

**Lernaufgaben im Unterricht:
Instruktionspsychologische Analysen
am Beispiel der Physik**

Dissertation

zur Erlangung des akademischen Grades eines

Doktors der Philosophie (Dr. phil.)

am Fachbereich Bildungswissenschaften

der Universität Duisburg-Essen

vorgelegt von

Katharina Schabram

aus Köln

Tag der Abgabe: 24.März 2007

Tag der Disputation: 22.August 2007

1. Gutachter

Prof. Dr. D. Leutner

Universität Duisburg-Essen

2. Gutachter

Prof. Dr. R. Brünken

Universität des Saarlandes

Für meine Eltern und Christoph.

Danksagung

An dieser Stelle möchte ich all jenen Personen danken, die mich auf ganz unterschiedliche Art und Weise bei der vorliegenden Arbeit unterstützt haben.

Mein besonderer Dank gilt Herrn Prof. Dr. D. Leutner für die inhaltliche Betreuung und fachliche Unterstützung sowie die vielfältigen Anregungen bei der Anfertigung dieser Arbeit.

Herrn Prof. Dr. R. Brünken danke ich dafür, dass er sich als 2.Gutachter meiner Arbeit zur Verfügung gestellt hat.

Bei meinen Kollegen Dr. Stefan Rumann und Paul Jatzwauk möchte ich mich für die gute Zusammenarbeit und die entgegengebrachte Unterstützung bedanken. Ebenfalls möchte ich Herrn Dipl.-Psych. Jens Fleischer für seine Hilfe bei der Rasch-Skalierung danken.

Mein ganz besonderer Dank gilt den insgesamt 685 Schülern und ihren 32 Lehrern, die sich bereit erklärt haben, an den Studien mitzuwirken. Ohne ihre Teilnahme hätte diese Arbeit nicht realisiert werden können.

Frau Dipl.-Psych. Birgit Spöth danke ich von ganzem Herzen für ihre großartige Unterstützung und kompetente Hilfe beim mehrmaligen Korrekturlesen und Formatieren der Arbeit.

Am meisten danke ich meiner Familie und meinen Freunden für die Unterstützung und Aufmunterung während des langen Wegs der Fertigstellung dieser Arbeit.

Inhaltsverzeichnis

Zusammenfassung	
1 Einleitung	1
2 Definition von Lernaufgaben	4
2.1 Was sind Lernaufgaben?	4
2.2 Abgrenzung von Aufgaben und Problemen	6
3 Forschungsergebnisse zur Effektivität von Lernaufgaben	8
3.1 Funktion und Effektivität von Lernaufgaben	9
3.2 Bestimmung der Schwierigkeit von Aufgaben.....	12
3.2.1 Low level- vs. high level-Aufgaben	13
3.2.2 Textbegleitende Fragen (adjunct questions)	17
3.2.3 Aufgaben mit Lösungsbeispielen (worked examples).....	19
3.2.4 Schwächen der bisherigen Untersuchungen.....	21
3.2.5 Fazit aus den bisherigen Forschungsergebnissen	25
3.3 Passung von Aufgabenschwierigkeit und Schülerfähigkeit	26
3.4 Diagnostische Kompetenz der Lehrer bei der Einschätzung der Aufgabenschwierigkeit und Schülerfähigkeit.....	30
4 Theorien und Modelle zur Aufgabenstellung und Aufgabenbearbeitung	35
4.1 Behavioristische Lerntheorie.....	35
4.2 Kognitive Lerntheorie	37
4.3 Konstruktivismus	38
4.4 Zone der proximalen Entwicklung nach Wygotsky	39
4.5 Das Prozess-Produkt-Paradigma	40
4.6 Blooms Modell des schulischen Lernens (1976).....	41
4.7 Das Modell der Direkten Instruktion	44
4.8 Fazit.....	46
5 Klassifikationen von Lernaufgaben	48
5.1 Nutzen von Taxonomien	48
5.2 Die Taxonomie von Lehrzielen im kognitiven Bereich (Bloom et al., 1956).....	49
5.3 Revision von Anderson et al. (2001)	51
5.4 Empirische Ergebnisse zur Struktur der Taxonomie	52
6 Konstruktion von Lernaufgaben	56
6.1 Schulleistungstests	57
6.2 Die Konstruktion von Lernaufgaben nach Klauer (1974, 1987).....	59

7	Schlussfolgerung für die Fragestellung der Studie 1	64
8	Studie 1: Fragestellung	66
9	Methode	70
9.1	Stichprobe.....	70
9.2	Design	71
9.3	Durchführung.....	71
9.4	Instrumente: Leistungstests	73
9.5	Beschreibung der kognitiven Prozesse für die Aufgabenkonstruktion	76
9.5.1	Wissen.....	76
9.5.2	Verstehen.....	78
9.5.3	Anwendung	80
9.5.4	Aufgabenbeispiele	83
9.6	Auswertung der Unterrichtsvideos und der Leistungstests	85
10	Ergebnisse	88
10.1	Videoauswertung.....	88
10.1.1	Interrater-Reliabilität der Videoauswertung – Aufgabenidentifikation und Aufgabenklassifikation	88
10.1.2	Repräsentativität der wiederholten Unterrichtsaufgaben.....	94
10.1.3	Aufgabeneinsatz im Physikunterricht	95
10.2	Deskriptive Auswertung der Leistungstests	106
10.3	Raschskalierung der Aufgaben und Personen.....	110
10.3.1	Schätzung der Itemparameter.....	113
10.3.2	Schätzung der Personenparameter.....	123
10.3.3	Passung zwischen Aufgabenschwierigkeit und Schülerfähigkeit	124
11	Zusammenfassung und Diskussion der Ergebnisse aus Studie 1 und Ableitung der Fragestellung für Studie 2	128
11.1	Zusammenfassung der Ergebnisse	128
11.2	Abgeleitete Fragestellung	133
12	Studie 2: Adaptives Testen und Unterrichten	135
12.1	Bedeutung der Adaptation für den Unterricht	135
12.2	Entwicklung des Konstrukts adaptiver Unterricht	136
12.3	Forschungsergebnisse zum adaptiven Testen und Unterricht.....	141
12.4	Schlussfolgerung für Fragestellung der Studie 2	144
13	Fragestellung	145
14	Methode	148

14.1	Stichprobe	148
14.2	Design und Durchführung	149
14.3	Instrumente – Adaptive Anpassung der Aufgabenschwierigkeit an die individuelle Schülerfähigkeit	151
14.4	Auswertung	153
15	Ergebnisse	155
15.1	Treatment-Check	155
15.2	Raschskalierung	159
15.2.1	Schätzung der Itemparameter	159
15.2.2	Schätzung der Personenparameter	164
16	Zusammenfassung und Diskussion der Ergebnisse der Studie 2	171
16.1	Zusammenfassung der Ergebnisse	171
16.2	Diskussion der Ergebnisse	172
17	Zusammenfassende Diskussion der Ergebnisse beider Studien und Forschungsausblick	176
17.1	Zusammenfassende Diskussion der Ergebnisse beider Studien	176
17.2	Ausblick und offene Fragen	181
17.3	Grenzen der Studien	182
	Literaturverzeichnis	184
	Tabellenverzeichnis	214
	Abbildungsverzeichnis	217
	Anhang	

Zusammenfassung

Eine angemessene Gestaltung von Lernaufgaben als Voraussetzung für effizientes Lernen ist von herausragender Bedeutung für den Unterricht. Lernaufgaben sind in dieser Arbeit als inhaltsbezogene und ergebnisorientierte Leistungsanforderungen des Lehrers an die Schüler¹ definiert. Sie bieten Schülern eine Möglichkeit zur intensiven Auseinandersetzung mit den zu lernenden Inhalten, sie fördern ein besseres Verständnis und haben einen positiven Lerneffekt (vgl. Häfele, 1995; Hamaker, 1986). Zwar ist der Nutzen von Lernaufgaben für die Lernleistung empirisch belegt, zu den vom Lehrer im Unterricht eingesetzten Aufgaben und deren Schwierigkeit gibt es jedoch bislang kaum Studien². Daher wird in der vorliegenden Arbeit der Zusammenhang zwischen der Schwierigkeit von Lernaufgaben und der Schülerleistung untersucht.

Eine bislang ebenfalls kaum erforschte Untersuchungsfrage bezieht sich auf die Passung zwischen der Schwierigkeit der eingesetzten Unterrichtsaufgaben und den Fähigkeiten der Schüler. Auch diesem Zusammenhang wird in dieser Arbeit nachgegangen. Dazu wurde zunächst in Studie 1 systematisch erfasst, was für Aufgaben im Physikunterricht eingesetzt werden und welche Schwierigkeit sie für die Schüler aufweisen. Im Anschluss daran stand die Frage, wie Lernaufgaben unterschiedlicher Schwierigkeit – operationalisiert über die für die Aufgabenbearbeitung erforderlichen kognitiven Prozesse – konstruiert werden können. Dazu wurden in je 15 Gymnasial- und Hauptschulklassen aus einstündigen Unterrichtsvideos die Unterrichtsaufgaben identifiziert. Eine Woche später wurde den Schülern ein Leistungstest vorgelegt. Dieser setzte sich aus wiederholten Unterrichtsaufgaben und neu konstruierten Aufgaben unterschiedlicher Schwierigkeit sowie TIMSS-Aufgaben zusammen. Es wurde eine Skalierung der Aufgaben und Schülerleistungen (Fähigkeitsparameter³) auf der Grundlage der *Item-Response-Theorie* (IRT) durchgeführt, um die Ergebnisse über die unterschiedlichen Unterrichtsinhalte hinweg miteinander vergleichen zu können. Im Ergebnis korreliert das mittlere Schwierigkeitsniveau der Unterrichtsaufgaben mit dem mittleren Fähigkeitsniveau der Schüler über die 20 Klassen hinweg moderat positive ($r=.29$). Jedoch zeigten sich beträchtliche Unterschiede bzgl. der Passung der Aufgabenschwierigkeit und

der Schülerfähigkeit zwischen den untersuchten Klassen. Des Weiteren konnte in Studie 1 gezeigt werden, dass die entwickelte Methode zur Konstruktion von Lernaufgaben unterschiedlicher Schwierigkeit somit eine valide Möglichkeit darstellt, a priori die Aufgabenschwierigkeit zu variieren.

Aufgrund der in Studie 1 nachgewiesenen z.T. großen Diskrepanzen zwischen der Aufgabenschwierigkeit der Unterrichtsaufgaben und der Schülerfähigkeit wurde in der anschließenden Studie 2 eine Möglichkeit untersucht, diese Diskrepanzen zu verringern. Dies erfolgte auf der Grundlage der Idee des adaptiven Testens. Den Schülern wurden zu zwei Übungszeitpunkten Aufgaben vorgelegt, welche individuell an die jeweilige Schülerfähigkeit angepasst waren. Dabei wurden drei Untersuchungsgruppen verglichen, die sich in der mittleren Lösungswahrscheinlichkeit der zu bearbeitenden Aufgaben unterschieden. Eine Gruppe erhielt Aufgaben mit einer 30%-igen Lösungswahrscheinlichkeit (LW), welche als mäßig schwierig definiert werden, die anderen Gruppen Aufgaben mit 50%iger (mittelschwierige Aufgaben) bzw. 80%iger Lösungswahrscheinlichkeit (leichte Aufgaben). Es zeigte sich, dass die Fähigkeitsparameter mit Hilfe mäßig schwieriger Lernaufgaben (30%ige LW) signifikant erhöht werden konnten, während die Bearbeitung der Aufgaben mit 50% und 80%iger LW zu keinen Veränderungen der Schülerfähigkeit führte.

Die Bearbeitung von mäßig schwierigen Aufgaben, welche individuell an die jeweilige Schülerfähigkeit angepasst wurde, scheint nach den vorliegenden Befunden eine sinnvolle Möglichkeit zu sein, die Leistungen bzw. Fähigkeiten der Schüler in besonderer Weise zu erhöhen.

Die Ergebnisse lassen sich hinsichtlich der Befunde zur Wirksamkeit der Aufgabenschwierigkeit auf die Leistung diskutieren. Andere Studien (vgl. Nussbaum & Leutner, 1986) konnten im Gegensatz zu der vorliegenden Studie zeigen, dass Personen mit sehr leichten Aufgaben (95%ige LW) die besten Lernergebnisse zeigen. Auch die Forschungsbefunde von Huber (1966) und Sweller (1977) sind konträr zu den vorliegenden Ergebnissen, daher sollte die Untersuchung der Lernwirksamkeit der Aufgabenschwierigkeit weiter vorangetrieben werden.

Erste theoretische und auch praktische Implikationen lassen sich jedoch bereits ableiten. So lässt sich die Wirksamkeit der mäßig schwierigen Lernaufgaben mit Hilfe Wy-

¹ In dieser Arbeit werden Personen-, Berufs- oder sonstige Bezeichnungen aus Gründen besserer Lesbarkeit in der männlichen Form verwendet. Diese schließen auch weibliche Personen-, Berufs- und sonstige Bezeichnungen ein.

² Die vom Lehrer im Unterricht eingesetzten Aufgaben werden im Folgenden als Unterrichtsaufgaben bezeichnet.

³ Die Schülerleistung wird im Folgenden in Anlehnung an die Begrifflichkeiten der IRT als Fähigkeitsparameter bezeichnet.

gotskys (1978) Zone der proximalen Entwicklung erklären. Demnach befindet sich die Aufgabenschwierigkeit etwas oberhalb des aktuellen Wissens- und Fähigkeitsniveaus des Schülers und hat eine herausfordernde und lernförderliche Wirkung auf die weitere Entwicklung der Schüler. Als schulpraktische Implikation kann die Schlussfolgerung gezogen werden, dass eine vermehrte Binnendifferenzierung innerhalb der Klasse bzgl. der Passung zwischen Aufgabenschwierigkeit und Schülerfähigkeit notwendig erscheint. Dabei sollte aufgrund des Ergebnisses, dass auch leistungsschwächere Schüler von individuell mäßig schwierigen Aufgaben profitieren, gewährleistet werden, dass anspruchsvolle, jedoch nicht wie sich z.T. in Studie 1 gezeigt hat, übermäßig schwierige und überfordernde Unterrichtsaufgaben eingesetzt werden.

1 Einleitung

Der Einsatz von Lernaufgaben an deutschen Schulen ist ein Thema, das in der Folge der TIMS- und PISA-Studien allgemeine Beachtung auf sich gezogen hat. Dabei zeigten sich erhebliche Schwierigkeiten der deutschen Schüler bei Aufgaben des naturwissenschaftlichen Verständnisses und der Wissensanwendung. Dies wird als Hinweis darauf angesehen, dass in Deutschland ein wenig problem- und anwendungsorientierter naturwissenschaftlicher Unterricht vorherrscht. „Nach wie vor gilt es, die in Deutschland erkennbare Neigung zum fragend-entwickelnden und fachsystematischen Unterricht zu überwinden und durch Anwendungsbezüge, Problemorientierung sowie Betonung mentaler Modelle [...] die Entwicklung eines tiefer gehenden Verständnisses und flexibel anwendbaren Wissens zu fördern“ (Prenzel, Rost, Senkbeil, Häußler & Klopp, 2001, S. 245). Der in Deutschland herrschende „Mangel an Aufgabenkultur“ wurde auf der Suche nach Gründen für das mittelmäßige Abschneiden der deutschen Schüler im internationalen Vergleich als eine der Ursachen identifiziert.

Die einflussreiche BLK-Expertise (BLK, 1997) konstatiert, dass „in der Weiterentwicklung von Aufgabenstellungen und der Form der Bearbeitung [...] ein beträchtliches Potential zur Verbesserung des mathematisch-naturwissenschaftlichen Unterrichts“ (S. 89) liegt. Die Bedeutsamkeit von Unterrichtsaufgaben macht Krumm (1985) deutlich: „Jeder Lehrer konfrontiert im Laufe eines Schultages seine Schüler direkt oder indirekt mit einer großen Zahl von Aufgaben, Fragen, Anweisungen. Man kann Unterricht als den systematischen Versuch betrachten, Schüler zu befähigen, Aufgaben unterschiedlichster Art zu bewältigen“ (S. 102). Dieser sowohl zeitliche als auch inhaltlich hohe Stellenwert macht den Einfluss von Aufgaben auf die Lernleistung der Schüler deutlich. Auch Arnold und Neber (2004) sehen es als evident an, dass das Stellen von Fragen ein Kernelement erfolgreichen Lernens und Lehrens darstellt. Doch wie genau Aufgaben bzw. deren Bearbeitung zum Lernen beitragen, so die Autoren, sei schon aufgrund der begrifflichen Unschärfe von Fragen und Aufgaben nicht klar zu beantworten.

In der vorliegenden Arbeit wurde auf Lernaufgaben im Physikunterricht fokussiert, da die Untersuchungen im Rahmen des DFG-geförderten Projekts „Aufgaben im naturwissenschaftlichen Unterricht“ durchgeführt wurden, welches wiederum Bestandteil der Forschergruppe „Naturwissenschaftlicher Unterricht“ an der Universität Duisburg-Essen ist. Ein weiteres entscheidendes Kriterium für die Untersuchung des Physikunterrichts war, dass bereits Forschungsergebnisse zu Unterrichtsskripten in der Physik vorliegen (vgl. Seidel, 2003),

jedoch keine Ergebnisse zum Aufgabeneinsatz im Physikunterricht, sowie zur Passung der Aufgabenschwierigkeit und Schülerfähigkeit bekannt sind.

Bevor auf den theoretischen Forschungshintergrund der hier vorliegenden Arbeit eingegangen wird, soll im Folgenden ein Überblick über die inhaltliche Struktur gegeben werden.

In Kapitel 2 wird mit Definitionsversuchen des Konstrukts „Lernaufgabe“ und dessen Abgrenzung vom Konstrukt „Problem“ in den Inhaltsbereich eingeführt.

In Kapitel 3 werden die wichtigsten Forschungsergebnisse zum Aufgabeneinsatz vorgestellt. Ein Schwerpunkt liegt auf der Effektivität verschiedener Aufgabentypen (*low level-* und *high level-*Aufgaben, *adjunct questions* und *worked examples*) auf die Schülerleistung (Kapitel 3.2). In Kapitel 3.3 wird die Passung zwischen Aufgabenschwierigkeit und Schülerfähigkeit thematisiert, und im anschließenden Unterkapitel (3.4) wird auf die Bedeutung der diagnostischen Kompetenz der Lehrer zur Einschätzung der Aufgabenschwierigkeit und Schülerfähigkeit als Voraussetzung für deren Passung eingegangen.

In Kapitel 4 wird auf wichtige Theorien bzw. Modelle zur Erklärung des Einflusses von Lernaufgaben auf die Leistung eingegangen. Die ersten Überlegungen des Behaviorismus werden kurz aufgezeigt und ebenso kognitionstheoretische Modelle sowie Wygotskys (1978) Überlegungen zur „Zone der proximalen Entwicklung“. Auf die Theorien des Konstruktivismus wird ebenfalls eingegangen. Im Anschluss daran wird das Modell des schulischen Lernens von Bloom (1976) sowie das Modell der „direkten Instruktion“, welches auf der Grundlage der Ergebnisse des Prozess-Produkt-Paradigmas entwickelt wurde, thematisiert.

Das Kapitel 5 beschäftigt sich mit den verschiedenen Klassifikationssystemen von Lehrzielen, da die im Unterricht eingesetzten Lernaufgaben eine Möglichkeit darstellen, diese Lehrziele zu erreichen. Einer der prominentesten und verbreitetsten Versuche stellt die Taxonomie von Bloom et al. (1956) dar. Die dieser Arbeit in Teilen zugrunde liegende Revision von Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, Raths und Wittrock (2001) wird ebenfalls beschrieben und auf empirische Ergebnisse zur Lehrzieltaxonomie eingegangen.

Die Aufgabenkonstruktion nach Klauer (1987), welche die Grundlage für die Konstruktion der in dieser Arbeit entwickelten Lernaufgaben darstellt, wird neben den verschiedenen Formen der Schulleistungstests in Kapitel 6 thematisiert.

In Kapitel 7 wird anhand der bisherigen Forschungsergebnisse und offen gebliebener Fragen die dieser Arbeit zugrunde liegenden Untersuchungsfrage ausformuliert.

Mit den Fragestellungen der Studie 1 beginnt in Kapitel 8 der empirische Teil dieser Arbeit.

In Kapitel 9 werden Methode, Stichprobe, Design, Durchführung und Instrumente erläutert. Schwerpunktmäßig wird in diesem Kapitel die Aufgabenkonstruktion beschrieben.

Die Darstellung der Untersuchungsergebnisse der Studie 1 erfolgt in Kapitel 10. Aus den Schlussfolgerungen der Diskussion der Studie 1 werden Überlegungen für die Studie 2 angestellt (Kapitel 11).

Im darauf folgenden Kapitel 12 wird auf die Idee des adaptiven Testens und Unterrichtens als zusätzlicher theoretischer Hintergrund für die Studie 2 eingegangen.

Die Fragestellungen der Studie 2 werden in Kapitel 13 dargelegt.

Im Anschluss daran erfolgt in Kapitel 14 die Erläuterung der Methode der Studie 2. Dabei wird ein besonderes Augenmerk auf die Adaptation der Aufgaben bzw. deren Schwierigkeit an die individuelle Schülerfähigkeit gelegt.

Die Ergebnisse und Schlussfolgerungen aus Studie 2 werden in Kapitel 15 dargestellt. In Kapitel 16 werden die Ergebnisse der Studie 2 mit bisherigen Befunden aus der Literatur in Beziehung gesetzt und diskutiert.

Im Anschluss daran werden die Ergebnisse aus beiden Studien zusammengefasst und diskutiert. Auch wird in Kapitel 17 auf Schwächen der vorgestellten Untersuchungen eingegangen. Darüber hinaus wird ein Forschungsausblick gegeben und noch offene Fragen werden erläutert.

2 Definition von Lernaufgaben

Im folgenden Kapitel wird der Versuch unternommen, das Konstrukt der „Lernaufgabe“ zu definieren (Kapitel 1.1) und von dem Konstrukt des „Problems“ abzugrenzen (Kapitel 1.2).

2.1 Was sind Lernaufgaben?

In der Literatur wird keine eindeutige Trennung zwischen schriftlichen Aufgaben und mündlichen Fragen vorgenommen (vgl. Kapitel 3.2.2 *adjunct questions*). Daher werden im Folgenden Definitionen zu beiden Aufgabenformen berücksichtigt.

Andre (1979) definiert eine Frage (question) als „a direction to a learner to examine instructional material or his memory of it and to produce some response“ (p. 281). Aus dieser Definition wird deutlich, dass eine Frage bzw. Aufgabe sowohl mit Hilfe des Gedächtnisses als auch durch die Auseinandersetzung mit neuen Inhalten beantwortet werden kann. Doyle (1983) betont zusätzlich den Prozess der Aufgabenbearbeitung: „Academic tasks, in other words, are defined by the answers students are required to produce, and the routes that can be used to obtain these answers“ [...]. In other words, accomplishing a task has two consequences. First, a person will acquire information - facts, concepts, principles, solutions - involved in the task that is accomplished. Second, a person will practice operations- memorizing, classifying, inferring, analyzing- used to obtain or produce the information demanded by the task“ (pp. 161-162). Doyle (1983) hebt drei wesentliche Aspekte hervor, die bei der Beschreibung von Lernaufgaben berücksichtigt werden sollten:

1. Produkte: Aufgaben resultieren in Produkten, wie z.B. die Antwort auf eine Testfrage.
2. Handlungen: Aufgaben verlangen Handlungen, die ausgeführt werden müssen, um ein Ergebnis (Produkt) zu erhalten, z.B. eine Berechnung durchführen, etwas konstruieren.
3. Ressourcen: Für die Behandlung einer Aufgabe stehen den Schülern bestimmte Methoden zur Verfügung, die die Bearbeitung der Aufgabe beeinflussen.

Auch Lienert und Raatz (1998) betonen, dass eine Aufgabe grundsätzlich aus zwei Elementen besteht:

1. „dem Problem (bei Intelligenz- und Leistungstests) oder der Frage bzw. der Aufforderung zu einer Stellungnahme (bei Fragebogen),
2. der Problemlösung (bei Intelligenz- und Leistungstests) oder Schlüsselantwort (bei Fragebogen)“ (S. 18).

Diese zwei Elemente einer Lernaufgabe werden auch bei der Definition von Klauer (1987) deutlich: „Eine Aufgabe ist die Verknüpfung einer Stimuluskomponente mit einer Response-Komponente. Die Stimuluskomponente besteht aus einem bestimmten Inhalt, der in einer bestimmten Art und Weise vorgelegt wird. Die Response-Komponente besteht aus der Handlung, die an der Stimuluskomponente ausgeübt werden soll“ (S. 15). In der Definition von Klauer (1987) werden sowohl die Art und Weise (d.h. die Itemform), in der die Aufgabe vorgelegt wird, als auch die Handlungen zur Aufgabenlösung berücksichtigt. Dies ermöglicht eine genauere Analyse der kognitiven Anforderungen, die an den Aufgabebearbeiter gestellt werden. Verschiedene Autoren (Klauer, 1987; Seel, 1981) unterscheiden weiter zwischen Lernaufgaben, die im Unterricht bearbeitet werden, und Testaufgaben, welche zur Lernerfolgskontrolle eingesetzt werden.

Renkl (1991) verwendet den Begriff „Aufgabe“ in einem auf kleinste Interventionen bezogenem Sinn: „'Aufgabe' im hier verwendeten Sinn bezieht sich in vielen Fällen auf Lehrerfragen. Da z.B. die Lösung einer Textaufgabe in aller Regel durch das Stellen mehrerer Lehrerfragen erarbeitet wird, bildet eine Textaufgabe nicht eine [Hervorhebung durch Renkl, 1991], sondern in der Mehrzahl der Fälle mehrere Aufgaben (entsprechend der Zahl der Episoden)“ (S. 89). Im Sinne von *soliciting* (kommt einer Übersetzung von „Aufforderung“ am nächsten) umfasst der Begriff „Aufgabe“ Fragen, Anforderungen etc.; damit sind nach Renkl (1991) „öffentliche Leistungsanforderungen an den Schüler“ (S. 63) gemeint. Unter Antworten werden alle erbrachten Leistungen zusammengefasst.

Insgesamt werden in den meisten Definitionen des Konstrukts „Aufgabe“ eine Aufforderung sowie deren Bearbeitung durch den Schüler hervorgehoben. Für die dieser Arbeit zugrunde liegende Definition wurde auf die von Renkl (1991) zurückgegriffen, um ein möglichst umfassendes Bild der von Lehrern eingesetzten Unterrichtsaufgaben zu erhalten. Daher werden, ebenso wie bei Renkl (1991), neben den schriftlichen Lernaufgaben auch mündlich gestellte Unterrichtsfragen unter dem Begriff „Lernaufgabe“ zusammengefasst. Denn im häufig zu beobachtenden fragend-entwickelnden Unterricht, welcher von Seidel (2003a) für den deutschen Physikunterricht bestätigt wurde, wird ein Großteil der Aufgaben mündlich gestellt. Um somit der Realität an deutschen Schulen gerecht zu werden, wird folgende Ar-

beitsdefinition zugrunde gelegt: *Jede inhaltsbezogene, ergebnisorientierte Leistungsanforderung, die der Lehrer den Schülern stellt, wird als Lernaufgabe definiert.*

2.2 Abgrenzung von Aufgaben und Problemen

Keitel, Otte und Seeger (1980) wiesen darauf hin, „dass das Verhältnis von Problem und Aufgabe für die Untersuchung der Aufgabe als Gegenstand der Schülertätigkeit von zentraler Bedeutung ist. Üblicherweise wird das Aufgabelösen als „algorithmisch“ charakterisiert, während das Problemlösen gerade durch die Abwesenheit oder die Nicht-Verfügbarkeit über einen Algorithmus definiert wird“ (S.110).

Seel (1981) unterscheidet „Probleme“ und „Lernaufgaben“ folgendermaßen: Lernaufgaben tragen mentale Anforderungen an den Lernenden heran, die er aufgrund seines verfügbaren Vorwissens und bekannter Lösungsverfahren meistern kann. Folglich entscheiden das Vorwissen und das Vermögen eines Individuums, effektive Operationen zur Bewältigung einer Aufgabenstellung anzuwenden, darüber, ob es sich um eine Lernaufgabe oder um ein Problem, speziell für diese Person, handelt. Diese Definition impliziert, dass ein Problem höhere Anforderungen an eine Person stellt, da vorhandenes Wissen und bekannte Lösungsverfahren nicht zur Problemlösung ausreichen. Dagegen ist demnach für die Bewältigung von Aufgaben lediglich der Abruf deklarativen und prozeduralen Wissens erforderlich.

Keitel et al. (1980) und Seel (1981) stehen beide in der Tradition einer in der Psychologie weit verbreiteten Definition. Hiernach besteht ein Problem aus dem Vorhandensein eines „unerwünschten Ausgangszustands s_a , eines erwünschten Endzustands s_e und einer Barriere, die die Transformation von s_a nach s_e im Moment verhindert“ (Dörner, 1976, S. 10). Dörner grenzt Aufgaben von Problemen derart ab, dass Aufgaben auch ohne Beteiligung von generalisierbarem heuristischem Wissen (unspezifische Problemlösestrategien) gelöst werden können, was bei Problemen nicht der Fall ist. Seeling (2005) weist jedoch darauf hin, dass die o.g. Abgrenzung zwar notwendig ist, aber im Alltagshandeln sehr schwer anwendbar zu sein scheint: Es gibt kaum Probleme, bei denen nicht auch algorithmisches Wissen und spezifische, routinemäßige Lösungshandlungen, die immer zum gleichen Ergebnis führen, helfen. Umgekehrt können viele Aufgaben nicht ausschließlich algorithmisch erledigt werden, fast immer ist auch heuristisches Wissen erforderlich.

Auch Renkl (1991) betont, dass die Art der Anforderung, die eine Aufgabe darstellt, vom Kenntnisstand des Schülers abhängt. Daher ist eine eindeutige Unterscheidung zwischen Aufgaben und Problemen nicht möglich; das Vorhandensein bzw. das Nicht-

Vorhandensein bestimmter Wissensinhalte stellt die entscheidende Größe dar, ob eine Aufgabe eine solche ist oder für den Schüler ein Problem darstellt.

Probleme können sich in Bezug auf die Genauigkeit des Ausgangs- und Zielzustand sowie die Verfügbarkeit von zur Lösung notwendigen Operationen unterscheiden. Diese verfügbaren Operationen hängen von zuvor gemachten Erfahrungen mit ähnlichen Problemen ab und variieren somit von Person zu Person. Aufgrund dessen bestimmen diese verfügbaren Operationen neben anderen Faktoren die individuelle Schwierigkeit einer Aufgabe bzw. eines Problems für die einzelne Person. Ist das Vorwissen nicht bekannt bzw. kontrollierbar, kann nicht definiert werden, ob eine bestimmte Aufgabenstellung ein Problem oder eine Aufgabe darstellt. Eine eindeutige und gültige Einschätzung, ob eine Leistungsanforderung ein Problem oder eine Aufgabe darstellt, kann also nur individuell und unter Berücksichtigung des jeweiligen Vorwissens erfolgen. Diese Unterscheidung zwischen Aufgabe und Problem kann in der hier beschriebenen Studie 1 aufgrund fehlender Informationen über das Vorwissen der Schüler in Bezug auf die zu bearbeitenden Aufgaben nicht getroffen werden (vgl. Kapitel 9.2).

3 Forschungsergebnisse zur Effektivität von Lernaufgaben

Lernaufgaben sind Realität in jedem Klassenraum und stellen im Rahmen des Unterrichts und Lehrens die vom Lehrer meist genutzte Lehrmethode dar (Foster, 1983; Schiever, 1991). Ihre Effektivität bzgl. des Lernens der Schüler wird seit der griechischen Klassik (Sokratischer Dialog) angenommen (vgl. Arnold & Neber, 2004; Dillon, 1982a;). Die wissenschaftliche Überprüfung erfolgt im Gegensatz dazu erst seit kürzerer Zeit. Dillon (1982a) fasst folgendermaßen zusammen: „Asking and answering questions are among the most common human activities, yet it is remarkable how little is known, in a systematic way, about the effects of questions on a respondent“ (p. 127).

Da es bislang, vor dem Hintergrund des physikunterricht-bezogenen Kontextes der hier vorliegenden Arbeit, kaum empirische physikspezifische Ergebnisse zu Lernaufgaben gibt (Baumert & Köller, 2000), werden in den folgenden Ausführungen auch Ergebnisse aus anderen Inhaltsbereichen, besonders aus der Mathematik, vorgestellt, da in diesem Fachgebiet weit mehr Arbeiten vorliegen und eine inhaltliche Nähe zwischen den Fächern gegeben ist. So stellt die Mathematik eine bedeutsame Hilfswissenschaft für die Physik dar. Auch ist der kumulative Wissenserwerb in den Naturwissenschaften ohne entsprechendes mathematisches Wissen nicht möglich (Walter, Senkbeil, Rost, Carstensen & Prenzel, 2006). Besonders in der Physik wird das Erlernen, Verstehen und Anwenden inhaltlicher Konzepte häufig durch mathematische Kompetenzen begünstigt. In den PISA-Ergebnissen zeigten sich entsprechend hohe Korrelationen ($r = .83$) zwischen diesen Domänen (Prenzel et al., 2001).

In diesem Kapitel soll ein Überblick über den bisherigen Forschungsstand zu Lernaufgaben gegeben werden. In Kapitel 3.1 wird auf die Funktion von Lernaufgaben im Unterricht eingegangen. Kapitel 3.2 thematisiert die Effektivität verschiedener Aufgabenarten: *low level*- und *high level*-Aufgaben, *adjunct questions* und *worked examples*. In Kapitel 3.3 wird das Konzept der „Passung“ zwischen Aufgabenschwierigkeit und Schülerfähigkeit thematisiert, bevor in Kapitel 3.4 die diagnostische Kompetenz von Lehrern bei der Einschätzung der Aufgabenschwierigkeit ausgearbeitet wird.

3.1 Funktion und Effektivität von Lernaufgaben

Lernaufgaben dienen im Unterrichtsprozess sowohl der Steuerung und Strukturierung des Unterrichts durch den Lehrer als auch zur Anregung der Schüleraktivitäten und um den Lernprozess zu initiieren, zu steuern und zu kontrollieren (Levin & Long, 1981; Seel, 1981).

Für Schüler haben Lernaufgaben die Funktion, eine Lernmöglichkeit herzustellen, die es ihnen erleichtert, sich aktiv am Unterricht zu beteiligen. Lernaufgaben beeinflussen, wie Schüler die präsentierten Informationen aufnehmen und in ihr Vorwissen integrieren. Wird durch die Lernaufgaben das Vorwissen aktiviert, so wirkt sich dies positiv auf die Lernleistung aus – insbesondere, wenn man den kumulativen und hierarchischen Aufbau der Lerninhalte betrachtet (Levin & Long, 1981). Insgesamt stellen Lernaufgaben im Unterricht eine Möglichkeit dar, Verständnis zu fördern und Wissen zu festigen. Zudem verdeutlichen Lernaufgaben, welche Inhalte und Fertigkeiten der Lehrer als relevant erachtet (Crooks, 1988). Die verschiedenen Funktionen von Lernaufgaben im Unterricht werden meistens heuristisch erfasst, eine empirische Überprüfung der meisten dieser Funktionen steht noch aus.

Es wurde in verschiedenen Studien gezeigt, dass Lehrer an einem typischen Schultag zwischen 300 und 400 Fragen stellen (z.B. Gall, 1970). Ebenfalls konnten Niegemann und Stadler (2001) in einer deskriptiven Analyse zeigen, dass Lehrer durchschnittlich 82 Fragen pro Unterrichtsstunde stellen. Das bedeutet, dass im Mittel der Lehrer alle 30 Sekunden eine Frage stellt. Die Streuung zwischen den untersuchten Lehrern ist allerdings beträchtlich, sie reicht von 60 bis 136 Fragen pro Unterrichtsstunde.

Dabei können Lernaufgaben nach Renkl (1991) mindestens anhand dreier Gesichtspunkten betrachtet werden:

1. Eine intensive Nutzung der Unterrichtszeit kann eine hohe Aufgabenhäufigkeit in einer bestimmten Klasse bewirken: Die lernförderliche Wirkung der intensiven Nutzung der Unterrichtszeit für fachliche Inhalte ist ein eindeutiges Ergebnis der Unterrichtsforschung (vgl. Brophy & Good, 1986; Rosenshine, 1979).
2. Eine große Anzahl von Aufgaben kann durch intensive Verwendung eines lehrerzentrierten Frage-Antwort-Unterrichts (Unterrichtsgespräch) oder aber durch Übungs- bzw. Stillarbeitsphasen zustande kommen. Die Forschungsergebnisse zu diesem Punkt sind weniger eindeutig. Es ist anzunehmen, dass es sowohl ein „Zuviel“ als auch ein „Zuwenig“ an Stillarbeit gibt (Renkl, 1991). So haben Soar und Soar (1979) einen kurvilinearen Zusammenhang zwischen dem zeitlichen Ausmaß der Stillarbeitsphasen und dem Leistungszuwachs festgestellt. Helmke und Schrader (1988) konnten keine bedeutsame Korrelation zwischen der Häufigkeit von Stillarbeitspha-

sen und Leistungszuwachs nachweisen. Die Autoren nehmen an, dass nicht die Quantität, sondern die Qualität der Stillarbeitsphasen die bedeutsame Variable sei. Diese Vermutung wird jedoch nicht weiter geprüft. Aufgrund des momentanen Stands der Forschung können keine gesicherten Aussagen über den optimalen Anteil der Stillarbeitsphasen an der Unterrichtszeit gemacht werden.

3. Ein Unterrichtsgespräch ist häufig durch den Einsatz von schnell-getakteten, „kleinschrittigen“ einzelnen Fragen gekennzeichnet. So zeigte Neubrand (2002) für die Reanalyse der TIMSS-Daten zum Mathematikunterricht in Deutschland, dass der fragend-entwickelnde Unterrichtsstil eine Aufgabe in sehr viele einzelne Teilschritte zerlegt. Diese Teilaufgaben werden meist als Fragen mündlich gestellt und im Klassenverband (ohne Stillarbeit) bearbeitet. Allerdings hat sich das Unterrichtsgespräch hauptsächlich in Untersuchungen, welche den Erwerb von Fertigkeiten in der Grundschule erforschten, bewährt und ist somit nicht ohne weiteres auf andere Bereiche des schulischen Lernens zu übertragen (z.B. Peterson, 1979; Rosenshine, 1987).

Im angelsächsischen Sprachraum stellen das darbietende Lehren (*expository teaching* im Sinne von Ausubel, 1974) bzw. die direkte Instruktion (im Sinne von Rosenshine, 1987; vgl. Kapitel 4.8) die zentralen Unterrichtsmethoden dar. Diese Methoden nutzen in hohem Maße Aufgaben, deren primäre Funktion nach Gall und Artero-Boname (1995) darin besteht, das Wissen der Schüler zu überprüfen. Auch in Deutschland werden der Frontalunterricht und das erarbeitende Klassengespräch, zumindest in der Sekundarstufe, als Unterrichtsmethode dominant eingesetzt (Hager, Patry & Brenzing, 2000); auch bei diesen beiden Methoden nimmt die vom Lehrer gestellte Lernaufgabe eine zentrale Stellung ein.

Die Bedeutung von Aufgaben für den Kompetenzzuwachs wurde im Rahmen der PISA-I-Plus-Erhebung untersucht (Kunter, Dubberke, Baumert, Blum, Brunner, Jordan, Klusmann, Krauss, Löwen, Neubrand & Tsai, 2006). Ein Teilziel dieser Analyse im Rahmen der COACTIV-Studie war die Identifikation von Faktoren, welche die mathematische Kompetenzentwicklung beeinflussen. Dazu wurden u.a. die im Mathematikunterricht den Schülern zur Verfügung gestellten Lerngelegenheiten als mögliches *Potential zur kognitiven Aktivierung* analysiert. Dieses Potential wird als relevant für den Lehr-Lernprozess im Unterricht angesehen (Baumert, Kunter, Brunner, Krauss, Blum & Neubrand, 2004; Ditton, 2006). Das Ausmaß des Potentials zur kognitiven Aktivierung, welches durch das Stellen und Bearbeiten von Aufgaben (Aufgabenimplementation) ermöglicht wird, variiert zwischen den Lehrern bzw. Unterrichtsstunden deutlich im Bezug auf die Komplexität (vgl. Neubrand, 2002).

Bei der Analyse von Aufgaben, die in Klassenarbeiten eingesetzt wurden, wurden von Kunter et al. (2006) folgende Aufgabenklassen unterschieden: (1) technische Aufgaben,

die das Abarbeiten einer bekannten Prozedur erfordern und (2) Modellierungsaufgaben. Modellierungsaufgaben wurden unterschieden in (2a) rechnerische Modellierungsaufgaben, bei denen prozedurales Denken während der Verarbeitung nötig ist, und (2b) begriffliche Modellierungsaufgaben, bei denen vorwiegend konzeptuelles Denken notwendig ist. Die Analyse von durchschnittlich 53 Klausuraufgaben pro Lehrer (N= 161) entspricht einer Gesamtstichprobe von etwa 8533 analysierten Aufgaben. Es zeigt sich, dass knapp die Hälfte aller Aufgaben (49%) lediglich technisches Wissen bzw. das Abarbeiten einer Prozedur verlangten (Gruppe 1) und weitere 43 Prozent rechnerisches Modellieren erforderten (Gruppe 2a). Aufgaben, für deren Lösung begriffliches, problembezogenes Modellieren und konzeptuelles Denken nötig war (Gruppe 2b), machen mit 8 % den geringsten Anteil der Aufgabenmenge aus. Daraus wurde der Schluss gezogen, „dass die gestellten Aufgaben nur in geringem Maße Potentiale zur kognitiv herausfordernden Auseinandersetzung mit dem Lerngegenstand eröffnen“ (Kunter et al., 2006, S. 179).

In der weiteren Analyse wurde das kognitive Potential - erfasst über die eingesetzten Aufgaben – von Kunter et al. (2006) mit anderen Unterrichtsmerkmalen in einem kovarianzanalytischen Mehrebenenmodell kombiniert, um die Annahme einer Längsschnittuntersuchung zu prüfen, ob Unterschiede in diesen Merkmalen (tatsächlich) Unterschiede in der Mathematikleistung erklären bzw. vorhersagen können. Dabei stellte das kognitive Potential einen signifikanten Prädiktor zur Vorhersage des Kriteriums *Mathematikleistung am Ende der 10. Klasse* dar ($b= 0.36$; $p < .05$). Dieses Ergebnis bedeutet, dass Schüler mit gleichen Eingangsvoraussetzungen, in deren Klassen Aufgaben mit hohem kognitiven Aktivierungspotential gestellt werden, demnach höhere Werte im Mathematiktest in der 10. Klasse erzielen als Schüler in Klassen mit Aufgaben mit niedrigerem kognitiven Potential (Kunter et al., 2006). Die Wirkrichtung, dass Lehrer möglicherweise erst dann vermehrt anspruchsvolle Aufgaben einsetzen, wenn ein hohes Leistungsniveau gegeben ist, wurde in einem weiteren Modell überprüft, wobei die mittlere Klassenleistung in Klasse 9 als Prädiktor für das Unterrichtsmerkmal *kognitives Potential* in Klasse 10 herangezogen wurde. Es konnte tatsächlich eine Anpassung des Unterrichts in Abhängigkeit vom Leistungsniveau der Klasse bestätigt werden. Dieses Ergebnis wird von den Autoren (Kunter et al., 2006) als Beleg für einen förderlichen Effekt eines kognitiv anregenden Unterricht auf die mathematische Leistung gesehen. Insgesamt konnte somit das *kognitive Potential* von Aufgaben für die Leistung bestätigt werden. Auch liefern die Ergebnisse einen Beleg für adaptives Verhalten der Lehrer, „die mit ihrer Unterrichtsgestaltung offensichtlich zu einem gewissen Grad auf das spezifische Leistungsniveau ihrer Klasse reagieren“ (Kunter et al., 2006, S. 186). Inwieweit dieses Ergebnis, welches anhand von Klausuraufgaben im Mathematikunterricht untersucht wurde,

auf die im regulären Physikunterricht eingesetzten Aufgaben übertragbar ist, muss noch untersucht werden.

