmittelwerte und Standardabweichungen des Konzepttests prä und post	95
Tabelle 4	Itemschwierigkeiten des Konzepttests prä- und post	96
Tabelle 5	Trennschärfen des Konzepttests prä- und post.....	98
Tabelle 6	Informationen zur Stichprobe je Bedingung und Schultyp	234
Tabelle 7	Übersicht über die Überarbeitung des Konzepttests Pilotstudie versus Hauptstudie	235
Tabelle 8	Itemmittelwerte und Standardabweichungen des Leistungstests je Messzeitpunkt	236
Tabelle 9	Trennschärfen und Lösungswahrscheinlichkeiten im Leistungstest je Messzeitpunkt	237
Tabelle 10	Mittelwerte und Standardabweichungen des Konzeptprätests	237
Tabelle 11	Mittelwerte und Standardabweichungen des Konzepttests je Stichprobe und Messzeitpunkt	238
Tabelle 12	Itemschwierigkeit für den Konzepttest: prä (Gesamt) sowie post und follow-up je Stichprobe	239
Tabelle 13	Korrigierte Trennschärfen für den Konzepttest je Stichprobe und Messzeitpunkt.....	240
Tabelle 14	Anzahl Individuen je Stichprobe und faktorenanalytische Methode	240
Tabelle 15	Mittelwerte und Standardabweichungen des Motivationsfragebogens je Messzeitpunkt	241
Tabelle 16	Trennschärfen und Lösungswahrscheinlichkeiten des Motivationsfragebogens je Messzeitpunkt	242
Tabelle 17	Deskriptive Ergebnisse zu den verwendeten Subskalen des I-S-T 2000 R	243
Tabelle 18	Kreuztabelle zur Verteilung der ausgewählten Schüler je Bedingung und Schulart	248
Tabelle 19	Kreuztabelle zur Verteilung der ausgewählten Schüler je Bedingung und Geschlecht.....	248
Tabelle 20	Überblick Einfluss der Kovariaten auf das konzeptuelle Verständnis	249
Tabelle 21	Überblick Einfluss der Kovariaten auf die Schülermotivation	250

Tabelle 22	Ergebnisse des Mehrebenenmodells zur Analyse von ATI-Effekten in Bezug auf die Physikleistung	251
Tabelle 24	Ergebnisse des Mehrebenenmodells zur Analyse von ATI-Effekten in Bezug auf die Schülermotivation	259
Tabelle 25	Ergebnisse des Mehrebenenmodells zur Analyse von ATI-Effekten in Bezug auf das konzeptuellen Verständnis Stichproben SV und Ko im Vergleich	262

Abbildungsverzeichnis

Abbildung 1:	Itemmittelwerte und Standardabweichungen des Konzepttests prä und post: Pilotstudie.....	97
Abbildung 2:	Itemmittelwerte und Standardabweichungen für den Leistungstest je Messzeitpunkt.....	236
Abbildung 3:	Jitter Plot zur grafischen Veranschaulichung des Zusammenhangs zwischen figural-räumlicher und verbaler Intelligenz.....	243
Abbildung 4:	Jitter Plot zur grafischen Veranschaulichung des Zusammenhangs zwischen figural-logischer und verbaler Intelligenz.....	244
Abbildung 5:	Jitter Plot zur grafischen Veranschaulichung des Zusammenhangs zwischen figural-logischer und figural-räumlicher Intelligenz	244
Abbildung 6:	Jitter Plot zur grafischen Veranschaulichung des Zusammenhangs zwischen Deutsch- und Mathematiknote	245
Abbildung 7:	Jitter Plot zur grafischen Veranschaulichung des Zusammenhangs zwischen Physik- und Mathematiknote	246
Abbildung 8:	Jitter Plot zur grafischen Veranschaulichung des Zusammenhangs zwischen Deutsch- und Physiknote	247