3.2 Bestimmung der Schwierigkeit von Aufgaben

Die Schwierigkeit der Unterrichtsaufgaben wird meist a posteriori durch den Prozentsatz der richtigen Antworten bestimmt, wobei die meisten Studien eine relativ hohe Rate an richtigen Antworten und somit leichte Aufgaben als effizient für die Schülerleistung ausweisen (70% bis 95%, vgl. Brophy & Good, 1986, Rosenshine, 1987). Diese Bestimmung der Schwierigkeit entspricht dem Vorgehen der klassischen Testtheorie, deren Definition wie folgt lautet: „Der Schwierigkeitsindex einer Aufgabe ist gleich dem prozentualen Anteil P der auf diese Aufgabe entfallenden richtigen Antworten in einer Analysestichprobe von der Größe N ; der Schwierigkeitsindex liegt also bei schwierigen Aufgaben niedrig, bei leichten hoch“ (Lienert & Raatz, 1998, S. 32). Wie aus der Definition deutlich wird, hängt bei dieser Berechnung die Schwierigkeit von der Personenstichprobe, welche die Aufgaben bearbeitet, ab. Williams und Clarke (1997) sowie Christiansen und Walther (1986) betonen, dass sich die tatsächliche Aufgabenschwierigkeit erst aus der Interaktion zwischen Schüler und Aufgabe ergibt. Auch nach Aebli (2006) kann die Schwierigkeit einer Aufgabe nicht aus ihren objektiven Zügen, sondern immer nur relativ zum Vorwissen des Menschen, der sie bearbeitet, bestimmt werden. Des Weiteren darf der empirische Schwierigkeitsgrad nicht mit dem kognitiven Anforderungsniveau verwechselt werden, denn die Lösungswahrscheinlichkeit wird von verschiedenen Faktoren (wie dem Vorwissen, durchgenommenem Inhaltsbereich usw.) beeinflusst (Klieme, 2000).

Da die Aufgabenschwierigkeit mit Hilfe der Lösungshäufigkeit erst nach der Bearbeitung berechnet werden kann, wird in einem Großteil der Forschung zur Effektivität unterschiedlicher Aufgabenarten versucht, die Schwierigkeit von Aufgaben a priori über verschiedene Aufgabenmerkmale zu variieren. Kennzeichnend für diesen Ansatz ist das Bemühen, Aufgaben je nach Komplexität der intendierten Denktätigkeit zur Aufgabenlösung zu unterscheiden (Seel, 1981).

Die wohl meist verwendete Klassifikation wurde auf der Grundlage der Bloom'schen Taxonomie (Bloom et al., 1956) vorgenommen. In der Regel werden in Studien die sechs Stufen der Taxonomie (1) Wissen, (2) Verstehen, (3) Anwenden, (4) Analyse, (5) Synthese und (6) Evaluation auf zwei Stufen reduziert (vgl. Crooks, 1988). In den meisten Untersuchungen erfolgt diese Unterteilung in Aufgaben des Wissensabrufs aus dem Gedächtnis (Wissens- bzw. *low level*-Aufgaben) und Aufgaben der Informationsverarbeitung (einige

bzw. alle weiteren kognitiven Prozesse) als *high level*-Aufgaben zusammengefasst (vgl. Brophy & Good, 1986). Diese Zweiteilung der Aufgaben wird in der Forschung vor allem aus pragmatischen Gründen genutzt. Dabei werden leichtere Aufgaben meist als solche definiert, bei denen lediglich die wörtliche oder in eigenen Worten erfolgende Wiedergabe des zuvor gelernten Materials erfolgt. Dafür müssen Schüler Fakten, Methoden, Prozesse, Definitionen etc. aus dem Gedächtnis abrufen (Hunkins, 1976; Winne, 1979). Diese Aufgabenart wird in der Literatur auch als Fakten-, direkte, *recall*- oder Wissensaufgabe bezeichnet. Im Gegensatz dazu werden kognitiv anspruchsvollere Aufgaben als solche definiert, bei denen der Schüler zuvor gelernte Informationen neu zusammenstellen oder aus den gelernten Informationen Schlussfolgerungen ziehen muss, um die Aufgabe beantworten zu können (Hunkins, 1976). Diese Aufgabenart wird in der Literatur auch als *high level*-, interpretative, Bewertungs-, Schlussfolgerungs- oder Synthesaufgabe bezeichnet. Jedoch unterscheiden sich die Definitionen der verschiedenen Autoren von *high level*-Aufgaben z.T. sehr (vgl. Crooks, 1988). Die allen Untersuchungen zum Effekt von *high level*-Aufgaben zugrunde liegende Annahme geht jedoch davon aus, dass dieser Aufgabentyp beim Schüler „höhere“ kognitive Prozesse auslöst und somit schwieriger als *low level*-Aufgaben ist (Hamaker, 1986).

3.2.1 Low level- vs. high level-Aufgaben

Einsatzhäufigkeit der beiden Aufgabenformen

Brophy und Good (1986) und Wilen (1991) konnten zeigen, dass der Großteil der vom Lehrer gestellten Aufgaben auf dem *low level*-Niveau klassifiziert wird. Je nach Studie sind dies zwischen 42% und 95% aller von Lehrern gestellten Fragen (Wilen & Clegg, 1986). Nur etwa 20 Prozent der Aufgaben fördern bzw. verlangen vom Schüler zur erfolgreichen Beantwortung höhere kognitive Prozesse, wobei meist einige bzw. alle fünf höheren kognitiven Prozesse der Bloom'schen Lehrzieltaxonomie zusammengefasst betrachtet werden (Arnold et al., 1973; Au & Mason, 1981; Crooks, 1988; Fleming & Chambers, 1983; Floyd, 1960; Gallagher, 1965; Goodlad, 1983; Lindfors, 1980; Wilson, 1969). Graesser und Person (1994) konnten lediglich 4% der Lehreraufgaben als *high level* klassifizieren. Renkl (1991) zeigte in seiner Arbeit, dass im Mittel mehr als zweimal so viele performanzorientierte (welche eine Nähe zu *low level*-Aufgaben zeigen) als strukturorientierte (Nähe zu *high level*-Aufgaben) Aufgaben im Mathematikunterricht an deutschen Grundschulen gestellt werden. „Diese Dominanz der weniger komplexen Aufgaben stimmt mit Ergebnissen zum Verhältnis zwischen *low level*- und *high level*-Fragen aus etlichen Untersuchungen überein“

(Renkl, 1991, S. 110). Des Weiteren konnte Renkl (1991) zeigen, dass die Häufigkeiten der beiden Aufgabenarten unkorreliert waren ($r = -.05$), so dass die Betonung der einen Aufgabenart nicht notwendigerweise zu einer Vernachlässigung der anderen führt.

Es werden verschiedene Überlegungen und mögliche Gründe zum schwerpunktmäßigen Einsatz von *low level*-Aufgaben durch Lehrer angeführt. Ellis (1993) geht davon aus, dass ein Lehrer hauptsächlich niedrig-kognitive Aufgaben einsetzt, um ein geringes Unterrichtstempo zu verhindern, welches durch eine längere Bearbeitungszeit der *high level*-Aufgaben entstehe, und um die Aufmerksamkeit der Schüler aufrecht zu halten. Ropo (1990) führt an, dass durch *low level*-Aufgaben fast alle Schüler aktiviert werden und sich somit am Unterricht beteiligen können, was bei *high level*-Aufgaben nur für eine Minderheit der Schüler gilt. Zum anderen sind *low level*-Aufgaben auch für den Lehrer leichter, sowohl in der Konstruktion als auch in der Überprüfung der gegebenen Antworten (Johnston, 1983). Als weiteren Grund nennt Ropo (1990) die grundlegende Tendenz von Lehrern, Faktenwissen, welches in der Folge aus dem Gedächtnis abgerufen werden kann, zu unterrichten.

Der sehr geringe Anteil von Unterrichtsaufgaben höheren kognitiven Niveaus wird von Niegemann und Stadler (2001) kritisch gesehen, da möglicherweise die Chance zur Förderung des Lernens von Zusammenhängen vergeben und kein tieferes Verständnis des Inhaltsbereichs gefördert wird (vgl. Brualdi, 1998). Dillon (1994) nimmt an, dass *low level*-Aufgaben weder das Denken der Schüler anregen, noch die aktive Teilnahme am Unterricht fördern. Gall (1984) kommt zu dem Schluss, dass der Aufgabeneinsatz an die jeweils zu unterrichtenden Schüler anzupassen sei. Sollen jüngere, leistungsschwächere Schüler grundlegende Fertigkeiten lernen, erscheint der Schwerpunkt auf niedrig kognitiven Aufgaben förderlich. Im Gegensatz dazu seien für leistungsstärkere Schüler kognitiv anspruchsvollere Aufgaben sinnvoll. Doyle (1988) nimmt an, dass schwierige Aufgaben, die mit Ambiguitäten verbunden sind, ein geeignetes Mittel zur Förderung von Verständnis und Problemlösekompetenzen sind. Aus diesen Annahmen wird die Forderung nach einer Kombination aus kognitiv niedrigeren und höheren Aufgaben abgeleitet (vgl. folgender Abschnitt „Einfluss auf die Schülerleistung“). Hierbei muss die Leistungsfähigkeit der Schüler berücksichtigt werden, um so diejenige Art von Aufgaben mehrheitlich einzusetzen, welche der Leistungsfähigkeit der Klasse, bzw. je nachdem wie individuell die Adaptivität der Aufgabenauswahl erfolgt, des einzelnen Schülers entspricht (Brualdi, 1998).

Einfluss auf die Schülerleistung

Auch wenn Leistungsunterschiede in Bezug auf den Einsatz von *low level*- und *high level*-Aufgaben seit einiger Zeit diskutiert werden, sind die Forschungsergebnisse dazu nicht

eindeutig. Diese wurden in verschiedenen Metaanalysen zusammengefasst und sollen im Folgenden kurz dargestellt werden. Der Aufbau der berücksichtigten Studie sah folgendermaßen aus: der Einfluss des Einsatzes der *low level*- und *high level*-Aufgaben des Lehrers im Unterricht wurde mit Hilfe eines Leistungstests, welchen die Schüler vor und nach dem Unterricht individuell bearbeiteten, untersucht.

Rosenshine (1971) schließt aus der Durchsicht von 10 Studien zum Zusammenhang zwischen dem Einsatz der beiden Aufgabenarten und den Schülerleistungen, dass kein eindeutiger linearer Zusammenhang zwischen der Frequenz des Einsatzes von *low level*- und *high level*-Aufgaben und der Schülerleistung besteht. Winne (1979) hat in seinem Review 18 experimenteller und quasiexperimenteller Studien zum Einfluss des Einsatzes der Aufgabenarten im Unterricht auf die Schülerleistung berücksichtigt und in drei Kategorien eingeteilt:

1. Studien, welche die Effektivität von *high level*-Aufgaben gegenüber *low level*-Aufgaben zeigen konnten,
2. Studien, welche den gegenteiligen Effekt erzielten und
3. Studien, die keine signifikanten Unterschiede in der Leistung der Schüler zwischen der Bearbeitung von *high level*- und *low level*-Aufgaben zeigen konnten.

Es wurde die Anzahl der Studien bzw. der prozentuale Anteil an der Gesamtanzahl der Studien zwischen den verschiedenen Kategorien miteinander verglichen (sog. *vote counts*). Die Studien der dritten Kategorie machten den größten Anteil aus (60% der Studien). Im Gegensatz dazu zeigten jeweils 15% einen signifikanten Effekt für *high level*-Aufgaben bzw. für *low level*-Aufgaben. Daraus zog Winne (1979) die Schlussfolgerung, dass *high level*-Aufgaben nicht effektiver sind als *low level*-Aufgaben.

Redfield und Rousseau (1981) kritisierten an der Methode der *vote counts*, dass so die Effektstärke nicht ermittelt werden kann. Sie haben in ihrer Metaanalyse 13 der von Winne (1979) berücksichtigten Studien sowie eine weitere Studie⁴ analysiert. Der Effekt der Schwierigkeit (*high level* vs. *low level*) der Unterrichtsaufgaben auf die Schülerleistung wurde mit Hilfe von Effektstärken verglichen. Es konnte eine relativ hohe Effektstärke ($d = 0.73$) zugunsten der *high level*-Aufgaben nachgewiesen werden. Das bedeutet, dass ein vorherrschender Einsatz von *high level*-Aufgaben während des Unterrichts einen positiven Effekt auf die Schülerleistung hat. Die Autoren schließen aus diesem Ergebnis, dass Aufgaben auf kognitiv höherem Niveau dann zu besseren Leistungen führen, wenn sie regelmäßiger Bestandteil des Unterrichts sind. Die Tatsache, dass Schüler, die es gewohnt sind, *high level*-Aufgaben zu bearbeiten, von diesen auch profitieren (Redfield & Rousseau, 1981), legt den

⁴ Diese zusätzlich berücksichtigte Studie wurde erst nach der Metaanalyse von Winne (1979) veröffentlicht.

Schluss nahe, dass man Schülern Gelegenheit geben sollte, sich mit solchen Aufgaben auseinander zu setzen. Auch sollten in den anschließenden Tests entsprechend *high level*-Aufgaben gestellt werden. Das Niveau der Unterrichtsaufgaben muss also mit dem Niveau der Testaufgaben korrespondieren (Levin, 2005). Die Ergebnisse der Metaanalyse sollten jedoch mit Vorsicht betrachtet werden, da lediglich in 2 der 14 Studien in der Experimentalgruppe mehr als 200 Personen teilnahmen. Die Stichprobengrößen der anderen Studien variieren zwischen 35 und 64 Teilnehmern, wodurch die Generalisierbarkeit der Ergebnisse gemindert ist (Redfield & Rousseau, 1981).

Die Metaanalyse von Samson, Strykowski, Weinstein und Walberg (1987), in der 14 Studien berücksichtigt wurden, ergab ebenfalls Hinweise auf eine Überlegenheit von *high level*-Aufgaben bezüglich der Schülerleistungen, jedoch verglichen mit den Ergebnissen der Metaanalyse von Redfield und Rousseau (1981) mit einer geringeren Effektstärke ($d=0.26$). Auch Hamaker (1986) konnte in seiner Metaanalyse nachweisen, dass *high level*-Aufgaben im Vergleich zu *low level*-Aufgaben einen generellen erleichternden Effekt mit einer Effektstärke von $mean ES=0.15^5$, haben.

Zu den Untersuchungen ist anzumerken, dass keiner der Autoren die alleinige Verwendung einer der beiden Aufgabenarten propagiert. Die Forschungsperspektive richtet sich vielmehr darauf, ein günstiges „Mischungsverhältnis“ zu finden. Burnstein (1980), Renkl (1991), Renkl und Helmke (1992) sprechen sich für eine Kombination aus beiden Aufgabenarten aus, so dass sowohl basale Fertigkeiten und Wissen als auch komplexere Prozesse, wie Schlussfolgerungen aus den gelernten Informationen ziehen, von den Schülern erlernt werden können. Dieses Mischungsverhältnis wird in mehreren Studien über den relativen Anteil der beiden Fragenarten operationalisiert (z.B. 75% *low level*- und 25% *high level*-Aufgaben). So untersuchten beispielsweise Gall, Ward, Berliner, Cahen, Winne, Elashoff und Stanton (1978) den Einfluss der Anzahl *high level*-Aufgaben in drei experimentellen Bedingungen. Der Anteil dieser Aufgaben an allen vom Lehrer gestellten Aufgaben variierte zwischen den Gruppen mit 25%, 50% oder 75%. Insgesamt konnte gezeigt werden, dass die Erinnerung von Fakten sowie anspruchsvolleren Informationen und Transfer für die 25% und 75% Bedingung signifikant höher war als für die 50%- und Kontrollbedingung. Dieser kurvilineare Zusammenhang zwischen den Untersuchungsgruppen und den verschiedenen Leistungsmessungen wird mit der Annahme zu erklären versucht, dass in der 50%-Gruppe dadurch, dass keine der beiden Aufgabenarten dominiert, weder Faktenlernen noch komplexeres Lernen

⁵ Hamaker (1986) benutzt in seiner Metaanalyse als Maß für die Effektstärke die Differenz des Prozentsatzes der richtigen Antworten im Leistungstest zwischen der Gruppe, die zuvor *high level*-Aufgaben bearbeitet hat, und der Gruppe, die zuvor *low level*-Aufgaben bearbeitet hat. In Anlehnung an Cohen (1977) wird eine Effektgröße nach dieser Berechnung zwischen .08 und .13 als mittlerer bis großer Effekt bezeichnet.

(wie Transfer) gefördert wird. Eine empirische Überprüfung dieser Hypothese steht noch aus.

Auch Brophy und Good (1986) nehmen an, dass eine Mischung von Aufgaben unterschiedlicher kognitiver Niveaus zu den besten Schülerleistungen führt. Vorgeschlagen wurde aufgrund der Ergebnisse von Tobin und Capie (1982) ein Verhältnis von 2/3 *high level*-Aufgaben zu 1/3 *low level*-Aufgaben. Renkl (1991) führt in diesem Zusammenhang an, dass solche Aussagen „Allgemeinplätze [bleiben], solange man aus der einschlägigen Forschung keine einigermaßen gesicherten Präskriptionen ableiten kann“ (Renkl, 1991, S. 73). Es besteht demnach noch Forschungsbedarf, bevor gesicherte Empfehlungen zum Verhältnis der verschiedenen Aufgabentypen gemacht werden können (vgl. Resnick & Ford, 1981).

Obwohl aus den bisherigen Forschungsergebnissen insgesamt keine eindeutigen Schlüsse gezogen werden können, gibt es seit Ende der 1970er Jahre kaum neue Untersuchungen, die zur Klärung dieser Frage beitragen (Klieme, 2000).

3.2.2 Textbegleitende Fragen (*adjunct questions*)

Es gibt eine Reihe von Bemühungen, das Verstehen von Lehrtexten mit Hilfe von Lernaufgaben zu erleichtern. Diese in die Texte eingeschobenen Fragen werden als *adjunct questions* bezeichnet (vgl. Andre, 1979; Hamaker, 1986; Rothkopf, 1972). Da die Forschung zu dieser speziellen Aufgabenform einige wichtige Hinweise auf die Wirksamkeit unterschiedlicher Aufgabentypen gibt, sollen die Ergebnisse dieser Forschungsrichtung im Folgenden skizziert werden.

Theoretische Grundlage der Verwendung textbegleitender Fragen ist meist die These von Rothkopf (1966), dass ein erfolgreiches Lernen mit Texten vor allem dann gewährleistet sei, wenn beim Leser sog. „mathemagene“, d.h. auf Erkenntnis und Lernen gerichtete Aktivitäten hervorgerufen werden. *Adjunct* Lernaufgaben sollen sowohl die Richtung als auch die Tiefe des Verarbeitungsprozesses beeinflussen.

Es wurden mehrere Metaanalysen (Anderson & Biddle, 1975; Crooks, 1988; Duchastel, 1981; Häfele, 1995; Hamaker, 1986; Hamilton, 1989) zum Effekt von verschiedenen Typen von *adjunct questions* auf die Leistung durchgeführt. Die meisten in den Metaanalysen berücksichtigten Primärstudien sind folgendermaßen aufgebaut: Die verschiedenen Gruppen bearbeiten einen Text und bekommen entweder vor, während oder nach der Textbearbeitung die zu bearbeitenden Aufgaben vorlegt, welche sich in einer zweiten Variable je nach Studie ebenfalls unterscheiden können: in *low level*- und *high level*-Aufgaben. Im Anschluss an die Bearbeitung bekommen alle Gruppen einen Leistungstest mit identischen

Aufgaben vorgelegt. Die Ergebnisse 50 berücksichtigter Studien bzgl. der *low level*-Aufgaben weisen eine Effektstärke von $d= 1.0$ im Vergleich zu einer Kontrollgruppe, welche den Text ohne Aufgaben bearbeitet hatte, in Bezug auf die Leistung auf. Dies gilt sowohl für Aufgaben, die vor, als auch für Aufgaben, welche nach dem Text gestellt wurden. Auch „verwandte“ Aufgaben, also solche Aufgaben, die inhaltlich nicht direkt im Text thematisiert wurden, jedoch zum gleichen Thema gehören, wurden mit einer Effektstärke von $d= 0.56$ für vorher und nachher gestellte Fragen ausgewiesen. Lediglich Aufgaben ohne inhaltlichen Bezug führten zu einer negativen Effektstärke ($d= -0.31$) (vgl. Hamaker, 1986). Insgesamt bestätigen diese Ergebnisse den erwartbaren Effekt, dass Schüler, die zuvor *adjunct questions* bearbeitet haben, signifikant bessere Testleistungen zeigen als Schüler, die zuvor keinerlei Aufgaben bearbeitet haben.

In einer Metaanalyse mit 21 berücksichtigten Studien zum Effekt von *adjunct high level*-Aufgaben auf die Schülerleistung wurden diese nicht - wie bei der Analyse der *low level*-Aufgaben - mit einer Kontrollgruppe verglichen, sondern mit einer Gruppe, die *adjunct low level*-Aufgaben bearbeitet hat (Hamaker, 1986). Im Vergleich dieser beiden Gruppen führen die *adjunct high level*-Aufgaben insgesamt zu besseren Leistungen, sowohl bei wiederholten ($d= 1.0$) als auch bei inhaltlich verwandten ($d= 0.3$) und Aufgaben ohne inhaltlichen Bezug ($d= 0.3$) (Hamaker, 1986).

Rothkopf und Bisbicos (1967) erklären das Ergebnis des lernförderlichen Effekts der *adjunct questions* mit der Aufmerksamkeitsfokussierung durch die Vorgabe von textbegleitenden Fragen (vgl. Anderson & Biddle, 1975; Hamaker, 1986; Wittrock, 1986). Der Lernende fokussiert demnach auf die Inhalte, welche für die Beantwortung der Aufgaben relevant sind. Lernende, denen *high level*-Aufgaben gestellt werden, haben dabei einen breiteren Aufmerksamkeitsfokus als Lernende, denen *low level*-Aufgaben gestellt wurden (Rothkopf & Bisbicos, 1967). Die dem jeweiligen Textsegment vorangestellten Fragen lenken wahrscheinlich bei der folgenden Verarbeitung die Aufmerksamkeit bevorzugt auf die fragenrelevanten Informationen, so dass hier vor allem ein intentionales Lernen stattfindet (Schnotz, 1994). Andre (1979) nimmt an, dass die präsentierten Informationen vom Schüler mehr oder weniger tief verarbeitet werden und eine tiefere Verarbeitung zu einer besseren Erinnerungsleistung führt (vgl. Anderson, 1970). Diese tiefere Informationsverarbeitung, soll durch *high level*-Aufgaben aufgrund der zur Beantwortung notwendigen höheren kognitiven Verarbeitung der Informationen im Vergleich zu den *low level*-Aufgaben eher gegeben sein. Die dem jeweiligen Textsegment nachgestellten Fragen hingegen bewirken ein Rekapitulieren und Durchdenken der zuvor gelesenen fragerlevanten Inhalte, wodurch ebenfalls intentionales Lernen stattfindet. Diese Prozesse beruhen auf dem Prinzip des *backward processing* (Ri-

ckards, 1979). Die nachgestellten Aufgaben scheinen auch die weitere Verarbeitung des Textes zu fördern, indem der Lernende ähnliche Textfragen erwartet und seinen Verarbeitungsprozess dementsprechend organisiert, indem er potentiell fragerrelevante Inhalte verstärkt beachtet (Fraser, 1968; Rickards & Denner, 1978). Auch hierbei ist aufgrund der zur Beantwortung der *high level*-Aufgaben notwendigen Prozesse zu erwarten, dass diese Aufgaben zu einer tieferen Textverarbeitung und in Folge dessen auch zu besseren Testleistungen führen. Eine Förderung des Gesamtverständnisses und übergreifender Zusammenhänge ist somit am ehesten durch nachgestellte *high level*-Aufgaben zu erzielen (Andre, 1979; Bean, 1985).

Insgesamt kann aufgrund der Forschungsergebnisse zu den *adjunct questions* ein größerer, lernförderlicher Effekt von *high level*-Aufgaben im Vergleich zu *low level*-Aufgaben angenommen werden.

3.2.3 Aufgaben mit Lösungsbeispielen (*worked examples*)

Eine weitere Aufgabenart stellen Aufgaben mit Lösungsbeispielen dar, sie werden auch englisch *worked examples* genannt. Sie bestehen aus einer Problem- oder Aufgabenstellung, Lösungsschritten und der Lösung selbst. Typischerweise findet man solche Aufgabenbeispiele in Mathematik- und Physikbüchern. Sie folgen meist der Erklärung eines Prinzips (z.B. Satz, Theorem, Gesetz). Ein Lösungsbeispiel soll aufzeigen, wie das eingeführte Prinzip angewendet werden kann. Nach dem Lösungsbeispiel folgen in der Regel Aufgaben, die von den Lernenden selbst gelöst werden sollen (Renkl, 2002).

Insbesondere die Studien aus der Arbeitsgruppe um Sweller (zusammenfassend Sweller, van Mörrienboer & Pass, 1998; Atkinson, Derry, Renkl & Wortham, 2000) zeigen, dass Lernen mit Lösungsbeispielen im Vergleich mit dem üblichen Vorgehen, Lernen durch Bearbeitung von Problemlöseaufgaben, überlegen ist. Das Lernen mit Lösungsbeispielen wird vor allem von Novizen als Lernform bevorzugt (z.B. Recker & Pirolli, 1995) und ist dem Lernen durch Problemlöseaufgaben besonders zu Beginn des Wissenserwerbs überlegen. Besonders effektive Lernprozesse treten auf, wenn die Phase der Auseinandersetzung mit Lösungsbeispielen verlängert und intensiviert wird (vgl. Atkinson et al., 2000). In späteren Phasen des Wissenserwerbs, wenn Lernende bereits Erfahrungen in einem Wissensbereich gesammelt haben, hat sich das Lernen durch Problemlösen im Vergleich zum Lernen durch Beispiellösungen als wirksamer herausgestellt (siehe Renkl & Atkinson, 2003; Kalyuga, Chandler, Tuovinen & Sweller, 2001). Die instruktionale Überlegenheit von Lösungsbei-

spielen gegenüber Problemlösen konnte auch im Bereich der Physik repliziert werden (Ward & Sweller, 1990).

Wiewohl Beispiele Handlungsanweisungen darstellen, laden sie jedoch viele Personen dazu ein, die präsentierten Informationen passiv und oberflächlich zu verarbeiten, was sich wiederum negativ auf den Lernerfolg auswirkt (Stark, Gruber, Renkl & Mandl, 2000). Diese Nebenwirkungen sind zumindest zum Teil auf Merkmale der Lernmethode zurückzuführen. Dadurch, dass lösungsrelevante Informationen in detaillierter, systematisch angeordneter Form bereitgestellt werden, bekommt der Lernende Handlungsanweisungen zur Verfügung gestellt. Es kann daraufhin der Eindruck erzeugt werden, dass es nicht notwendig ist, sich intensiv mit den einzelnen Lösungsschritten auseinanderzusetzen. In der Untersuchung von Stark et al. (2000) konnten gezeigt werden, dass durch die Kombination von Lösungsbeispielen und Problemlöseaufgaben sowohl die Häufigkeit der Elaborationsaktivität als auch die Qualität der Elaborationen gesteigert werden konnte. Außerdem erzielten die Versuchspersonen der kombinierten Lernbedingung signifikant und substantiell höhere Transferleistungen mit einer Effektstärke von bis zu $d = 1.6$ als Versuchspersonen, die ausschließlich Aufgaben mit Lösungsbeispiele bearbeitet hatten. Dieser Befund weist darauf hin, dass nicht eine Lernmethode - Lernen durch Lösungsbeispiele oder Problemlöseaufgaben - insgesamt gesehen effektiver ist, sondern vielmehr resultieren aus einer Kombination der Vorteile beider Bedingungen positive Effekte (Stark et al., 2000).

In der einschlägigen Literatur werden Aufgaben mit Lösungsbeispielen mit verschiedenen Funktionen in Beziehung gebracht, die ihnen beim Lernen zukommen können - beispielsweise eine Hinweisfunktion, indem sie auf abstrakte Prinzipien oder Formeln verweisen. Auch wird diesen Aufgaben eine Klassifikationsfunktion zugeschrieben, womit gemeint ist, dass Beispiele benötigt werden, um Aufgaben zu klassifizieren und dann das Lösungsprinzip zu aktivieren (Ross, 1987). Eine Lernfunktion ordnet Reimann (1997) in diesem Zusammenhang Beispielen zu, indem sie Informationen enthalten können, die in anderer Form (z.B. im Lehrtext) nicht bereitgestellt werden. Da für die vorliegende Arbeit die Forschung zu Aufgaben mit Lösungsbeispielen eine untergeordnete Rolle spielt, sei für weitere Funktionen auf die Literatur verwiesen (z.B. Stark, 1999).

Zur Erklärung der Effektivität des Lernens aus Lösungsbeispielen wird derzeit bevorzugt die *Cognitive Load*-Theorie herangezogen, wonach der wichtigste Vorteil im Vergleich zum Lernen durch Problemlösen darin liegt, dass Lernen aus Lösungsbeispielen einen geringeren *Extraneous Load* verursacht und daher mehr kognitive Kapazität für die eigentlichen Lernprozesse (*Germane Load*) übrig bleibt (Renkl & Atkinson, 2003). Diese theoretische Erklärung der Effektivität des Lernens aus Lösungsbeispielen wurde von Renkl, Gru-

ber, Weber, Lerche und Schweizer (2003) empirisch überprüft und bestätigt. Für ausführliche Erklärungen der Effektivität von *worked examples* mit Hilfe der *Cognitive Load*-Theorie sei an dieser Stelle auf Sweller (1988) und Sweller et al. (1998) verwiesen.

Zum tatsächlichen Einsatz von Lösungsbeispielen in der gängigen Unterrichtspraxis gibt es dagegen bislang kaum empirische Studien. Eine Ausnahme ist die Bedarfsanalyse im mathematisch-naturwissenschaftlichen Unterricht von Renkl, Schworm und Hilbert (2004). Dazu wurden 181 Unterrichtseinheiten im Mathematikunterricht der 8. Jahrgangsstufe sowie 34 im Physikanfangsunterricht der 7. und 8. Jahrgangsstufe auf beispielbasierte Lehrsequenzen hin analysiert. Die Ergebnisse zeigen für den Mathematikunterricht, dass 42% der auf Video aufgenommenen Unterrichtssequenzen keinerlei beispielbasierte instruktionale Maßnahmen enthielten. Im Physikunterricht zeigte sich ein noch deutlicheres Ungleichgewicht: In den insgesamt 34 Physik-Unterrichtsstunden ergaben sich lediglich neun beispielbasierte Unterrichtssequenzen. Die Autoren vermuten, dass die geringe Anzahl von Unterrichtssequenzen, die Beispiele beinhalteten, darin begründet ist, dass es sich bei den analysierten Unterrichtsstunden um Physikanfangsunterricht handelte, welcher sehr phänomenorientiert war und kaum Berechnungen erforderte. Des Weiteren müssten die Ergebnisse wegen der geringen Stichprobengröße mit Vorsicht betrachtet werden.

Inwieweit diese Ergebnisse auf den Physikunterricht der Jahrgangsstufen 9 und 10, welcher in der vorliegenden Arbeit im Mittelpunkt des Forschungsinteresses steht, übertragen werden kann, ist fraglich, da in den höheren Jahrgängen im Physikunterricht mehr mit Formeln und dem Berechnen von Größen gearbeitet wird.

Insgesamt kann festgehalten werden, dass Lernen aus Lösungsbeispielen zu größeren Leistungszuwächsen im Vergleich zum Problemlösen und zu Vorteilen bei der Bildung kognitiver Schemata führen kann, welche dann den Transfer des Erlernenen auf neue Problemstellungen erlauben. Somit stellt das Lernen aus Lösungsbeispielen eine Möglichkeit dar, die Entstehung von „trägem Wissen“ zu verhindern (Renkl, 1996). In Kapitel 9.5.3 wird auf die Bedeutung dieses Problems im Detail eingegangen.

3.2.4 Schwächen der bisherigen Untersuchungen

Für die bisherigen, zum Teil widersprüchlichen und inkonsistenten Untersuchungsergebnisse zu den verschiedenen Aufgabenarten und ihrer Effektivität werden verschiedene Gründe angeführt. So nennen Gall und Rhody (1987) drei Faktoren, welche für die Ergebnisse verantwortlich sind:

1. Die Definitionen von kognitiv niedrigen und höheren Aufgaben unterscheiden sich zwischen den einzelnen Studien stark (Crooks, 1988). Einige Autoren legen die Taxonomie von Bloom et al. (1956) zugrunde, andere das Klassifikationssystem von Guilfords Intelligenz-Struktur-Modell (Aschner, 1961). Dabei stellt sich die Frage, inwieweit die Ergebnisse der verschiedenen Klassifikationssysteme miteinander verglichen werden können.
2. Da die meisten Studien in den USA durchgeführt wurden, spielt der amerikanische Lehrplan eine wichtige Rolle, in dem kognitiv niedrigere Lehrziele betont werden (Goodlad, 1983), wodurch der große Anteil der *low level*-Aufgaben im regulären Unterricht z.T. erklärt werden kann (Blosser, 1979; Gall, 1970; Hare & Pulliam, 1978).
3. Die unterschiedlichen Schülerstichproben haben ebenfalls einen Einfluss auf die inkonsistenten Ergebnisse (Andre, 1979; Brophy & Good, 1986). Clegg (1987) erklärt die Ergebnisse damit, dass Faktoren wie das Alter der Schüler, der sozioökonomische Status, die Komplexität des Inhaltsbereichs und das Lehrziel den Zusammenhang zwischen der Aufgabenart und der Schülerleistung beeinflussen. So weisen Gage und Berliner (1996) darauf hin, dass der Prozentsatz der richtigen Antworten mit dem sozioökonomischen Status variiert. Interindividuelle Unterschiede wurden in den meisten Untersuchungen zur Effektivität der verschiedenen Aufgabenarten ignoriert bzw. nicht erhoben (Andre, 1979).

Eine Schwäche, die allen Untersuchungen zu kognitiven Prozessen gemeinsam ist, bezieht sich auf die indirekte Erhebung der interessierenden Prozesse. Diese können nicht direkt beobachtet, sondern lediglich aus der Bearbeitung der verschiedenen Aufgabenarten erschlossen werden (z.B. Gall, 1970). Bereits Bloom et al. (1956) wiesen auf dieses Problem hin. Es kann nicht sichergestellt werden, dass ein Schüler eine als Anwendungsaufgabe definierte Aufgabe nicht auch durch die Erinnerung von Fakten beantwortet. Andre (1979) geht in seiner Kritik darüber hinaus und weist darauf hin, dass die Effekte von Aufgaben auf das Lernen kaum bekannt sind, da die Bedingungen, unter denen Aufgaben die Leistung beeinflussen, ungenau und allgemein bleiben. Weiter kritisiert Andre (1979) an unterrichtsbezogenen Studien, dass mündliche Fragen des Lehrers meist nur von *einem* Schüler bzw. wenigen Schülern beantwortet werden. In einem typischen Klassenzimmer mit 25-35 Schülern kann jeder Schüler nur einen geringen Prozentsatz der Unterrichtszeit für seine Antworten eingeräumt bekommen. Der für die gesamte Lerngruppe eintretende Lernzuwachs erklärt sich nach Levin (2005) daher sowohl aus der Rezeption der Antwort, die von den Mitschülern oder dem Lehrer selbst gegeben wird, als auch aus dem Nachdenken über die zu gebende Antwort. Jedoch bestehe bei der mündlichen Beantwortung von Aufgaben die Gefahr,

dass die Schüler sich nicht mit der Aufgabenbeantwortung beschäftigen bzw. erst dann, wenn sie vom Lehrer aufgerufen werden. Die Wissensvertiefung und damit auch ein möglicher Leistungszuwachs ist nach Anderson und Faust (1973) daher erst gewährleistet, wenn eine individuelle Beantwortung der Aufgabe durch jeden einzelnen Schüler erfolgt. Weiter ist der Lernzuwachs größer, wenn die Schüler die Aufgaben schriftlich bearbeiten, da dann eher sichergestellt werden kann, dass alle Schüler die Aufgabe bearbeiten und nicht nur einige wenige sich am Unterrichtsgespräch beteiligen (vgl. Wilen, 1987). Außerdem sei eine verbesserte Erinnerungsleistung durch das schriftliche Festhalten möglich (Anderson & Faust, 1973). In einer Untersuchung von Anderson und Biddle (1975) konnte nachgewiesen werden, dass ausführliche Antworten, besonders in schriftlicher Form, die Testleistungen für bekannte Aufgaben verbesserten im Vergleich zu einer Bedingung, in der die Schüler ihre Antwort in Gedanken formulieren sollten. Antworten Schüler nicht aktiv auf eine Frage, besteht die Gefahr, dass sie sich mit anderen Dingen, welche nicht relevant für den Unterricht sind, beschäftigen. Berliner und Fisher (1985) konnten zeigen, dass dieses nicht-unterrichtsrelevante Verhalten mit Leistungsabnahme in Zusammenhang steht. Interessiert das individuelle Antwortverhalten des einzelnen Schülers auf die Unterrichtsaufgaben, so müssen die Aufgaben schriftlich bearbeitet und individuell ausgewertet werden, was in den meisten der bisherigen Untersuchungen nicht systematisch untersucht wurde. Diese offene Frage nach dem individuellen Antwortverhalten der einzelnen Schüler in Bezug auf unterschiedlich schwierige Aufgaben soll in Studie 2 (Kapitel 13-16) erörtert werden.

Crooks (1988) führt als weiteren Kritikpunkt an der bisherigen Forschung an, dass die meisten Untersuchungen nur von kurzer Dauer sind, und nimmt an, dass der Einsatz von *high level*-Aufgaben erst im Laufe einiger Wochen ihren Effekt zeigt, da die Schüler zunächst den Umgang mit dieser im Unterricht eher selten eingesetzten Aufgabenart lernen und üben müssen (vgl. Redfield & Rousseau, 1981). Diese Annahme wird von Samson et al. (1987) gestützt, die in ihrer Metaanalyse zum Effekt von *high level* bzw. *low level*-Unterrichtsaufgaben auf die Schülerleistung gezeigt haben, dass die Effektstärke in Abhängigkeit von der Untersuchungsdauer variiert. Für Studien zum Vergleich von *high level*- und *low level*-Aufgaben auf die Schülerleistung, welche 5 Tage oder kürzer dauern, ergab sich eine Effektstärke von $d = 0.05$. Im Gegensatz dazu zeigten die Studien, welche länger als 20 Tage dauerten, eine Effektstärke von $d = 0.83$ zugunsten des Effekts von *high level*-Aufgaben auf die Schülerleistungen. Da dieses Ergebnis auf der Grundlage von lediglich 4 Studien beruht, sollte es in weiteren Untersuchungen repliziert werden. Die in den Kapiteln 13-16 vorgestellte Studie 2 wird aufgrund dieser Ergebnisse über eine Unterrichtsreihe von 3 Monaten durchgeführt.

Weiterhin verwendeten die Autoren nur in wenigen Studien Aufgaben höheren kognitiven Niveaus als Lern- und Testaufgaben. Auch wurden Anwendungs- und Transferleistungen selten überprüft (Reimann, 1998; Häfele, 1995). Erst mit solchen komplexen Aufgaben höheren kognitiven Niveaus kann jedoch kontrolliert werden, ob die Inhalte auch wirklich verstanden wurden, da nicht davon auszugehen ist, dass ein Schüler, der bestimmte *low level*-Aufgaben zu einem zuvor gelerntem Text bzw. unterrichteten Inhaltsbereich beantworten kann, den Inhalt notwendigerweise auch verstanden haben muss und somit in der Lage ist, das Gelernte richtig einordnen und anwenden zu können (Anderson, 1972; Andre, 1979). Daher sollten in den Leistungstests auch *high level*-Aufgaben gestellt werden, um somit das Verständnis bzw. die Wissensanwendung erfassen zu können.

Des Weiteren muss bei den bisherigen Untersuchungen berücksichtigt werden, dass die meisten nicht die Antworten der Schüler auf die vom Lehrer gestellten Aufgaben, sondern auf von Forschern konstruierte Aufgaben untersuchen (Levin & Long, 1981). Diese können sich aber entscheidend von denen des Lehrers, die der Schüler normalerweise beantwortet, unterscheiden. Die curriculare Validität ist also nicht notwendig gegeben und lässt somit nur begrenzte Rückschlüsse auf den Zusammenhang zwischen den Leistungen der Schüler und der Aufgabenart bei der Beantwortung von Lehrerfragen zu. Diese methodische Schwäche soll in der hier in den Kapiteln 8-11 beschriebenen Studie 1 vermieden werden, indem die Schüler zum Teil wörtlich übernommene, vom Lehrer gestellte Aufgaben bearbeiten sollen, um generalisierbare Aussagen über die Wirksamkeit der von Lehrern im Unterricht eingesetzten Aufgaben und Fragen machen zu können.

Auch Crooks (1988) zweifelt an der curricularen Validität der Ergebnisse von experimentellen Studien für den regulären Unterricht. Es sei fraglich, inwieweit die Ergebnisse ohne Einschränkungen auf den Unterricht übertragen werden können, da dort eine Reihe von Variablen wirken, welche nicht - wie in einem experimentellen Setting - kontrolliert bzw. konstant gehalten werden können. Im Gegensatz zu kontrollierten Laboruntersuchungen, die nur einen Gegenstandsbereich kontrolliert untersuchen können, gibt es im Klassenraum eine Reihe an unkontrollierten Einflussfaktoren, welche die Untersuchungen erschweren und daher die Ergebnisse häufig nicht eindeutig erscheinen lassen und ihre Interpretation schwierig machen (Andre, 1979). Somit müssen die Vorteile eines experimentellen Settings (wie z.B. Variablenkontrolle) und die einer Feldstudie (z.B. curriculare Validität) bei der Untersuchungsplanung abgewogen werden. Da es in der vorliegenden Studie 1 um den Zusammenhang zwischen der Schwierigkeit der vom Lehrer eingesetzten Unterrichtsaufgaben und den Schülerfähigkeiten geht, bietet sich eine feldexperimentelle Untersuchung zur Beantwortung dieser Fragestellung an. Um sicherzustellen, dass sich alle Schüler mit den Aufgabeninhalten

beschäftigen, und um so das individuelle Antwortverhalten zu erfassen, wurden die Aufgaben schriftlich bearbeitet. Außerdem werden nicht nur *low level*-Aufgaben zur Bearbeitung vorgelegt, sondern auch Aufgaben, welche zur Bearbeitung kognitiv höhere Prozesse erfordern. Diese häufig lediglich implizit angenommene und selten empirisch kontrollierte unterschiedliche Schwierigkeit der verschiedenen Aufgabenarten (vgl. Crooks, 1988) wird in den vorliegenden Studien mit Hilfe klassischer wie probabilistischer Testverfahren erfasst.

3.2.5 Fazit aus den bisherigen Forschungsergebnissen

Seel (1981) schlussfolgerte, dass die Wirksamkeit von Lernaufgaben des konventionellen Unterrichts bei weitem nicht so gut erforscht ist wie die Wirksamkeit experimentell bzw. von wissenschaftlicher Seite konstruierter Lernaufgaben. Diese Kritik hat heute noch ihre Berechtigung. Daher wird in der vorliegenden Arbeit der Versuch unternommen, die Lernwirksamkeit der tatsächlich vom Lehrer im Unterricht eingesetzten Aufgaben empirisch zu untersuchen. Auch soll überprüft werden, inwieweit im Physikunterricht ein tieferes Verständnis sowie anwendbares Wissen mit Hilfe des Aufgabeneinsatzes gefördert wird. Aufgrund der bisherigen Ergebnisse zu Aufgaben im Unterricht (Doyle, 1986; Wilen & Clegg, 1986; Graesser & Person, 1994; Niegemann & Stadler, 2001) wird für die in den Kapiteln 8-11 beschriebene Studie 1 erwartet, dass im Physikunterricht die Lehrer die meisten Aufgaben auf dem *low level*-Niveau (auch Wissensniveau genannt) stellen und kaum Aufgaben formuliert werden, welche eine Verstehensleistung des Schülers erfordern und/oder auf die Anwendung (bzw. den Transfer) des Gelernten auf neue Inhaltsbereiche abzielen. Dieser Schwerpunkt der vom Lehrer im Unterricht eingesetzten Aufgaben auf wenig anspruchsvollen *low level*-Aufgaben wurde in jüngerer Zeit von Neubrand (2002) für den Mathematikunterricht bestätigt. Die Autorin fasste die Ergebnisse der TIMSS-Mathematik-Videos in Bezug auf den Aufgabeneinsatz wie folgt zusammen: „Variationen in der Wissensart, den Aufgabentypen sowie der Aufgabenart und damit Flexibilität beim Einsatz mathematischer Fähigkeiten sind in den reinen Seatwork-Aufgaben in den untersuchten deutschen Mathematikstunden also tendenziell eher wenig zu erkennen. [...] Dies ist auch zu erwarten, wenn man davon ausgeht, dass ein fragend-entwickelnder Unterrichtsstil vorherrscht. Selbstständigkeit bei einer Vielfalt von Wissensarten und Aufgabentypen ist damit aber nicht realisiert“ (Neubrand, 2002, S. 352). Diese Einschätzung kann möglicherweise auf den in der vorliegenden Studie 1 untersuchten Physikunterricht übertragen werden.

Levin (2005) fasst die Forschungsergebnisse zum Zusammenhang von Aufgabenart und Schülerleistung folgendermaßen zusammen: „In der Mehrheit der Experimente führen

Fragen auf einem höheren kognitiven Niveau zu besseren Lernergebnissen als Fragen auf kognitiv niedrigerem Niveau“ (S. 41). Im Gegensatz dazu sieht Renkl (1991) wegen der bisher inkonsistenten Ergebnisse diese Forschungsfrage noch nicht abschließend beantwortet, stattdessen bedarf es weiterer Untersuchungen. Auch besteht weiterhin Unklarheit bzgl. des Zusammenhangs zwischen Aufgabenart und den dadurch ausgelösten kognitiven Prozessen (Gavelek & Raphael, 1985; Seel, 1981). Werden die Ergebnisse zum Effekt von *adjunct questions* herangezogen, so ist anzunehmen, dass *high level*-Aufgaben einen im Vergleich zu den *low level*-Aufgaben größeren Effekt auf die Lernleistung haben (vgl. Hamaker, 1986).

Aufgrund der bislang inkonsistenten Ergebnisse zum Einsatz und zum lernförderlichen Mischungsverhältnis der verschiedenen Aufgabentypen (*low level* vs. *high level*-Aufgaben als schwierigkeitsunterscheidendes Merkmal) soll in der hier vorgestellten Arbeit, in Studie 1, der Einsatz unterschiedlicher Aufgaben an zwei sehr leistungsunterschiedlichen Schulformen (Gymnasium und Hauptschule) untersucht werden.

3.3 Passung von Aufgabenschwierigkeit und Schülerfähigkeit

Seel (1981) schließt aus der Analyse der empirischen Studien von Garten (1977), Kretschmann und Malburg (1976) und Tobias (1973), dass die Passung zwischen Schülerfähigkeiten und Lehrhandlungen bisher nicht gelungen bzw. nur wenige empirische Studien insgesamt dazu vorliegen (vgl. Überblicksreferate von Snow & Salomon, 1968; Treiber & Petermann, 1976). Die Strukturen und die Schwierigkeit von Lernaufgaben im konventionellen Unterricht sowie deren Wirksamkeit sind noch weitgehend unerforscht (Seel, 1981). Außerdem wird die Aufgabenschwierigkeit in der Regel mit Hilfe der klassischen Testtheorie festgestellt, wobei deren Abhängigkeit von individuellen Voraussetzungen berücksichtigt werden muss (vgl. auch Klieme, 2000; Williams & Clarke, 1997).

Die Aufgabenschwierigkeit wurde in verschiedenen Forschungsrichtungen untersucht. Von lerntheoretischen Überlegungen ausgehend hat sich in Experimenten zum programmierten Unterricht das „Prinzip der kleinen Schritte“, welches Lernen möglichst ohne Misserfolge und mit einer hohen Lösungswahrscheinlichkeit vorsieht, als allen anderen Alternativen überlegen erwiesen (z.B. Huber, 1966). Misserfolge wirken sich demnach, insbesondere zu Beginn des Lernens, hinderlich auf den Lernerfolg aus. Personen, die eine Serie von Lernaufgaben mit unlösbaren (Sweller, 1976) oder mittelschwierigen (Fürntratt, 1978; Kloep & Weimann, 1982) Aufgaben begannen, konnten die nachfolgenden leichten Aufgaben überhaupt nicht bzw. schlechter lösen als Personen, die mit leichten Aufgaben begonnen hatten. Aus diesen Ergebnissen zieht Kloep (1985) die Schlussfolgerung, dass es nicht sinn-

voll für den Lernerfolg ist, Schülern Aufgaben mit einer geringen Lösungswahrscheinlichkeit (unter 50%) vorzulegen, sondern vielmehr leichte Aufgaben eingesetzt werden sollten, da diese u.a. dazu führen, dass Fertigkeiten schneller gelernt und sicherer beherrscht werden. Im Gegensatz dazu gehen andere Autoren (Heckhausen, 1980; Herber, 1979; Schiefele, 1974) aufgrund motivationspsychologischer Ergebnisse davon aus, dass ein nicht-monotoner, umgekehrt U-förmiger bzw. U-förmiger Zusammenhang zwischen Schwierigkeit und Lernleistung, welcher in Zusammenhang mit den Konstrukten „Hoffnung auf Erfolg“ bzw. „Furcht vor Misserfolg“ untersucht wurde, besteht. Dürfen Versuchspersonen die Aufgabenschwierigkeit frei wählen, zeigt sich bei Erfolgsmotivierten ein umgekehrt U-förmiger Zusammenhang (Hoffnung auf Erfolg), d.h. die meisten dieser Probanden wählen Aufgaben mit mittlerer Schwierigkeit. Im Gegensatz dazu wählen Misserfolgsorientierte entweder sehr leichte oder sehr schwere Aufgaben aus, wodurch ein U-förmiger Zusammenhang zustande kommt (Furcht vor Misserfolg). Inwieweit diese motivationspsychologischen Befunde auf den Zusammenhang zwischen der vom Lehrer ausgewählten Aufgabenschwierigkeit und Schülerleistungen im Unterricht übertragen werden können, ist fraglich.

Manche didaktische Überlegungen legen im Gegensatz zu einem konstant schwierigen Lehrstoff eine kontinuierliche Veränderung des Anforderungsniveaus zu Grunde, d.h. die Aufgabenschwierigkeit variiert von schwer zu leicht oder umgekehrt (z.B. Roth, 1966). Burnstein (1980) schließt aus seinen Studienergebnissen, dass ein Mischungsverhältnis aus Aufgaben mit hohen und mittleren Lösungswahrscheinlichkeiten sowie wenigen Aufgaben mit niedrigen Lösungswahrscheinlichkeiten zu optimalen Schülerleistungen führt. Nussbaum und Leutner (1986) untersuchten in einem Experiment zum entdeckenden Lernen (nach Bruner, 1961) von Aufgabenlösungsregeln fünf verschiedene Bedingungen, in denen die Aufgabenschwierigkeit systematisch variiert wurde. Die Ergebnisse legen nahe, dass bei entdeckend-lernenden Lernbedingungen es (zumindest für die von den Autoren verwendeten Matrizenaufgaben) von Vorteil ist, dem Lernenden das Erschließen von Lösungsregeln auf möglichst einfache Weise zu ermöglichen. Des Weiteren war eine streng monotone Beziehung zwischen der Schwierigkeit der Entdeckungsbedingung und der Lernleistung zu beobachten (Nussbaum & Leutner, 1986). Ob diese anhand eines relativ einfachen und gut strukturierten Lehrstoffs erzielten Ergebnisse ohne Einschränkungen auf Inhalte des Schulunterrichts zu übertragen sind, muss noch empirisch überprüft werden.

Bislang gibt es nur wenig Forschung zu den von Lehrern im Unterricht eingesetzten Aufgaben. Einige Untersuchungen zeigen jedoch, dass Schüler meist Aufgaben erhalten, welche viel zu schwierig für sie sind (z.B. Fisher, Berliner, Filby, Marliave, Cahen & Dishaw, 1980; Jorgenson, 1977). Im Gegensatz dazu konnte Renkl (1991) zeigen, dass die

Schwierigkeit der Aufgaben, die im Mathematikunterricht an Grundschulen gestellt werden, im Allgemeinen gering ist, die Gesamtlösungshäufigkeit war mit 83.3% sehr hoch.

Insgesamt kann keine eindeutige Schlussfolgerung aus der bisherigen Forschung zu dem Zusammenhang zwischen Aufgabenschwierigkeit und Lernleistung gezogen werden. Daher soll in der in den Kapiteln 12-16 berichteten Studie 2 die Aufgabenschwierigkeit, in Anlehnung an die Konzeption von Nussbaum und Leutner (1986), individuell angepasst und in drei Schwierigkeitsstufen variiert werden.

Indirekte Hinweise auf die Schwierigkeit der im Unterricht eingesetzten Aufgaben ergeben sich aus den Untersuchungen zur Passung bzw. kognitiven Korrespondenz zwischen eingesetzten Aufgaben und den darauf bezogenen Schülerantworten, auf die im Folgenden eingegangen wird.

Heckhausens (1968) Prinzip der Passung geht davon aus, dass bei mittlerer Schwierigkeit die höchste Lernleistung zu erwarten ist. In diesem Prinzip stützt sich Heckhausen (1968) auf das Motivationsmodell von Atkinson (1957), wonach bei Erfolgsmotivierten für Lernbedingungen mit mittlerer Erfolgswahrscheinlichkeit die höchste Leistungsmotivation vorherzusagen ist. Umgekehrt gilt für Misserfolgsmotivierte, dass bei mittlerer Erfolgswahrscheinlichkeit die niedrigste Lernmotivation resultiert. Inwieweit diese Ergebnisse aus der Motivationspsychologie auf die Schülerleistungen im Unterricht übertragbar sind, erscheint fraglich. Brophy (1983) betont die Bedeutung der Passung zwischen dem Schwierigkeitsniveau des gesamten Unterrichts und den Schülerfähigkeiten, wobei die Bedeutung der Aufgaben als passend bzw. angemessen für die jeweiligen Schüler hervorgehoben wird. Diese Passung ist in den meisten untersuchten Klassen jedoch nicht gegeben (Brophy, 1983). Besonders akzentuiert wird das Prinzip der Passung auch in den Konzeptionen des adaptiven Unterrichts, als ein Beispiel dessen das Modell des zielerreichenden Lernens (vgl. Bloom, 1976) gilt. Dabei erfolgt die Anpassung des Unterrichts an die unterschiedlichen Lernvoraussetzungen der Schüler. Auf dieses Modell wird in Kapitel 4.6 und auf die Konzeption des adaptiven Unterrichts in Kapitel 12 näher eingegangen.

Die fortlaufende unterrichtsbegleitende Diagnose des Leistungsstands stellt eine wichtige Voraussetzung für die erfolgreiche Korrespondenz zwischen Aufgabenschwierigkeit und Schülerfähigkeit dar (Corno & Snow, 1986). „To learn efficiently, students must be engaged in activities that are appropriate in difficulty level and otherwise suited to their current achievement levels and needs“ (Brophy & Good, 1986, p. 360). Weiter nehmen die Autoren an, dass der Lehrer über eine ausgeprägte Diagnosefähigkeit verfügen muss, um die Schwierigkeit der Unterrichtsaufgaben an die Schülerfähigkeiten anzupassen. Denn nur so

könne Unterricht ohne größere Probleme ablaufen und könnten Schüler Aufgaben mit hoher Lösungswahrscheinlichkeit bearbeiten. Diesen Zusammenhang zwischen der Antwortrichtigkeit der Schüler auf die Lehrerfragen und den Leistungen der Schüler konnten Brophy und Evertson (1976) nachweisen. Die Ergebnisse zeigten, dass diejenigen Klassen die besten Schülerleistungen erzielten, in denen Lehrer geeignete Lernaufgaben angepasst an die Schülerfähigkeiten stellen. Auf die Diagnosefähigkeit von Lehrern wird in Kapitel 3.4 eingegangen.

Schrader und Helmke (1987) nehmen an, dass, wenn die Passung zwischen Unterrichtsschwierigkeit und Lernvoraussetzungen der Schüler stark verringert ist, die Unterrichtseffektivität deutlich Einbußen erleidet. Auch Helmke (2004) greift das Konzept der Passung auf. Er definiert es als Optimierung der Balance zwischen Anforderungen des Unterrichts und Voraussetzungen der Schüler und stellt es als ein übergeordnetes Merkmal jedes kompetenten unterrichtlichen Handelns dar. Ein wesentlicher Gesichtspunkt, der sich auf die Passung bezieht, ist die Schwierigkeit der Unterrichtsinhalte. Im Idealfall sollten die unterrichtlichen Anforderungen in einer Schwierigkeitszone liegen, die oberhalb des aktuellen Wissensstandes der Schüler liegt, aber auch nicht zu weit davon entfernt sein darf (Helmke, 2004). Anforderungen in dieser „Zone der proximalen Entwicklung“ (Wygotsky, 1978) sind dadurch gekennzeichnet, dass mit Unterstützung der Lehrkraft neues Wissen erworben wird, ohne dass dies durch Unterforderung (zu einfache Aufgaben, im Extremfall „sattsam bekannt“) oder Überforderung (zu schwierige Aufgaben, welche als Bedrohung oder als aussichtslos in der Lösung empfunden werden und Angst bewirken können) beeinträchtigt wird (Helmke, 2004). Auf die theoretischen Überlegungen von Wygotsky (1978) und deren Bedeutung für die vorliegenden Studien wird in Kapitel 4.3 näher eingegangen.

Obwohl die Bedeutung der Passung zwischen Aufgabenschwierigkeit und Schülerfähigkeit von verschiedenen Autoren betont wurde, weisen die empirischen Ergebnisse nicht darauf hin, dass diese im Unterricht gegeben ist. Dillon (1982a) konnte Belege dafür finden, dass das kognitive Niveau von Lehrerfragen und Schülerantworten relativ häufig divergiert. In 50% aller untersuchten Fälle wurden Antworten von den Schülern auf einem anderen Niveau gegeben als die Frage gestellt wurde (auch Willson, 1973). Ob die Schüler dabei häufiger ein höheres oder niedrigeres Antwortniveau als die Lehrerfrage nutzten, kann nicht berichtet werden, da das vom Frageniveau divergierende Antwortniveau in der Auswertung lediglich zusammengefasst angegeben wurde. Der besonders kritische Punkt dabei ist, dass durch die Fragen oftmals nicht die Art von kognitiven Prozessen - indiziert durch die Art der Antwort - ausgelöst wird, die vom Lehrer intendiert wurde (vgl. Winne, 1987). Dillon (1982b) fasst die Ergebnisse der Studien von Willson (1973), Mills, Rice, Berliner und

Rousseau (1980) und Dillon (1982a) zusammen, in denen das Niveau der Antwort in der Analyse berücksichtigt wurde, und kommt zu dem Ergebnis, dass über die drei Studien gemittelt 57% der Antworten dem kognitiven Niveau der Frage entsprechen, während 24% der Antworten auf einem niedrigeren und 19% auf einem höheren Niveau klassifiziert werden als die Lehrerfrage. Mills et al. (1980) konnten zeigen, dass die Korrespondenz zwischen Lehrerfragen und Schülerantworten für *low level*-Aufgaben (nach Bloom et al., 1956) höher war (62.3%) als für *high level*-Aufgaben, mit einer Übereinstimmung von 46.9%. Prinzipiell ist Dillons (1982a, vgl. auch Winne, 1987) Kritik berechtigt, dass man nicht davon ausgehen kann, dass bestimmte Aufgaben notwendigerweise die intendierten Reaktionen bei den Schülern bewirken. Winne und Marx (1977) nehmen als Gründe für die mangelnde Korrespondenz an, dass die Schüler entweder den Inhaltsbereich nicht gelernt haben; oder sie verstehen nicht, mit Hilfe welcher kognitiven Prozesse die Aufgabe erfolgreich gelöst werden kann. Es handelt sich also um keinen deterministischen Zusammenhang zwischen dem Aufgabenniveau und dem Niveau der kognitiven Prozesse, sondern um einen probabilistischen (vgl. Bromme, Seeger & Steinbring, 1990; Renkl, 1991). Diese unbefriedigende kognitive Korrespondenz kann auch als Hinweis darauf gesehen werden, dass die Schwierigkeit der Lehrerfragen dem Leistungsniveau der Schüler nicht anpasst bzw. nicht angemessen ausgewählt ist.

Insgesamt liegen nur wenige Forschungsergebnisse zur Passung zwischen Aufgabenschwierigkeit und Schülerfähigkeit vor (vgl. Seel, 1981). Inwieweit die oben erwähnten Ergebnisse zur Passung zwischen Aufgaben- und Antwortniveau diesen Zusammenhang ausreichend und detailliert genug beleuchten, ist fraglich. Dieser mangelnden empirischen Absicherung soll in der vorliegenden Studie Rechnung getragen werden, indem der Zusammenhang zwischen den Fähigkeiten der Schüler und der Aufgabenschwierigkeit untersucht wird.

3.4 Diagnostische Kompetenz der Lehrer bei der Einschätzung der Aufgabenschwierigkeit und Schülerfähigkeit

Mit diagnostischer Kompetenz, auch unter dem engl. Begriff *accuracy of judgement* bekannt, wird die Fähigkeit eines Urteilers bezeichnet, Personen zutreffend zu beurteilen (Schrader, 2001). Ganz allgemein wird unter der diagnostischen Kompetenz des Lehrers die Fähigkeit verstanden, die zur Optimierung pädagogischer Entscheidungen erforderlichen Diagnoseaufgaben sach- und zielgerecht auszuführen. Die informellen Diagnoseleistungen des Lehrers spielen für die Steuerung des unterrichtlichen Handelns eine wichtige Rolle (Schrader, 1989). Die Angemessenheit oder Qualität dieser Urteilsbildung bemisst sich letztlich am

Beitrag zur Optimierung des jeweiligen Entscheidungsproblems. Verschiedene Autoren (Hosenfeld, Helmke & Schrader, 2002; Wang & Lindvall, 1984; Brophy, 1983b) betonen die Bedeutung der Diagnosekompetenz für eine effektive Unterrichtsgestaltung, da eine möglichst gute Passung zwischen dem Unterrichtsangebot und den Lernvoraussetzungen der Schüler relevant ist, wie bereits in Kapitel 3.3 dargelegt wurde. Lehrkräfte, die über Fähigkeitsunterschiede, über Stärken und Schwächen ihrer Schüler nicht Bescheid wissen, dürften Schwierigkeiten bei der Passung zwischen Lernvoraussetzungen und Aufgabenschwierigkeit haben (Brophy, 1983b). Individuelle Förderung setzt ausreichendes diagnostisches Wissen der verantwortlichen Lehrkraft nach Helmke (2004) zwingend voraus. Schnotz (1971) und Schrader (1989) konnten zeigen, dass eine hohe diagnostische Kompetenz für eine angemessene Gestaltung von Aufgabenstellungen und deren Schwierigkeit wichtig ist.

Schrader (1989) kommt nach der Durchsicht der Literatur zu dem Schluss, „dass Lehrer zumindest unter bestimmten Bedingungen in der Lage sind, relativ genaue Einschätzungen abzugeben“ (S. 67). Allerdings schränkt der Autor ein, dass das Ergebnis, dass Lehrerurteile etwa 50% der Leistungsvarianz aufklären, nicht unbedingt als Indikator für die „bemerkenswerte Diagnosefähigkeit“ (S. 67) der Lehrer angesehen werden kann. Im Hinblick auf die Genauigkeit der personenbezogenen Lehrerurteile bzgl. der Testleistungen ihrer Schüler zeigten sich z.T. deutliche Unterschiede. Schrader konnte in seiner Untersuchung zeigen, dass Personen- und aufgabenbezogene Diagnosekompetenz nicht in einer kompensatorischen Beziehung stehen, sondern in einem Kopplungsverhältnis. Der Lehrer muss demnach sowohl in der Lage sein, Aufgabenschwierigkeiten und damit Anforderungen zutreffend einzuschätzen als auch Leistungen einzelner Schüler beurteilen zu können. Beide Prozesse sind nach Schrader (1989) für die Realisierung adaptiver Unterrichtsprozesse erforderlich.

Es scheint jedoch so, dass Lehrer die generelle Tendenz besitzen, die tatsächliche Schwierigkeit von Aufgaben für ihre Schüler zu unterschätzen (Schnotz, 1971). So konnte Schnotz zeigen, dass Lehrer unabhängig von der tatsächlichen Schwierigkeit der Aufgaben dazu neigen, die Aufgabenschwierigkeit im Allgemeinen zu unterschätzen und nur einen relativ geringen Anteil an Aufgaben (10-25% der Aufgaben) in ihrem Schwierigkeitsgrad richtig einzustufen. „Nur bei etwa einem Fünftel bis einem Viertel der Aufgaben kann man erwarten, dass Lehrer im Allgemeinen die Schwierigkeit richtig beurteilen. Ein knappes Zehntel der Aufgaben wird systematisch überschätzt, die überwiegende Mehrzahl der Aufgaben - nämlich 7/10 von allen - werden unterschätzt“ (S. 114). Hosenfeld et al. (2002) konnten zeigen, dass Lehrer im Durchschnitt ihre Klasse bei der Einschätzung der Lösungshäufigkeit von zu bearbeitenden Aufgaben im Vergleich zur tatsächlichen Lösungshäufigkeit ihrer

Klasse um 18% überschätzen. Das Maximum der Überschätzung der Lösungshäufigkeit betrug 45.5%.

Auch Schrader und Helmke (1987) zeigten, dass Lehrer eher zu einer Über- als einer Unterschätzung des Leistungsniveaus neigen. Im Extremfall wurde das Gesamtniveau der eigenen Klasse um 42.5% bezüglich der zu zeigenden Testleistung – im Vergleich zur tatsächlichen Testleistung - überschätzt, wohingegen die größte Unterschätzung lediglich 13.5% betrug. Des Weiteren zeigten sich erhebliche Unterschiede zwischen den Lehrern in der jeweiligen Korrelation der Lehrerurteile und der tatsächlichen Schülerleistung, diese variierten zwischen $r = .04$ bis $r = .88$ (vgl. Hoge & Coladarci, 1989). Auch Helmke und Schrader (1988) konnten große Differenzen in der diagnostischen Kompetenz zwischen Lehrern nachweisen.

Schnotz (1971) nennt als mögliche Gründe für die Überschätzung von Lehrern die Selbsteinschätzung des eigenen Unterrichtserfolgs, Prestige Gründe und die Unkenntnis der Fähigkeiten und des Vorwissens der Schüler. Als weitere mögliche Erklärung wird angeführt, dass die Lehrer sich in ihrer Einschätzung stark an normativen Vorstellungen darüber orientieren, zu welchen Leistungen Schüler aufgrund des Lehrplans und des Unterrichts eigentlich in der Lage sein sollten. Außerdem wird vermutet, dass die Lehrerurteile eher kompetenz- statt performanzorientiert sind, also ausdrücken, welche Leistungen von Schülern unter optimalen Bedingungen erbracht werden können, und nicht, was die Schüler unter den gegebenen Voraussetzungen tatsächlich zu leisten im Stande sind (Hosenfeld et al., 2002; Schrader & Helmke, 1987). Hosenfeld et al. (2002) nehmen weiter an, dass die Überschätzung der Klassenleistung durch die Lehrer möglicherweise dadurch zustande kommt, dass leistungsmindernde Faktoren, wie z.B. Vergessen des bereits durchgenommenen und früher beherrschten Stoffs, Begrenzung der Bearbeitungszeit in einem Test und mangelnde Anstrengungsbereitschaft, nicht hinreichend in Rechnung gestellt werden.

Helmke, Schrader und Weinert (1987) konnten ebenfalls zeigen, dass die diagnostische Kompetenz der Lehrer hinsichtlich der Einschätzung der Leistungsfähigkeit ihrer Schüler individuell stark variiert, insgesamt eher schlecht ist und nicht mit den tatsächlichen Schülerleistungen übereinstimmt. Ein von den Autoren angeführter Grund dafür, dass Lehrer den aktuellen Wissensstand ihrer Schüler unzutreffend einschätzen, sei die unzureichende diagnostische Sensibilität des Lehrers. Dies führe wiederum dazu, dass Lehrer ihren Schülern zu schwierige oder auch zu leichte Aufgaben geben. Weiter zeigte die Untersuchung, dass es erhebliche Unterschiede zwischen den Lehrern hinsichtlich ihrer diagnostischen Kompetenz gab, wenn sie voraussagen sollten, wie gut jeder einzelne Schüler der eigenen Klasse Testaufgaben lösen werde. Dabei konnten einige Lehrer den Leistungsstand ihrer

Schüler nahezu perfekt vorhersagen, während andere Lehrer diesen kaum vorhersagen konnten. Ein optimaler Aufgabeneinsatz und damit auch deren Effektivität setzen jedoch voraus, dass der Lehrer abschätzen kann, welcher Schüler welchen Wissenstand hat und welche Aufgaben - bzgl. Schwierigkeit und Bearbeitungsaufwand - für diesen Schüler sinnvoll sind (Helmke et al., 1987).

In letzter Zeit wurde durch PISA 2000 die diagnostische Kompetenz von Lehrkräften erfasst, indem die Deutschlehrer mit Hilfe einer vorgegebenen Definition die „schwachen Leser“ (auf dem Niveau der Kompetenzstufe I) ihrer Klasse benennen sollten. Diese wurden dann mit den PISA-Ergebnissen der Schüler verglichen. Dabei gibt es zwei Typen von Urteilsfehlern: (a) schwache Leser (laut PISA) werden von den Lehrern nicht erkannt und werden somit in ihrer Fähigkeit überschätzt; in der Entscheidungstheorie spricht man von *missing*, (b) Leser, die nach PISA nicht schwach sind, werden seitens der Lehrkräfte jedoch als schwache Leser eingeschätzt, diesen Fehlertyp bezeichnet man als *false alarm*. Insgesamt wurden 88.6% der Schüler, die sich laut PISA-Testergebnis auf und unterhalb der Kompetenzstufe 1 befanden, nicht als solche von den Lehrern erkannt, sondern wurden vielmehr als unauffällig eingeschätzt. Dieses Ergebnis ist in Hinblick auf die Förderungsbedürftigkeit dieser Schüler von Bedeutung: Da die geringe Lesefähigkeit der Schüler nicht erkannt wurde, kann diesen Schülern, die weiterer Hilfe bedürfen, diese Förderung nicht zuteil werden. Der zweite Diagnosefehler (*false alarm*) kam sehr viel seltener vor: Nur 2.8% der Schüler wurden in dieser Form falsch diagnostiziert (Baumert, Klieme, Neubrand, Prenzel, Schiefele, Schneider, Stanat, Tillmann & Weiss, 2001). Einschränkend zu diesen Ergebnissen muss beachtet werden, dass die Lehrer zum einen mit der der Bildung von Kompetenzstufen zugrunde liegenden Skalierung nicht vertraut waren, die Definition von „schwachen Lesern“ z.T. relativ unbestimmt war und die Lehrkräfte die Fähigkeit der Schüler einschätzen sollten, jedoch die resultierenden Testergebnisse zugrunde gelegt wurden.

Nicht nur Lehrern fällt es offensichtlich schwer, die Aufgabenschwierigkeit im Bezug auf das Leistungsniveau der Schüler richtig einzuschätzen. Im Rahmen der TIMS III-Studie (Baumert, Bos & Lehmann, 2000) haben Experten der Fachdidaktiken die Lösungshäufigkeiten der Schüler sowohl in Grund- wie in Leistungskursen in den Fächern Mathematik und Physik signifikant überschätzt. Die Resultate im Physikgrundkurs wurden um 10% überschätzt und die im Leistungskurs um 26%. Die Korrelation zwischen beobachteter und geschätzter Lösungshäufigkeit betrug insgesamt in der Physik $r = .40$ ($p < .001$). Je höher das Anspruchsniveau der Aufgabe und je mehr fachliche Kenntnisse vorausgesetzt wurden, desto unrealistischer fiel die Einschätzung aus. Besonders im oberen Leistungsbereich neigen

die Experten also dazu, Kenntnisse und Fähigkeiten von Schülern zu überschätzen (Klieme, 2000).

Insgesamt lässt sich aus den bisherigen Forschungsergebnissen schlussfolgern, dass Lehrer meist über eine unzureichende diagnostische Kompetenz sowohl bzgl. der Einschätzung der Schülerfähigkeiten als auch der Aufgabenschwierigkeit verfügen, obschon interindividuelle Unterschiede zwischen den einzelnen Lehrern bestehen. Jedoch weist Schrader (1989) darauf hin, dass „der Forschungsstand zur Diagnosegenauigkeit von Lehrern noch als zu wenig entwickelt angesehen werden [muss], um hinlänglich gesicherte Informationen zur unterrichtlichen Bedeutung diagnostischer Kompetenz zu liefern“ (S. 69). Diese Einschätzung hat auch heute noch Gültigkeit. Aufgrund bisheriger Ergebnisse zur diagnostischen Kompetenz ist zu erwarten, dass es lediglich eine unzureichende Passung zwischen der Schwierigkeit der vom Lehrer eingesetzten Aufgaben und den Schülerfähigkeiten gibt.

4 Theorien und Modelle zur Aufgabenstellung und Aufgabenbearbeitung

Theorien über die Art der mentalen Repräsentation von Wissen und dem Wissenserwerb sind von großer Bedeutung für das Verständnis der Aufgabenbearbeitung bzw. der dazu notwendigen kognitiven Prozesse (Gardner, 1989). Renkl und Helmke (1992) gehen – wie andere Modelle des Wissenserwerbs auch (z.B. Siegler, 1989) - davon aus, dass schulisches Lernen auch durch die Interaktion mit Lernaufgaben stattfindet. Eine Beschreibung der im Unterricht gestellten Aufgaben sowie der zur Bearbeitung notwendigen kognitiven Prozesse setzt voraus, dass die möglichen zugrunde liegenden Modelle des Lernens und Wissens identifiziert werden. Es gibt verschiedene theoretische Überlegungen, die den Zusammenhang zwischen Aufgaben und Lernprozessen zu erklären versuchen. Dabei werden die Aufgaben mit der jeweiligen Orientierung des Unterrichts in Zusammenhang gebracht. Im Folgenden wird auf behavioristische Ansätze (Kapitel 4.1), kognitionspsychologische Theorien (4.2), konstruktivistische Theorien (4.3) und Wygotskys (1978) Überlegungen zur *Zone der proximalen Entwicklung* (4.4) zur Gestaltung von Lernumgebungen durch Aufgaben überblicksartig eingegangen. Ein weiterer Versuch, Faktoren zu identifizieren, die für das schulische Lernen relevant sind, wird mit Hilfe des Prozess-Produkt-Paradigmas verfolgt (Kapitel 4.5). In der Folge dieses Ansatzes wurden verschiedene Modelle des schulischen Lernens und des Aufgabeneinsatzes entwickelt, auf Blooms Modell des schulischen Lernens (1976) wird in Kapitel 4.6 und auf das Modell der direkten Instruktion in Kapitel 4.7 eingegangen.

4.1 Behavioristische Lerntheorie

Behavioristische Ansätze gehen davon aus, dass sich Lernen im allgemeinen und damit auch in der Schule über die Etablierung und Stärkung einfacher Assoziationen bzw. Assoziationsketten vollzieht. Der wohl einflussreichste Vertreter des Behaviorismus war B.F. Skinner (1958), der seinerzeit die Vorstellungen des Lernens durch operante Konditionierung auf das schulische Lernen übertrug und den Begriff des *programmierten Unterrichts* bzw. der *programmierten Unterweisung* prägte. Dieser Unterricht wird auf der Grundlage der Überlegungen zum *shaping* realisiert, wonach sich ein komplexes Verhalten durch die Verkettung einfacher Reaktionen aufbauen lässt. Dementsprechend wird der Lerner in vorbestimmten, sehr kleinen und aufeinander aufbauenden Lernschritten auf das Lehrziel hinge-

führt. Der Schüler bearbeitet dabei eine Reihe zuvor geplanter Lernschritte bzw. Aufgaben und erhält sofort eine Rückmeldung über die Richtigkeit der Bearbeitung. Diese positive Rückmeldung soll als Verstärker wirken und somit die Auftretenswahrscheinlichkeit des jeweils verstärkten Verhaltens des Schülers erhöhen, so dass gelernt wird. Im Gegensatz dazu bedeutet jede falsche Antwort Zeitverschwendung, da keine Verstärkung erfolgt und somit auch die Auftretenswahrscheinlichkeit des richtigen Verhaltens nicht erhöht wird. Insgesamt muss der Lehrstoff so dargeboten werden, dass operantes Konditionieren möglich ist, die Verstärkung muss planmäßig erfolgen, und jeder Schüler muss das für ihn angemessene Lerntempo wählen können (Cohen, 1963). Da diese Lernform mit den traditionellen Unterrichtsformen kaum zu vereinbaren ist, wurde der Programmierete Unterricht entwickelt. Maßgebliches Kennzeichen der von Skinner (1958) entwickelten Programme ist der Einsatz sehr leichter Aufgaben, durch deren Bearbeitung und die Verstärkung der Antwort die Schüler lernen sollen. Nach dieser Unterrichtsform lernen Schüler demnach mit Hilfe geringer Aufgabenschwierigkeit besonders gut. Inwieweit dieser Annahme auf den regulären Unterricht, in dem die positive Verstärkung der richtigen Antwort nicht immer unmittelbar gegeben ist, übertragbar ist, bedarf der empirischen Überprüfung.

Verschiedene Metaanalysen zur Wirksamkeit des Programmiereten Unterrichts konnten eine durchschnittliche Leistungsverbesserung nach dem Einsatz des Programmiereten Unterrichts im Vergleich zum konventionellen Unterricht von 0.11 bzw. 0.24 Standardabweichungen nachweisen (Hartley, 1977; Kulik, Cohen & Ebeling, 1980). Nach Fletcher (1992) legen diese Ergebnisse nahe, dass „the positive impact of programmed instruction is genuine, but limited“ (p. 16).

Weitere behavioristische Vorstellungen zur Unterrichtsgestaltung zeichnen sich oft durch das Stellen gleichartiger Aufgaben zu bestimmten Themen oder Verfahren aus. Die Aufgaben sind dabei auf einen zu automatisierenden Algorithmus hin ausgewählt, und deren Bearbeitung hat die Funktion des Einübens (Neubrand, 2002). Zu ihrer Beantwortung bedarf es weniger des Erkennens konzeptueller Zusammenhänge, also Verstehensleistungen. Hierbei wird also eher Wissen eingeübt als Verstehen oder Anwendung der erlernten Inhalte entwickelt.

Zusammenfassend kann festgehalten werden, dass sich Lernprozesse in den behavioristischen Theorien prinzipiell über die Ausführung von Verhalten und dessen Verstärkung vollziehen. Aus Platzgründen kann an dieser Stelle keine eingehende Kritik des behavioristischen Ansatzes erfolgen. Es sei jedoch angemerkt, dass dieser Ansatz sich in erster Linie auf „Input-Output“- bzw. Stimulus-Reaktions-Beziehungen beschränkt, die meist in relativ einfachen Lernsituationen zu beobachten sind. Komplexere Formen des Lernens, Verstehens

oder Problemlösens, bei denen zwischen Input und Output mediierende Prozesse von grundlegender Bedeutung sind, bleiben in der Regel ausgespart (Mietzel, 2003).

4.2 Kognitive Lerntheorie

Im Gegensatz zu den behavioristischen Ansätzen betonen die im Zuge der kognitiven Wende entwickelten kognitionstheoretischen Modelle die Lernprozesse und -bedingungen des Lerners in der Gesamtheit der Prozesse, die mit der Informationsaufnahme, ihrer Verarbeitung und Speicherung im Gedächtnis sowie – relevant für die vorliegende Arbeit - deren Anwendung bei der Aufgabenbearbeitung verbunden sind. In den kognitiven Ansätzen (Ausubel, 1974; Gagné, 1965) wird das Lernen als besondere Form der Informationsaufnahme und –speicherung betrachtet, dessen Güte vor allem von der Art der Informationsaufbereitung und Informationsdarbietung einerseits und den kognitiven Aktivitäten des Schülers andererseits abhängig ist. Dem Lernenden wird dabei eine aktive Rolle im Gegensatz zu der passiven Vorstellung des Behaviorismus zugestanden. Eine wichtige Errungenschaft besteht zudem in der Erkenntnis der bedeutsamen Rolle des (Vor-)Wissens sowohl für die kognitiven Leistungen als auch für das weitere Lernen (vgl. Glaser, 1984; Voss, 1987). Lernen bedeutet die Veränderung kognitiver Strukturen und Prozesse.

Das Hauptinteresse kognitionspsychologischer Untersuchungen bezieht sich nach Resnick und Ford (1981) auf die explizite Einbeziehung des Lerninhalts. Dabei geht es um die Frage, wie die Inhalte strukturiert und Unterrichtssequenzen organisiert werden sollten, um die Schüler zu aktivieren. Darin enthalten ist die Frage, welche Bedingungen für Lernaufgaben daraus resultieren. Hier geht es um die Identifikation der Strukturen von Lernaufgaben, um Lernprozesse bzgl. der Informationsaufnahme und –verknüpfung mit bereits gelernten Inhalten bei den Schülern zu initiieren. Lernaufgaben werden dazu von Gagné (1968) in Sequenzen angeordnet, wobei leichtere (bzw. weniger komplexe) Aufgaben als Voraussetzung für schwierigere (bzw. komplexere) Aufgaben eingesetzt werden. Durch die Bearbeitung dieser Aufgaben sollen Wissensnetzwerke aufgebaut werden, statt lediglich richtige Assoziationen einzuüben.

Eine eingehende empirische Überprüfung der kognitionstheoretischen Überlegungen bzgl. der Aufgabengestaltung steht bislang noch aus.

4.3 *Konstruktivismus*

In neueren Theorien spielt der Begriff des konstruktivistischen Lernens eine wichtige Rolle, dem unterliegt die Annahme, dass sich menschliches Denken und Erkennen durch Prozesse des Konstruierens definieren lassen. Demnach ist Wissen keine Kopie der Wirklichkeit, sondern eine Konstruktion durch den Menschen (Reinmann-Rothmeier & Mandl, 1998). Denk- und Handlungsstrukturen entstehen immer aus der Interaktion bereits vorhandener Wissensstrukturen und neuer Wahrnehmungen (vgl. Cobb, Perlwitz & Underwood, 1996; von Glaserfeld, 1994). Wissen wird demnach nicht einfach rezipiert, sondern es wird von den Lernenden aktiv in einem bestimmten Handlungskontext konstruiert (Bednar, Cunningham, Duffy & Perry, 1992). Lernen ist damit ein aktiver Prozess, bei dem Lernaktivitäten wie experimentieren, Erfahrungen einbringen, Hypothesen bilden und überprüfen eine wichtige Rolle im Wissenserwerb spielen. Das konstruktivistische Paradigma hat seine Wurzeln vor allem in Untersuchungen zum Erwerb mathematisch-naturwissenschaftlichen Wissens.

In konstruktivistischen Lernumgebungen haben Aufgaben für Schüler die Funktion, neues Wissen selbstständig zu erarbeiten und nicht nur Wissen zu wiederholen und zu festigen. Der Begriff Lernumgebung meint die Möglichkeiten, anhand derer gelernt wird, und umfasst somit auch Unterrichtsaufgaben. Dem Lerner wird dabei eine aktive, selbstgesteuerte Rolle zugesprochen. Bedeutungsvolle, authentische Lernumgebungen, die Lernhandlungen initiieren und bei allen Schüler Lernaktivitäten hervorrufen, sind daher ein zentrales Merkmal konstruktivistischer Ansätze zum Wissenserwerb. Wissen steht hier in einem Anwendungs- und Handlungskontext (Reinmann-Rothmeier & Mandl, 2001). Die Aufgabe des Lehrers besteht darin, den Aufbau der Wissenskonstruktion der Schüler zu initiieren, zu unterstützen und zu begleiten, indem er Lernumgebungen in strukturierter und dem Vorwissen der Schüler angemessener Form bereitstellt. Wissenserwerb ist demnach besonders effektiv, wenn anwendbares Wissen als bedeutungsvolles Wissen im Handlungskontext gelernt wird.

Wichtig wären empirische Untersuchungen, die es derzeit kaum gibt, um Aussagen über Lernprozesse zu treffen und damit auch konkrete Kriterien für konstruktivistisch gestaltete Lernaufgaben abzuleiten. Da sich konstruktivistische Unterrichtsansätze vor allem dadurch von regulärem Unterricht unterscheiden, dass der konstruktivistische Unterricht inhaltlich auf das Lösen von komplexen Problemen und den Aufbau von Kompetenzen zum Transfer der gelernten Sachverhalte auf neue Problemsituationen gerichtet ist, ist die Übertragbarkeit dieses Konstrukts auf den in der vorliegenden Studie untersuchten regulären Unterricht als begrenzt anzusehen.

4.4 Zone der proximalen Entwicklung nach Wygotsky

Die Überlegungen zur kognitiven Entwicklung aus entwicklungspsychologischer, speziell der soziokulturellen, Perspektive hat besonders Wygotsky (1978, 1993) geprägt. Der Autor beschäftigte sich mit dem Zusammenwirken von Unterricht und kognitiver Entwicklung auf den verschiedenen Altersstufen. Dabei stellt die *zone of proximal development* des Kindes (Zone der nächstmöglichen Entwicklung) ein zentrales Konstrukt dar. Wygotsky (1978) definierte diese Zone als „the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined under adult guidance or in collaboration with more capable peers” (p. 86). Es wird demnach unterschieden zwischen der erfolgten und der potentiellen Entwicklung eines Kindes. Die erfolgte Entwicklung bzw. das aktuelle Entwicklungsniveau bestimmt, was ein Kind ohne Hilfe Anderer leisten kann, während mit der potentiellen Entwicklung zum Ausdruck gebracht wird, was ein Kind unter Anleitung bzw. mit Unterstützung, meist des Lehrers, leisten kann (Mietzel, 2003). Die Größe der Zone lässt sich durch die Distanz zwischen der unteren Grenze (das mentale Alter, angegeben durch die Aufgaben, welche vom Schüler selbstständig bearbeitet werden) und der oberen Grenze (das mentale Alter wird geschätzt über die Leistung, welche mit Hilfestellungen des Lehrers erzielt wird) bestimmen. Die Zone spezifiziert also, was ein Lernender mit Hilfe anderer tun *könnte*. Weiter führt Wygotsky aus, dass das optimale Niveau schulischer Anforderungen gegeben ist, wenn der Unterricht und damit auch die Schwierigkeit der eingesetzten Aufgaben anhand des aktuellen mentalen Alters der Klasse ausgewählt wird. Dass das mentale Alter der Schüler einer Klasse interindividuell variiert, müssen Lehrer berücksichtigen. Dies kann durch den Einsatz von Aufgaben unterschiedlicher Schwierigkeit Berücksichtigung finden.

Die Zone stellt ein grundlegendes Konstrukt der kognitiven Entwicklung dar und kann für die Messung der Schülerleistung genutzt werden, um Unterrichtsmaßnahmen dem zeitlichen Verlauf der Entwicklung und der individuellen Zone der jeweiligen Schüler anzupassen. Dieses Konstrukt verdeutlicht zudem die Bedeutung des Lehrers und dessen angemessene Auswahl von Aufgaben für die Entwicklung von Schülern. Nach dem Konstrukt wird im Laufe des schulischen Lernprozesses die zu Beginn gebotene Unterstützung in Abhängigkeit der vom Schüler gemachten Fortschritte nach und nach reduziert. Wer Unterstützung bietet, muss das aktuelle Lern- und Verständnisniveau des Schülers kennen. Somit muss der Lehrer über ausreichende Diagnosekompetenz verfügen, welche - wie in Kapitel 3.4 ausgeführt wurde - nicht immer gegeben ist (Tudge & Rogoff, 1989). Die Funktion der Unterstützung ist es dabei, die Aufgabenschwierigkeit so zu verringern, dass der Schüler die

Aufgabe, welche sich am oberen Rand der Zone der nächstmöglichen Entwicklung befindet, erfolgreich lösen kann (Mietzel, 2003). Wegen der individuellen Unterschiede bzgl. der benötigten Unterstützung sind in diesem Zusammenhang die Möglichkeiten des adaptiven Unterrichts von Bedeutung, auf die in Kapitel 12 detaillierter eingegangen wird. In nachfolgenden Arbeiten, in denen Wygotskys Vorstellungen weiterentwickelt worden sind, wird stärker die Abhängigkeit der Zone vom dem jeweils verfügbaren Wissen in einem Lernbereich herausgestellt (z.B. Greenfield, 1984; Wood, Bruner & Ross, 1976).

4.5 Das Prozess-Produkt-Paradigma

Das Prozess-Produkt-Paradigma ist ein Forschungsansatz der Unterrichtsforschung, der auf Merkmale effektiven Unterrichtshandelns von Lehrern fokussiert. Im Rahmen dessen wurden Versuche zur Identifikation relevanter Faktoren des schulischen Lernens und der Aufgabenbearbeitung unternommen (vgl. Brophy & Good, 1986; Weinert, Schrader & Helmke, 1989). Die zugrunde liegende Frage bezieht sich auf die Beziehung zwischen einem bestimmten Lehrerverhalten als Prozess-Variable und spezifischen Schülerleistungen als Produktvariablen. Im Kontext dieser Arbeit interessiert besonders der Zusammenhang zwischen dem Einsatz verschiedener Aufgabenarten (als Prozessvariable auf der Lehrerseite) und den Schülerleistungen als Produktvariable.

Als Untersuchungsergebnisse der Prozess-Produkt-Forschung resultieren meist statistische Zusammenhänge in Form von Korrelationen. Dieses Vorgehen erscheint intuitiv plausibel, wenn es darum geht, Gesetzmäßigkeiten des Lehrens und Lernens zu erforschen. Mit den korrelativen Ergebnissen, wie z.B. dem Zusammenhang zwischen Aufgabenart (*low level*- vs. *high level*-Aufgaben) und Schülerleistung, wird – in z.T. unzulässiger Weise - versucht, für den Leistungszuwachs kausal bedeutsame Unterrichtscharakteristika zu identifizieren. Ein weiterer Nachteil ist gegeben, wenn Korrelationen als einzige Methode eingesetzt werden, da dann das lehrende bzw. lernende Individuum mit seinen spezifischen Merkmalskonfigurationen außer Acht bleibt. Helmke (2004) schlägt daher vor, diesen variablenorientierten Forschungsansatz durch einen personenzentrierten Ansatz zu ergänzen, bei dem diejenigen Lehrer identifiziert werden, die einen besonders „guten Unterricht“ zeigen, welcher dann in einem weiteren Schritt untersucht wird. Abgesehen von der kausalen Interpretation korrelativer Befunde wird häufig kritisiert, dass Korrelationen gewöhnlich lediglich lineare Zusammenhänge erfassen, jedoch auch nichtlineare Zusammenhänge vorhanden sein können. Solche Zusammenhänge werden nach Helmke (2004) sehr selten geprüft. Weitere Kritik an diesem Ansatz bezieht sich auf das eher atheoretische Vorgehen, welches meist mit

ad-hoc-Anleihen aus anderen Forschungssträngen oder Plausibilitätsannahmen arbeitet und keine eigenständige, elaborierte Theorie hervorgebracht habe (vgl. Winne, 1987). Auch ist zu kritisieren, dass bei der Untersuchung der Effekte von Lehrervariablen, welche Schülerleistungen beeinflussen, medierende Faktoren, wie z.B. die Informationsverarbeitungsprozesse der Schüler, meist nicht berücksichtigt werden (Brophy & Good, 1986).

Ausgehend von den oben genannten Kritikpunkten, besonders der fehlenden Frage nach dem *Warum* korrelativer Zusammenhänge in der bisherigen Prozess-Produkt-Forschung, wurde das Paradigma zum einen um vermittelnde und interpretative Prozesse auf Schülerseite, *mediating processes*, erweitert (Doyle, 1986; Winne & Marx, 1979), und zum anderen wurden komplexere Wirkbeziehungen, wie z.B. auch nichtlineare Effekte und Wechselwirkungen, berücksichtigt. Ebenfalls aus dieser Kritik entstanden verschiedene theoretische Modelle, die als Erweiterung des Prozess-Produkt-Paradigmas gesehen werden können. Hierzu zählt z.B. eine auf Carroll (1963) zurückgehende Tradition von Modellen des Schullernens, welche der Quantität an Lern- bzw. Unterrichtszeit eine entscheidende Rolle zuspricht (Bloom, 1976; vgl. Kapitel 4.6), sowie Modelle der direkten Instruktion (z.B. Peterson, 1979; Rosenshine, 1979; vgl. Kapitel 4.7). In beiden Modellen wird die Funktion von Lernaufgaben für das schulische Lernen erstmals explizit berücksichtigt, weshalb diese Modelle im Folgenden skizziert werden.

4.6 Blooms Modell des schulischen Lernens (1976)

Hintergrund des Modells von Bloom (1976) sind optimistische Erwartungen über die Wirksamkeit des schulischen Lehrens und Lernens, welche unter dem Begriff des *mastery learnings* Eingang in die pädagogische Psychologie der 1960er und 1970er Jahre gefunden hat: „What any person in the world can learn, almost all persons can learn *if* provided with appropriate prior and current conditions of learning“ (Bloom, 1976, p. 7). In dem Zitat wird die Annahme deutlich, dass fast alle Schüler lernen können, was in der Schule unterrichtet wird, wenn sie angemessen unterrichtet werden und ausreichend Zeit zur Bearbeitung der Lernaufgaben erhalten. Damit wird die Nähe zum Modell des schulischen Lernens von Carroll (1963) deutlich.

Das Modell des schulischen Lernens von Bloom (1976) stellt einen Versuch dar, mit Hilfe einiger relevanter Variablen die Unterschiede im schulischen Lernen zu erklären. Im Mittelpunkt der Theorie stehen drei sich wechselseitig beeinflussende Variablengruppen: auf der Seite der Schüler sind die kognitiven Voraussetzungen und das Ausmaß der Motivation, am Lernprozess teilzunehmen, von Bedeutung. Die Qualität des Unterrichts und die darin

gestellten Lernaufgaben stellen die zweite Variablengruppe dar. Die dritte, zentrale Variablengruppe steht unter der Kontrolle des Lehrers und bezieht sich auf die Angemessenheit des Unterrichts in Bezug auf die Lernergebnisse der Schüler. Insgesamt beschäftigt sich die Theorie somit mit Schülermerkmalen, Unterricht und Lernergebnissen und den Wechselwirkungen dieser Variablen (vgl. Abbildung 1).

Bloom (1976) sieht die Beteiligung des Schülers am Lernen als eine Funktion kognitiver und affektiver Merkmale an. Die affektiven Merkmale sind Lernmotivation, Einstellung zur Schule und das Selbstkonzept des Schülers als Lerner (Anderson, 1973; Hecht, 1977). Dem Vorwissen und der bisherigen Lerngeschichte des Schülers als kognitiven Merkmalen wird besondere Bedeutung beigemessen, da diese die Art der Interaktion zwischen Schüler und Lernaufgabe festlegen (Bloom, 1976). Denn eine Lernaufgabe, so einfach sie sein mag, kann nur dann gelöst werden, wenn ein gewisses Vorwissen gegeben ist und der Schüler die Aufgabenstellung verstanden hat.

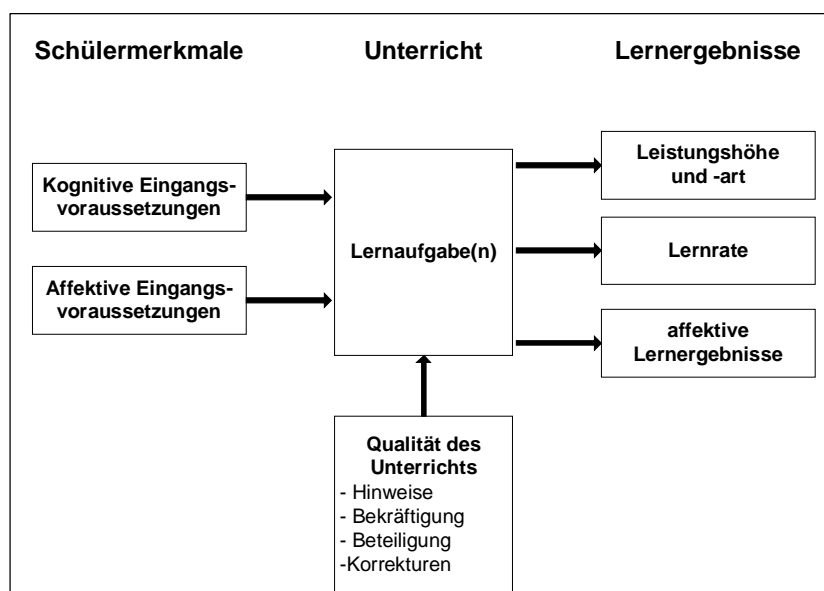


Abbildung 1: Modell des schulischen Lernens nach Bloom (1976)

Die Qualität des Unterrichts hat einen einflussreichen Effekt auf die Lernergebnisse der Schüler. Dabei werden verschiedene bedeutsame Variablen berücksichtigt: Hinweise und Anweisungen, die dem Schüler zur Verfügung gestellt werden, Verstärkung, welche den Schüler in seinem Lernverhalten bekräftigen sollen, Teilnahme des Schülers an den Lernaktivitäten sowie Rückmeldung und Korrektur bzgl. der Aufgabenbearbeitung und der Lernergebnisse (vgl. Bloom, 1976).

Unterschiede in den Schülermerkmalen und in der Qualität des Unterrichts determinieren in dem Modell Art und Ausmaß der Lernergebnisse. Diese werden in dem Modell ebenfalls differenziert nach Leistungshöhe und –art, Lernrate und den affektiven Lernergebnissen betrachtet. Bloom (1976) postuliert, dass die Lernergebnisse positiv sein werden, wenn eine vorteilhafte Kombination aus Schülermerkmalen und Qualität des Unterrichts besteht. Lernschwierigkeiten treten dagegen auf, wenn eine oder mehrere dieser Variablen nicht optimal ausgeprägt sind. Um die Unterschiede des Vorwissens der Schüler im Unterricht zu berücksichtigen, schlägt Bloom (1976) vor, die Lernaufgaben so zu verändern, dass sie jeweils unterschiedliches Vorwissen verlangen. Somit wird in diesem Modell die Idee einer adaptiven Anpassung verfolgt, auf welche in Kapitel 12 näher eingegangen wird.

Wie zu Beginn des Unterkapitels bereits erwähnt, bildet das Modell für Bloom (1976) die Grundlage für das Konzept des zielerreichenden Lehrens und Lernens (*mastery learning*). Damit ist die unterrichtliche Strategie gemeint, durch Gewährung und optimale Nutzung zusätzlicher Lernzeit (20 bis 30%) möglichst vielen Schülern einer Klasse (z.B. 90 bis 95% aller Schüler) das Erreichen anspruchsvoller Lernziele, z.B. 90 bis 95% richtiger Lösungen bei curricular validen Schulleistungstests, zu ermöglichen. Zugleich wird angenommen, dass sich bei längerfristiger Anwendung dieser Strategie die kognitiven und motivationalen Differenzen zwischen den Schülern und damit auch die Unterschiede in der benötigten Lernzeit zunehmend reduzieren lassen.

Kulik und Kulik (1987) konnten die Effektivität des *mastery learnings* im Vergleich zu regulärem Unterricht für die Schülerleistung mit einer Effektstärke von $d = 0.47$ nachweisen. Jedoch zeigen andere Befunde, dass bei unterschiedlichen Lernvoraussetzungen der Schüler kontinuierlich große Unterschiede in der benötigten Lernzeit bestehen bleiben, wenn man gleich anspruchsvolle Leistungen anstrebt; oder es treten erhebliche Differenzen in den Lernleistungen auf, wenn gleiche Lernzeiten gewährt werden (vgl. Arlin, 1984; Slavin, 1987). Relativiert man die unter der *mastery*-Bedingung erzielten Lernerträge am benötigten Zeitaufwand, so scheint der traditionelle Unterricht den Ansätzen zielerreichenden Lernens überlegen zu sein (Arlin & Webster, 1983). Helmke und Weinert (1997) schließen daraus, dass die Egalisierung interindividueller Schülerdifferenzen nur ein spezifisches, aber kein generelles Unterrichtsziel im Klassenzimmer sein kann. Auch Mietzel (2003) sieht es im alltäglichen Unterricht als problematisch an, wo die zusätzliche Unterrichtszeit, die schwächere Schüler benötigen, hergenommen werden soll.

Ein Verdienst des Autors (Bloom, 1976) stellt, trotz der ernüchternden empirischen Überprüfung, die Entwicklung eines Modells des schulischen Lernens dar, in dem explizit auf die Bedeutung von Lernaufgaben für den Unterricht und das Lernen der Schüler einge-

gangen wird. Des Weiteren wird der Einfluss interindividueller Unterschiede in den kognitiven Voraussetzungen für die Bearbeitung der Lernaufgaben und als Folge dessen des Lernens betont. Diese individuellen Unterschiede zwischen den Schülern, welche die Aufgabenbearbeitung beeinflussen können, sind in der in den Kapiteln 12-16 vorgestellten Studie 2 von großer Bedeutung. Auch haben sich verschiedene Autoren in der Folgezeit um eine Weiterentwicklung dieses Modells bemüht (vgl. Creemers, 1994; Slavin, 1987, 1994).

4.7 Das Modell der Direkten Instruktion

Diejenigen Lehrerverhaltensweisen, die sich im Forschungsansatz des Prozess-Produkt-Paradigmas auf der Suche nach erfolgreichen Lehrerhandeln eines effektiven Unterrichtstils als lernförderlich erwiesen haben, wurden zum Modell der „*Direkten Instruktion*“ zusammengefasst (Peterson, 1979; Rosenshine, 1979), welches in verschiedenen Varianten existiert: z.B. *active teaching* (Brophy & Good, 1986) oder *explicit teaching* (Rosenshine, 1987). „There are times when the most effective and efficient way to teach students is for the teacher to present information, skills, or concepts in a direct fashion. The term direct instruction is used to describe lessons in which the teacher transmits information directly to students, structuring class time to reach a clearly defined set of objectives as efficiently as possible“ (Slavin, 1997, p. 265). In diesem Zitat wird das Vorgehen des direkten Unterrichts deutlich. Es handelt sich alles in allem um einen lehrerzentrierten Unterrichtsstil, der auf fachliche Inhalte fokussiert, welche durch viele, relativ leicht zu bewältigende Aufgaben gekennzeichnet sind (vgl. Rosenshine, 1979, 1987). Es wird implizit ein Fertigungsansatz vertreten, der annimmt, dass sich guter Unterricht additiv aus Einzelkomponenten zusammensetzt (vgl. Shulman, 1986; Weinert, Schrader & Helmke, 1990). Burns (1984) gliedert den Unterricht nach drei Aspekten: dem Zweck, den Tätigkeitsmerkmalen und der inhaltlichen Aufgabe, welche wiederum durch ihren Inhalt, ihre „prozedurale Kompliziertheit“ und das Niveau der kognitiven Anforderung charakterisiert wird.

Nach der Durchsicht der Ergebnisse der Prozess-Produkt-Forschung haben sich für das Stellen von Aufgaben die folgenden leistungsrelevanten Variablen herausgestellt: Nutzung der Unterrichtszeit durch eine hohe Anzahl von Aufgaben bzw. ausgeprägte Aufgabenorientierung, die Schwierigkeit der Aufgaben und ihr kognitives Niveau, die Passung des Unterrichts an die Lernvoraussetzungen der Schüler und die kontinuierliche Diagnose ihres Lernfortschritts (Brophy & Good, 1986). Diese Variablen wurden durch neuere Studien im deutschsprachigen Raum weitgehend bestätigt (Helmke & Weinert, 1997). Auf die Effekte dieser Variablen wurde bereits in den Ausführungen des Kapitels 3 eingegangen.

Die bisherigen empirischen Überprüfungen der Methode der direkten Instruktion konnten keine generelle Überlegenheit bzgl. des Lernerfolgs oder der Leistung im Vergleich zu anderen Lehrmethoden nachweisen. Wenn das Unterrichtsziel ist, alle Schüler einer Klasse so gut wie möglich zu fördern und hohe Durchschnittsleistungen mit geringen interindividuellen Varianzen zu erzielen, so hat sich die Methode der *direkten Instruktion* bewährt (Peterson, 1979; Rosenshine, 1979). Jedoch muss berücksichtigt werden, dass die bisher durchgeführten Studien auf grundlegende Fertigkeiten wie Lesen und Grundrechenarten fokussierten und meist mit Grundschulern durchgeführt wurden. Für andere Inhaltsbereiche und Altersgruppen gibt es bislang keine empirische Überprüfung der Effektivität (Slavin, 1997). Shulman (1982) weist darauf hin, dass dieses Unterrichtsmodell im Widerspruch zu den aktuellen Modellen des Wissenserwerbs steht, welche die geistigen Eigenaktivitäten der Schüler als entscheidende Bedingung für einen effizienten Wissensaufbau identifizieren. Es wird kritisiert, dass der lehrerzentrierte Unterricht Schüler zu passiven und unmotivierten Lernern mache (vgl. Gardner, 1993). Es wird angenommen, dass die direkte Instruktion die Kinder symbolisch an die Hand nimmt und die kognitiven Prozesse auf Schülerseite initiiert, die zu effizientem Lernen führen (Shulman, 1982; Weinert & Helmke, 1987). Slavin (1997) weist darauf hin, dass die Methode der direkten Instruktion wenig geeignet erscheint, um *conceptual change*, entdeckendes Lernen und tiefes Verständnis auf Schülerseite zu initiieren. Renkl (1991) nimmt spekulativ an, dass ein Unterricht im Sinne der direkten Instruktion durch sein fehlervermeidendes Vorgehen besonders zur Automatisierung von Prozeduren geeignet ist, während ein Verständnis fördernder Lehrstil in erster Linie die Problemlösekompetenz fördert. Gruber, Mandl und Renkl (2000) weisen darauf hin, dass die alleinige Verwendung des Konzepts der direkten Instruktion im Gegensatz zu instruktionalen Versuchen steht, bei denen Lerninhalte mit Hilfe von multiplen Perspektiven und in unterschiedlichen Kontexten betrachtet werden, um so die Entstehung von „trägem Wissen“ einzudämmen und die Möglichkeit zum Transfer zu erhöhen. Vielmehr seien für unterschiedliche Kompetenzaspekte verschiedene Unterrichtsformen anzuwenden: Für den Erwerb inhaltlichen Wissens erweisen sich – so die Autoren - Formen der direkten Instruktion als geeignet, für den Erwerb von Anwendungswissen Formen des situierten Lernens. Über das Bereitstellen komplexer Lernumgebungen kann die Entwicklung verstehenden, auf Erfahrung beruhenden Lernens in authentischen Situationen realisiert werden (Baumert, 1998).

Insgesamt kann die Bedeutung des Modells der direkten Instruktion darin gesehen werden, dass dieses einen Versuch darstellt, die Variablen, welche sich in der Prozess-Produkt-Forschung des regulären Unterrichts als relevant für das schulische Lernen herausgestellt haben, in einem gemeinsamem Rahmen zu betrachten. Inwieweit das Modell auf den

in den vorliegenden Studien untersuchten Inhaltsbereich der Physik und höhere Schulklassen angewendet werden kann, muss die empirische Überprüfung noch zeigen (vgl. Slavin, 1997).

4.8 Fazit

Der Nutzen der vorgestellten Theorien und Überlegungen für die Erklärung der unterrichtlichen Effekte von Aufgaben aus Kapitel 3 ist z.T. begrenzt. Da bislang jedoch keine eigenen theoretischen Modelle zum Lernen durch die Bearbeitung von Aufgaben bzw. zum Effekt unterschiedlicher Aufgabenarten existieren, muss zur Erklärung dieser Effekte auf vorhandene Modelle zurückgegriffen werden.

Beim behavioristisch orientiertem Lernen und dem programmierten Unterricht handelt es sich meist um sehr einfachere Aufgaben, wobei die Bedeutung der direkt folgenden Verstärkung eine besondere Rolle spielt, deren Funktion für das Lernen nicht ohne Einschränkungen auf die komplexen Lernprozesse im Unterricht übertragen werden kann. Auch wird von einem passiven Lerner in diesen Überlegungen ausgegangen, was so nicht mehr haltbar ist. Die kognitionspsychologischen und konstruktivistischen Überlegungen gehen im Gegensatz dazu von einer aktiven Rolle des Schülers aus, welche für die Lösung komplexer Aufgaben besonders relevant erscheint. Meist wurden diese Hypothesen in experimentellen Studien überprüft. Inwieweit die Ergebnisse auf das Unterrichtsgeschehen übertragbar sind, ist zumindest teilweise fraglich. Um jedoch die Ergebnisse zur Aufgabenschwierigkeit bzw. die unterschiedlichen Effekte verschiedener Aufgabentypen erklären zu können, reichen diese Ansätze nicht aus. Auch die alleinige Fokussierung auf die benötigte Lernzeit in den Untersuchungen zum *mastery learning* ist nicht ausreichend, um den individuellen Unterschieden zwischen Lernern gerecht zu werden. Wygotskys (1978) Konzept erscheint dabei eine bedeutsame Rolle zu spielen, da hier die individuellen Unterschiede der einzelnen Lerner betont werden und die Idee der adaptiven Anpassung der Aufgabenschwierigkeit an die individuellen Fähigkeiten der Schüler herangezogen werden kann, um die inkonsistenten Ergebnisse aus Kapitel 3 mit zu erklären.

Die empirischen Befunde zur Schwierigkeit von Unterrichtsaufgaben kann jedoch keine der genannten Theorien ausreichend erklären. Ein Teil der Schwierigkeiten, diese Theorien auf die Unterrichtssituation anzuwenden, kommt dadurch zustande, dass diese und die zugehörigen Studien nicht zur unmittelbaren praktischen Verwendung angelegt und deshalb nicht direkt übertragbar sind (Shull, 1980). Befunde über Lernen und Kognitionen sind freilich notwendig, um effektive Unterrichtsmethoden und –materialien zu entwickeln, doch sind diese Kenntnisse qualitativ unterschiedlich von den Kenntnissen, welche zur Entwick-

lung von Unterrichtsmethoden essentiell sind (Gagné, 1962). Auch gehen die meisten theoretischen Überlegungen zur Aufgabenbearbeitung von einem genau definierten Zustand aus, von dem aus der Lerner startet. Diese Annahme ist jedoch nicht haltbar, da die kognitiven Strukturen verschiedener Lerner nicht als homogen angenommen werden können. Es wird zudem nicht darauf eingegangen, wie dieser Zustand überhaupt erreicht wird, d.h. wie die Wissensbasis, die für die Vorgänge beim Bearbeiten von Aufgaben heranzuziehen ist, aufgebaut wird. Insgesamt können die vorliegenden Forschungsergebnisse zur Lernwirksamkeit von Aufgaben nicht umfassend mit Hilfe der vorhandenen theoretischen Überlegungen erklärt werden. Es bedarf folglich weiterer theoriegeleiteter Forschung bzgl. der Wirkmechanismen von Aufgaben auf die Lernleistung der Schüler.

5 Klassifikationen von Lernaufgaben

In diesem Kapitel wird zunächst eine bestehende, in der Unterrichtsforschung bereits vielfältig angewandte Taxonomie von Lehrzielen und Lernaufgaben skizziert, die Bloom'sche Lehrzieltaxonomie (Bloom et al., 1956; Kapitel 5.2). Darauf aufbauend wird anschließend das in dieser Arbeit berücksichtigte System zur Klassifikation der kognitiven Anforderungen von Aufgaben (Anderson et al., 2001) beschrieben (Kapitel 5.3). Im letzten Unterkapitel (5.4) werden bisherige Forschungsergebnisse zu der Taxonomie dargestellt.

5.1 Nutzen von Taxonomien

Schule und Unterricht verfolgen bekanntlich eine Vielzahl von Zielen und haben zahlreiche -intendierte oder auch unbeabsichtigte - Wirkungen auf die Schüler. Die Klassifikation solcher Ziele und Effekte ist keineswegs nur akademisch, sondern für die Frage nach der Unterrichtsqualität von großer Wichtigkeit, denn Aussagen über die Güte und Angemessenheit ausschlaggebender Unterrichtsstile und -methoden lassen sich immer nur im Hinblick auf spezifische Ziele machen (Helmke, 2004). Eine Lehrzieltaxonomie stellt eine Klassifikation des Verhaltens dar, das die beabsichtigten Ergebnisse einer Bildungsmaßnahme repräsentiert (Bloom et al., 1956) und der genaueren Beschreibung eines Lehrziels hinsichtlich des Anforderungsniveaus an die Lernenden dient.

In der Literatur wird eine Fülle von Begriffen - wie z.B. Lehrziel, Lernziel, Bildungsziel usw. - synonym verwendet (Mietzel, 2003). Im Folgenden wird von Lehrziel gesprochen. Lehrziele beschreiben das Verhalten und die Prozesse, welche die Schüler durch den Unterricht erwerben sollen (Levin & Long, 1981). Die Funktion der Kategorisierung von Lehrzielen ist die Optimierung der durch Lernaufgaben zu initiierten Lernprozesse im Unterricht (Gall, 1970; Seel, 1981). Dabei ist eine weit verbreitete Unterteilung von Lehrzielen die nach dem Schwierigkeits- und Komplexitätsgrad der kognitiven Prozesse (Bloom et al., 1956).

Das seit seiner Entwicklung gebräuchlichste Dimensionierungskonzept und Hierarchisierungsinstrument zur Klassifizierung des Anforderungsniveaus von Lehrzielen stellt die Taxonomie von Lehrzielen im kognitiven Bereich nach Bloom et al. (1956) dar, auf welche im folgenden Kapitel genauer eingegangen wird.

5.2 Die Taxonomie von Lehrzielen im kognitiven Bereich (Bloom et al., 1956)

Die Lehrzieltaxonomie von Bloom et al. (1956) ordnet Lehrziele nach der Komplexität der geforderten Verhaltensweisen im kognitiven Bereich. Sie stellt eine eindimensionale Klassifikation kognitiver Prozesse dar. Es wird davon ausgegangen, dass die Schwierigkeit einer Aufgabe von den für ihre Lösung erforderlichen kognitiven Prozessen abhängt. Diese kognitiven Prozesse, die von einfachen Reproduktionsleistungen bis zu komplexen Problemlöse- und Argumentationstätigkeiten reichen, werden in folgende sechs Kategorien eingeteilt:

1. *Wissen*: Erinnern spezifischer Fakten,
2. *Verstehen*: Übersetzen, Interpretieren und Extrapolieren von Sachverhalten innerhalb eines Inhaltsbereichs,
3. *Anwenden*: Generalisierung und Gebrauch abstrakter Informationen in konkreten Situationen,
4. *Analysieren*: Zerlegen eines Problems in seine grundlegenden Teile und Entdecken von Beziehungen zwischen den Teilen,
5. *Synthetisieren*: Zusammenfügen von Teilen zu einem Ganzen,
6. *Evaluiieren*: Nutzen von Kriterien, um Urteile über den Wert eines Sachverhalts zu fällen.

Bei der Taxonomie handelt es sich um ein hierarchisches Modell: Es wird postuliert, dass jede ranghöhere Kategorie den Inhalt aller ihr untergeordneten Kategorien einschließt. Somit kann ein Lehrziel oder das Beherrschen einer Fertigkeit nur dann erreicht werden, wenn die untergeordneten kognitiven Ebenen auch beherrscht werden (Bloom et al., 1956; Messner, 1973).

Das pädagogische Ziel dieser Taxonomie bestand vor allem darin, die Lehrer zu kognitiv anspruchsvolleren, über bloßes Faktenwissen hinausgehenden Fragen, Anforderungen und Aufgaben anzuregen (Helmke, 2004). So weisen auch Lienert und Raatz (1998) darauf hin, dass eine Einteilung nach der inhaltlichen Komplexität von Lernaufgaben auf Grundlage dieser Taxonomie erfolgen sollte, wenn es um die praktische Bedeutsamkeit, d.h. um die Anwendung auf Lernaufgaben in Schulleistungstests geht.

Für empirische Anwendungszwecke kam häufig eine Zusammenfassung der von Bloom et al. (1956) zur Lehrzieloperationalisierung entwickelten sechsstufigen Taxonomie zum Einsatz. In der Regel werden in Studien, die Aufgaben auf der Grundlage der Taxono-

mie klassifizieren und konstruieren, die sechs Stufen auf zwei (bzw. drei) Stufen reduziert (vgl. Kapitel 3; Brophy & Good, 1986). Nach Levin (2005) ist zu bedenken, inwieweit durch eine solche Vereinfachung mögliche Unterschiede in den Wirkungsweisen von Aufgaben, welche zur Bearbeitung, z.B. den kognitiven Prozess der *Bewertung* bzw. *Anwendung* erfordern, verdeckt werden. Es wäre denkbar, dass diese unterschiedlichen Aufgabentypen andere Formen der Wissensverarbeitung und –vertiefung begünstigen. Dubs (1978) kritisiert, dass die Niveaustufen nicht hinreichend trennscharf formuliert sind.

Die Taxonomie wird dennoch vorzugsweise herangezogen, da sie sich in der Unterrichtsforschung für die lehrzieloperationalisierte Unterrichtsplanung als bedeutsam erwiesen hat. Sie bietet eine relativ einfach zu handhabende Möglichkeit, Aufgaben gemäß ihrer kognitiven Niveaustufen einzuordnen. Spätere Versuche, die im Unterricht gestellten Aufgaben mit Hilfe eines Klassifikationssystems bzgl. der kognitiven Prozesse zu kategorisieren (z.B. Graesser & Person, 1994; West & Pearson, 1994), konnten weder in Bezug auf die empirische Überprüfung, noch auf die Praktikabilität die Bedeutung der Bloom'schen Taxonomie erreichen (Levin, 2005).

Nach der Veröffentlichung der Taxonomie von Bloom et al. (1956) sind eine ganze Reihe von weiteren Klassifikationssystemen - meist auf Grundlage von Bloom - entwickelt worden, um diese zu verbessern und zu erweitern. Eine umfassende Beschreibung der vielseitigen Weiterentwicklungen kann im Rahmen dieser Arbeit aus Platzgründen nicht erfolgen und ist auch nicht die Fragestellung der hier vorgestellten Studien. Daher sei an dieser Stelle auf die Arbeiten von Biggs und Collins (1982), Gagné (1968, 1972) Gerlach und Sullivan (1967), Marzano (1992) Quellmalz (1987), Sanders (1966) und Williams (1977) verwiesen. Auch in neueren Untersuchungen wird häufig auf die Bloom'sche Taxonomie zurückgegriffen. So wurde beispielsweise in der dritten internationalen Mathematik- und Naturwissenschaftsstudie der IEA, der TIMS-Studie, eine Taxonomie von Lernzielen verwendet, die auf Blooms Taxonomie (Bloom et al., 1956) aufbaut (Robitaille & McKnight, 1993).

Insgesamt hatte und hat die Bloom'sche Lehrzieltaxonomie einen bedeutsamen Einfluss auf die weitere Entwicklung von Lehrzielen und im Rahmen dessen auch auf die Beschreibung und Konstruktion von Lernaufgaben. Im folgenden Kapitel (4.3) soll lediglich auf die dieser Arbeit zugrunde gelegte Revision von Anderson et al. (2001) näher eingegangen werden.

5.3 Revision von Anderson et al. (2001)

Anders als bei Bloom et al. (1956) ist die Revision der Taxonomie durch Anderson et al. (2001) zweidimensional angelegt. Die Dimensionen beziehen sich zum einen auf die *kognitiven Prozesse* und zum anderen auf das *Wissen*, dargestellt in Tabelle 1.

Tabelle 1: Lehrzieltaxonomie nach Anderson et al. (2001)

Kognitive Prozess-Kategorien	Wissensdimensionen			
	A. Faktenwissen	B. Konzeptuelles Wissen	C. Prozedurales Wissen	D. Metakognitives Wissen
1. Wissen (Relevantes Wissen aus dem Langzeitgedächtnis abrufen)				
2. Verstehen (Bedeutung/Relevanz von Wissen erkennen und herstellen, indem z.B., neues mit altem Wissen verknüpft wird.)				
3. Anwenden (Bestimmte Verfahren in bestimmten Situationen ausführen/verwenden)				
4. Analysieren (Gliederung eines Materials in seine konstituierenden Teile und Bestimmung ihrer Interrelation und/oder Relation zu einer übergeordneten Struktur)				
5. Bewerten (Urteile anhand von Kriterien und Standards fällen)				
6. (Er)Schaffen, konstruieren, entwickeln (Elemente zu einem neuen, kohärenten, funktionierenden Ganzen zusammenführen/reorganisieren.)				

Die *Wissensdimension* wird in der vorliegenden Arbeit nicht weiter berücksichtigt, wird aber der Vollständigkeit halber mit aufgeführt. Sie umfasst die vier Wissenskategorien Fakten-, konzeptuelles-, prozedurales- und metakognitives Wissen. Das der Wissensdimension zugrunde liegende Kontinuum verläuft von konkretem (Faktenwissen) zu abstraktem (metakognitivem) Wissen.

Bei der *kognitiven Prozessdimension* platzieren die Autoren (Anderson et al., 2001) im Gegensatz zur Bloom'schen Taxonomie die Stufe der Evaluation vor der Stufe der Synthese, welche als *create* bezeichnet wird. Übersetzt bedeutet dies kreieren, erschaffen und entwickeln.

Auf der Grundlage der ursprünglichen Taxonomie Blooms et al. (1956) und anderer Klassifikationssysteme (z.B. DeLandsheere, 1977; Metfessel, Michael & Kirsner, 1969; Mosenthal, 1998; Royer, Ciscero & Carlo, 1993; Sternberg, 1998) wurden neben den sechs kognitiven Hauptprozessen (Wissen, Verstehen, Anwendung, Analyse, Evaluation und Entwicklung) 19 weitere kognitive Prozesse von den Autoren (Anderson et al. 2001) berücksichtigt. Diese wurden den Hauptkategorien als Untertypen zugeordnet, wodurch eine detaillierte Beschreibung der kognitiven Prozesse erreicht werden soll. In der Revision von Anderson et al. (2001) stehen an einem Ende des Kontinuums Grundfertigkeiten und Fakten, am anderen Ende höherrangige Fertigkeiten und fortgeschrittenes Wissen, womit kritische Denkfähigkeit, Problemlösungsfähigkeit und Verstehen gemeint sind. Beide Wissensarten werden als relevant und notwendig für das schulische Lernen betrachtet (Anderson et al., 2001).

Für die Aufgabenklassifikation und –konstruktion werden in der vorliegenden Arbeit ausschließlich die drei kognitiven Prozesskategorien Wissen, Verstehen und Anwenden berücksichtigt und im Rahmen der Beschreibung der Instrumente (Kapitel 9.4) genauer dargestellt. Eine Dreiteilung der Aufgaben hat sich bereits in den Untersuchungen von Ballstaedt, Mandl, Schnotz und Tergan (1981) sowie Levin (2005) als sinnvoll herausgestellt. Auf die Wissensdimension und die höheren kognitiven Prozesse wurde aufgrund mehrerer Faktoren verzichtet: Zum einen sind dies Gründe der Aufgabenkonstruktion, da es bislang keine trennscharfen Konstruktionshinweise gibt (vgl. Kropp & Stoker, 1966) und die Sequenz der höheren Prozesse bislang empirisch nicht bestätigt wurde (vgl. Kapitel 5.4). Zum anderen hätte die Berücksichtigung der vier Wissenskategorien bei jedem der drei kognitiven Prozesse zu insgesamt 12 Kategorien geführt, was die Schüler, besonders Hauptschüler, möglicherweise überfordert hätte (vgl. Studie 1).

Die revidierte Taxonomie von Anderson et al. (2001) wird in der vorliegenden Studie genutzt, um die von Lehrern im Unterricht eingesetzten Aufgaben zu kategorisieren, und sie stellt den Ausgangspunkt für die Konstruktion unterschiedlich schwieriger Aufgaben dar. Daher wird in Kapitel 9.5 genauer auf die bei der Aufgabenkonstruktion berücksichtigten kognitiven Prozesse eingegangen.

5.4 Empirische Ergebnisse zur Struktur der Taxonomie

Da die Originaltaxonomie auf der Grundlage von heuristischen Annahmen entwickelt wurde, waren Forscher seit der Veröffentlichung daran interessiert zu überprüfen, ob die gemachten Aussagen empirisch bestätigt werden können. Besonderes Forschungsinteresse

galt der Annahme der aufsteigenden Schwierigkeit über den Verlauf der sechs Niveaustufen. Dabei nehmen die Autoren (Anderson et al., 2001; siehe auch Kottke & Schuster, 1990) an, dass sich die Kategorien in ihrer Schwierigkeit nicht immer klar voneinander unterscheiden lassen, sondern dass es Überschneidungen in ihrer Schwierigkeit gibt. Dies wird bei den Kategorien „Verstehen“ und „Anwenden“ besonders deutlich (vgl. Ausubel & Robinson, 1969). Die bisherigen Forschungsergebnisse wurden in einem Review von Kreitzer und Madaus (1994) zusammengefasst und werden daher an dieser Stelle nur kurz dargestellt.

Eine aufsteigende Schwierigkeit der kognitiven Prozesse sollte sich in einer sinkenden Lösungshäufigkeit der dazugehörigen Aufgaben widerspiegeln. Kottke und Schuster (1990) konnten diese abfallenden Lösungshäufigkeiten für die sechs Prozesse nachweisen, wobei die Lösungswahrscheinlichkeiten für die vier höheren Prozesse alle im Bereich zwischen 14% und 18% lagen und somit die vier höheren Prozesse nicht mit Hilfe der Lösungswahrscheinlichkeiten voneinander zu unterscheiden sind, da diese (fast) identisch sind. Eine andere Methode zur empirischen Überprüfung der Taxonomie besteht darin, Schülern Aufgaben aller Kategorien vorzulegen und die Leistungen in den einzelnen Kategorien mit Hilfe von Korrelationen miteinander zu vergleichen. Dabei sollte die Korrelation zwischen zwei direkt „benachbarten“ Kategorien, z.B. Wissen und Verstehen, am höchsten sein, und je weiter die Kategorien in der Taxonomie voneinander entfernt sind, desto geringer sollten diese miteinander korrelieren, d.h. es sollte sich eine sog. Simplex-Struktur zeigen. Derartige Studien kommen meistens zu dem Ergebnis, dass die Korrelationen zwischen den niedrigeren Kategorien Wissen, Verstehen, Anwendung und Analyse erwartungsgemäß so sind, dass die Korrelationen zwischen den benachbarten Kategorien jeweils am höchsten und mit steigender Entfernung immer geringer sind. Die höchsten Kategorien - Synthese und Evaluation – können in den meisten Studien jedoch nicht eindeutig in diese Rangreihe gebracht werden (vgl. Anderson et al., 2001; Hancock, 1998; Kreitzer & Madaus, 1994) Möglicherweise beruht die nicht nachgewiesene Anordnung der Schwierigkeit der höheren Niveaustufen auf Konstruktionsproblemen, denn Kropp und Stoker (1966) konnten zeigen, dass sich die Itemkonstruktion für die kognitiven Prozesse „Synthese“ und „Evaluation“ als sehr schwierig erwies. Diese Items führten ausschließlich zu Bodeneffekten.

Eine Analyse mit Hilfe von Strukturgleichungsmodellen (Hill & McGaw, 1981) kam zu dem Schluss, dass ein fünffaktorielles Modell die Taxonomie angemessen wiedergibt, und zwar mit den Prozessen in der folgenden Reihenfolge: Verstehen, Anwendung, Analyse, Synthese und Evaluation. Der Prozess Wissen wurde von den Autoren aus folgendem Grund nicht in das Modell aufgenommen: Bloom et al. (1956) unterscheiden klar zwischen „Wissen“ als Wissensabruf auf der einen Seite und den anderen, kognitiv höheren Prozessen, als

Formen der Wissensverarbeitung auf der anderen Seite. Damit wird mit dem kognitiven Prozess des Wissens etwas eindeutig anderes erfasst als mit den anderen Prozessen. Kottke und Schuster (1990) konnten mit Hilfe einer Faktorenanalyse eine zweifaktorielle Lösung für die sechs Prozesse der Taxonomie nachweisen, wobei der erste Faktor durch den Wissensprozess charakterisiert wird und alle weiteren kognitiven Prozesse auf dem zweiten Faktor hoch laden.

Kreitzer und Madaus (1994) fassen die empirischen Untersuchungen zur Bloom'schen Taxonomie dahingehend zusammen, dass die Reihenfolge der ersten Kategorien Wissen, Verstehen, Anwendung und Analyse meist bestätigt wurde und es empirische Evidenz für die Annahme der ansteigenden Schwierigkeit für diese Kategorien gibt, jedoch nach den bisherigen Ergebnissen die Reihenfolge von Synthese und Evaluation noch nicht abschließend beantwortet werden kann (Anderson et al., 2001; Helmke, 2004; Kreitzer & Madaus, 1994; Mietzel, 2003). Gleichwohl haben sich zahlreiche Forscher (z.B. De Landsheere, 1977; Lienert & Raatz, 1998) zur kognitiven Taxonomie als einem ausreichendem Instrumentarium der Konstruktion von Aufgaben unterschiedlicher Schwierigkeit bekannt. Sie stelle eine nützliche Strukturierungshilfe beim Ausarbeiten von Unterrichtszielen dar.

Es muss bei den Studien zur Überprüfung der Taxonomie berücksichtigt werden, dass die meisten Studien mit kleinen Stichproben gearbeitet und häufig kleine Unterschiede zwischen den Kategorien gefunden haben, welche möglicherweise auf Mess- bzw. Stichprobenfehler zurückgeführt werden können. Des Weiteren gibt es nur sehr wenige Studien, die alle sechs Kategorien in ihrer Analyse berücksichtigen bzw. die komplexeren Kategorien Synthese und Evaluation einbeziehen. Ein weiterer Nachteil vieler Studien ist, dass auch die komplexeren Prozesse ausschließlich mit *multiple-choice*-Items erfasst wurden, wobei die Frage der Gütekriterien meist nicht untersucht wurde. Selbst wenn sie untersucht werden, haben sie lediglich begrenzte Aussagekraft für die Erfassung kognitiver Prozesse wie Verstehen und Anwenden, da Antworten vorgegeben sind und nicht selbst hergeleitet werden müssen (vgl. Lienert & Raatz, 1998).

Obwohl in den Jahrzehnten seit der Veröffentlichung der Originaltaxonomie große Fortschritte in der *Item-Response*-Theorie, dem computergestützten Testen und der Technologie der Normierung gemacht wurden, führte dies kaum bzw. lediglich zu kleinen Fortschritten bei der angemessenen Konstruktion von Testaufgaben (Sternberg, 1997). Daher sollten die Ergebnisse insgesamt mit Vorsicht betrachtet bzw. diese Schwäche berücksichtigt werden. Insgesamt zeigt sich also, dass die Originaltaxonomie seit langem und häufig zur Erforschung von Lernaufgaben genutzt wird. Gerade deshalb sollte die Taxonomie weiterhin empirisch überprüft werden (Anderson & Sisniak, 1994).

Vor dem Hintergrund der empirischen Ergebnisse zur Aufgabenkonstruktion mit Hilfe der Bloom'schen Taxonomie werden in der vorliegenden Arbeit die drei kognitiven Prozesse Wissen, Verstehen und Anwenden berücksichtigt. Diese Dreiteilung hatte sich auch in den Arbeiten von Ballstaedt et al. (1981) und Levin (2005) als sinnvoll erwiesen. Dabei wird versucht, einen Mittelweg zu gehen, wobei zum einen aus ökonomischen Gründen der Aufgabenkonstruktion nicht alle sechs Stufen der Taxonomie berücksichtigt werden können (Kropp & Stoker, 1966). Zum anderen soll die Frage der Aufgabenschwierigkeit bzw. die Bedeutung der verschiedenen kognitiven Prozesse differenzierter untersucht werden, als es mit Hilfe der häufig in der Literatur zu findenden Zusammenfassung der sechs kognitiven Prozesse der Taxonomie, welche meist zu einer Dichotomisierung in *low level*- und *high level*-Aufgaben führt, möglich ist.

Da die Klassifikation von Aufgaben allein nicht ausreicht, um Aufgaben für Forschungszwecke zu konstruieren, wird im folgenden Kapitel 6 der Prozess der Konstruktion von Lernaufgaben thematisiert.

6 Konstruktion von Lernaufgaben

An vielen Studien wird kritisiert, dass nicht erläutert wurde, wie und auf welcher theoretischen Grundlage die jeweiligen Lern- oder Testfragen konstruiert wurden. Außerdem werden zur Überprüfung der Behaltenseffekte oft lediglich Fakten- bzw. *low level*-Aufgaben verwendet. Mit solchen kann jedoch nicht überprüft werden, ob die Lernenden den Inhalt verstanden haben und ihn anwenden können. Dies ist nur mit *high level*-Aufgaben möglich (Häfele, 1995).

Interviewstudien mit Lehrern haben gezeigt, dass Lehrer Aufgaben und Tests eher intuitiv und ohne Klassifikationsschema entwickeln (Graesser, 1981; van Dijk & Kintsch, 1983). Seel (1981) kommt zu dem Schluss, dass im Bereich des konventionellen Unterrichts die Ableitung von Fragen und Aufgaben, besonders im Bezug auf die Erarbeitung von Sachverhalten in schriftlicher und mündlicher Form, dem Zufall überlassen wird.

Diese Argumente sprechen für eine rationale Konstruktion von Aufgaben, um die Unzulänglichkeiten eher intuitiv erzeugter Aufgabenmengen zu überwinden. Rationale Aufgabenkonstruktion „bedeutet, dass für die Erstellung von Items zu einem bestimmten Bereich so weitgehende Vorschriften entwickelt werden, dass voneinander unabhängig arbeitende Testautoren zu gleichartigen, d.h. weitgehend übereinstimmenden Items gelangen, und zwar geschieht dies in einem ersten Schritt und nicht erst durch eine Auswahl im Anschluss an irgendwelche Analysen“ (Fricke & Lühmann, 1983; nach Feger, 1984, S. 24). Diese rationale Itemgenerierung ist vor allem im Zusammenhang mit der kriteriums- oder lehrzielorientierten Leistungsmessung entwickelt worden (vgl. Klauer, 1987; Feger, 1984).

In der vorliegenden Arbeit wurde auf den Ansatz zur Klassifikation von Aufgaben nach Anderson et al. (2001) zurückgegriffen, welcher mit dem Vorgehen von Klauer (1987) zur Konstruktion kontentvalider Tests (siehe Kapitel 6.2) kombiniert wurde. Damit können sowohl Lern- als auch Testaufgaben mit unterschiedlichen kognitiven Anforderungen und so auch mit unterschiedlicher, postulierter Aufgabenschwierigkeit systematisch und theoriegeleitet konstruiert werden. Ferner wird auf unterschiedliche Formen von Schulleistungstests eingegangen (Kapitel 6.1). Diese stellen den übergeordneten Rahmen dar, in dem die Aufgabenkonstruktion stattfindet.

6.1 Schulleistungstests

„Schulleistungstests sind Verfahren der Pädagogischen Diagnostik, mit deren Hilfe Ergebnisse geplanter und an Curricula orientierter Lernvorgänge möglichst objektiv, zuverlässig und gültig gemessen und durch Lehrende (z.T. auch durch Lernende) oder Beratende ausgewertet, interpretiert und für pädagogisches Handeln nutzbar gemacht werden“ (Ingenkamp, 1997, S. 117). Bei Schulleistungstests wird meist das Ergebnis eines Einzelnen mit den Ergebnissen aller Tests aus einer Bezugsgruppe verglichen. Diese Art von Tests wird meist als „normorientiert“ oder „standardisiert“ bezeichnet und wird mit Hilfe der klassischen Testtheorie ausgewertet (Klauer, 1987).

Im Gegensatz dazu wurden im Zusammenhang mit dem *Programmiertem Unterricht* Tests entwickelt, bei denen das Ergebnis mit einem zuvor festgelegtem Lehrziel bzw. einem Kriterium in Beziehung gesetzt wird, welches meist sachliche Bezugsnormen sind. Diese Art von Tests wird als „lehrziel- oder kriterienorientiert“ bezeichnet (Ingenkamp, 1997). Klauer (1987) definiert wie folgt: „Kriteriumsorientiert ist ein Test, der die Gesamtheit einer wohl definierten Menge von Aufgaben enthält oder repräsentiert und der zu dem Zweck konstruiert ist; die Fähigkeit des Probanden zur Lösung der Aufgaben der definierten Menge zu schätzen oder/und ihn gemäß dieser Fähigkeit einer Klasse von Probanden zuzuordnen“ (S. 11). Ein kriteriumsorientierter Test wird als Verfahren zur Untersuchung der Frage, ob und wie gut ein Schüler ein bestimmtes Lehrziel erreicht hat, eingesetzt (z.B. Fricke, 1974). Klauer (2001) spricht in diesem Zusammenhang von lehrzielorientierter *Messung*. Dabei ist es uninteressant zu wissen, wie andere Schüler im gleichen Test abgeschnitten haben (Fricke & Lühmann, 1983). Die Zuordnung von Schülern in z.B. Könner und Nichtkönner bezeichnet man als lehrzielorientierte *Klassifikation* (Klauer, 2001). Die in kriteriumsorientierten Tests eingesetzten Aufgaben dienen dazu, den individuellen Fähigkeitsgrad eines Schülers mit einem gewünschten Fähigkeitsgrad zu vergleichen. Für diesen Vergleich sind folgende Schritte erforderlich: die Quantifizierung des Lehrziels, eine quantitative Erfassung der Schülerleistung und ein Messmodell für die Entscheidung darüber, ob das Lehrziel erreicht worden ist. Eine besondere Schwierigkeit bei der Konstruktion kriteriumsorientierter Tests ist die Entwicklung von Testaufgaben, welche den gesamten Lerninhalt möglichst gültig repräsentieren (Ingenkamp, 1997). Die in Kapitel 5.2 beschriebene Bloom'sche Taxonomie (Bloom et al., 1956) wurde dabei als Möglichkeit eingesetzt, um kriteriumsorientierte Aufgaben bzw. Tests zu konstruieren (Fricke & Lühmann, 1983).

Ein weiterer Unterschied kriteriumsorientierter Schulleistungstests gegenüber normorientierten Tests liegt in der Auswertung. Bei kriteriumsorientierten Tests muss definiert

werden, a) wann das durch mehrere Aufgaben repräsentierte Lehrziel erreicht ist und wann nicht, b) welchen Anteil der Aufgaben eine Person gelöst hat und c) welcher Leistungsgruppe bzgl. des Abstandes vom Lehrziel die Person zuzuordnen ist (Ingenkamp, 1997). Ein geeignetes quantitatives Modell zur Auswertung von kriteriumsorientierten Tests ist das Modell von Rasch (1960), da es unter anderem den Vorteil hat, Aufgaben mit unterschiedlichen Lösungswahrscheinlichkeiten (Itemschwierigkeiten) zuzulassen. Es ordnet die verschiedenen schwierigen Testaufgaben und die verschiedenen fähigen Personen auf einer Skala so an, dass man aus den beiden Parametern jeweils auf die Aufgabenlösungswahrscheinlichkeit schließen kann. Ohne Bezug auf eine Personenstichprobe erlaubt die Zahl der gelösten Aufgaben dann eine hinreichende Aussage über die Kompetenz des Geprüften (Fricke & Lüthmann, 1983).

Ein weiterer Vorteil des Rasch-Modells liegt in der Annahme, dass das Antwortverhalten von latenten Eigenschaften (z.B. der Fähigkeit) abhängig ist. Danach beeinflusst deren Ausprägungsgrad das Antwortverhalten. Für eine kriteriumsorientierte Interpretation von Testergebnissen werden die Ergebnisse daher nach inhaltlichen Kriterien verglichen. Diese Modelleigenschaft wird heutzutage für die Auswertung internationaler Studien wie TIMSS und PISA genutzt, um somit Schülerkompetenzen zu modellieren und Kompetenzstufen aufgrund der Lösungswahrscheinlichkeit der Testaufgaben festzulegen. Dies ist aufgrund der Eigenschaft des Modells möglich, dass sich die aus den Daten geschätzten Aufgabenschwierigkeits- und Personenparameter auf einer gemeinsamen Dimension darstellen lassen, wodurch eine direkte inhaltliche Ergebnisinterpretation möglich ist. Inhaltlich lassen sich unterschiedliche Kompetenzstufen durch die Aufgabenschwierigkeit beschreiben, die dem Anforderungsniveau einer bestimmten Stufe entspricht. Aufgaben, welche vergleichbare Eigenschaften messen, werden zusammengefasst, und das Anforderungsniveau wird anhand inhaltlicher Kriterien beschrieben und bildet die jeweilige Kompetenzstufe. Personen werden dann aufgrund ihres Antwortverhaltens den verschiedenen Kompetenzstufen zugeordnet, wodurch inhaltlich beschrieben werden kann, über welche Fähigkeiten bzw. Kompetenzen die Schüler mit einem entsprechenden Fähigkeitsparameter mit hinreichender Sicherheit verfügen. Für genauere Beschreibungen der Entwicklung solcher Kompetenzmodelle sei auf Fleischer, Wirth und Leutner (in Druck) verwiesen, da sie im Rahmen der vorliegenden Arbeit nicht von maßgeblicher Bedeutung sind. Die Rasch-Skalierung wird in erster Linie aufgrund ihrer Eigenschaft der „spezifischen Objektivität“ eingesetzt, um Aufgaben unterschiedlicher Inhaltsbereiche auf einer Dimension skalieren zu können und die Fähigkeiten der Schüler, welche die unterschiedlichen Inhalte in den Aufgaben bearbeitet haben, miteinander vergleichen zu können, da diese auf einer gemeinsamen Skala abgetragen werden.

An kriteriumsorientierten Tests wird häufig kritisiert, dass Lehrer nicht in der Lage sind bzw. nicht über die erforderlichen Kenntnisse verfügen, um reliable und valide kriteriumsorientierte Tests zu konstruieren (Gage & Berliner, 1996). Lehrer setzen vielmehr „informelle“ Schulleistungstests, auch *teacher made tests* genannt, im Unterricht ein. Ihre Konstruktion durch den Lehrer stellt ihr wichtigstes Unterscheidungsmerkmal gegenüber den oben erwähnten formellen Schulleistungstests dar. Ein weiteres Unterscheidungsmerkmal ist die Orientierung des Tests an den pädagogischen Bedürfnissen einer oder weniger Schulklassen. Allein die curricularen Ziele und das Informationsbedürfnis des Lehrers bestimmen die Testentwicklung. Von den formellen Schulleistungstests unterscheiden sich diese vom Lehrer konstruierten Tests darin, dass eine Überprüfung der Gütekriterien üblicherweise nicht stattfindet (Ingenkamp, 1997; Lienert & Raatz, 1998).

Nach Baumert, Bos und Waterman (1998) muss ein Schulleistungstest, der kumulative und additive Komponenten des Wissenserwerbs erfassen soll, mindestens zwei Bedingungen erfüllen:

1. Er muss über hinreichend komplexe Aufgaben verfügen, die ein vertieftes Verständnis des Sachgebiets erfassen, und
2. er muss in der Korrespondenz von unterrichteten und geprüften Stoffen spezifische curriculare Validität besitzen.

Mit dem in der vorliegenden Arbeit gewählten Vorgehen wird versucht, die von Baumert et al. (1998) genannten Bedingungen zu erfüllen, indem Aufgaben konstruiert werden, deren Bearbeitung unterschiedliche kognitive Prozesse, so auch das Verständnis des Sachgebiets, erfordert. Die curriculare Validität wird dadurch gewährleistet, dass der unterrichtete und der geprüfte Inhaltsbereich identisch bzw. aufeinander bezogen ist und die Aufgabenkonstruktion mit Hilfe der Vorschriften von Klauer (1987) - auf die im Folgenden eingegangen wird - erfolgt.

6.2 Die Konstruktion von Lernaufgaben nach Klauer (1974, 1987)

Nach Seel (1981) ist es das Verdienst von Klauer (1974), die Zusammenhänge zwischen Lehren und Lernen grundlegend geklärt zu haben. In Überwindung der behavioristischen Tradition, Lernen als Prozess der Verhaltensänderung zu definieren (vgl. Kapitel 4.2), plädiert Klauer (1974) dafür, Lernen als eine besondere Klasse der Konstruktänderung zu definieren, die über Verhaltensänderungen indirekt erschlossen werden kann. Klauer (1974) hat die Bedingungen für das Lernen funktional dargestellt:

$$\text{Lernen} = f(\text{Lernender, äußere Situation, Lehrstoff})$$

Danach wird Lernen als abhängig von individuellen Lernvoraussetzungen, von externalen Bedingungen der Lernsituation und vom Lehrstoff definiert. Eine besondere Bedeutung nimmt im Rahmen dieser Definition die Gestaltung der externalen Bedingungen ein, zu denen die Lernaufgaben gehören. Deren Konstruktion wird von Klauer (1987) explizit mit dem jeweiligen Lehrziel in Verbindung gebracht. Um wiederum einen lehrzielvaliden Test konstruieren zu können, sind die in der Abbildung 2 verdeutlichten Schritte vom Lehrziel zum Lerninhalt notwendig.

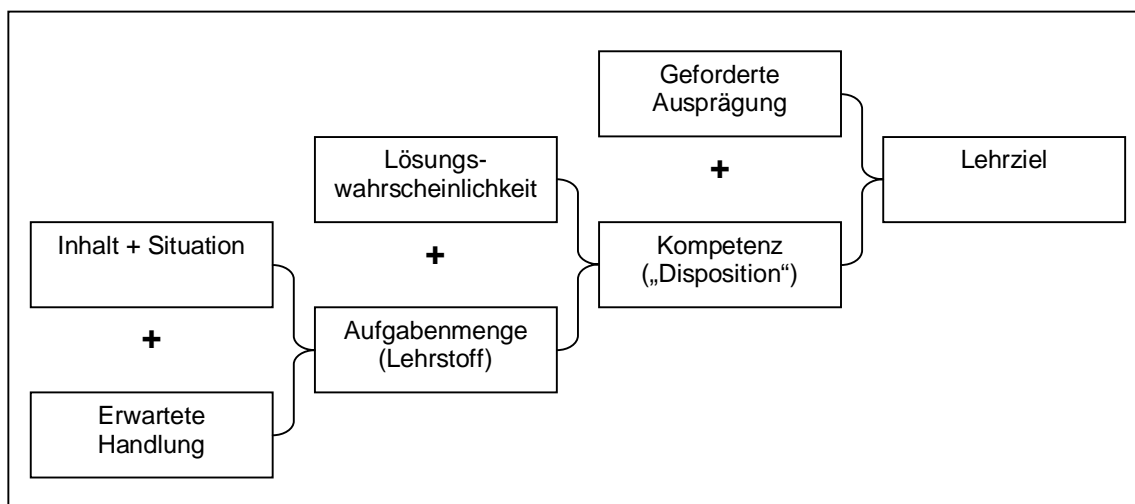


Abbildung 2: Vom Lehrinhalt zum Lehrziel nach Klauer (1987)

Am Anfang einer Unterrichtsplanung und Aufgabenkonstruktion steht die Präzisierung von Lehr- und Lernzielen. Nach Klauer (1974, 1987, 2001) müssen zur Bestimmung der Lehrziele das angestrebte Persönlichkeitsmerkmal, d.h. die zur Aufgabenlösung benötigten Fähigkeiten und Fertigkeiten, sowie der angestrebte Kompetenzgrad präzisiert werden. Klauer (1987) definiert ein Lehrziel als „die geforderte Ausprägung einer Kompetenz“ (S. 14). Die Kompetenz, welche auch als Fähigkeit bezeichnet werden kann, stellt den qualitativen Aspekt des Lehrziels dar und die geforderte Ausprägung den quantitativen. Die Kompetenz wird bestimmt durch „die Wahrscheinlichkeit eines Menschen, mit der er die Aufgaben einer Aufgabenmenge löst“ (Klauer, 1987, S. 15). Die Höhe der Lösungswahrscheinlichkeit gibt Auskunft über den Ausprägungsgrad der gemessenen Kompetenz. Ist die Lösungswahrscheinlichkeit hoch, ist die Kompetenz stark ausgeprägt. Ist im Gegensatz dazu die Lösungswahrscheinlichkeit null, so ist die Kompetenz nicht existent. Dabei ist zu beachten, dass mit dem Wahrscheinlichkeitsbegriff keine deterministische, sondern eine stochastische Beziehung zwischen Kompetenz und beobachtbarem Verhalten angenommen wird (Klauer,

1987). Empirisch wird die Kompetenzausprägung mit Hilfe der auf die Aufgabenmenge gegebenen Antworten bestimmt. Zuvor wird der geforderte Kompetenzgrad meist normativ festgesetzt, indem beispielsweise entschieden wird, dass jeder Schüler 100% der Aufgaben abzüglich einer gewissen Fehlermarge beherrschen können soll. Orientiert man sich an den konventionellen Signifikanzniveaus von 5 % (oder 10 %), so kommt man auf Zielmarken forderbarer Lösungswahrscheinlichkeit von $p = .95$ (oder $p = .90$), in anderen Zusammenhängen wird es von Klauer (1987, 2001) als sinnvoll erachtet, die Zielmarke auf etwa $p = .60$ festzulegen.

Lernaufgaben sowie Testaufgaben werden als Verknüpfung einer Situations- und einer Verhaltenskomponente definiert (Klauer, 1987). Mit der Verhaltenskomponente wird das Verhalten angegeben, das als Reaktion auf die situativen Anforderungen erwartet wird. Die Itemform entscheidet in Verbindung mit dem Inhalt darüber, welche Anforderung an den Probanden gestellt, welches Verhalten von ihm erwartet wird (Klauer, 1987). Aufgrund dieses Weges vom Lehrziel zu konkreten Aufgaben ist der Lehrstoff mit der Aufgabenmenge identisch und nicht mit einem bestimmten Inhalt. Der Schüler soll lernen, mit einem gewissen Inhalt in bestimmter Weise umzugehen. Damit ist auch gemeint, den Inhalt zu verstehen, anzuwenden und zu bewerten, was bei der Beantwortung von Lernaufgaben gezeigt werden kann bzw. muss. Ein auf diese Weise entwickelter Test repräsentiert also den Lehrstoff des Lehrziels und ist damit nach Klauer (1987) inhalts- bzw. kontentvalide.

Ein zentrales Charakteristikum eines so konstruierten Tests stellt demnach die *Kontentvalidität* dar. Für einen Schulleistungstest bedeutet das, dass dieser dann kontentvalide ist, wenn seine Aufgaben inhaltlich eine repräsentative Auswahl aus dem Unterrichtsstoff (Curriculum) darstellen (Lienert & Raatz, 1998). Klauer (1987) betont die enge Beziehung zwischen Kontent- und Lehrzielvalidität. So klar und eindeutig in vielen Fällen einem Test Kontentvalidität zugeschrieben werden kann, so ist eine gewisse Gefahr der Übergeneralisierung gegeben, welche bei Schulleistungstest besonders groß ist.

Ein Test ist also dann kontentvalide, wenn er repräsentativ für die Gesamtheit aller möglichen Aufgaben in einem Gegenstandsbereich ist. Um jedoch eine repräsentative Stichprobe zu erzeugen, muss zunächst über eine Grundmenge verfügt werden, aus der man diese Stichprobe ziehen kann. Unabdingbare Voraussetzung für die Konstruktion eines kontentvaliden Tests ist deshalb die präzise Definition der Menge an möglichen Aufgaben. Es sind zwei Möglichkeiten bekannt, wie eine Menge definiert werden kann: die vollständige Aufzählung aller ihrer Elemente (Realdefinition) oder die Angabe eines Oberbegriffs und der mengenbildenden spezifischen Merkmale (Nominaldefinition). In den meisten Themenbereichen, wie auch dem der vorliegenden Untersuchung, ist eine vollständige Aufzählung aller

zu diesem Bereich zählenden Aufgaben nicht realisierbar. Daher bietet sich hier die Definition durch die Angabe der mengenbildenden Merkmale an.

Um zu gewährleisten, dass ein Test kontentvalide ist, müssen des Weiteren die Items rational nach einem vorher festgelegten Verfahren bzw. nach einer Vorschrift konstruiert werden. Die eigentliche Aufgabenkonstruktion besteht bei Klauer (1987) aus drei Schritten. In einem ersten Schritt wird der Inhalt, der in Aufgaben umgewandelt werden soll, in einer präzisen Weise möglichst vollständig dargestellt. Im zweiten Schritt wird eine geeignete Aufgabenform gewählt. Zuletzt werden Regeln für die Transformation des Inhalts in die Aufgabenform formuliert. Diese Transformationsregeln gewährleisten nach Klauer (1987, 2001) eine objektive und kontentvalide Konstruktion von Testaufgaben. Um einen Test zusammenzustellen, muss anschließend die Grundmenge von Items in Teilmengen zerlegt werden. Eine repräsentative Itemstichprobe führt dann in ihrer Zusammenstellung gemäß den Testkonstruktionsvorschriften zu einem kontentvaliden Test.

Zur Aufgabengenerierung wird von Klauer (1974) die Verwendung der sog. „Tyler-Matrix“ (Tyler, 1973) empfohlen. Jede Aufgabe kann dabei als Verknüpfung des betreffenden Sachverhalts mit einer Itemform betrachtet werden. In den Zeilen der Tyler-Matrix sind bestimmte Sachverhalte, die nach rein inhaltlichen Gesichtspunkten in Teilkomplexe untergliedert werden, in den Spalten Verhaltenskategorien (z.B. kognitive Prozesse) angeordnet. In jeder Zelle dieser Matrix findet sich eine Aufgabenmenge, die aus allen Aufgaben besteht, die sich durch die Verknüpfung des fraglichen Inhaltes mit der jeweiligen Verhaltenskategorie erzeugen lassen. Für die Ausdifferenzierung der Verhaltenskategorien wird häufig – wie im Beispiel (vgl. Tabelle 2) - auf die Bloom'sche Lehrzieltaxonomie zurückgegriffen (Klauer, 2001; vgl. Gage & Berliner, 1996; Lienert & Raatz, 1998).

Tabelle 2: Beispiel einer Lehrzielmatrix nach Klauer (2001)

Inhalte	Wissen	Verstehen	Anwendung
Elektrischer Strom; Stromkreis	A	B	C
Stromstärke und Spannung	D	E	F
Das Ohmsche Gesetz	G	H	I

Tabelle 2 macht deutlich, dass für jede Verbindung des fraglichen Inhalts mit den kognitiven Prozessen (nach Bloom et al., 1956) unterschiedliche Aufgaben erzeugt werden müssen, dargestellt in der Tabelle durch die unterschiedlichen Buchstaben. Hierbei müssen beispielsweise bei drei Inhaltsbereichen und drei berücksichtigten kognitiven Prozessen be-

reits neun verschiedene Aufgabentypen konstruiert werden. Für eine detaillierte Beschreibung der Aufgabenkonstruktion sei auf Klauer (1987) verwiesen.

Eine Auswertung kriteriumsorientierter Tests nach der klassischen Testtheorie wird von Klauer (1987, 2001) wegen der populationsabhängigen Schätzung der Aufgabenschwierigkeit und der Personenparameter als problematisch angesehen. Daher wird eine Auswertung auf Basis der probabilistischen bzw. *Item-Response*-Theorie empfohlen, da die Ergebnisse stichprobenunabhängig sind (vgl. Rasch, 1980; Rost, 2004). Das von Klauer präferierte Binomialmodell kann als „Spezialfall des Raschmodells“ (Klauer, 1987, S. 139) angesehen werden. Dabei wird aus der Annahme linearer Itemcharakteristikkurven (ICCs) und der Annahme stochastisch bzw. lokal stochastisch unabhängiger Itemantworten gefolgert, dass die relative Anzahl gelöster Items einen Schätzer für die Personenfähigkeit darstellt. Dazu müssen konstante Schwierigkeiten und Trennschärfen für alle Items vorausgesetzt werden. Für eine genauere Beschreibung des Binomialmodells sei auf Klauer (1987) und Rost (2004) verwiesen.

Die Konstruktion eines kriteriumsorientierten Tests nach Klauer (1987) wurde in der hier vorliegenden Arbeit in Kombination mit der revidierten Lehrzieltaxonomie nach Anderson et al. (2001) für den physikalischen Inhaltsbereich durchgeführt. Bevor auf die im Rahmen dieser Arbeit durchgeführten Studien im Detail eingegangen wird, soll zunächst in Kapitel 7 aus den offenen Fragen bzw. aus der bisherigen Forschung die Untersuchungsfrage der Studie 1 abgeleitet werden.

7 Schlussfolgerung für die Fragestellung der Studie 1

Seel (1981) schlussfolgert aus der bisherigen Forschung, dass insgesamt die Strukturen und die Wirksamkeit von Lernaufgaben des regulären Unterrichts bei weitem nicht so gut erforscht sind, wie die Wirksamkeit von Lernaufgaben, welche in experimentellen Untersuchungen analysiert wurde (vgl. Levin & Long, 1981). „Deshalb muss ein allgemeines Forschungsdefizit für diesen Bereich von Unterricht festgestellt werden“ (Seel, 1981, S. 74). Diese Kritik hat heute - mehr als 25 Jahre später - immer noch Berechtigung, und in der hier vorliegenden Arbeit wird ein Versuch unternommen, die Lernwirksamkeit der vom Lehrer im Unterricht eingesetzten Aufgaben empirisch zu untersuchen.

Aufgrund der bisherigen Ergebnisse zu Aufgaben im Unterricht (Doyle, 1986; Graesser & Person, 1994; Niegemann & Stadler, 2001; Wilen & Clegg, 1986) wird für die erste im Rahmen der vorliegende Arbeit durchgeführte Studie erwartet, dass im Physikunterricht die Lehrer die meisten Aufgaben auf dem Wissensniveau stellen und kaum Aufgaben formulieren, welche eine Verstehensleistung des Schülers erfordern und/oder auf die Anwendung bzw. den Transfer des Gelernten auf neue Inhaltsbereiche abzielen.

Aufgrund der bislang inkonsistenten Ergebnisse zum Einsatz (Hamaker, 1986; Levin, 2005; Renkl, 1991) und möglichen Mischungsverhältnis der verschiedenen Aufgabenarten (*low level-* vs. *high level-*Aufgaben als schwierigkeitsunterscheidendes Merkmal) soll in der vorgestellten Arbeit der Einsatz unterschiedlicher Aufgabenarten untersucht werden. Angesichts beträchtlicher Schulformunterschiede zwischen Gymnasien und Hauptschulen, welche in den PISA-Erhebungen deutlich wurden (so betrug die Diskrepanz zwischen den mittleren Kompetenzwerten bei PISA 2003 183 Punkte, was fast zwei Standardabweichungen entspricht; vgl. Rost, Walter, Carstensen, Senkbeil & Prenzel, 2004), wurde die Studie 1 an Gymnasien und Hauptschulen durchgeführt. Auch gibt es bislang kaum Untersuchungen, die an der Hauptschule durchgeführt wurden. Inwieweit die Ergebnisse amerikanischer *High-schools* auf das deutsche Schulsystem übertragen werden können, ist fraglich. Es werden somit mögliche Unterschiede in der mittleren Schwierigkeit der Unterrichtsaufgaben zwischen den Schulformen erwartet.

Die Analyse des Forschungsstandes hat gezeigt, dass die Passung zwischen Unterrichtsmerkmalen bzw. Aufgabenschwierigkeit und Schülerfähigkeit bislang insgesamt nicht hinreichend erforscht ist (vgl. Seel, 1981). Daher erachtet Seel (1981) es als notwendig, diese Forschung mit dem Ziel zu intensivieren, die externalen Bedingungen der Lehr-

Lernprozesse, speziell durch das Stellen von Lernaufgaben, zu identifizieren und deren Zusammenhang mit den individuellen Voraussetzungen der Schüler zu erfassen. Von den meisten Autoren wird ein angemessenes Schwierigkeitsniveau postuliert, ohne dass dieses empirisch überprüft wurde. Dieser bislang fehlenden empirischen Absicherung soll in der vorliegenden Arbeit Rechnung getragen werden, indem die Passung zwischen der Fähigkeit von Schülern und den Anforderungen von Lernaufgaben untersucht wird. Aus der Forschung zur Diagnosekompetenz von Lehrern gibt es bislang lediglich indirekte Hinweise auf eine mangelnde Übereinstimmung zwischen Aufgabenschwierigkeit und Schülerfähigkeit. So unterschätzen Lehrer die Aufgabenschwierigkeit (Schnotz, 1971) und auch die Fähigkeits- bzw. Kompetenzeinschätzung der Schüler gelingt häufig nur unzureichend (Baumert et al., 2001; Hosenfeld et al., 2002; Schrader, 1989). Auch die unbefriedigende Passung zwischen dem kognitiven Niveau der vom Lehrer im Unterricht gestellten Aufgaben und dem Niveau der darauf folgenden Schülerantwort (Dillon, 1982; Mills et al., 1980) kann als Hinweis dafür gesehen werden, dass die Schwierigkeit der Lehreraufgaben dem Leistungsniveau der Schüler nicht anpasst bzw. nicht angemessen ausgewählt ist. In der im Folgenden beschriebenen Studie 1 wird daher das Forschungsinteresse im Besonderen auf die Schwierigkeit der Unterrichtsaufgaben und ihrer Passung mit der Schülerfähigkeit gelegt. Aufgrund der bisherigen Hinweise wird dabei eine unzureichende Passung zwischen den vom Lehrer eingesetzten Aufgaben bzgl. ihrer Schwierigkeit und den Schülerfähigkeiten angenommen.

8 Studie 1: Fragestellung

Bislang sind die Forschungsergebnisse zum Einsatz verschiedener Aufgabenarten (*low vs. high level*-Aufgaben) inkonsistent (Hamaker, 1986; Levin, 2005; Renkl, 1991). Daher wurde in Studie 1 zunächst deskriptiv-explorativ untersucht, wie häufig verschiedene Aufgabenarten im regulären Physikunterricht am Ende der Sekundarstufe 1 an Gymnasien und Hauptschulen eingesetzt werden.

Da der Bearbeitung von Lernaufgaben hohe Bedeutsamkeit für die Lernleistung zugeschrieben wird, wurde zudem ermittelt, durch welche schwierighkeitsrelevanten Merkmale sich die von Lehrern eingesetzten Unterrichtsaufgaben charakterisieren lassen. Diese vom Lehrer im Unterricht eingesetzten Aufgaben werden im Folgenden als Unterrichtsaufgaben bezeichnet. Im Gegensatz dazu werden die in der vorliegenden Arbeit konstruierten Aufgaben als Lernaufgaben definiert. Dabei wurde in Anlehnung an Anderson et al. (2001) der Fokus auf die kognitiven Prozesse während der Aufgabebearbeitung gelegt und somit zwischen den Prozessen Wissen, Verstehen und Anwenden unterschieden (vgl. Kapitel 4).

Das Hauptinteresse der Untersuchung liegt indes auf dem Zusammenhang zwischen der Schwierigkeit der Unterrichtsaufgaben und den Schülerleistungen. Inwieweit passen Lehrer die Schwierigkeit ihrer Unterrichtsaufgaben den Schülerfähigkeiten an? Um diesen Zusammenhang untersuchen zu können, bekamen die Schüler einige Unterrichtsaufgaben eine Woche, nachdem sie Teil einer aufgezeichneten Unterrichtsstunde waren, nochmals zur Bearbeitung vorgelegt.

Um Effekte unterschiedlich schwieriger Lernaufgaben auf die Leistung der Schüler differenzierter untersuchen zu können, wurden zusätzlich neue Aufgaben unterschiedlicher Schwierigkeit zu dem unterrichteten Inhaltsbereich konstruiert und den Schülern, neben den wiederholten Unterrichtsaufgaben, eine Woche später ebenfalls zur Bearbeitung vorgelegt. Um die Aufgabenschwierigkeiten und die Schülerfähigkeiten über die verschiedenen Inhaltsbereiche hinweg miteinander vergleichen zu können, erfolgt die Schätzung dieser beiden Variablen auf der Grundlage der Rasch-Skalierung, auf deren Vorteile in Kapitel 6 eingegangen wurde.

Die allgemeine Fragestellung der Studie 1 besteht also in der Überprüfung der Auswirkungen von Aufgabencharakteristika, welche die Schwierigkeit der Aufgaben bestimmen, auf die Lernleistungen der Schüler.

Folgende spezifische Fragestellungen sollen mit der Studie beantwortet werden:

1. Welches kognitive Anspruchsniveau haben Aufgaben, die von Lehrern im Physikunterricht am Ende der Sekundarstufe I eingesetzt werden? Welche der kognitiven Prozesse - Wissen, Verstehen und Anwendung - sind bei der Bearbeitung der Unterrichtsaufgaben erforderlich?
2. Welche Itemschwierigkeiten lassen sich für die eingesetzten Unterrichtsaufgaben mit Hilfe der Rasch-Skalierung ermitteln?
3. Welcher Zusammenhang („Passung“) besteht zwischen der durchschnittlichen Aufgabenschwierigkeit der von den Lehrern in ihren Klassen eingesetzten Unterrichtsaufgaben und der durchschnittlichen Fähigkeit der Schüler in den jeweiligen Klassen?

In Bezug auf die drei dargestellten Fragestellungen werden schulformspezifische Unterschiede zwischen Gymnasien und Hauptschulen dahingehend angenommen, dass die Ergebnisse an den Gymnasien bzgl. der Lösungshäufigkeit der konstruierten Aufgaben höhere Werte annehmen als an den Hauptschulen. Auch werden Unterschiede in den skalierten Personen- und Itemparametern der Unterrichtsaufgaben zwischen den Schulformen erwartet, die Skalierung sollte zu numerisch größeren Parametern für die Schülerfähigkeit und für die Unterrichtsaufgabenschwierigkeit an Gymnasien führen.

Eine weitere Fragestellung bezieht sich auf die neu konstruierten Aufgaben unterschiedlicher Schwierigkeit:

4. Ist es möglich, basierend auf den für die Aufgabenbearbeitung erforderlichen kognitiven Prozessen a priori unterschiedlich schwierige Lernaufgaben zu konstruieren?

Im Folgenden werden die Hypothesen zu den einzelnen Fragestellungen formuliert.

Hypothese 1

Auf Grundlage der beschriebenen Forschungsergebnisse wird erwartet, dass die Unterrichtsaufgaben zum großen Teil der Wissenskatgorie zuzuordnen sind und eine geringe Variation der kognitiven Prozesse vorherrscht (vgl. Kapitel 3.2). Das bedeutet, dass sich die meisten Aufgaben auf das Wiederholen von Wissen und Fakten (*low level*-Aufgaben) beschränken und nicht auf das Verstehen und Anwenden des gelernten Wissens oder auf Problemlösen (*high level*-Aufgaben), wie u.a. in der TIMS-Studie (1997) gezeigt wurde (Neubrand, 2002).

Hypothese 2

Folgend aus der Hypothese 1 wird angenommen, dass eine Rasch-Skalierung der Unterrichtsaufgaben eher niedrige Schwierigkeitsparameter für diese Aufgaben aufweisen wird.

Hypothese 3

Da die empirischen Befunde nicht darauf hinweisen, dass eine Passung zwischen Aufgabenschwierigkeit und Schülerfähigkeit im Unterricht gegeben ist, und Forschungsergebnisse zeigen, dass die diagnostische Kompetenz von Lehrern gering ist (vgl. Kapitel 3.3. und 3.4), wird eine unzureichende Passung zwischen Aufgabenschwierigkeit und Schülerfähigkeit erwartet. In der vorliegenden Studie wird die Passung mit Hilfe der Aufgaben- (s) und Schülerparameter (f) der Rasch-Skalierung definiert. Sind die Werte beider Parameter bekannt, kann mit Hilfe der Formel: $p = \frac{e^{f-s}}{1+e^{f-s}}$ die zu erwartende Lösungswahrscheinlichkeit für eine zu bearbeitende Aufgabenmenge berechnet werden. Eine mittlere Lösungswahrscheinlichkeit von 50% der Unterrichtsaufgaben wird in der vorliegenden Arbeit als ausreichende Passung bezeichnet. Ist die mittlere Lösungswahrscheinlichkeit jedoch deutlich geringer als 50%, so wird dies als geringe bzw. unzureichende Passung definiert.

Für die ersten drei Fragestellungen bzw. die dazugehörigen Hypothesen werden schulformspezifische Unterschiede der Ergebnisse angenommen. Insgesamt wird davon ausgegangen, dass sich die eingesetzten Unterrichtsaufgaben in ihrer Schwierigkeit zwischen den Schulformen, aufgrund der Schülerfähigkeitsverteilungen der naturwissenschaftlichen Kompetenz an den beiden Schulformen bei den PISA Erhebungen, unterscheiden (Prenzel et al., 2001, Rost et al., 2004). Beispielsweise hat PISA 2003 bei der naturwissenschaftlichen Kompetenz einen sehr großen Leistungsabstand (1.8 SD) zwischen Hauptschulen und Gymnasien aufgezeigt (Rost et al., 2004). Einen indirekten Hinweis auf die unterschiedliche Leistungsfähigkeit der Schüler an den verschiedenen Schulformen kann die Häufigkeit, mit der die unterschiedlichen Aufgabenarten im jeweiligen Unterricht eingesetzt werden, geben. Aufgrund der Annahme von Ropo (1990), dass durch niedrig-kognitive Aufgaben fast alle Schüler aktiviert werden und sich somit am Unterricht beteiligen können, was bei schwierigen Aufgaben nur für eine Minderheit der Schüler gilt, kann erwartet werden, dass besonders in den aufgezeichneten Hauptschulstunden überwiegend leichtere Wissensaufgaben Einsatz finden.

Auch sollten bedeutsame Unterschiede in den Schülerfähigkeiten zwischen den Schulformen deutlich werden, denn in PISA 2003 wurden 52.4% der Hauptschule der „Risiko-Gruppe“ zugeordnet (< 421 Punkte). Damit sind Schüler gemeint, welche lediglich die un-

terste Kompetenzstufe - nominelle naturwissenschaftliche Grundbildung (I) - erreichen, d.h. sie verfügen über Faktenwissen und können unter Verwendung von Alltagswissen Schlussfolgerungen ziehen, sind jedoch nicht in der Lage, ihr Wissen auszubauen bzw. auf andere Wissensbereiche sinnvoll und erfolgreich anzuwenden (vgl. Rost et al., 2004).

Hypothese 4

Aufgrund der bisherigen Ergebnisse zur Aufgabenkonstruktion (vgl. Kapitel 5.5) ist es mit Hilfe der in der vorliegenden Studie berücksichtigten Dreiteilung der kognitiven Prozesse in Wissen, Verstehen und Anwendung möglich, unterschiedlich schwierige Aufgaben zu konstruieren. Somit wird erwartet, dass sich die gemittelte Schwierigkeit der verschiedenen neu konstruierten Aufgabenarten – erfasst über die Itemparameter und die Lösungshäufigkeit – signifikant unterscheiden.

9 Methode

Die Beschreibung der Methode erfolgt in den folgenden Unterkapiteln. Zunächst wird die Stichprobe der Untersuchung beschrieben (Kapitel 9.1), dann werden das Design (Kapitel 9.2) und die Durchführung der Studie (Kapitel 9.3) erläutert. Im Anschluss wird der Leistungstest mit dem Schwerpunkt auf der Beschreibung der Aufgabenkonstruktion dargestellt (Kapitel 9.4). Abschließend wird im letzten Unterkapitel auf die Datenauswertung eingegangen (Kapitel 9.5).

9.1 Stichprobe

Die Gesamtstichprobe setzt sich aus 642 Schülern aus Nordrhein-Westfalen zusammen; die demographischen Daten sind in Tabelle 3 dargestellt.

Tabelle 3: Beschreibung der Untersuchungsstichprobe

	Gesamtstichprobe	Hauptschule	Gymnasium
Anzahl der Schüler	642	251	391
%- Angaben	100%	39.1%	60.9%
Durchschnittsalter	15.9 J.	16.1 J.	15.7 J.
Anzahl der Mädchen	333	116	217
Prozent der Mädchen	52%	46.4%	55.6%
Anzahl der Jungen	307	134	173
Prozent der Jungen	48%	53.6%	44.4%

Die untersuchte Lehrerstichprobe bestand aus 30 Physiklehrern, davon waren 27 männlich (90%) und 3 weiblich (10%). Das durchschnittliche Alter der Lehrer betrug 48.4 Jahre, und sie waren im Durchschnitt seit 20.4 Jahren im Schuldienst.

Da sich beispielsweise in den PISA-Erhebungen im Vergleich der verschiedenen Schulformen zwischen Gymnasien und Hauptschulen am Ende der Pflichtschulzeit der größte Unterschied in den Schülerfähigkeitsverteilungen bzw. im Leistungsstand der naturwissenschaftlichen Kompetenz gezeigt hatte (Prenzel et al., 2001, Rost et al., 2004), werden in der vorliegenden Studie diese beiden Schulformen untersucht. Somit sollen größtmögliche schulformspezifische Unterschiede im Aufgabeneinsatz des Physikunterrichts, besonders in

der Schwierigkeit der Unterrichtsaufgaben, identifiziert werden. Die Klassenstufe war am Gymnasium das 10. Schuljahr. An den Hauptschulen war die ausschließliche Berücksichtigung des 10. Schuljahrs nicht möglich, da sich die Rekrutierung der Hauptschulen als recht schwierig erwies, so dass auch vier Klassen des 9. Hauptschuljahrgangs an der Untersuchung teilnahmen. Die Schulen wurden nach einer Zufallsstichproben-Ziehung aus allen Gymnasien und Hauptschulen in NRW ausgewählt und telefonisch kontaktiert. Bei der Stichprobenziehung wurde ein Radius von 60 km um Essen gezogen und eine Gleichverteilung der Stadt-Land Repräsentation der Schulen in NRW berücksichtigt. Die Teilnahme an der Studie war für die Schulen freiwillig. Jede Klasse bekam für die Studienteilnahme 50 € für ihre Klassenkasse. Diese Vergütung war zuvor nicht angekündigt worden, so dass mögliche Motivationseffekte aufgrund des Geldes ausgeschlossen werden können.

9.2 Design

Die vorliegende Untersuchung wurde im Rahmen eines durch die DFG geförderten Projekts zum Aufgabeneinsatz im naturwissenschaftlichen Unterricht durchgeführt (vgl. Tiemann, 2003). Im Mittelpunkt dieses Projekts stand die Untersuchung der Fragestellung 1 die *Deskription des Aufgabeneinsatzes im Physikunterricht* (vgl. Kapitel 8). Die Studie wurde aufgrund dessen als *Ein-Gruppen-Nachtest-Design* konzipiert (vgl. Cook & Campbell, 1979).

Grundlage der Analyse war jeweils eine isolierte Unterrichtsstunde pro Lehrer, wobei auf die darin eingesetzten und bearbeiteten Aufgaben fokussiert wurde. Die unabhängige Variable stellt die Aufgabenart dar, wobei die Unterrichtsaufgaben des Lehrers sowie auf den jeweiligen Unterrichtsinhalt bezogene, neu konstruierte Aufgaben unterschieden werden. Die Schwierigkeit der Aufgaben wurde mit Hilfe der IRT-Skalierung erfasst (vgl. Kapitel 10.3). Die Testleistung der Schüler wurde zur Schätzung der Personenparameter herangezogen, welche die abhängige Variable darstellt.

9.3 Durchführung

Die Studie wurde im 2. Halbjahr des Schuljahrs 2004/2005 durchgeführt. Aus organisatorischen Gründen erfolgte die Erhebung in dem Zeitraum Januar bis März 2005 an den Gymnasien und im Anschluss von März bis Juni 2005 an den Hauptschulen. Die Studierendurchführung bestand für jede untersuchte Schulkasse aus zwei Terminen. An dem ersten Untersuchungstermin wurde eine reguläre Unterrichtsstunde (Dauer: 45 Minuten) pro Lehrer

auf Video aufgezeichnet. Dabei wurden dem Lehrer keine Vorgaben in Bezug auf den Unterrichtsinhalt bzw. hinsichtlich der Planung und Art der Durchführung des Unterrichts gemacht. Vielmehr wurde der Lehrer gebeten, dass er seinen Unterricht möglichst so planen und halten solle, wie er es normalerweise tue, da das Ziel der Untersuchung die realistische Deskription des Physikunterrichts sei. Aufgrund dieser „freien“ Unterrichtsgestaltung waren in den einzelnen Unterrichtsstunden der verschiedenen Lehrer sehr unterschiedliche Inhaltbereiche zu erwarten. Um den Unterricht und das Verhalten der Lehrer und Schüler möglichst wenig zu verfälschen bzw. zu beeinflussen, wurden die Untersuchungsteilnehmer in Bezug auf die Forschungsfragestellung, sprich die eingesetzten Unterrichtsaufgaben, nicht aufgeklärt.

Technisch erfolgte die Videoaufzeichnung mit Hilfe zweier Kameras, wovon eine manuell bedient wurde, um die Bewegungen des Lehrers verfolgen zu können (sog. Aktionskamera). Mit der zweiten Kamera wurde vor allem eine Sicherheitskopie erstellt, falls es technische Probleme mit der Aktionskamera geben sollte. Hierbei wurde das Unterrichtsgeschehen aus der „Totalen“ aufgenommen und der Unterricht statisch, d.h. ohne manuelle Bedienung aufgezeichnet. Der Ton wurde über insgesamt vier Funkmikrofone aufgezeichnet: Eins trug der jeweilige Lehrer am Körper, die anderen drei Mikrofone wurden im Raum und an den Schülertischen angebracht. Für eine genaue Beschreibung des technischen Aufbaus der Videoaufnahme sei auf Jatzwauk (in Druck) verwiesen.

Der zeitliche Abstand zwischen den beiden Untersuchungsterminen betrug in allen Klassen genau eine Woche. Zum zweiten Untersuchungszeitpunkt bekamen die Schüler einen zuvor nicht angekündigten Leistungstest zur Bearbeitung vorgelegt. Die Lehrer wurden gebeten, ihren Schülern zuvor nichts von dem Leistungstests zu sagen, damit keine Unterschiede zwischen den Klassen bzw. Übungs- und Vorbereitungseffekte bzgl. des Tests entstehen konnten. Um den Tests zu bearbeiten, hatten die Schüler 45 Minuten Zeit. Diese Bearbeitungszeit hatte sich in einer Pilotstudie als ausreichend bewährt. Testleiterin war dabei die Autorin.

Um die Untersuchungsbedingungen zwischen den einzelnen Schulklassen konstant zu halten und so eine Vergleichbarkeit zwischen den Klassen zu gewährleisten, wurde die Unterrichtszeit zwischen den beiden Untersuchungsterminen vereinheitlicht, indem in allen Klassen zwischen der Videoaufzeichnung und dem Leistungstest lediglich eine weitere Einzelstunde Physikunterricht vom Lehrer gehalten wurde. Diese Einzelstunde ist nicht Teil der vorliegenden Studie und konnte aus organisatorischen Gründen nicht aufgezeichnet werden, so dass keine Informationen über den physikalischen Inhalt dieser Stunde vorliegen.

9.4 Instrumente: Leistungstests

Aufgabenarten

Die Leistungstests setzten sich aus den folgenden Aufgabenarten zusammen:

1. Drei wörtlich wiederholte Unterrichtsaufgaben, welche der Lehrer in der aufgezeichneten Unterrichtsstunde eine Woche zuvor bereits eingesetzt hatte.
Dabei wurden sowohl schriftliche als auch mündliche Aufgaben berücksichtigt, da z.T. das fragend-entwickelnde Unterrichtsgespräch (vgl. Seidel, 2003) vorherrschte und aufgrund dessen in einigen der aufgezeichneten Unterrichtsstunden keine schriftlichen Aufgaben vom Lehrer gestellt wurden.
2. Zehn neu konstruierte Aufgaben auf unterschiedlichem lehrziel-taxonomischem Niveau, welche den Inhaltsbereich der aufgezeichneten Unterrichtsstunde abbildeten. Zur Kategorisierung der Aufgaben und zur Vorhersage der Aufgabenschwierigkeit wurde die Revision der Lehrzieltaxonomie von Bloom (Anderson et al., 2001, vgl. Kapitel 5) herangezogen, wobei die drei kognitiven Prozesse Wissen, Verstehen und Anwenden der Aufgabenbearbeitung unterschieden wurden.
3. Zusätzlich wurden neun TIMSS-Items (Martin & Kelly, 1998) als Ankeritems eingesetzt, um die Leistungen in den verschiedenen Klassen über die unterschiedlichen Unterrichtsinhalte hinweg mit Hilfe der Rasch-Skalierung vergleichen zu können. Es bekamen alle Schüler sechs TIMSS-Items vorgelegt, wobei jeweils alle Gymnasialisten und alle Hauptschüler dieselben Items bearbeitet haben. Zwischen den Schulformen bestand des Weiteren eine Schnittmenge von 3 identischen Items, die drei anderen Items waren schulformspezifisch anhand ihrer Schwierigkeit ausgewählt worden. Die eingesetzten TIMSS-Items sind in Anhang F, welcher sich auf der beiliegenden CD befindet, aufgeführt.

Insgesamt bestand jeder Leistungstest somit aus 13 inhaltsbezogenen Aufgaben und sechs TIMSS-Items. Die inhaltsbezogenen Aufgaben setzen sich wie folgt zusammen: drei wiederholte Unterrichtsaufgaben, vier neu konstruierte Wissensaufgaben, davon waren zwei mit offenem Antwortformat und zwei multiple-choice Aufgaben mit einer richtigen Antwort und drei Distraktoren, vier neu konstruierte Verstehensaufgaben und zwei neu konstruierte Anwendungsaufgaben.

Aufgabenanalyse

Die Konstruktion der Aufgaben erfolgte in Anlehnung an Klauers (1987) Empfehlungen zur Konstruktion lehrzielvalider Tests (vgl. Kapitel 6.2). Eine wichtige Voraussetzung hierfür ist die Aufgabenanalyse, die in der vorliegenden Studie nach Jonassen, Tessmer und Hannum (1999) erfolgte. Der Prozess der Aufgabenanalyse umfasst fünf Schritte, die beispielhaft anhand einer aufgezeichneten Unterrichtsstunde zum Thema „Bestandteile und Funktionsweise eines Elektromotors“ skizziert wird:

1. *Bestandsaufnahme der im Unterricht vorkommenden Aufgaben und des Inhaltsbereichs:*

In diesem ersten Schritt werden die Unterrichtsaufgaben aus der aufgezeichneten Unterrichtsstunde identifiziert und zusammengefasst.

Beispiel: Es werden alle Aufgaben aufgezählt, welche zum Inhaltsbereich bzw. dem Thema der Stunde „Bestandteile und Funktionsweise des Elektromotors“ gestellt wurden.

2. *Auswahl von Lernaufgaben für die Analyse:*

Da es meist eine Vielzahl von möglichen Lernaufgaben zur Zielerreichung gibt bzw. im vorliegenden Fall nicht alle Unterrichtsaufgaben im Leistungstest eingesetzt werden können, muss eine Auswahl getroffen werden (vgl. Klauer, 1987). Hierbei muss berücksichtigt werden, welche Aufgaben von besonderer Bedeutung für das Verständnis und somit das Erlernen des Inhaltsbereichs sind. Dazu müssen jeweils Auswahlkriterien entwickelt werden, um eine valide Menge der Gesamtheit aller möglichen Aufgaben auszuwählen.

Beispiel: Mit Hilfe einer richtig beschrifteten Abbildung eines Elektromotors kann überprüft werden, ob ein bestimmter Teil des Inhaltsbereichs, z.B. die „Kenntnis der Bestandteile des Elektromotors“, von den Schülern gelernt wurde.

3. *Aufgabenbeschreibung:*

Dies ist das zentrale Merkmal der Aufgabenanalyse. Dabei wird die Aufgabe in ihre einzelnen Komponenten zergliedert, um so zu identifizieren, was der Schüler im Detail lernen muss (Briggs, 1977; Dick & Carey, 1996; Jonassen et al., 1999). Im Mittelpunkt der Analyse steht die Frage, welche kognitiven Prozesse der Schüler durchlaufen muss, um die Lernaufgabe lösen zu können. Eine weitere wichtige Frage, welche beantwortet werden muss, bezieht sich auf das notwendige Vorwissen und auf Voraussetzungen in Bezug auf die Fertigkeiten, über die der Schüler zur erfolgreichen Aufgabenbearbeitung verfügen muss.

Beispiel: Um die Abbildung eines Elektromotors richtig beschriften zu können, muss der Schüler die einzelnen Bestandteile wiedererkennen und die dazugehörigen Fachbegriffe aus dem Gedächtnis abrufen können. Entsprechend der Taxonomie von Anderson et al. (2001) handelt es sich demnach bei der Aufgabenbearbeitung um den kognitiven Prozess des Wissens.

4. *Sequenzierung der Lernaufgaben:*

Der 4. Schritt erfolgt über die Beschreibung der Lernaufgaben als sinnvolle Sequenz, in der diese im Unterricht eingesetzt werden sollten, um das Lernen des Inhaltsbereichs für die Schüler zu erleichtern. Nach den Taxonomien (vgl. Anderson et al., 2001; Bloom et al., 1956), sollte der Unterricht zu einem neuen Inhaltsbereich – aufgrund der hierarchischen Struktur der kognitiven Prozesse – mit Wissensaufgaben begonnen werden. Erst wenn die dazugehörigen Lernaufgaben von den Schülern erfolgreich beantwortet werden, sollten Verstehensaufgaben eingesetzt werden, da diese kognitiv höhere Prozesse zur Beantwortung erfordern und dabei den Prozess des Wissens mit einschließen (vgl. Kapitel 5).

Beispiel: Die Schüler sollten zunächst die Bestandteile des Elektromotors richtig nennen und der Abbildung zuordnen können, was entsprechend der Taxonomie (Anderson et al., 2001) als Wissensaufgabe klassifiziert wird, bevor sie die Verstehensaufgabe, die Funktionsweise des Elektromotors zu erklären, gestellt bekommen.

5. *Klassifikation der Lernergebnisse:*

Die definitive Funktion der Schritte der Aufgabenanalyse besteht in der Festlegung und Klassifikation der notwendigen Lernergebnisse, um das Lehrziel als erreicht ansehen zu können. Damit sind die Inhalte gemeint, welche der Schüler im Laufe des Unterrichts gelernt haben muss. Dies kann mit Hilfe von Aufgaben, welche verschiedene kognitive Prozesse bei der Bearbeitung erfordern, überprüft werden. Bearbeitet ein Schüler die Aufgaben korrekt, so kann das Lehrziel, welches dem Unterricht zugrunde liegt, als erreicht angesehen werden. Somit wird sichergestellt, dass mit den Wissensvoraussetzungen bzw. den zugrundeliegenden Fertigkeiten begonnen wird, also zunächst Wissen erworben wird, bevor das Verständnis gelehrt und gelernt wird.

Beispiel: Der Unterricht beginnt mit den Bestandteilen des Elektromotors, und wenn diese erfolgreich von den Schülern gelernt wurden, wird die Funktionsweise eingeführt.

9.5 Beschreibung der kognitiven Prozesse für die Aufgabenkonstruktion

Nachdem der relevante Inhaltsbereich und die dazu vom Lehrer gestellten Aufgaben in Bezug auf die kognitiven Prozesse, welche zur Bearbeitung notwendig sind, anhand des Unterrichtsvideos des jeweiligen Lehrers analysiert wurden, wurden jeweils zehn zu diesem Inhaltsbereich zugehörige Aufgaben neu konstruiert. Auch die Aufgabenkonstruktion soll im Folgenden anhand des Unterrichtsbeispiels „Bestandteile und Funktionsweise eines Elektromotors“ verdeutlicht werden.

Die drei kognitiven Prozesse Wissen, Verstehen und Anwendung lassen sich bei der Aufgabenkonstruktion über den Inhaltsbereich und die Aufgabenformulierung operationalisieren. Dafür sollen die kognitiven Prozesse im Folgenden zunächst näher beschrieben werden.

9.5.1 Wissen

In der Originaltaxonomie (Bloom et al., 1956) wird der kognitive Prozess des Wissens als „little more than remembering of the idea or phenomenon in a form very close to that in which it was originally encountered“ (pp. 28-29) definiert. Damit ist der Abruf von gelehrten Inhalten aus dem Gedächtnis gemeint: die Fähigkeit, Begriffe, Fakten usw. zu erinnern, sie aus dem Gedächtnis zu reproduzieren, sie in einer Situation wieder zu erkennen, in der bestimmte Stichwörter gegeben werden, mit dem Ziel, festzustellen, welche Kenntnisse gespeichert worden sind.

In der Psychologie wird eine ganze Reihe unterschiedlicher Formen des Wissens angeführt. „Über das, was man unter Wissen versteht, gibt es in der Wissenspsychologie zwar keine einheitliche Theorie, aber eine Art von Grundkonsens, der bestimmte Unterscheidungen von Wissen postuliert. Deren bekannteste ist die von Ryle (1969) übernommene Unterscheidung von prozeduralem (knowing how) und deklarativem Wissen (knowing that)“ (Gerstenmaier & Mandl, 2000, S. 291). Diese Unterscheidung gilt inzwischen als etabliert (Renkl, 1991).

Nach McCarthy und Hayes (1969) benötigt ein symbolverarbeitendes kognitives System jeweils Daten über die repräsentierte Realität und Prozeduren, um mit diesen Daten umzugehen. Die gespeicherten Datenstrukturen über die Realität bilden das deklarative Wissen des Systems, welches Fakten, allgemeine Zusammenhänge, Prozesse usw. repräsentiert. Das prozedurale Wissen hingegen bildet Datenstrukturen, die vom kognitiven System als Anweisungen bzw. Befehle interpretiert werden und von ihm unmittelbar ausführbar sind. Die

betreffenden Prozeduren dienen dazu, anhand der gespeicherten Daten neue Daten zu generieren, d.h. aus vorhandenem Wissen neues Wissen abzuleiten, und sie spielen eine bedeutende Rolle bei der Ausführung von Prozessen. Nur wenn das System über spezifische Prozeduren zum Vollzug dieser Prozesse verfügt, kann man von einem entsprechenden prozeduralen Wissen sprechen (Schnotz, 1994). Die Unterscheidung zwischen deklarativem und prozeduralem Wissen ist keine Differenzierung nach dem Wissensinhalt, sondern eine Differenzierung nach der Funktion der betreffenden Datenstrukturen innerhalb des kognitiven Systems. Die verschiedenen Wissensarten werden zwar aus analytisch-wissenschaftlichen Gründen unterschieden, in den meisten Fällen der Bearbeitung von Aufgaben dürften jedoch mehrere Komponenten zugleich beteiligt sein (Renkl, 1991).

Neben dieser Unterscheidung zwischen deklarativem und prozeduralem Wissen gibt es eine Vielzahl weiterer Wissensformen, auf die im Rahmen dieser Arbeit nicht näher eingegangen werden kann. Es soll lediglich die Fülle an Konstrukten mit Hilfe der Unterteilung von Reinmann-Rothmeier und Mandl (1998) verdeutlicht werden, wobei domänenspezifisches Wissen (deklaratives Wissen), prozedurales Wissen, strategisches Wissen, metakognitives Wissen, verbale Fähigkeiten sowie soziale Fertigkeiten und Kompetenzen unterschieden werden (vgl. Law, 2000).

Basierend auf der kognitiven Psychologie wird angenommen, dass Wissen domänenspezifisch und kontextabhängig ist. Außerdem spielen soziale Erfahrungen und der Kontext bei der Konstruktion und Entwicklung des Wissens eine wichtige Rolle (vgl. Kapitel 4.4; auch Bereiter & Scardamalia, 1998; Bransford, Brown & Cocking, 1999; Case, 1998; Keil, 1998; Mandler, 1998; Wellman & Gelman, 1998).

Der Erwerb von Wissen als Lehrziel wird oft kontrovers diskutiert. Auf der einen Seite wird der Wert des „reinen“ Wissens oder der Fähigkeit, sich Fakten, Begriffe und Formeln zu merken, und des schematischen Auswendiglernens bezweifelt. Die Hauptkritik beispielsweise am verbreiteten Einsatz von Wissensfragen bezieht sich auf die mangelnde Anschlussfähigkeit dieser Wissensart (z.B. Gall, 1984). Schüler machen oft die wichtigen Verbindungen zwischen gelernten Fakten nicht und können häufig keine Transferleistungen erbringen. Dies wird häufig als Problem des „trägen Wissens“ bezeichnet (Renkl, 1996), worauf in Kapitel 9.5.3 näher eingegangen wird. Schüler erwerben demnach häufig viel Faktenwissen, haben dieses aber oft nicht tiefer gehend verstanden bzw. können es nicht integrieren oder systematisch organisieren und anwenden (Bereiter & Scardamalia, 1998; Bransford, Brown & Cocking, 1999). Helmke (2004) sieht beim Üben im Unterricht eine Schiefelage, indem das Üben häufig als mechanisch, als passiv, stures Pauken, für den Lernprozess eigentlich überflüssig und dem aktiven Lernen zuwiderlaufend angesehen wird. Doch ohne

quantitativ ausreichendes Üben, ohne vielfältige Wiederholungen, den lernpsychologischen Gesetzmäßigkeiten folgend auf die Vervollkommnung von Fertigkeiten und zugleich auf den Wissenstransfer abzielend, also welches nicht bloß repetitiv ist, ist keine „Flüssigkeit grundlegender Fertigkeiten“ (Helmke, 2004, S. 38) erreichbar, welche jedoch eine zwingende Voraussetzung für anspruchsvollere Kompetenzen darstellt. Auch Gruber, Mandl und Renkl (2000) betonen, dass „nicht infrage gestellt werden [soll], dass rezeptives Lernen in Bezug auf eine ganze Reihe von Unterrichtszielen erfolgreich sein kann“ (S. 147). Darauf wies auch Weinert (1998) hin. Das Erinnern von Wissen ist somit wesentlich für bedeutungsvolles Lernen und Problemlösen, da es als Voraussetzung für die Bearbeitung komplexer Aufgaben erforderlich ist (Bratman, 1987). Nach Helmke (2004) ist die wichtigste Voraussetzung für kumulative und anspruchsvolle Lernprozesse eine solide und gut organisierte Wissensbasis, d.h. ein in sich vernetztes, in verschiedenen Situationen erprobtes und flexibel anzupassendes Wissen, das Fakten, Konzepte, Theorien und Methoden gleichermaßen umfasst. Daher sollte nach Anderson et al. (2001) Lernen über das Auswendiglernen von isolierten Wissensfragmenten hinausgehen und als integrierter Teil einer komplexeren Aufgabe angesehen werden.

Da die verschiedenen Wissensarten bei der Aufgabenkonstruktion nicht eindeutig voneinander zu trennen sind, wird in den Studien 1 und 2 der hier vorliegenden Arbeit eine differenzierte Betrachtung der unterschiedlichen Wissensarten nicht geleistet, sondern es wird allgemein von „Wissensaufgaben“ gesprochen.

9.5.2 Verstehen

Der zweite kognitive Prozess der Taxonomie entspricht dem Niveau des Verstehens, wobei in eigenen Worten eine Zusammenfassung des Gelernten gegeben werden muss, aber noch keine Beziehungen zu anderen Informationen und Implikationen herzustellen sind. Nach Bloom et al. (1956) geht es um die Fähigkeit, das, was kommunikativ vermittelt wird, aufzunehmen und zu verarbeiten, ohne es notwendigerweise auf andere Bereiche zu übertragen bzw. seine Implikationen zu erkennen. Ohlsson und Rees (1991) unterstellen zudem, dass „understanding of a procedure must involve both knowledge of its domain and of its technology“ (p. 118) und betonen damit die Bedeutung von Vorwissen für das Verstehen. Außerdem müssen Lernende z.B. eine Prozedur zunächst verstanden haben, bevor sie diese auf ein Problem erfolgreich anwenden können (Van Gog, Paas & van Merriënboer, 2004). Somit wird wiederum die Vorbedingung des Verstehens für die Wissensanwendung deutlich.

Nach Stern (1999) zeigt sich Verständnis in der Anwendung elaborierter Strategien und in der Verfügbarkeit von Konzepten. Auch Greeno (1987) definiert: „Understanding occurs when the information taken from specific problems is fitted into one of these general structures“ (p. 67). Verstehen bedeutet nach Brophy (2000), dass Schüler lernen, Elemente in ein Netzwerk zusammengehöriger Inhalte, mit Beziehungen zwischen den verschiedenen Bereichen, zu erkennen und dieses Wissen in eigenen Worten ausdrücken zu können und es mit ihrem Vorwissen zusammenzuführen. Dabei spielen die im Gedächtnis zu diesem Inhaltsbereich gespeicherten Informationen eine wichtige Rolle, da dort die verschiedenen Wissensarten integriert werden, die zur Beantwortung der Frage benötigt werden (Galambos & Black, 1985). Für weitere Definitionen und theoretische Überlegungen des Verstehens sei auf Johnson-Laird (1983), Kintsch (1988), Rinck (2000) sowie van Dijk und Kintsch (1983) verwiesen.

Wenn man im Verstehen den Aufbau einer in sich stimmigen, kohärenten mentalen Repräsentation sieht, so bedeutet dies allerdings nicht, dass die Repräsentation des Gegenstands immer adäquat ist. Sowohl beim Verstehen von Situationen, Prozessen und Handlungen als auch beim Verstehen von Texten gibt es die Möglichkeit des Missverstehens sowie des Nichtverstehens. Im Fall des Missverstehens wird eine mentale Repräsentation gebildet, die zwar in sich konsistent und kohärent ist, den betreffenden Sachverhalt jedoch inadäquat abbildet (Schnotz, 1994). Die Grenze zwischen dem richtigen Verstehen und dem Missverstehen ist fließend. Auch Missverstehen ist ein Verstehen und basiert auf grundsätzlich gleichen kognitiven Prozessen wie richtiges Verstehen (vgl. Scriven, 1972). Ein Sachverhalt wird im Gegensatz dazu nicht verstanden, wenn der Schüler zum Aufbau einer kohärenten und in sich stimmigen mentalen Repräsentation nicht in der Lage ist. Beispielsweise bezeichnet man einen Text als nicht verstanden, wenn der Schüler zwar die einzelnen Sätze für sich genommen versteht, jedoch den Gesamtzusammenhang nicht erfasst und die mentale Repräsentation nur aus unzusammenhängenden Teilen besteht. Die Abgrenzung von Verstehen und Nichtverstehen ist ebenfalls fließend, da es letztlich eine Ermessensfrage ist, ab welchem Grad von mentaler Kohärenz man von Verstehen oder Nichtverstehen spricht (Schnotz, 1994). Gerade bei der Einschätzung des Schülerverständnisses durch die Lehrer gibt es große Unterschiede, ab wann etwas als verstanden angesehen wird.

Es hat sich gezeigt, dass Übungsphasen im deutschen Mathematikunterricht dadurch gekennzeichnet sind, dass die gelernten Inhalte meist in Form von Wissensabruf aufgegriffen werden, ohne jedoch ein tieferes Verständnis zu gewährleisten, da kein Versuch gemacht wird, explizit auf Aspekte des Verstehens und des Durcharbeitens einzugehen (Aebli, 2006; Steiner, 1995). Der Inhalt wird dann automatisiert, bevor er verstanden ist. Durch diese Art

des Übens werden isolierte Wissensteile ohne innere Zusammenhänge aufgebaut. Damit kann das Wissen auch kaum auf andere Aufgaben angewendet werden, und es besteht die Gefahr, dass „träges Wissen“ (vgl. Renkl, 1996) entsteht. Die kognitive Funktion dieser Übungsart ist hauptsächlich das Enkodieren von Informationen (Neubrand, 2002).

Aufgrund der bisherigen Ergebnisse zur Förderung des Verständnisses im Unterricht (vgl. Neubrand, 2002; Steiner, 1995) wird für die vorliegende Studie 1 davon ausgegangen, dass im Physikunterricht ebenfalls wenig tieferes Verständnis gefördert wird und die Schüler Probleme im Verstehen physikalischer Sachverhalte haben, was auch die TIMSS-Studie gezeigt hat (Klieme, 2000).

9.5.3 Anwendung

Mit dem kognitiven Prozess der Wissensanwendung ist nach Anderson et al. (2001) die Auswahl und Anwendung einer Methode, Regel oder Idee zur Lösung eines Problems in einer gegebenen Situation gemeint. In den Naturwissenschaften ist dies oft die Anwendung von Untersuchungsmethoden und Denkweisen zur Lösung eines Problems.

Es ist eine Nähe zwischen den Konzepten „Anwendung“ und „Transfer“ gegeben, welche durch die gängige Definition von Transfer als Anwendung von gelerntem Wissen auf neue Inhaltsbereiche deutlich wird (vgl. Law, 2000). Diese Beschreibung zeigt Ähnlichkeit mit der von Kognitionspsychologen (z.B. Anderson, Reder & Simon, 1996; Mayer & Wittrock, 1996) gebrauchten Definition. Demgemäß besteht Transfer in der Anwendung relativ stabiler, dekontextualisierter Fähigkeiten in neuen Situationen. Eine Voraussetzung der erfolgreichen Wissensanwendung bzw. des Transfers stellt das Verständnis des thematisierten Inhaltsbereichs dar (Bynes, 1996; Mayer & Wittrock, 1996). Kurz gesagt bedeutet Transfer nicht nur die Erinnerung von zuvor Gelerntem, sondern auch dass das Gelernte Sinn macht und deren Anwendung auf andere Bereiche bzw. Inhalte stattfindet (Bransford, Brown & Cocking, 1999; Detterman & Sternberg, 1993; Mayer, 1995; McKeough, Lupart & Marini, 1995; Phye, 1997). Von verschiedenen Autoren (vgl. Ausubel & Robinson, 1969; Salomon & Perkins, 1987; Seel, 2003) werden unterschiedliche Formen des Transfers genannt, so z.B. naher und weiter Transfer, lateraler, sequentieller und vertikaler, sowie positiver und negativer Transfer, auf die im Rahmen dieser Arbeit nicht näher eingegangen werden kann.

Nach Roth (1963) ist die Leitidee pädagogischen Handelns, Schüler zu befähigen, die in der Schule erworbenen Kenntnisse und Fertigkeiten auf Situationen des Alltags anzuwenden. Auch in der Konzeption der naturwissenschaftlichen Grundbildung in PISA betonen die Autoren (Prenzel et al., 2001), dass dieses Wissen von den Schülern in unterschiedlichen

Situationen flexibel genutzt und angewendet werden sollte. Weiter wird hervorgehoben, dass naturwissenschaftliches Wissen anschlussfähig sein muss, d.h. Kompetenzen umfasst, die für die Aneignung neuer Wissensbestände bzw. für eine Auseinandersetzung mit den Naturwissenschaften erforderlich sind. Insgesamt sind die Kompetenzanforderungen „bestimmt durch hohe Erwartungen an die Anwendbarkeit des Wissens bzw. an den Transfer“ (Prenzel et al., 2001, S. 195). Gleichwohl stellt Seel (2003) fest, dass im schulischen Kontext das meiste Wissen durch Assimilation erworben wird, d.h. durch Ausweitung und Verfeinerung bestehender Wissensstrukturen. Radikale Veränderungen der Wissensstrukturen sind in der Schule selten, Schüler erwerben deklaratives Wissen über Konzepte, werden aber nicht in die Lage versetzt, dieses zu interpretieren und flexibel anzuwenden (Seel, 2003).

Das Phänomen, dass theoretisches Wissen, das in der Schule erlernt wird, in außerschulischen Kontexten vielfach nicht angewandt wird, wurde bereits 1929 von Whitehead unter den Begriff des „trägen Wissens“ (*inert knowledge*) beschrieben. Brown (1989) zeigte auf, dass die meisten Schüler daran scheitern, ihr Wissen auf neue Situationen und Bereiche anzuwenden. Wenn es um mathematische Anforderungen geht, zeigt sich immer wieder, dass es Versuchspersonen nicht bzw. nur unzureichend gelingt, Formeln zu verstehen. Sie werden oft auswendig gelernt und automatisch bzw. mechanisch angewendet, ein solcher Umgang mit mathematischen Formeln kann als eher *mindless* (Salomon & Globerson, 1987) bezeichnet werden. Dieser Effekt kann auf den Umgang mit physikalischen Formeln übertragen werden. Führt ein entsprechender Einsatz von Formeln zufällig zum Ziel, können Verständnisprobleme kaschiert werden, unter dem Begriff „Kompetenzillusion“ eingeführt (Stark, 1999), und es treten Probleme der Wissensanwendung und des Transfers auf (Stark et al., 2000). Weitere Befunde zeigen, dass die wenig anwendungsbezogene, oft abstrakte und systematisierte Form der Wissensvermittlung, welche der Komplexität des Alltags nur selten gerecht wird, dazu beiträgt, dass „träges Wissen“ entsteht (Renkl, Gruber & Mandl, 1996). In komplexen Problemsituationen gelingt die Wissensanwendung oft nur unvollständig oder überhaupt nicht, wodurch es zu einer Kluft zwischen Wissen und Handeln kommt (Mandl & Gerstenmaier, 2000). Es ist empirisch hinreichend gesichert, dass es Schülern oft schwer fällt, im Unterricht erworbenes Wissen auf ähnliche Aufgabenstellungen desselben Inhaltsbereichs anzuwenden (Lave, 1992; Renkl, 1996; Saxe, 1988; Van Gog et al., 2004). Renkl und Helmke (1992) konnten in einer Untersuchung zu Lehrerfragen im Mathematikunterricht zeigen, dass die bloße Automatisierung von Rechenfertigkeiten, was dem Wissensniveau zugehörig ist, nicht zu einer erfolgreichen Anwendung führt. Erst das Verständnis der Prozesse und der zugrunde liegenden Prinzipien ermöglichte die Anwendung auf Problemstellungen. Damit wird die Bedeutsamkeit des Wissens und Verstehens als Vorbedingung für

deren erfolgreiche Anwendung auf andere Aufgaben deutlich. Hamilton (1989) hat nachgewiesen, dass Studierende, die in einer Übungsphase sowohl Definitions- als auch Anwendungsaufgaben bearbeiteten, signifikant bessere Testleistungen zeigten als Studierende, die zuvor ausschließlich Anwendungsaufgaben bearbeitet hatten. Der Autor führte als Erklärung an, dass mit Hilfe der Definitionsaufgaben das notwendige Faktenwissen zur Verfügung gestellt wurde, um die nachfolgenden Anwendungsaufgaben angemessen bearbeiten zu können. Es ist allerdings zu überlegen, inwieweit das Vorwissen die Effekte von Anwendungsaufgaben beeinflusst: Personen, die über ein gefestigtes Faktenvorwissen verfügen und in der Lage sind, dieses selbstständig zu aktivieren, profitieren möglicherweise von Anwendungsaufgaben stärker als Personen, die wenig und kaum gefestigtes Faktenvorwissen haben und somit zunächst einer Unterstützung zur Bereitstellung des notwendigen Vorwissens bedürfen (Levin, 2005). Eine solche Interaktion zwischen Vorwissen und genutzten Aufgaben bedarf noch der empirischen Überprüfung.

Die Übertragung und Anwendung des Gelernten auf neuartige Situationen ist offenkundig der neuralgische Punkt des Handlungsfeldes Schule (Seel, 2003). Diese Beurteilung der im Unterricht stattfindenden Transferbemühungen wird durch neuere Befunde wie z.B. TIMSS (vgl. Baumert, Lehmann, Lehrke, Schmitz, Clausen, Hosenfeld, Köhler & Neubrand, 1997) belegt. Im Gegensatz dazu kommt die Metaanalyse von Semb und Ellis (1994) zu dem Ergebnis, dass das langzeitige Behalten von Schulwissen beachtlich und substantiell ist: Schüler behalten demnach vieles von dem, was sie im Unterricht lernen, und zwar über einen längeren Zeitraum. Allerdings geht die Behaltensleistung mit der Zeit zurück, aber die Vergessenskurven für Schulwissen nehmen nicht so schnell ab wie die Vergessenskurven in psychologischen Lernexperimenten. Dies wird von Seel (2003) damit erklärt, dass das in der Schule erworbene Wissen eben nicht aus isolierten Einheiten bestehe, sondern kumulativ aufgebaut werde und vielseitige Beziehungen unter Wissenseinheiten aufweise.

Gagné (1965) nahm an, dass ein erfolgreicher Transfer einer Aufgabenlösung auf eine andere Lernaufgabe davon abhängig ist, inwieweit es dem Lernenden gelingt, die Inhalte der verschiedenen Aufgabenstellungen miteinander in Beziehung zu setzen, und ob der Lernende eine generalisierbare Fähigkeit zur Aufgabenlösung durch die Erfahrungen mit der ersten Lernaufgabe aufgebaut hat und anwenden kann. Dazu muss der Lernende jedoch erkennen, dass die Lösung, die er bei einer Aufgabe erfolgreich entwickelt hat, auf die strukturell äquivalenten Aufgaben in einem anderen Bereich übertragbar ist, was den Lernenden in verschiedenen Untersuchungen nicht gelang (Gick & Holyoak, 1983, 1987).

Renkl (1996) führt folgende drei Typen von Erklärungen für das Phänomen des „trägen Wissens“ aus der Literatur an:

1. *Metaprozesserklärungen* gehen davon aus, dass das notwendige Wissen vorhanden ist, aber nicht genutzt wird. Es wird dabei angenommen, dass Metaprozesse (z.B. metakognitive Steuerungsprozesse), die gewissermaßen über dem anzuwendenden Wissen ablaufen, defizitär sind.
2. *Strukturdefiziterklärungen* sehen die Defizite im anzuwendenden Wissen selbst, d.h. das Wissen ist nicht in einer Form vorhanden, die eine Anwendung erlauben würde. Häufig wird die fehlende Wissensanwendung auf Defizite im Verständniswissen oder auf eine mangelnde Überführung von Faktenwissen in Handlungswissen zurückgeführt.
3. *Situiertheitserklärungen* stellen den traditionellen Wissens- und Transferbegriff der kognitiven Psychologie in Frage und gehen von der Grundannahme aus, dass Wissen prinzipiell situativ gebunden sei. Deshalb gilt mangelnder Wissenstransfer weniger als „Defizitfall“ denn als Normalfall.

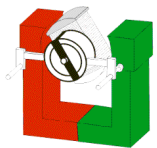
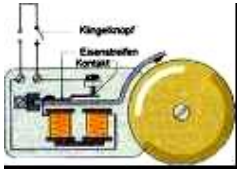
Den Erklärungsansätzen ist trotz ihrer Unterschiedlichkeit gemeinsam, dass an komplexen, authentischen oder zumindest realitätsnahen Problemstellungen gelernt werden soll, um die Entstehung von trägem Wissen zu vermeiden, da Wissen dann direkt in einem Anwendungskontext erworben wird (Renkl, 1996). Ein Defizit dieser Ansätze liegt jedoch in der noch mangelnden empirischen Fundierung der Lerneffekte, denn traditionelle Leistungstests erfassen die Anwendungsqualität des Wissens meist nicht. Dies soll in der vorliegenden Arbeit berücksichtigt werden, indem explizit Anwendungsaufgaben in allen Leistungstests eingesetzt werden und somit die Wissensanwendung der Schüler im Physikunterricht empirisch überprüft werden kann.


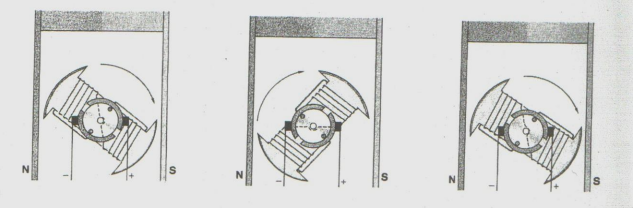
9.5.4 Aufgabenbeispiele

Mit Hilfe der auf der Grundlage der Taxonomie von Anderson et al. (2001) definierten kognitiven Prozesse sowie der Konstruktionsvorschriften von Klauer (1987) wurden die Ergebnisse der Aufgabenanalyse zugrunde gelegt, um die verschiedenen schwierigen Aufgaben zu konstruieren.

In der folgenden Tabelle 4 sind die Aufgaben, welche in dem Leistungstest zu der Unterrichtsstunde „Bestandteile und Funktionsweise eines Elektromotors“ gestellt wurden, mit der Klassifikation der Aufgabenart aufgelistet.

Tabelle 4: Aufgabenstellung und Aufgabenart am Beispiel einer Unterrichtsstunde zu "Bestandteile und Funktionsweise eines Elektromotors"

Aufgabenstellung	Aufgabenart
1 Welche Funktion hat der Kommutator (Stromwender) bei dem Elektromotor?	Wiederholte Unterrichtsaufgabe (Verstehensaufgabe)
2 Welche der folgenden Aussagen zu den Magnetpolen ist richtig? A. Jedes Magnetfeld hat einen Plus- und einen Minuspol. B. Der Nordpol wird in Abbildungen immer grün und der Südpol immer rot dargestellt. C. Jedes Magnetfeld hat einen Nord- und einen Südpol. D. Ungleiche Pole stoßen sich gegenseitig ab, gleiche Pole ziehen sich an.	MC-Wissensaufgabe
3 Aus einem Stabmagnet als Rotor und einer Spule mit Eisenkern als Stator kann man einen einfacher „Elektromotor“ bauen. Wie kannst du einen solchen Magneten zum Drehen bringen? Welche Nachteile besitzt dieser „Motor“ und wie kannst du diese Nachteile beheben?	Verstehensaufgabe
4a Beschrifte die wichtigen Elemente des Elektromotors in der Abbildung.	Wissensaufgabe
	
4b Erkläre die Funktionsweise des Elektromotors.	Verstehensaufgabe
5a Eine elektrische Klingel hat den dargestellten Aufbau. Beschreibe diesen und erkläre die Wirkungsweise einer solchen Klingel. Nutze dafür dein Wissen über Elektromotoren.	Anwendungsaufgabe
	
5b Welche anderen Geräte kennst du, die von Elektromotoren betrieben werden?	Wiederholte Unterrichtsaufgabe (Wissensaufgabe)
6 Was passiert, wenn du bei einem Elektromotor an die senkrecht zur Achse gesteckte Spule mit Eisenkern eine Gleichspannung von 4 Volt anlegst. Wie verhält sich die Spule? Und wie verhält sich die Spule, wenn du die Stromzuleitung an der Gleichspannungsquelle umpolst?	Wiederholte Unterrichtsaufgabe (Wissensaufgabe)

Aufgabenstellung	Aufgabenart
<p>7 In manchen Motoren findet man einen Anker mit 3 Spulen (Abbildung). Welchen Vorteil hat er gegenüber einem Anker mit nur zwei gegenüberliegenden Spulen?</p> 	Anwendungsaufgabe
<p>8 Zeichne jeweils einen Stabmagneten, einen Hufeisenmagneten und einen Elektromagneten mit dem dazugehörigen Magnetfeld!</p>	Wissensaufgabe
<p>9 Welche der folgenden Aussagen zum Elektromotor ist richtig?</p> <p>A. Sowohl die gleichnamigen und die ungleichnamigen Pole von Rotor und Stator ziehen sich gegenseitig an.</p> <p>B. Die gleichnamigen Pole von Rotor und Stator ziehen sich an, die ungleichnamigen stoßen sich ab.</p> <p>C. Sowohl die gleichnamigen, wie die ungleichnamigen Pole von Rotor und Stator stoßen sich gegenseitig ab.</p> <p>D. Die gleichnamigen Pole von Rotor und Stator stoßen sich gegenseitig ab, die ungleichnamigen ziehen sich an.</p>	MC-Wissensaufgabe
<p>10a Die unten stehenden Bilder zeigen einen Elektromotor mit verschiedenen Stellungen des Rotors. Erkläre, warum sich der Rotor jeweils weiterdreht. Kennzeichnen die Pole des Elektromagneten in allen drei Bildern.</p> 	Verstehensaufgabe
<p>10b Wenn der Rotor waagrecht steht, fließt kein Strom durch ihn. Welches Problem tritt dadurch auf?</p>	Verstehensaufgabe

9.6 Auswertung der Unterrichtsvideos und der Leistungstests

Die Videos der aufgezeichneten Unterrichtsstunden wurden in Bezug auf die durch den Lehrer eingesetzten Aufgaben ausgewertet. Das dafür entwickelte Auswertungssystem ist in Anhang A dargestellt. Für eine detaillierte Beschreibung der Videoauswertung sei an dieser Stelle auf Jatzwauk (in Druck) verwiesen, da dieser Auswertungsteil für die vorliegende Arbeit nicht von maßgeblicher Bedeutung ist. Die Aufgaben wurden, nachdem sie in

der aufgezeichneten Unterrichtsstunde identifiziert worden waren, im Hinblick auf die zur Aufgabenbeantwortung notwendigen kognitiven Prozesse klassifiziert.

Die Leistungstests wurden mit Hilfe von Lösungsbeispielen ausgewertet, um so die Auswertungsobjektivität des offenen Antwortformats zu erhöhen (siehe Lienert & Raatz, 1998). Dabei wurde der Wert „Null“ für eine falsche Antwort bzw. Nichtbearbeitung vergeben. Der Wert „Eins“ bzw. ein Punkt wurde für eine richtige bzw. ausreichende Antwort im Sinne der Beispiellösungen vergeben. Aus allen im Leistungstest erhaltenen Punkten wurde der Summenscore gebildet, der somit das Gesamtergebnis des Tests darstellt. Um die Lösungshäufigkeiten der einzelnen Aufgabenarten nach den klassischen Testverfahren berechnen zu können, wurde auf die Ergebnisse aller Schüler, welche eine bestimmte Aufgabe bearbeitet hatten, zurückgegriffen.

Die Ergebnisse der Leistungstests wurden zugrunde gelegt, um sowohl Itemparameter als auch Personen- bzw. Fähigkeitsparameter mit Hilfe der IRT-Skalierung berechnen zu können. In der sog. *Item-Response-Theorie* (IRT) bzw. den probabilistischen Testtheorien⁶ stehen zur Testauswertung Verfahren zur Verfügung, die einen Zusammenhang zwischen Personenfähigkeit, Aufgabenschwierigkeiten und Lösungswahrscheinlichkeiten herstellen. Das logistische Raschmodell für Aufgaben mit dichotomen Antwortausprägungen ist dabei das einfachste Modell der IRT und zugleich dasjenige, das der üblichen klassischen Auswertung über die Summe der Aufgabenlösungen am nächsten kommt (vgl. Andrich, 1988; Rasch, 1980; Rost, 2004). Es wird auch in den PISA-Berechnungen eingesetzt (siehe den technischen Bericht; Adams & Wu, 2002).

Grundlage für die Messung der Personenfähigkeiten ist dabei ein Aufgabensatz aus einem Pool von Aufgaben, die alle dieselbe Personenfähigkeit f (auch als ξ oder Θ dargestellt) messen. Die Aufgaben unterscheiden sich lediglich hinsichtlich ihrer Schwierigkeit (s , auch als σ_i dargestellt), sind aber ansonsten homogen (vgl. Fischer, 1968). Das bedeutet, dass - auch wenn Schüler unterschiedliche Aufgaben aus diesem Itempool bearbeiten - ihre Fähigkeiten auf einer gemeinsamen Skala darstellbar und somit direkt vergleichbar sind.

Auch für die Ermittlung der Aufgabenschwierigkeiten gilt, dass sie aufgrund ihrer Homogenitätseigenschaft anhand beliebiger Personenstichproben skaliert werden können und Unterschiede in den ermittelten Schwierigkeiten der Aufgaben unabhängig von der getesteten Personengruppe interpretierbar sind. Verfügt ein Testmodell über diese Stichprobenunabhängigkeit, so wird dies als spezifische Objektivität bezeichnet (vgl. Fischer, 1974). Rost (2004) spricht in diesem Zusammenhang von der Theorie der spezifisch objektiven

⁶ Der Begriff der probabilistischen Testverfahren bzw. Testtheorien wird in dieser Arbeit synonym mit dem vor allem im englischen Sprachraum genutzten Begriff der *Item-Response-Theorie* (IRT) verwendet.

Messung. Das bedeutet, dass die Schülerfähigkeit unabhängig von den durch die Schüler bearbeiteten Aufgaben bestimmt wird, und auch umgekehrt, dass die Bestimmung der Aufgabenschwierigkeit unabhängig von der Schülerstichprobe, die diese Aufgaben bearbeitet hat, erfolgt (vgl. Rost & Spada, 1978). Diese „wechselseitige Stichprobenunabhängigkeit“ erlaubt es, dass unterschiedliche Schülergruppen teilweise unterschiedliche Unterrichtsaufgaben bearbeiten und trotzdem die Schülerfähigkeiten bzw. Schwierigkeiten der Aufgaben miteinander verglichen werden können, da diese auf einer gemeinsamen Skala abgetragen werden. Die Vergleichbarkeit ist bei der Auswertung nach der klassischen Testtheorie nicht möglich. Dieser Vorteil der IRT-Skalierung ist für die Ergebnisinterpretation der vorliegenden Studie von immenser Bedeutung, da die Schüler unterschiedliche Lernaufgaben aus verschiedenen Inhaltsbereichen bearbeitet haben.

10 Ergebnisse

In den folgenden Unterkapiteln werden die Ergebnisse der Videoauswertung (Kapitel 10.1) sowie die der Leistungstests (Kapitel 10.2) dargestellt. Mit Hilfe der Videoauswertung wird die Interrater-Reliabilität in Bezug auf die Identifikation und Klassifikation der Unterrichtsaufgaben bestimmt (Kapitel 10.1.1). Ferner wird untersucht, ob die zur Wiederholung ausgewählten Unterrichtsaufgaben als eine hinreichend repräsentative und zufällige Auswahl der Gesamtheit der in den aufgezeichneten Unterrichtsstunden gestellten Aufgaben angesehen werden können (Kapitel 10.1.2). Des Weiteren wurden die aufgezeichneten Unterrichtsstunden zur Beantwortung der Fragestellung 1 zugrunde gelegt: Was für Aufgaben setzen Lehrer im Physikunterricht ein (Kapitel 10.1.3)? Die Auswertung der Leistungstests beginnt mit den deskriptiven Ergebnissen in Bezug auf die Lösungshäufigkeiten der verschiedenen Aufgabenarten, welche mit Hilfe der klassischen Testverfahren ausgewertet wurden, dargestellt in Kapitel 10.2. Im Anschluss daran werden die Ergebnisse der IRT-Skalierung sowohl für die Itemparameter der verschiedenen Aufgabenarten (Kapitel 10.3.1) als auch für die Personenparameter in den beiden Schulformen (Kapitel 10.3.2) beschrieben. Um die Frage nach der Passung zwischen Aufgabenschwierigkeit und Schülerfähigkeit beantworten zu können, werden die beiden Parameter, innerhalb der untersuchten Schulklassen gemittelt, im Anschluss gemeinsam betrachtet (Kapitel 10.3.3).

10.1 Videoauswertung

Die Überprüfung der Interrater-Reliabilität bei der Identifikation und Klassifikation der Unterrichtsaufgaben ist für die nachfolgenden Analysen bedeutsam, da diese Aufgaben Grundlage für die weiteren Berechnungen sind. Gleiches gilt für die Analyse der Repräsentativität der Aufgaben: Nur wenn die Unterrichtsaufgaben, die im Leistungstest wiederholt werden, eine repräsentative Auswahl der Gesamtheit der eingesetzten Aufgaben darstellen, können Aussagen über die Gesamtheit der Aufgaben sinnvoll getroffen werden.

10.1.1 Interrater-Reliabilität der Videoauswertung – Aufgabenidentifikation und Aufgabenklassifikation

Die grundlegende und gebräuchlichste Methode, die Güte von Auswertungssystemen zu überprüfen, stellt die Berechnung von Beurteiler- bzw. Beobachterübereinstimmungen dar (vgl. Bortz & Döring, 2002; Wirtz & Caspers, 2002). Die als Übereinstimmungsmaße

gebräuchlichsten Koeffizienten sind die prozentuale Übereinstimmung und Cohens κ (Kappa). Die prozentuale Übereinstimmung berechnet als einfachstes Maß der Übereinstimmung den prozentualen Anteil der Fälle, in denen zwei oder mehrere Beurteiler ein Objekt gleich beurteilen. Sie ist jedoch als Übereinstimmungsmaß nur beschränkt aussagekräftig, da sie nicht angibt, inwiefern die berechnete Übereinstimmung größer als eine im Falle des Zufalls erwartete Übereinstimmung ist. Cohens κ wird als standardisierte Maßzahl zwischen -1 und +1 angegeben und gibt das Ausmaß an, in welchem die tatsächlich beobachtete Übereinstimmung positiv von der Zufallserwartung differiert (Wirtz & Caspar, 2002). Bezüglich der Richtwerte für die Güte von Koeffizienten bietet die Literatur keine einheitlichen Vorgaben. Wirtz und Caspar (2002) geben als Faustregel an, dass ein $\kappa > .75$ eine sehr gute, ein κ zwischen $.60$ und $.75$ eine gute Übereinstimmung anzeigt. Voraussetzung für die Berechnung der Beurteilerübereinstimmung ist die unabhängige Beurteilung einer Teilstichprobe der aufgezeichneten Unterrichtsstunden (6 Stunden; entspricht 20% der Stichprobe), in Bezug auf die Identifikation und Klassifikation der Unterrichtsaufgaben durch zwei trainierte Kodierer.

Konkretisierung des Begriffs „Lernaufgabe“ für die vorliegende Untersuchung

Für eine reliable Identifikation der eingesetzten Unterrichtsaufgaben muss klar abgegrenzt werden, was als Lernaufgabe und was nicht als Lernaufgabe angesehen wird. Die Leistungsanforderungen als Aufgaben (vgl. Kapitel 2.1) im Unterricht lassen sich dabei in drei Gruppen differenzieren:

1. Aufforderungen zum Unterrichtsmanagement

Diese von Scholz (1980) als „arbeitsorganisierende Aufgabenstellungen“ bezeichneten Aufforderungen dienen der Steuerung des Unterrichtsgeschehens. Sie sind nicht unmittelbar auf den Umgang mit einem fachlichen Inhalt bezogen und sind somit keine Lernaufgaben im eigentlichen Sinne. Mit ihrer Hilfe organisiert der Lehrer den geordneten Unterrichtsverlauf, arrangiert unter anderem die Bearbeitung von Lernaufgaben und leitet den Austausch von Ergebnissen.

2. Aufgaben im Unterrichtsgespräch (mündliche Fragen)

In diese von Dubs (1978) ausführlich betrachteten Gruppe von Aufgaben gehören sämtliche Fragen und Aufforderungen, die vom Lehrer in einem Unterrichtsgespräch gestellt und von den Schülern innerhalb dieses Unterrichtsgesprächs auch unmittelbar bearbeitet und gelöst werden. Sie zeichnen sich dadurch aus, dass sie mündlich gestellt und nicht im Wort-

laut schriftlich fixiert sind. Aufgaben im Unterrichtsgespräch haben eine sehr kurze, meist nur wenige Sekunden umfassende Bearbeitungszeit, und die Schüler erhalten somit nicht die Gelegenheit, sich intensiv und individuell mit der Aufgabenstellung auseinanderzusetzen (Brophy & Good, 1986; Tobin & Capie, 1982). Die Lösung solcher Aufgaben wird nur von einem Teil der Schüler einer Klasse genannt, weswegen auch häufig lediglich ein Teil der Klasse aktiviert wird. Beispiele für Aufgaben dieser Gruppe sind folgende im Unterrichtsgespräch gestellte Fragen: „Und welche Funktion hat der Elektromotor?“, „Beschreibe ´mal die Abbildung!“.

3. Aufgaben in Schülerarbeitsphasen (schriftliche Aufgaben)

In diese Gruppen gehören die Lernaufgaben im eigentlichen Sinne. Sie werden in Schülerarbeitsphasen (vgl. Seidel, 2003) in Einzel-, Partner- oder Gruppenarbeit bearbeitet und haben in irgendeiner Form eine schriftliche Fixierung (vgl. Häußler & Lind, 1998). Bei diesen Aufgaben wird allen Schülern stets eine längere Bearbeitungszeit zur Verfügung gestellt, sie können sich individuell mit der Aufgabenstellung auseinandersetzen. Die Schüler erhalten jedoch meist kein unmittelbares Feedback auf während der Bearbeitung gemachte Fehler oder eingeschlagene Lösungswege. Aufgaben in Schülerarbeitsphasen sind beispielsweise: „Beschrifte die Abbildung auf dem Arbeitsblatt!“, „Bearbeite die Aufgaben 1-4 auf S. 214 des Buches!“

Nur die Aufgaben im Unterrichtsgespräch und in den Schülerarbeitsphasen sind als Lernaufgaben im engeren Sinne anzusehen, da sie im Gegensatz zu den Aufforderungen zum Unterrichtsmanagement einen konkreten Inhaltsbezug aufweisen. Daher werden diese in der folgenden Analyse als Lernaufgaben klassifiziert und somit in der Auswertung berücksichtigt. Dabei ist es von besonderer praktischer Relevanz für die Studie, auch die mündlichen Unterrichtsgesprächsaufgaben zu berücksichtigen: In einigen der aufgezeichneten Unterrichtsstunden wurden keinerlei schriftliche Aufgaben gestellt, sondern lediglich mündliche Fragen. Um auch in diesen Stunden eine Auswahl an Unterrichtsaufgaben für den Leistungstest berücksichtigen zu können, muss demnach auch auf die mündlichen Fragen zurückgegriffen werden.

Identifikation und Klassifikation der Analyseeinheit „Lernaufgabe“

Bevor Analyseeinheiten entsprechend ihrer Anforderungen und Charakteristika mit einem Kategoriensystem eingeschätzt werden können, stellt sich die notwendige Frage nach einer objektiven und reliablen Identifikation dieser Analyseeinheiten (vgl. Bortz & Döring,

2002; Greve & Wentura, 1997). In der vorliegenden Arbeit soll ein möglichst breites Spektrum an im Unterricht eingesetzten Aufgaben analysiert werden, so dass jede inhaltsbezogene Leistungsanforderung als Aufgabe definiert wird (vgl. Kapitel 2.1).

Die formale Anordnung von Leistungsanforderungen in Form von Aufgaben ist abhängig von der Intention derjenigen Person, welche die Aufgaben und die zugehörigen Materialien konstruiert und zusammenstellt. Eine Orientierung an formalen Ordnungsmerkmalen, wie z.B. alpha-numerische Kennzeichnungen oder räumliche Trennungen, gewährleisten daher nicht die notwendige Objektivität und Generalisierbarkeit der Analyse. Für die Identifikation von Aufgaben müssen andere Indikatoren definiert und verwendet werden. Die in der vorliegenden Studie verwendete Möglichkeit, Aufgaben zu identifizieren, ist die Verwendung von Operatoren als Indikatoren. Operatoren sind Handlungsanweisungen bzw. allgemeine Handlungsprogramme (vgl. Seel, 2003), die in Form von Verben in der Imperativform, z.B. „*Nenne...*“, „*Beschrifte...*“, „*Gib...*“, „*Finde heraus...*“, „*Erörtere...*“, „*Nimm Stellung...*“ oder in äquivalenten Formulierungen in Form von Substantiven oder Fragewörtern z.B. „*Beobachtung:*“, „*Erklärung:.....*“, „*Welche...*“, „*Wie...*“ vorkommen.

Unter der Verwendung von Operatoren als Indikatoren wird als Untersuchungseinheit die „Operatoraufgabe“, in Anlehnung an die Aufgabendefinition, definiert. Als Operatoraufgabe zählt jede inhaltsbezogene Leistungsanforderung, die genau einen eigenständigen Operator oder eine einem eigenständigen Operator äquivalente Formulierung enthält. Der Terminus „inhaltsbezogen“ verdeutlicht dabei die notwendige Fokussierung der Aufgabe auf einen fachlichen Inhalt. „Eigenständig“ heißt, dass es sich nicht um eine organisatorische Aufforderung zur Lösung der eigentlichen Aufgabe handelt.

Beispiele für Operatoraufgaben:

- „*Nenne* die Begriffe der einzelnen Bestandteile des Elektromotors.“
- „*Welche* Bedeutung hat der Kommutator für die Funktionsfähigkeit des Elektromotors?“
- „*Informiert* euch im Internet über den Atomunfall in Tschernobyl.“

Beispiele für Aufforderungen, die nicht als Operatoraufgaben eingeschätzt werden:

- „Gehe bitte an die Tafel!“
- „Räumt die Materialien weg!“
- „Ihr solltet in einer halben Stunde fertig sein.“
- „Habt ihr Probleme?“

Für die genaue Vorgehensweise der Identifikation und Klassifikation der Aufgaben sei auf Anhang A und Jatzwauk (in Druck) verwiesen.

Interrater-Reliabilität bei der Aufgabenidentifikation und -klassifikation

Die Zuverlässigkeit der Identifikation von Aufgaben wurde mittels der Interrater-Übereinstimmung überprüft. Hierzu wurden 20% der Videos (entspricht 6 Unterrichtsstunden) von zwei trainierten Kodierern bzgl. der Aufgabenarten (Wissen-, Verstehens-, und Anwendungsaufgaben) doppelt kodiert. Anschließend wurde überprüft, welche Aufgaben von beiden Kodierern identifiziert wurden und welche Aufgaben (oder „scheinbare Aufgaben“) nur von einem der beiden Kodierer erkannt wurden. Aus der Zahl der von beiden gemeinsam identifizierten Aufgaben und der Anzahl der insgesamt identifizierten Aufgaben konnte dann die prozentuale Übereinstimmung als Kennwert berechnet werden. Die Anwendung eines zufallsbereinigten Koeffizienten, wie beispielsweise Cohens Kappa war bei diesem Schritt der Analyse nicht möglich, da bei der Identifikation der Aufgaben die zu vergleichenden Fälle erst mit dem Prozess der Identifizierung erzeugt wurden.

Die Werte der prozentualen Übereinstimmung beinhalten bei einer solchen Vorgehensweise zwangsläufig einen systematischen Fehler, welcher sich jedoch konservativ auswirkt. Da die bei der Berechnung der prozentualen Übereinstimmung berücksichtigten Fälle (Lernaufgaben) erst mit dem Prozess der Identifizierung erzeugt wurden, wurden diejenigen Fälle nicht erfasst, in welchen beide Kodierer sich für „Es liegt keine Aufgabe vor.“ entschieden haben. Somit sind diese nicht in die Berechnung des Kennwertes eingegangen. Daraus folgt, dass die berechnete prozentuale Übereinstimmung bedeutend geringer ausfällt als sie tatsächlich ist.

Die prozentuale Übereinstimmung wurde sowohl getrennt für die Identifikation von Aufgaben in Schülerarbeitsphasen (schriftliche Aufgaben) und von Aufgaben im Unterrichtsgespräch (mündliche Aufgaben) als auch gemeinsam für beide Aufgabengruppen berechnet. Die Kennwerte der Interrater-Übereinstimmung sind in Tabelle 5 angegeben.

Tabelle 5: Interrater-Reliabilität der Aufgabenidentifikation

	Anzahl der Aufgaben	Übereinstimmung in %
Aufgaben in Schülerarbeitsphasen	51	94.1
Aufgaben im Unterrichtsgespräch	123	78.9
Aufgaben in Schülerarbeitsphasen und im Unterrichtsgespräch (zusammen)	174	83.3

Die Übereinstimmung bei der Identifikation der Aufgaben in Unterrichtsgesprächen ist mit ca. 79% geringer als die Übereinstimmung der Aufgaben in Schülerarbeitsphasen

(94%). Dies liegt nicht zuletzt darin begründet, dass bei den Aufgaben in Unterrichtsgesprächen sehr viele Wiederholungen und Ergänzungen auftreten und sich weiterhin insbesondere bei Nachfragen des Lehrers die Abgrenzung der Aufgaben zur Gruppe der „Aufforderungen zum Unterrichtsmanagement“ schwierig gestaltet.

Unter Berücksichtigung des Umstandes, dass es sich um konservativ berechnete Kennwerte handelt, kann festgestellt werden, dass Operatoraufgaben mit hinreichender Objektivität und Reliabilität identifiziert werden können und somit die Grundlage gelegt werden konnte, um die Klassifikation der Aufgaben in Schülerarbeitsphasen anhand der kognitiven Prozesse der Aufgabenbearbeitung vornehmen zu können.

Auf der Grundlage der oben beschriebenen Aufgabenidentifikation erfolgte im Anschluss die Klassifikation der Unterrichtsaufgaben nach den drei kognitiven Prozessen Wissen, Verstehen und Anwenden bei der Aufgabenbearbeitung. In Tabelle 6 sind verschiedene Reliabilitätsmaße für die sechs doppelt kodierten Physikstunden der Stichprobe dargestellt.

Tabelle 6: Verschiedene Interrater-Reliabilitätsmaße zur Aufgabenklassifikation (N = 174 Unterrichtsaufgaben)

Reliabilitätsmaß	Wert	Signifikanz
Cohens Kappa	.850	.001***
Spearman's Rho	.887	.001***
Gamma	.928	.001***
Prozentuale Übereinstimmung	90%	

Anmerkung:

*** Der Wert ist auf dem Niveau von 0.001 signifikant

Als Maß der Beurteilerübereinstimmung wird Cohens Kappa mit 0.850 angegeben, womit dies nach Wirtz und Caspers (2002) als „sehr gute“ Übereinstimmung bezeichnet werden kann. Die Korrelation nach Spearman's Rho ist mit 0.887 ebenfalls als gut bis sehr gut zu beschreiben. Dieses Maß der Beurteiler-Reliabilität wurde in der vorliegenden Untersuchung eingesetzt, da ein ordinales Skalenniveau der „kognitiven Aufgabenanforderung“ angenommen wird. Die kognitiven Prozesse können nur dahingehend abgegrenzt werden, dass eine Rangfolge der Merkmalsausprägungen geben ist, jedoch nicht - wie im Falle intervallskalierteter Merkmale – dass die Unterschiede zwischen den Merkmalsausprägungen gleich groß sind. Als weiteres Maß für ordinalskalierte Daten wird Gamma angegeben, welches ein Maß für schwach monotone Zusammenhänge und mit einem Wert von 0.928 ebenfalls sehr hoch ist. Zur genaueren Unterscheidung zwischen Beurteiler-Reliabilität und -

übereinstimmung sei an dieser Stelle auf Wirtz und Caspers (2002) verwiesen. Auch die prozentuale Übereinstimmung der Rater, als weiteres Reliabilitätsmaß, ist mit 90% ebenfalls ausreichend hoch.

Insgesamt ist somit die Auswertungskategorie *kognitives Anforderungsniveau* der Unterrichtsaufgaben mit Hilfe der hier gewählten Einteilung der kognitiven Prozesse reliabel auszuwerten.

10.1.2 Repräsentativität der wiederholten Unterrichtsaufgaben

Bei diesem Teil der Videoauswertung geht es darum, sicherzustellen, dass die zufällig für die Leistungstests ausgewählten Unterrichtsaufgaben in Bezug auf die kognitiven Prozesse der Aufgabenbearbeitung repräsentativ, für die Gesamtstichprobe der von den Lehrern in der aufgezeichneten Stunde eingesetzten Aufgaben sind.

Tabelle 7 zeigt die prozentualen Anteile der kognitiven Prozesse an der Gesamtheit aller Unterrichtsaufgaben sowie den in den Leistungstests wiederholten Unterrichtsaufgaben.

Tabelle 7: Verteilung der Aufgabenarten der im Unterricht gestellten Aufgaben und der in den Leistungstests wiederholten Unterrichtsaufgaben in Bezug auf die kognitiven Prozesse (für die Gesamtstichprobe von 30 Unterrichtsstunden)

	Aufgabenart		
	Wissensaufgabe	Verstehensaufgabe	Anwendungsaufgabe
Prozent der Unterrichtsaufgaben (100%)	61.4%	32.9%	5.7%
Prozent der wiederholten Unterrichtsaufgaben (100%)	55.7%	32.9%	11.4%

Anhand der Prozentwerte wird deskriptiver betrachtet deutlich, dass es sich bei den in den Leistungstests wiederholten Unterrichtsaufgaben nicht um eine perfekt repräsentative Auswahl der von den Lehrern im Unterricht gestellten Aufgaben handelt. Daher wurde mit einem t-Test für eine Stichprobe überprüft, ob sich die Aufgabenarten im Unterricht signifikant von denen im Leistungstest unterscheiden. Die Aufgabenart wurde dazu folgendermaßen kodiert: Wissensaufgaben wurden mit einer 1, Verstehensaufgaben mit 2 und Anwendungsaufgaben mit einer 3. Hierbei zeigte sich ein signifikanter Unterschied ($T_{(91)} = 2.021$; $p = .046$). Der Mittelwert der Aufgabenart in der Gesamtstichprobe beträgt 1.41 ($SD = .579$) und der Mittelwert der in den Leistungstests wiederholten Aufgaben liegt bei 1.55 ($SD =$

.685). Die wiederholten Unterrichtsaufgaben repräsentieren die Häufigkeit des Einsatzes der verschiedenen Aufgabenarten durch die Lehrer also nicht, in den Leistungstests wurden mehr Aufgaben kognitiv höheren Niveaus, insbesondere Anwendungsaufgaben, wiederholt.

10.1.3 Aufgabeneinsatz im Physikunterricht

Im Folgenden soll deskriptiv untersucht werden, wie viele und welche Art von Aufgaben in den aufgezeichneten Physikstunden von den Lehrern eingesetzt werden (Fragestellung 1).

Einsatz der verschiedenen Aufgabenarten

Die für beide Schulformen gemeinsame Videoauswertung hat gezeigt, dass Lehrer im Mittel 32.4 Aufgaben innerhalb einer 45-minütigen Unterrichtsstunde stellen ($SD= 14.62$). Zwischen den einzelnen Klassen variiert die Anzahl beträchtlich, es werden zwischen 15 und 67 Aufgaben pro Unterrichtsstunde gestellt. Auf die verschiedenen Aufgabenarten verteilt sieht der Einsatz im Unterricht wie folgt aus (Abbildung 3):

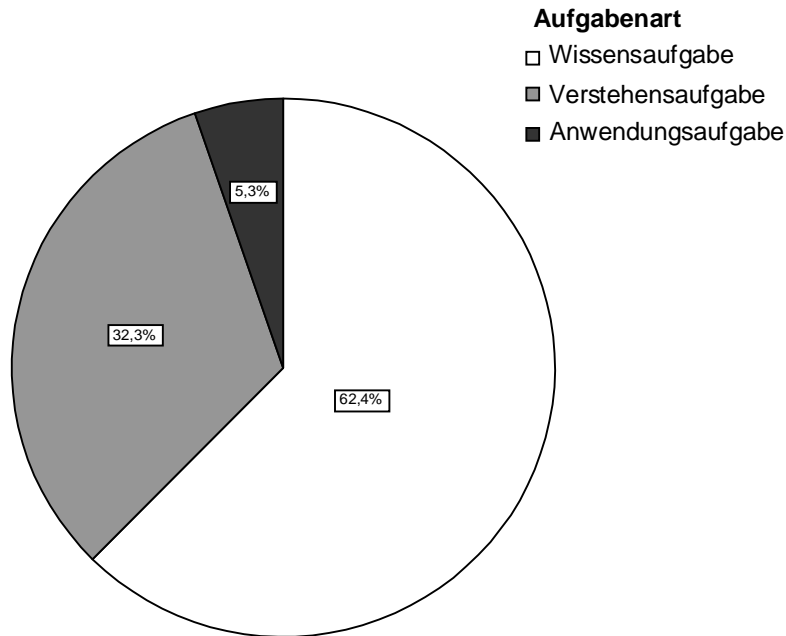


Abbildung 3: Anteil der verschiedenen Aufgabenarten im Unterricht ($N = 872$ Unterrichtsaufgaben)

Deskriptiv betrachtet werden von den untersuchten Lehrern mit Abstand am häufigsten Wissensaufgaben gestellt, im Mittel 20 Aufgaben je Unterrichtsstunde. Wie sich in Tabelle 8 insbesondere anhand der Minima von 4 und Maxima von 50 Wissensaufgaben pro Unterrichtsstunde zeigt, unterscheiden sich die einzelnen Lehrer ganz erheblich im Gebrauch dieser Aufgabenart. Dies wird auch durch die Standardabweichung (SD) von 11.5 verdeutlicht. In Prozent ausgedrückt werden im Mittel 62.4% der in einer Unterrichtsstunde gestellten Aufgaben auf dem Wissensniveau klassifiziert. Im Mittel werden im Vergleich dazu 10.6 Verstehensaufgaben pro Unterrichtsstunde gestellt (SD= 6.1), dies entspricht 32.3% aller Unterrichtsaufgaben. Im Minimum werden 2 und im Maximum 26 Verstehensaufgaben pro Unterrichtsstunde den Schülern gestellt. Am seltensten werden Anwendungsaufgaben gestellt, im Mittel sind dies 1.9 Aufgaben bzw. 5.3% der gesamten Unterrichtsaufgaben (SD= 1.6; min= 0; max= 5). Aufgrund der z.T. beträchtlichen Unterschiede in der Häufigkeit des Aufgabeneinsatzes sind in Anhang B die Anteile der verschiedenen Aufgabenarten pro Lehrer aufgeführt.

Insgesamt werden im Mittel also fast zweimal so viele Wissensaufgaben wie Verstehensaufgaben in den untersuchten Physikstunden gestellt. Zudem werden Anwendungsauf-

gaben in einer fast schon zu vernachlässigenden Anzahl verwendet, mit einem Minimum von 0 Aufgaben in der untersuchten Stichprobe.

Eine übersichtliche Darstellung der statistischen Werte zum Einsatz der verschiedenen Aufgabenarten bezogen auf die Gesamtstichprobe und getrennt für Gymnasien und Hauptschulen findet sich in Tabelle 8.

Tabelle 8: Einsatz der verschiedenen Aufgabenarten im Physikunterricht

	Gesamtanzahl der Aufgaben	Wissens- aufgaben	Verstehens- aufgaben	Anwendungs- aufgaben
Gesamt (Gymnasien und Hauptschulen)				
Mittelwert	32.4	20	10.6	1.9
Anteil in %	100% ¹	62.4%	32.3%	5.3%
Standardabweichung	14.6	11.5	6.1	1.6
Minimum	15	4	2	0
Maximum	67	50	26	5
Gymnasium				
Mittelwert	25.4	15.3	7.5	2.6
Anteil in %	100%	60.2%	29.5%	10.3%
Standardabweichung	7.3	6.9	3.2	1.7
Minimum	17	4	2	0
Maximum	40	26	11	5
Hauptschule				
Mittelwert	39.4	24.7	13.6	1.1
Anteil in %	100%	62.7%	34.5%	2.8%
Standardabweichung	17.0	13.5	6.8	1.1
Minimum	15	8	5	0
Maximum	67	50	26	3

Anmerkung:

1 Die 100% in Klammern beziehen sich auf die folgenden Zeilenangaben der Aufgabenarten

Schulformspezifische Unterschiede im Einsatz der Aufgabenarten

In der nach Schulformen getrennten deskriptiven Betrachtung sind die Einsatzhäufigkeiten der unterschiedlichen Aufgabenarten ähnlich. Jedoch zeigen sich schulformspezifische Unterschiede in Bezug auf die Gesamtanzahl der in einer Stunde gestellten Aufgaben. An den Hauptschulen werden im Mittel 39.4 Aufgaben gestellt und an den Gymnasien im Vergleich dazu 25.4 Aufgaben. In Bezug auf die Aufgabenarten zeigen sich Unterschiede zwischen den Schulformen bei den Verstehensaufgaben dahingehend, dass in den Gymnasien im Mittel 29.5% der Aufgaben auf diesem Niveau gestellt werden, an den Hauptschulen 34.5%. Im Gegensatz dazu werden an den Gymnasien deutlich mehr Anwendungsaufgaben gestellt als an den Hauptschulen (10.3% bzw. 2.8% aller Unterrichtsaufgaben).

Rechnet man die mittlere Anzahl der gestellten Unterrichtsaufgaben, 32.4 in der Gesamtstichprobe, auf die Unterrichtszeit von 45 Minuten um, so bleibt für die Bearbeitung und Beantwortung jeder Aufgabe im Mittel 1.41 Minuten. Diese Berechnung wurde auch von Neubrand (2002) in ihrer Arbeit angestellt und als Indiz für die Aufgabenschwierigkeit dahingehend herangezogen, dass die eingesetzten Aufgaben insgesamt nicht sehr schwierig sein dürften, wenn zu ihrer Bearbeitung und Beantwortung den Schülern lediglich so kurze Zeit zugestanden wird.

Inferenzstatistische Absicherung der Unterschiede

Da bei der Berechnung einer ANOVA mit Messwiederholung zu den Unterschieden in der Häufigkeit des Einsatzes der unterschiedlichen Aufgabenarten an den beiden Schulformen die Voraussetzung der Sphärizität verletzt ist ($\chi^2_{(2)} = 17.244$, $p = .001$) und auch der Box-Test auf Gleichheit der Kovarianzmatrizen signifikant ist ($F_{(6, 4897.8)} = 2.598$, $p = .016$) und daher nicht anzunehmen ist, dass die beobachteten Kovarianzmatrizen der abhängigen Variablen über die Gruppen hinweg gleich sind, werden die Analysen mit Hilfe nicht-parametrischer Verfahren durchgeführt. Zunächst wurde ein möglicher Haupteffekt der Aufgabenart auf Signifikanz geprüft, d.h. inwieweit unterscheiden sich die Mediane der Wissensaufgaben, der Verstehensaufgaben und der Anwendungsaufgaben signifikant voneinander. Dieser nicht-parametrische Vergleich zwischen den drei Aufgabenarten wird mit Hilfe des Friedman-Tests durchgeführt, welcher signifikant ist: Friedman's $\chi^2_{(2)} = 48.881$; $p < .001$, d.h. es liegen signifikante Unterschiede in der Anzahl der eingesetzten Aufgaben zwischen den drei Aufgabenarten vor. Die im Anschluss durchgeführten Schaich-Hamerle-Einzelvergleiche (vgl. Schaich & Hamerle, 1984) zeigen, dass zwischen allen drei Aufgabenarten jeweils signifikante Unterschiede in der Anzahl der im Unterricht eingesetzten Aufgaben bestehen. Diese sind in Tabelle 9 dargestellt.

Tabelle 9: Schaich-Hamerle-Einzelvergleiche zur Anzahl der Aufgabenarten

	Wissensaufgaben	Verstehensaufgaben	Anwendungsaufgaben
Wissensaufgaben	1	D= 0.64; p= 0.049*	D= 1.77; p< 0.001***
Verstehensaufgaben		1	D= 1.13; p< 0.001***
Anwendungsaufgaben			1

Anmerkungen:

* Die Unterschiede sind auf dem 0.05-Niveau signifikant

*** Die Unterschiede sind auf dem 0.001-Niveau signifikant

In einem nächsten Auswertungsschritt wurde der mögliche Effekt der Schulform bzgl. der Häufigkeit des Einsatzes der verschiedenen Aufgabenarten mit Hilfe eines Mann-Whitney-U-Tests überprüft. Die Ergebnisse sind in den Tabellen 10 und 11 dargestellt. Dabei zeigen sich aufgrund der deskriptiven Ergebnisse erwartungskonforme, signifikante Unterschiede in der Anzahl der eingesetzten Verstehensaufgaben im Unterricht dahingehend, dass in den Hauptschulen signifikant mehr Verstehensaufgaben gestellt werden als an den Gymnasien ($Z = -2.061$; $p = .041$). Bei der Anzahl der Anwendungsaufgaben zeigt sich ein gegenteiliger Effekt: An den Gymnasien werden signifikant mehr dieser Aufgaben gestellt als an den Hauptschulen ($Z = -2.588$; $p = .011$).

Tabelle 10: Ränge der verschiedenen Aufgabenarten an Hauptschulen und Gymnasien

	Schultyp	N	Mittlerer Rang	Rangsumme
Anzahl der Wissensaufgaben	Gymnasien	15	13.63	204.50
	Hauptschulen	15	17.37	260.50
Anzahl der Verstehensaufgaben	Gymnasien	15	12.20	183.00
	Hauptschulen	15	18.80	282.00
Anzahl der Anwendungsaufgaben	Gymnasien	15	19.53	293.00
	Hauptschulen	15	11.47	172.00

Tabelle 11: Unterschiede in der Anzahl der verschiedenen Aufgabenarten an Hauptschulen und Gymnasien

	Anzahl der Wissensaufgaben	Anzahl der Verstehens- aufgaben	Anzahl der Anwendungs- aufgaben
Mann-Whitney-U	84.500	63.000	52.000
Wilcoxon-W	204.500	183.000	172.000
Z	-1.165	-2.061*	-2.588**
Asymptotische Signifikanz (2-seitig)	.244	.039	.010
Exakte Signifikanz [2*(1-seitig Sig.)]	.250	.041	.011

Anmerkungen:

** Der Wert ist auf dem Niveau von 0.01 (2-seitig) signifikant.

* Der Wert ist auf dem Niveau von 0.05 (2-seitig) signifikant.

Die nach den Schulformen getrennte Berechnung des Friedman-Tests führt ebenfalls zu signifikanten Unterschieden. An den Gymnasien ist der Wert mit Friedman's $\chi^2(2) = 23.724$; $p < .001$ signifikant, d.h. es liegen Unterschiede in der Anzahl der Aufgabenarten vor. Auch an den Hauptschulen konnten signifikante Unterschiede in der Anzahl der Aufgabenart nachgewiesen werden: Friedman's $\chi^2(2) = 25.200$; $p < .001$. Die im Anschluss durchgeführten Einzelvergleiche sind in Tabelle 12 getrennt nach Schulform dargestellt. Da die Stichprobengrößen gering sind, pro Schulform $N = 15$ Klassen, werden in diesem Fall keine Schaich-Hamerle-Einzelvergleiche, sondern Wilcoxon-Wilcox-Tests durchgeführt.

Tabelle 12: Wilcoxon-Wilcox-Tests zur Anzahl der Aufgabenarten pro Schulform

	Wissensaufgaben	Verstehensaufgaben	Anwendungsaufgaben
Gymnasium			
Wissensaufgaben	1	D= 10.00; $p > 0.05$	D= 26.00; $p < 0.01^{**}$
Verstehensaufgaben		1	D= 16.00; $p < 0.01^{**}$
Anwendungsaufgaben			1
Hauptschule			
Wissensaufgaben	1	D= 9.00; $p > 0.05$	D= 27.00; $p < 0.01^{**}$
Verstehensaufgaben		1	D= 18.00; $p < 0.01^{**}$
Anwendungsaufgaben			1

Anmerkung:

** Die Unterschiede sind auf dem 0.01-Niveau signifikant

Sowohl im Unterricht der untersuchten Gymnasien als auch der Hauptschulen bestehen signifikante Unterschiede in der Häufigkeit des Einsatzes der Wissensaufgaben und der Anwendungsaufgaben, sowie der Verstehens- und Anwendungsaufgaben. Im Gegensatz dazu bestehen keine signifikanten Unterschiede in der Häufigkeit des Einsatzes der Wissens- und Verstehensaufgaben. Die mittleren Ränge der Einsatzhäufigkeiten der Aufgabenarten für die Schulformen werden mit Hilfe der Abbildung 4 verdeutlicht.

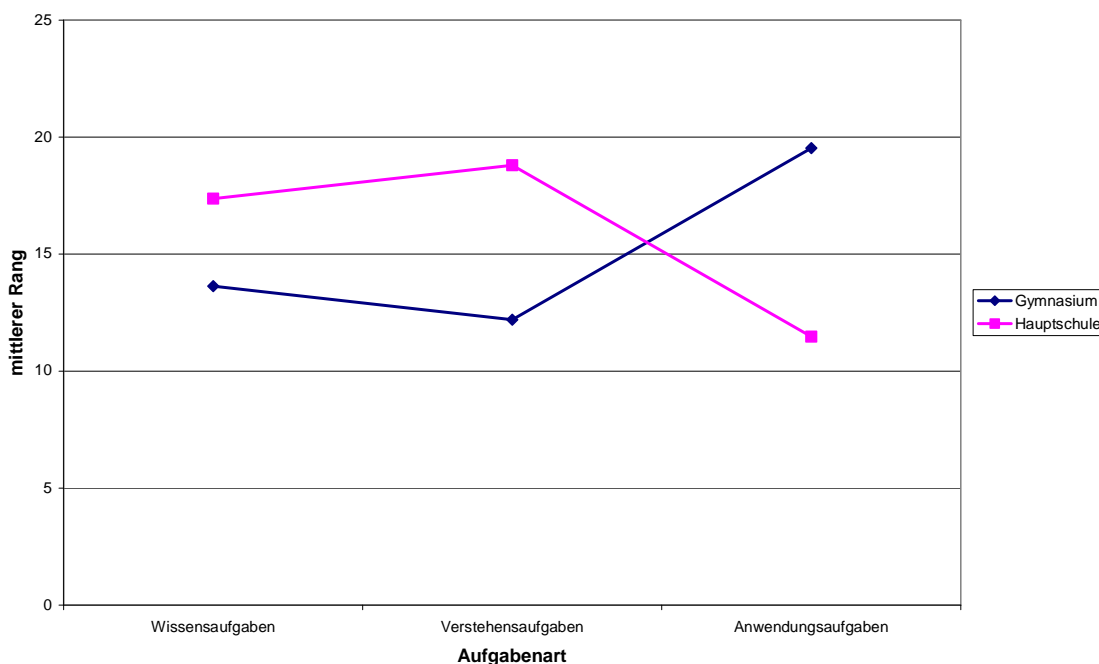


Abbildung 4: Mittlere Ränge des Einsatzes der Aufgabenarten an den Schulformen

Mündliche vs. schriftliche Aufgaben

Tabelle 13 zeigt, dass im Vergleich zwischen mündlichen und schriftlichen Unterrichtsaufgaben ein deutlicher Unterschied in der Häufigkeit dieses Aufgabenformats besteht: Die allermeisten gestellten Unterrichtsaufgaben werden vom Lehrer mündlich während des Unterrichtsgesprächs gestellt. Diese Aufgabenart macht 90.4% aller Unterrichtsaufgaben aus. Im Gegensatz dazu werden lediglich 9.6% der Aufgaben in schriftlicher Form gestellt (Tabelle 13).

Tabelle 13: Anteil der mündlichen und schriftlichen Aufgaben an Unterrichtsaufgaben

	Gesamtanzahl der Aufgaben	Wissensaufgaben	Verstehens- aufgaben	Anwendungs- aufgaben
Gesamt (Gymnasien und Hauptschulen)				
mündliche A. (%)	90.4% (100%) ¹	62.1%	32.3%	5.6%
schriftliche A. (%)	9.6% (100%)	55.0%	38.3%	6.7%
Gymnasium				
mündliche A. (%)	89.5% (100%)	63.1%	25.8%	11.1%
schriftliche A. (%)	10.5% (100%)	43.5%	51.4%	5.1%
Hauptschule				
mündliche A. (%)	93.1% (100%)	61.5%	36.1%	2.4%
schriftliche A. (%)	6.9% (100%)	73.9%	17.4%	8.7%

Anmerkung:

1 Die 100% in Klammern beziehen sich auf die folgenden Zeilenangaben der Aufgabenarten

Werden die jeweils schriftlichen und mündlichen Lernaufgaben in Bezug auf die kognitiven Prozesse der Aufgabenbearbeitung differenziert betrachtet, ergibt sich ein ähnliches Bild (Abbildung 5 und 6) wie bei der gemeinsamen Betrachtung. Mit Abstand am häufigsten werden sowohl mündlich wie schriftlich Wissensaufgaben (mündlich: 62.1% bzw. schriftlich: 55%) gestellt, gefolgt von Verstehensaufgaben (32.3% bzw. 38.3%). Lediglich ein kleiner Anteil beider Aufgabenformen stellt Anwendungsaufgaben dar (5.6% bzw. 6.7%).

Beim Vergleich der mündlichen und schriftlichen Aufgaben zwischen den Schulformen zeigt sich, dass an den Gymnasien mehr Verstehens- (51.4%) als Wissensaufgaben (43.5%) in schriftlicher Form von den Lehrern gestellt werden. Im Gegensatz dazu zeigt sich an den Hauptschulen, dass überwiegend Wissensaufgaben als schriftliche Aufgaben (73.9%) im Unterricht gegeben werden. Es werden selten Verstehens- (17.4%) oder Anwendungsaufgaben (8.7%) den Hauptschülern der untersuchten Stichprobe in schriftlicher Form zur Bearbeitung vorgelegt.

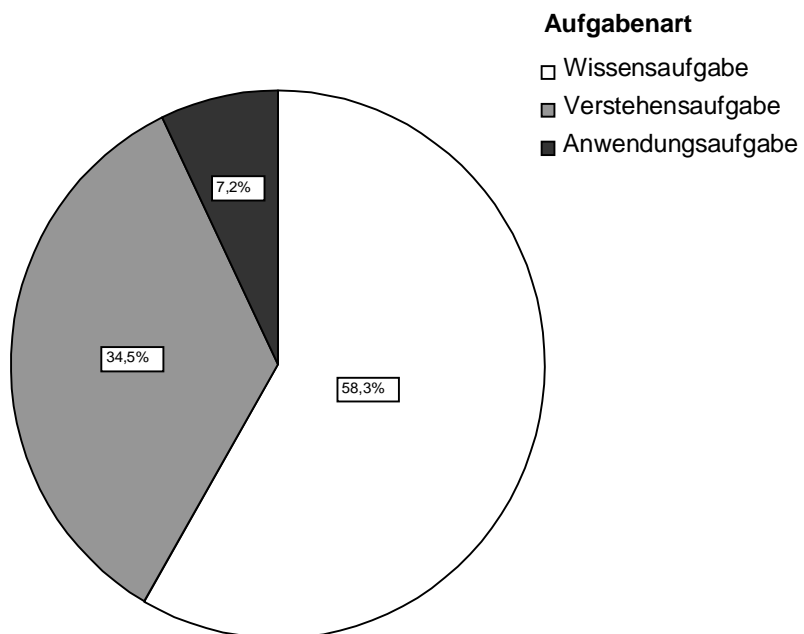


Abbildung 5: Anteile der unterschiedlichen Aufgabenarten bei schriftlichen Unterrichtsaufgaben

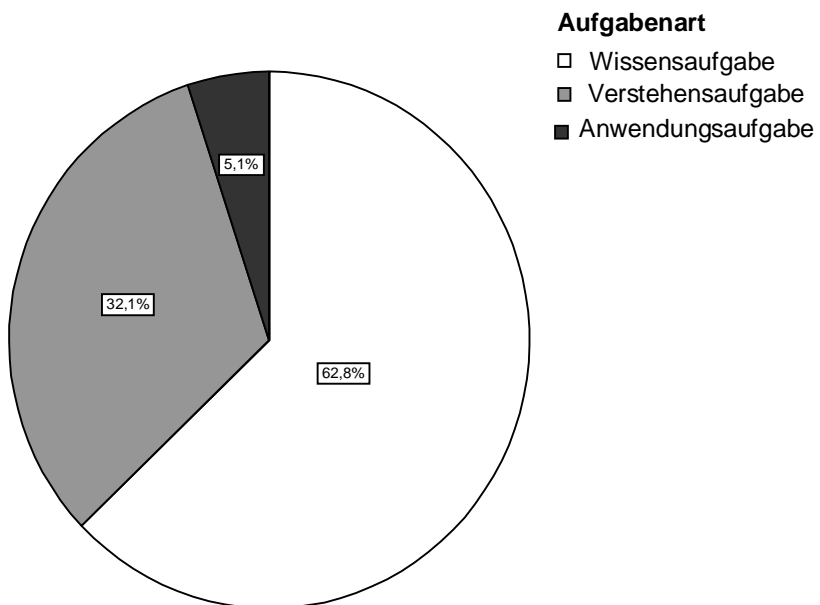


Abbildung 6: Anteile der unterschiedlichen Aufgabenarten bei mündlichen Unterrichtsaufgaben

Die deskriptiven Ergebnisse wurden mit für die Schulformen getrennt berechneten t-Tests für unabhängige Stichproben überprüft. Die deskriptiven Ergebnisse sind in der Tabelle 14 dargestellt. Die Aufgabenart wurden folgendermaßen kodiert: Wissensaufgaben erhalten eine 1, Verstehensaufgaben eine 2 und Anwendungsaufgaben eine 3. In der Analyse zeigt sich, dass sich weder an den Gymnasien ($T_{(373)} < 1$), noch an den Hauptschulen ($T_{(34.55)} < 1$) die Häufigkeiten der eingesetzten Aufgabenarten signifikant zwischen mündlich und schriftlich gestellten Aufgaben unterscheiden. An den Hauptschulen sind die Varianzen der Aufgabenarten ungleich.

Tabelle 14: Häufigkeiten der Aufgabenarten nach Schulform und Aufgabenformat

Schulform	Aufgabenformat	Mittelwert	Standardabweichung	N
Gymnasium	schriftlich	1.51	.579	51
	mündlich	1.46	.664	324
Hauptschule	schriftlich	1.45	.711	33
	mündlich	1.40	.528	464

Die Mittelwerte des Einsatzes der Aufgabenarten für schriftliche und mündliche Aufgaben sind für die Gymnasialklassen in Abbildung 7, für die Hauptschulen in Abbildung 8 dargestellt. In beiden Abbildungen wird deutlich, dass die Mittelwerte der verschiedenen Aufgabenarten sowohl für die Schulformen, als auch für die Aufgabenformate (schriftlich vs. mündlich) ähnlich verlaufen. Wissensaufgaben werden am häufigsten gestellt, gefolgt von den Verstehensaufgaben. In allen Fällen werden fast keine Anwendungsaufgaben gestellt.

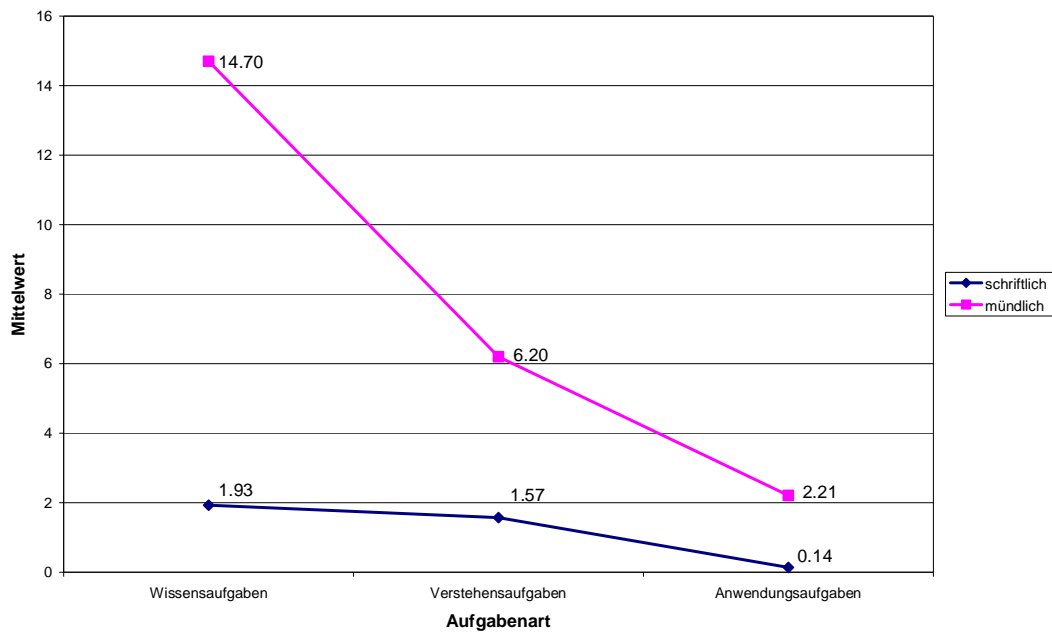


Abbildung 7: Mittelwerte der Aufgabenart bei mündlichen und schriftlichen Aufgaben an Gymnasien

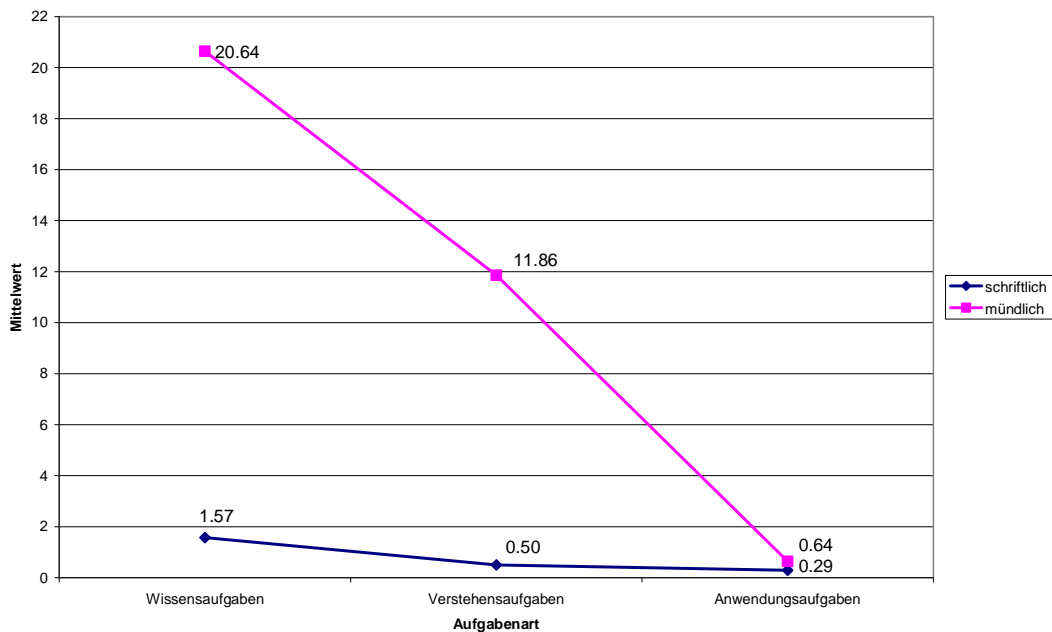


Abbildung 8: Mittelwert der Aufgabenart bei mündlichen und schriftlichen Aufgaben an den Hauptschulen

Zusammenfassung der Ergebnisse

Die Ergebnisse der Videoauswertung bzgl. der im Unterricht eingesetzten Aufgaben lassen sich für beide Schulformen dahingehend zusammenfassen, dass weit mehr der Unterrichtsaufgaben in mündlicher (90.3%) als in schriftlicher Form (9.7%) gestellt werden. Insgesamt herrscht in den untersuchten Klassen der fragend-entwickelnde Unterricht vor.

Die Aufgaben werden des Weiteren meist auf kognitiv niedrigem Niveau gestellt: Es hat sich für die Gesamtstichprobe gezeigt, dass Wissensaufgaben mit Abstand am häufigsten im Unterricht (62.4%) eingesetzt werden, gefolgt von den Verstehens- (32.3%) und Anwendungsaufgaben (5.3%). Weiter wird deutlich, dass es zwischen den einzelnen Lehrern große Unterschiede sowohl in der Gesamtanzahl der in einer Unterrichtsstunde eingesetzten Aufgaben als auch in der Anzahl der unterschiedlichen Aufgabenarten gibt (vgl. Anhang B).

Neben diesen Ergebnissen der Gesamtstichprobe wurden die Unterrichtsvideos schulformspezifisch ausgewertet, wobei sich ähnliche Ergebnisse zeigen wie bei der Betrachtung der Gesamtstichprobe. Es gibt jedoch folgende bedeutsame schulformspezifische Unterschiede: An den Hauptschulen werden insgesamt signifikant mehr Aufgaben im Unterricht gestellt ($M= 39.4$) als in Gymnasien ($M= 25.4$). Was den Einsatz der unterschiedlichen Aufgabenarten bzgl. der kognitiven Prozesse betrifft, werden an Gymnasien signifikant mehr Anwendungsaufgaben eingesetzt als an den Hauptschulen; mit 10.3% der Gesamtanzahl der Unterrichtsaufgaben an den Gymnasien und 2.8% an den Hauptschulen. An Hauptschulen werden demgegenüber im Physikunterricht von den Lehrern signifikant mehr Verstehensaufgaben (34.5% aller Unterrichtsaufgaben) eingesetzt als an Gymnasien (29.5%).

Die dargestellten Ergebnisse weisen empirische Evidenz für die, zu der Fragestellung 1 aufgestellten, Hypothese nach, da der Großteil der von den Lehrern eingesetzten Unterrichtsaufgaben erwartungsgemäß dem Wissensniveau zugeordnet werden kann und deutlich weniger Aufgaben auf dem Verstehens- und Anwendungsniveau gestellt werden.

10.2 Deskriptive Auswertung der Leistungstests

Die folgende Analyse dient der Überprüfung, wie hoch die Schwierigkeit der in den Leistungstests eingesetzten Aufgaben ist. Dabei stellt die relative Häufigkeit, mit der eine Aufgabe gelöst wurde, den Indikator dar. Ferner wird untersucht, ob die neu konstruierten Aufgaben in Abhängigkeit von der Aufgabenart tatsächlich die postulierten Schwierigkeitsunterschiede aufweisen. Hierüber kann geprüft werden, ob die a priori zugewiesene Aufgabenschwierigkeit empirisch bestätigt werden kann. Dazu wurden mit Hilfe der klassischen Testverfahren die durchschnittlichen Lösungshäufigkeiten über alle Aufgaben, welche zu

einer Aufgabenart gehören, gemittelt. Die Ergebnisse der Berechnung dieses klassische Schwierigkeitsmaß (vgl. Kapitel 3.2 und 9.6) sind in Tabelle 15, getrennt nach Schulform, dargestellt.

Tabelle 15: Durchschnittliche Lösungshäufigkeit der verschiedenen Aufgabenarten in den Leistungstests (N = 340 Aufgaben)

Aufgabenart	Gymnasien			Hauptschulen		
	Anzahl der Aufgaben	M	SD	Anzahl der Aufgaben	M	SD
Unterrichtsaufgaben	45	0.30	0.33	45	0.35	0.28
Wissensaufgaben	45	0.47	0.33	56	0.41	0.32
Verstehensaufgaben	38	0.27	0.25	48	0.29	0.29
Anwendungsaufgaben	26	0.28	0.25	28	0.17	0.29
Verstehens- und Anwendungsaufgaben	64	0.27	0.23	79	0.23	0.23
TIMSS-Items	6	0.60	0.27	3	0.72	0.21

Anmerkungen:

M= mittlere Lösungshäufigkeit als Prozentsatz richtiger Antworten

SD= Standardabweichung

Die in den Leistungstests wiederholten Unterrichtsaufgaben werden an Gymnasien im Mittel von 30% und an Hauptschulen von 34% der Schüler im Leistungstest richtig gelöst (Tabelle 15). Die TIMSS-Items sind für die Schüler die leichtesten Aufgaben; dies ist erwartungsgemäß, da sie Inhalte des 8. Schuljahrs thematisieren, welche die Schüler der untersuchten Stichprobe bereits gelernt hatten. Die drei ausschließlich in den Hauptschulen eingesetzten TIMSS-Items weisen im Übrigen eine numerisch höhere Lösungshäufigkeit auf und sind dementsprechend leichter als die drei ausschließlich in den Gymnasien eingesetzten TIMSS-Items. Dadurch lässt sich die höhere, relative Lösungshäufigkeit der TIMSS-Items an den untersuchten Hauptschulen im Vergleich zu den Gymnasien erklären. Die konstruierten Wissensaufgaben sind die im Mittel am zweithäufigsten gelösten Testaufgaben. Am schwierigsten sind erwartungsgemäß die konstruierten Verstehens- und Anwendungsaufgaben. Diese beiden Aufgabenarten werden in den weiteren Berechnungen zusammengefasst betrachtet, da die Anwendungsaufgaben nur einen sehr geringen Anteil an der Gesamtheit der Testaufgaben, pro Leistungstest zwei Anwendungsaufgaben, ausmachen und lediglich von einer sehr geringen Anzahl an Schülern richtig gelöst werden.

In Abbildung 9 werden die für die beiden Schulformen gemittelten Lösungshäufigkeiten dargestellt, wobei der ähnliche Verlauf der mittleren Lösungshäufigkeiten über die verschiedenen Aufgabenarten für die beiden Schulformen deutlich wird.

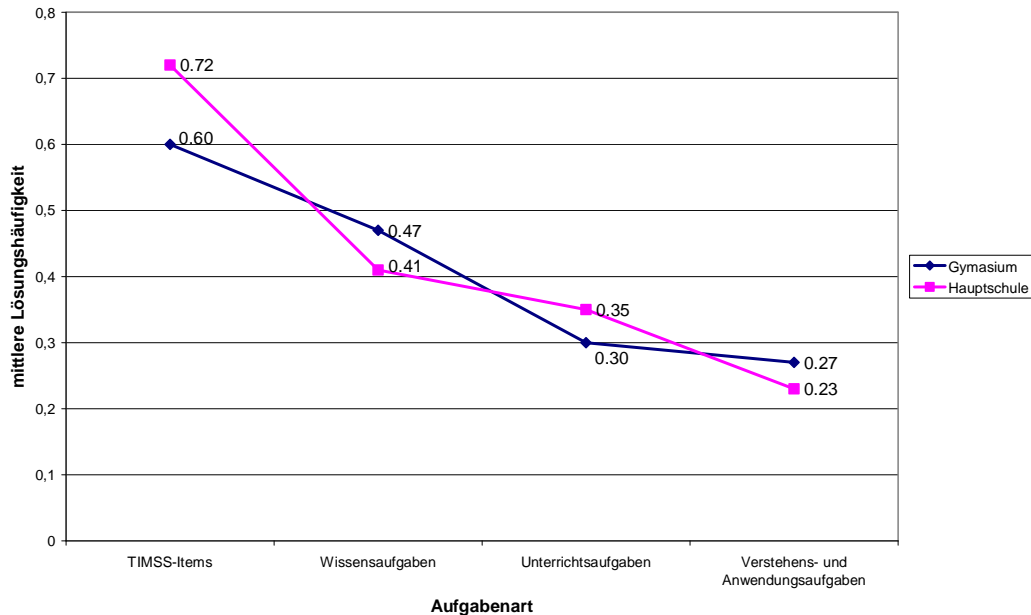


Abbildung 9: Relative Lösungshäufigkeiten der verschiedenen Aufgabenarten für beide Schulformen

Es wurde eine zweifaktorielle Varianzanalyse berechnet. Die relativen Lösungshäufigkeiten der vier Aufgabenarten (TIMSS-Items, Wissens-, Unterrichtsaufgaben, sowie Verstehens- und Anwendungsaufgaben zusammengefasst) stellen die abhängige Variable dar, somit sind die 201 nach Aufgabenart und Schulform klassifizierten Aufgaben die Analyseinheit. Die Aufgabenart und die Schulform dabei die unabhängigen Variablen. Die Analyse bestätigt die bedeutsamen Unterschiede in der Lösungshäufigkeit zwischen den Aufgabenarten. Es zeigt sich ein signifikanter Effekt der Aufgabenart ($F_{(3, 193)} = 24.293$; $p < .001$; $\eta^2 = .274$). Im Gegensatz dazu zeigt sich kein signifikanter Effekt der Schulform ($F_{(1, 193)} = 2.193$; $p = .140$; $\eta^2 = .011$). Auch die Interaktion zwischen Aufgabenart und Schulform ist nicht signifikant: $F_{(3, 193)} = 1.633$; $p = .183$; $\eta^2 = .025$, vgl. Tabellen 16 und 17. Somit konnten die in Abbildung 9 ähnlichen Verläufe der mittleren Lösungshäufigkeiten der verschiedenen Aufgabenarten für die beiden Schulformen bestätigt werden.

Tabelle 16: Relative Lösungshäufigkeiten der Testaufgaben in Abhängigkeit von Aufgabenart und Schulform

Schultyp	Aufgabenart	Mittelwert	Standardabweichung	N
Gymnasium	TIMSS	.601	.204	6
	Wissensaufgaben	.468	.201	25
	Unterrichtsaufgaben	.304	.276	31
	Verstehens-, Anwendungsaufgaben	.270	.223	52
	Gesamt	.414	.258	114
Hauptschule	TIMSS	.724	.061	3
	Wissensaufgaben	.413	.221	20
	Unterrichtsaufgaben	.349	.205	30
	Verstehens-, Anwendungsaufgaben	.232	.145	34
	Gesamt	.430	.232	87

Tabelle 17: Effekte der Schulform und Aufgabenart auf die relative Lösungshäufigkeit

Quelle	Quadratsumme vom Typ III	df	Varianz	F	Signifikanz	Partielles Eta-Quadrat
Schulform	.099	1	.099	2.193	.140	.011
Aufgabenart	3.292	3	1.097	24.293***	.001	.274
Schulform * Aufgabenart	.221	3	.074	1.633	.183	.025
Fehler	8.717	193	.045			
Gesamt	42.846	201	.021			

Anmerkung:

*** Der Wert ist auf dem Niveau von 0.001 (2-seitig) signifikant

Die Berechnung eines polynomialen Kontrasts weist den in Abbildung 9 ersichtlichen linearen Trend der relativen Lösungshäufigkeit als statistisch signifikant aus: Der Wert des Kontrasts beträgt $-.325$; er ist signifikant mit $p < .001$ und liegt in einem Konfidenzintervall zwischen $-.422$ und $-.227$.

Insgesamt lässt sich festhalten, dass die relative Lösungshäufigkeit der wiederholten Unterrichtsaufgaben mit einem Mittelwert von 0.30 an den Gymnasien bzw. 0.35 an den

Hauptschulen unerwartet gering ist. Diese Lösungshäufigkeiten unterscheiden sich in bedeutsamer Weise von den Lösungshäufigkeiten der anderen Aufgabenarten im Leistungstest. Die Ergebnisse können als erste Hinweise für die Gültigkeit der hier eingesetzten Methode zur Konstruktion von Aufgaben mit a priori festgelegter, unterschiedlicher Schwierigkeit angesehen werden.

Des Weiteren zeigte sich kein signifikanter Effekt der Schulform bzgl. der relativen Lösungshäufigkeiten der verschiedenen Aufgabenarten. Dieser Befund ist im Vergleich zu den Ergebnissen zur naturwissenschaftlichen Kompetenz der PISA-Untersuchungen (vgl. Prenzel et al., 2001; Rost et al., 2004) überraschend, da der dort berichtete Unterschied in der Leistungsfähigkeit der Schüler der Gymnasien im Vergleich zu der Leistungsfähigkeit der Hauptschüler fast zwei Standardabweichungen (1.8 SD) beträgt. Daher muss das Ergebnis mit Hilfe der Itemparameter der Rasch-Skalierung weitergehend betrachtet werden. Möglicherweise zeigen sich in Bezug auf die Rasch-Schwierigkeitsparameter bedeutsame Unterschiede dahingehend, dass die Itemparameter der Unterrichtsaufgaben an den Gymnasien höher sind als die der Hauptschulen, was in Einklang mit den oben erwähnten Ergebnissen der PISA-Erhebungen stehen würde. Um also Vergleiche der Aufgaben bzw. der Aufgabenschwierigkeiten, trotz der unterschiedlichen in den Aufgaben thematisierten Inhaltsbereiche vornehmen zu können, erfolgte in einem zweiten Auswertungsschritt der Leistungstests eine Rasch-Skalierung der eingesetzten Lernaufgaben und Schülerfähigkeiten. Dabei ist die Eigenschaft der „spezifischen Objektivität“ besonders relevant. Denn diese Modelleigenschaft ermöglicht es, dass unterschiedliche Schülergruppen, obwohl diese teilweise unterschiedliche Aufgaben bearbeiten haben, in ihren Fähigkeiten miteinander verglichen werden können. Auch die Schwierigkeiten der verschiedenen Aufgaben können so verglichen werden.

10.3 Raschskalierung der Aufgaben und Personen

Um die zentrale Frage nach dem Zusammenhang zwischen der Schwierigkeit der verschiedenen Aufgabenarten und den Schülerfähigkeiten beantworten zu können, werden zunächst die Ergebnisse der Raschskalierung der neu konstruierten Aufgaben sowie der Unterrichtsaufgaben (Kapitel 10.3.1) dargestellt. Dann werden die geschätzten Personenparameter (10.3.2) beschrieben, bevor die Zusammenführung der beiden Parameter in Kapitel 10.3.3 erfolgt. Bevor auf diese Ergebnisse eingegangen wird, soll die in der Skalierung nicht berücksichtigte Stichprobe skizziert werden.

Exkurs: Die nicht-berücksichtigte Stichprobe

Aufgrund der angewendeten Skalierungskriterien, welche noch genauer beschrieben werden, konnte nicht die gesamte Personen- und Aufgabenstichprobe in den weiteren Berechnungen berücksichtigt werden. Im Verlauf des iterativen Prozesses der Modellschätzung mussten einige Aufgaben und Klassen aus der Berechnung herausgenommen werden. Da für diese Klassen keine Item- und Personenparameter vorliegen, kann diese nicht-berücksichtigte Teilstichprobe lediglich hinsichtlich der klassischen Lösungshäufigkeiten bei den verschiedenen Aufgabenarten sowie hinsichtlich des Aufgabeneinsatzes im Unterricht beschrieben werden. Die Beschreibung ist relevant, um mögliche Unterschiede in den oben beschriebenen Merkmalen zwischen der in der Skalierung berücksichtigten Stichprobe und der nicht-berücksichtigten Stichprobe aufdecken zu können.

Die nicht-berücksichtigte Stichprobe besteht aus insgesamt 10 der 30 untersuchten Klassen, jeweils fünf am Gymnasium und fünf an der Hauptschule. Diese Teilstichprobe setzt sich aus insgesamt 155 Schülern zusammen. Die mittlere Klassengröße beträgt somit 15.5 Schüler und unterscheidet sich von der durchschnittlichen Klassengröße der in der Skalierung berücksichtigten Stichprobe ($M= 24.4$). Diese geringe Klassengröße der nicht-berücksichtigten Stichprobe kann als erster Hinweis darauf angesehen werden, warum diese Klassen nicht in der Skalierung berücksichtigt werden konnten, denn zur Rasch-Skalierung sind große Personenstichproben notwendig, welche die einzelnen Aufgaben bearbeitet haben. Dieses Kriterium scheint in der nicht-berücksichtigten Stichprobe nicht gegeben zu sein.

Der Einsatz der verschiedenen Aufgabenarten im Unterricht der beiden Stichproben ist in Bezug auf die Aufgabenart (Wissens-, Verstehens- und Anwendungsaufgaben) und das Aufgabenformat (mündlich und schriftlich) in Tabelle 18 für die berücksichtigte und die nicht-berücksichtigte Stichprobe dargestellt. Dabei zeigt sich, dass sich der Aufgabeneinsatz in den beiden Stichproben nicht bedeutsam unterscheidet.

Tabelle 18: Prozentuale Anteile der Aufgabenart und des Aufgabenformats im Unterricht bei der berücksichtigten und der nicht-berücksichtigten Stichprobe

	Berücksichtigte Stichprobe	Nicht-berücksichtigte Stichprobe
Aufgabenart		
Wissensaufgaben	61.7%	64.3%
Verstehensaufgaben	32.6%	31.7%
Anwendungsaufgaben	5.7%	4.0%
Aufgabenformat		
schriftlich	90.7%	89.3%
mündlich	9.3%	10.7%

Abbildung 10 zeigt die mittleren, relativen Lösungshäufigkeiten der verschiedenen Aufgabenarten, welche in den Leistungstests gestellt wurden, pro Schulform (Gymnasium und Hauptschule) jeweils für die 5 nicht-berücksichtigten Klassen sowie die 10 berücksichtigten Klassen. Dabei wird deutlich, dass die nicht-berücksichtigte Gymnasialstichprobe höhere Lösungshäufigkeiten aufweist als die in der Skalierung berücksichtigte Stichprobe am Gymnasium. Im Gegensatz dazu zeigen sich bei der nicht-berücksichtigten Hauptschulstichprobe geringere Lösungshäufigkeiten als in den 10 berücksichtigten Hauptschulklassen.

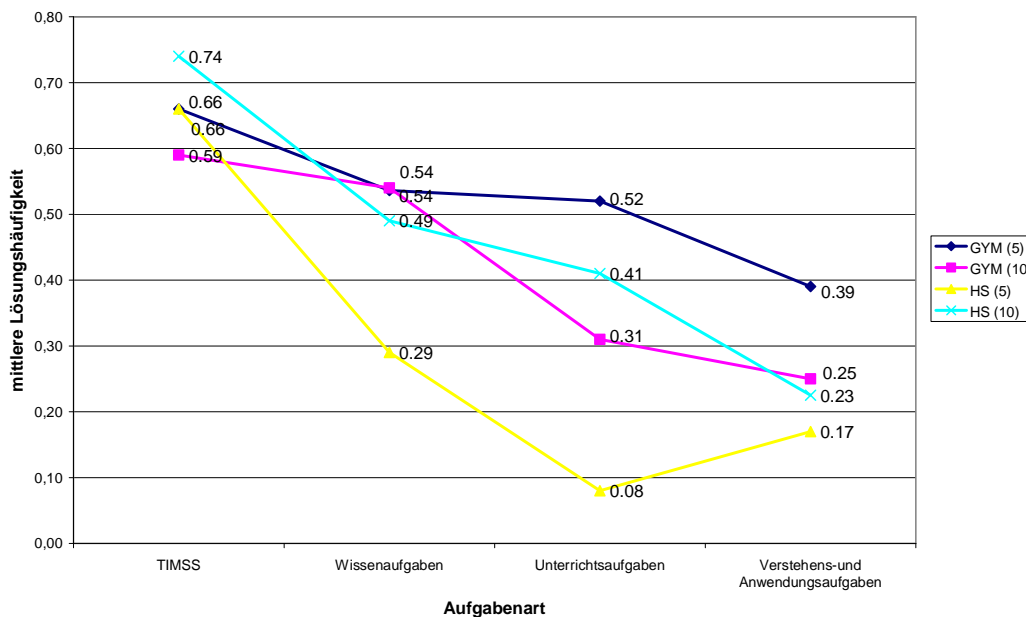


Abbildung 10: Relative Lösungshäufigkeit der verschiedenen Aufgabenarten für die verschiedenen Stichproben

Insgesamt kann aufgrund der Beschreibung der nicht-berücksichtigten Stichprobe und der in der Skalierung berücksichtigten Stichprobe davon ausgegangen werden, dass diese sich in Bezug auf den Aufgabeneinsatz im Unterricht und die relativen Lösungshäufigkeiten der verschiedenen Aufgabenarten nicht bedeutsam unterscheiden.

10.3.1 Schätzung der Itemparameter

Für die Auswertung der Daten kam, neben der statistischen Standardsoftware *SPSS*, das Spezialprogramm *ACER ConQuest* zum Einsatz, das z.B. auch bei der Skalierung der PISA-Daten verwendet wird (vgl. Wu, Adams & Wilson, 1998). Bei diesem Programm handelt es sich um eine Skalierungssoftware zur Modellierung von *Item-Response*-Modellen, womit in der vorliegenden Arbeit die Item- und Personenparameter geschätzt wurden. Die Rasch-Skalierung der Aufgaben wird in zwei Schritten durchgeführt: Zunächst erfolgt die Schätzung der Aufgabenschwierigkeiten. Im Anschluss wurden die Fähigkeitsparameter der Schüler geschätzt.

Bei der Skalierung der Items wurde so vorgegangen, dass mit einer kleinen Stichprobe von Items und Personen, welche diese bearbeitet hatten, die IRT-Skalierung begonnen wurde. In einem zweiten Schritt wurden die Itemparameter auf die Werte fixiert, welche sie in der vorangegangenen Skalierung erhalten hatten, und es wurden dann weitere Items in der neuen Skalierung berücksichtigt, wobei lediglich die Parameter dieser neu berücksichtigten

Items frei geschätzt wurden. Dieses iterative Schätzverfahren wurde so lange wiederholt, bis möglichst viele der Aufgaben, sowie ein Großteil der Schülerstichprobe in der Skalierung Berücksichtigung gefunden hatten. Nach der Itemskalierung konnten den Schülern basierend auf den jeweils bearbeiteten Aufgaben Personenfähigkeitswerte zugewiesen werden.

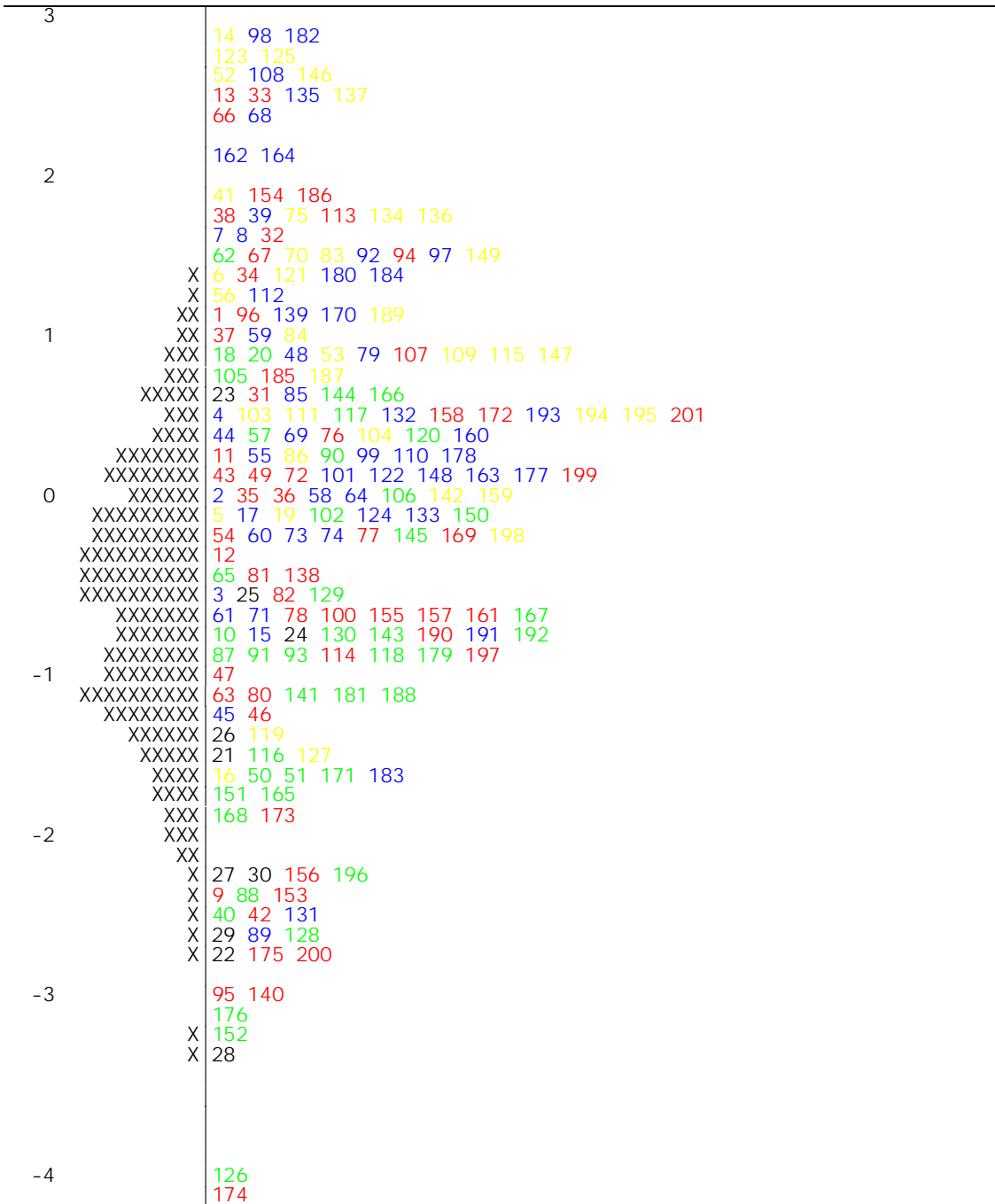
Die Skalierung erbrachte für 20 der insgesamt 30 untersuchten Klassen hinreichend viele eindimensional IRT-skalierbare Aufgaben, um die Fähigkeitsparameter der Schüler bestimmen und über unterschiedliche Inhaltsbereiche hinweg vergleichen zu können. Im Ergebnis konnten immerhin 201 Items (von insgesamt 340 Items) und 487 Schüler (von insgesamt 642 Schülern der Gesamtstichprobe) berücksichtigt werden. Das ist bemerkenswert angesichts der kleinen Anzahl von Schülern, welche die einzelnen Items bearbeitet haben. In der Literatur werden zur Parameterschätzung teilweise Stichprobengrößen von 1000 Personen als minimal und 2000 als adäquat empfohlen (z.B. Green et al., 1984). Lord (1980) konnte im Gegensatz dazu jedoch zeigen, dass kleinere Stichproben ($N = 150-200$) hinreichend für die Schätzung eines eindimensionalen Rasch-Modells sind. Doch auch diese Stichprobengrößen sind in der vorliegenden Untersuchung aufgrund der großen Variation bzgl. des im Unterricht behandelten Inhalts nicht realistisch. Insbesondere an den Hauptschulen wurden kleine Klassen vorgefunden, im Mittel waren 17 Schüler in einer Hauptschulklasse. Somit wurden die inhaltspezifischen Aufgaben z.T. nur von wenigen Schülern bearbeitet.

Zur Schätzung der Fähigkeitsparameter wurden lediglich Klassen berücksichtigt, in denen mindestens drei Unterrichtsaufgaben der jeweiligen Klasse skaliert werden konnten. Des Weiteren mussten mindestens sieben neu konstruierte Aufgaben skalierbar sein, so dass die Fähigkeitsparameter der Schüler in der jeweiligen Klasse auf der Grundlage von insgesamt mindestens 10 Aufgaben geschätzt werden konnten. Die inhaltspezifischen Aufgaben wurden in den meisten Fällen lediglich von den Schülern einer Klasse bearbeitet, so dass diese Aufgaben im Mittel von 24 Schülern bearbeitet wurden.

Von den in der skalierten Stichprobe verbliebenen Schülern besuchten 312 (64.1%) das Gymnasium und 175 (35.9%) die Hauptschule. Diese etwas ungleichen Anteile der Schulformen sind in ähnlicher Form in der Gesamtstichprobe gegeben (Gymnasium: 60.9%; Hauptschule: 39.1%). 43.8% der Schüler am Gymnasium sind Jungen und 56.2% sind Mädchen. An der Hauptschule sind 55.1% der Schüler Jungen und 44.9% Mädchen. Das durchschnittliche Alter der Gymnasiasten betrug 15.8 Jahre und das der Hauptschüler 16.2 Jahre. Auch diese Verhältnisse bzw. das Durchschnittsalter spiegeln die Gesamtstichprobe ausreichend wider (vgl. Kapitel 9.1).

Der Großteil der 20 Physiklehrer der untersuchten Stichprobe ist männlich (17 bzw. 85%), lediglich 3 Lehrerinnen (15%) sind in der Stichprobe vertreten. Das Durchschnittsalter beträgt 46 Jahre ($SD= 10.0$). Im Mittel waren die Lehrer zum Zeitpunkt der Untersuchung seit 19.1 Jahre im Schuldienst.

Abbildung 11 zeigt das Ergebnis des Skalierungslaufs. Am linken Abbildungsrand ist die Skalierungsmetrik in Logit-Einheiten (logit-transformierte Lösungswahrscheinlichkeiten) dargestellt, auf der die geschätzten Personenfähigkeiten und Aufgabenschwierigkeiten abgetragen sind. Die latente Verteilung der Personenfähigkeitsparameter ist durch die ‚X‘-Symbole dargestellt; die einzelnen Aufgaben sind durch Zahlen von 1-201 gekennzeichnet und entsprechend ihres Itemparameters angeordnet.

**Anmerkung:**

Jedes X repräsentiert 2.8 Fälle.

Abbildung 11: Skalierung der Aufgaben- und Personenstichprobe

Die Aufgabenschwierigkeiten aller in der Skalierung berücksichtigten Aufgaben streuen zwischen -4 und knapp 3 Logits, mit einem auf 0.000 fixierten Mittelwert und einer Standardabweichung von 1.49. Zur besseren Anschaulichkeit sind die verschiedenen Aufgabenarten mit Hilfe unterschiedlicher Farben kenntlich gemacht worden. Bei der Betrachtung der Verteilung der Aufgabenarten nach ihrer Schwierigkeit zeigt sich, dass die Verteilung insgesamt erwartungskonform ist. Die a priori als schwer eingeschätzten Anwendungsaufgaben (gelb) finden sich vermehrt im oberen Bereich der Verteilung, versetzt zu diesen zeigen die meisten Verstehensaufgaben (blau) ebenfalls mittlere bis höhere Logit-Werte. Die grünen Wissensaufgaben befinden sich eher im unteren Bereich der Verteilung. Die wiederholten Unterrichtsaufgaben (rot) streuen stark in ihrer Schwierigkeit. Dies wird durch die über die gesamte Verteilung variierende Lokalisation verdeutlicht. Ein Großteil der Unterrichtsaufgaben befindet sich jedoch eher im unteren und mittleren Bereich der Schwierigkeitsverteilung.

Die Personenparameter der untersuchten Schüler variieren ebenfalls stark; sie liegen im Bereich zwischen -3.2 der leistungsschwächsten Schüler und 1.3 Logits der leistungstärksten Schüler. Der Mittelwert beträgt -0.61 Logits mit einer Standardabweichung von 1.07.

Zur Beurteilung der Modellgüte der Skalierung lassen sich zwei Statistiken heranziehen, die zeigen, wie gut die empirischen Daten und die Annahmen des theoretischen Modells zueinander passen. Für das Gesamtmodell wird von dem Programm *ACER ConQuest* zu diesem Zweck die *final deviance* als Maß der globalen Anpassungsgüte ausgegeben. Diese Statistik nimmt für das hier gerechnete einparametrische logistische Modell einen Wert von 9306.62 an und kann im Vergleich mit anderen Modellschätzungen und deren *Deviance*-Werten betrachtet werden. Die *Deviance*-Werte der vorherigen Modellschätzungen im Verlauf des iterativen Skalierungsprozesses betragen 9350.29 bzw. 9418.89. Da diese Werte im oben beschriebenen Vorgehen der Modellschätzungen geringer werden, kann die hier dargestellte Schätzung als die mit der besseren Modellgüte angesehen werden.

Aufschlussreicher sind dem globalen Maß der *final deviance* gegenüber die so genannten Fit-Statistiken auf Aufgabenebene. Diese Statistiken zeigen an, ob die einzelnen Testaufgaben mit den Annahmen des Modells übereinstimmen. Die hier berichteten Fit-Statistiken (*infit*) zählen zu den so genannten Residualmaßzahlen und beruhen auf der Differenz zwischen den aus den empirischen Daten berechneten Lösungshäufigkeiten und den nach dem Modell zu erwartenden Lösungswahrscheinlichkeiten (vgl. Rost, 2004). Für die Beurteilung der Modellkonformität einer Aufgabe hat sich vor allem das *Infit*-Maß (*infit mean square*) bewährt. Diese Statistik hat einen Erwartungswert von 1. Beidseitige Abwei-

chungen von 1 zeigen Probleme mit der Modellkonformität der Aufgabe an. Positive Abweichungen weisen auf eine zu geringe Trennschärfe der Aufgabe hin, d.h. die Funktionsverläufe der Aufgaben fallen eher zu flach aus. Negative Abweichungen, die auf eine im Vergleich mit den Modellannahmen zu hohe Trennschärfe der Aufgabe hinweisen, zeichnen sich durch einen zu steilen Funktionsverlauf der Aufgabe aus.

Als Anhaltspunkt dafür, ab welchem Schwellenwert der *Infit* als Indikator für eine problematische Aufgabe anzusehen ist, werden die Vorgaben der PISA-Untersuchung herangezogen. Ein *Infit*-Wert größer als 1.20 und kleiner als 0.80 wurde bei PISA als Indikator für problematische Aufgabeneigenschaften gewertet (vgl. Adams, 2002). Im Bereich zwischen 1.20 und 0.80 liegen 75.6% der 201 in der hier vorliegenden Studie skalierten Aufgaben. Lediglich 24.4% der Aufgaben liegen damit außerhalb dieses Schwellenbereichs, wobei dies auf die z.T. sehr kleinen Schülerstichproben zurückzuführen ist, welche bestimmte Aufgaben bearbeitet haben. Folgt man zudem Adams und Khoo (1996), so liegt der noch akzeptable Bereich des *Infits* zwischen 0.60 und 1.60. In diesen Bereich fallen insgesamt 98.5% der skalierten Aufgaben, so dass nur 1.5% der Aufgaben nach diesen Schwellenwerten als problematisch betrachtet werden müssen. Eine genaue Beschreibung der skalierten Itemparameter ist in Anhang C zu finden, welcher sich auf der beiliegenden CD befindet.

Zusammengefasst ist festzustellen, dass die Skalierung eines großen Teils der in den Leistungstests eingesetzten Aufgaben für den hier verfolgten Zweck hinreichend gut möglich ist und die Ergebnisse des Skalierungslaufs als Grundlage für weitere Datenanalysen genutzt werden können.

Treatment-Check

Als Beleg für die Modellgültigkeit der oben beschriebenen Itemskalierung werden die Ergebnisse mit den mit Hilfe der klassischen Testverfahren berechneten Lösungshäufigkeiten (vgl. Kapitel 10.2) verglichen. Dazu werden gemittelte Werte der Aufgaben- (s) und Personenparameter (f) in die Funktionsgleichung $p = e(f-s)/(1+e(f-s))$ - des einparametrischen Raschmodells eingesetzt, so dass man die theoretisch erwartbare mittlere Lösungswahrscheinlichkeit erhält. Diese berechnete Lösungswahrscheinlichkeit wird dann mit der tatsächlichen relativen Lösungshäufigkeit verglichen. Die Berechnung der theoretischen Lösungswahrscheinlichkeiten wird für die vier Aufgabenarten der skalierten Gesamtstichprobe sowie getrennt für die beiden Schulformen in der Tabelle 19 dargestellt.

Tabelle 19: Mittlere Lösungswahrscheinlichkeiten und Lösungshäufigkeiten der verschiedenen Aufgabenarten

	Gesamtstichprobe		Gymnasium		Hauptschule	
	LW	LH	LW	LH	LW	LH
alle Aufgabenarten	33.0%	33.9%	37.0%	34.7%	30.0%	33.0%
Unterrichtsaufgaben	37.2%	32.5%	34.1%	30.0%	43.7%	35.0%
Wissensaufgaben	53.4%	44.0%	59.5%	47.0%	48.5%	41.0%
Verstehens- und Anwendungsaufgaben	20.3%	25.0%	21.8%	27.0%	20.2%	23.0%

Anmerkungen:

LW = theoretisch berechnete Lösungswahrscheinlichkeit

LH = relative Lösungshäufigkeit als Prozentsatz richtiger Antworten

Aufgrund der geringen Differenzen zwischen den tatsächlichen Lösungshäufigkeiten und den berechneten Lösungswahrscheinlichkeiten kann dieser Vergleich als Hinweis für die Modellgültigkeit angesehen werden. Dabei ist zu beachten, dass die Werte der relativen Lösungshäufigkeiten auf der Grundlage aller in den 30 Leistungstests eingesetzten Aufgaben (N= 340 Aufgaben) berechnet wurden (vgl. Kapitel 10.2), während die theoretischen Lösungswahrscheinlichkeiten auf Grundlage der Itemparameter einer skalierbaren Teilmenge der eingesetzten Aufgaben (N= 201) berechnet wurde, wodurch möglicherweise die Unterschiede zwischen Lösungswahrscheinlichkeit und relativer Lösungshäufigkeit erklärbar sind.

Zur Absicherung der Schwierigkeitsskalierung können auch die Korrelationen zwischen der klassisch berechneten relativen Lösungshäufigkeit und der IRT-skalierten Aufgabenschwierigkeit herangezogen werden. Dabei zeigt sich eine hoch signifikante erwartungskonforme Korrelation aller in der Skalierung berücksichtigten Aufgaben von $r = -.954$ ($p < .001$). Der negative Wert kommt dadurch zustande, dass eine leichte Aufgabe, d.h. eine Aufgabe mit hoher relativer Lösungshäufigkeit, durch einen negativen Itemparameter charakterisiert ist. Gleichfalls zeigt eine Korrelation zwischen den IRT-skalierten Itemparametern der Unterrichtsaufgaben und den korrespondierenden relativen Lösungshäufigkeiten der Unterrichtsaufgaben in die gleiche Richtung ($r = -.461$; $p < .001$). Auch die Korrelationen zwischen den Lösungshäufigkeiten und den Itemschwierigkeiten der jeweiligen neu konstruierten Aufgabenarten sind erwartungskonform signifikant: für die Wissensaufgaben beträgt die Korrelation $r = -.330$; ($p < .001$), für die zusammengefasst betrachteten Verstehens- und Anwendungsaufgaben $r = -.350$ ($p < .001$).

Schwierigkeit der unterschiedlichen Aufgabenarten

Um zu überprüfen, ob sich die verschiedenen Aufgabenarten in ihrem mittleren Schwierigkeitsparameter der IRT-Skalierung in bedeutsamer Weise, möglicherweise auch zwischen den Schulformen, unterscheiden, wird eine zweifaktorielle Varianzanalyse gerechnet. Die Aufgabenart und die Schulform stellen die unabhängigen Variablen dar und die Itemparameter die abhängige Variable dar. Die Ergebnisse sind in den Tabellen 20 und 21 dargestellt. Wie in Tabelle 20 deutlich wird, sind die Itemparameter der verschiedenen Aufgabenarten erwartungskonform: Die Wissensaufgaben sind die leichtesten Aufgaben, gefolgt von den Unterrichtsaufgaben. Die schwierigsten Aufgaben stellen die als eine Aufgabenart zusammengefassten Verstehens- und Anwendungsaufgaben dar. Bei der Berechnung der ANOVA zeigt sich ein signifikanter Haupteffekt der Aufgabenart $F(2, 186) = 20.241$; $p < .001$; $\eta^2 = .180$), d.h. der mittlere Itemparameter unterscheidet sich in bedeutsamer Weise zwischen den Aufgabenarten. Weder die Schulform $F(1, 186) = 1.523$; $p = .219$; $\eta^2 = .008$), noch die Interaktion zwischen Schulform und Aufgabenart sind im Gegensatz dazu signifikant: $F(2, 186) = 1.338$; $p = .265$; $\eta^2 = .014$ (vgl. Tabelle 21).

Tabelle 20: Mittlere Schwierigkeitsparameter der Aufgaben in Abhängigkeit von Schulform und Aufgabenart

Schulform	Aufgabenart	Mittelwert	Standardabweichung	N
Gymnasium	Wissensaufgaben	-.894	1.264	24
	Unterrichtsaufgaben	.149	1.780	31
	Verstehens- und Anwendungsaufgaben	.768	1.368	50
	Gesamt	.205	1.606	105
Hauptschule	Wissensaufgaben	-.741	1.160	19
	Unterrichtsaufgaben	-.550	1.107	30
	Verstehens- und Anwendungsaufgaben	.571	1.014	37
	Gesamt	-.057	1.224	86

Tabelle 21: Effekte der Schulform und Aufgabenart auf den mittleren Schwierigkeitsparameter

Quelle	Quadrat- summe vom Typ III	df	Varianz	F	Signifi- kanz	Partielles Eta-Quadrat
Schulform	2.630	1	2.630	1.523	.219	.008
Aufgabenart	69.900	2	34.950	20.241***	.001	.180
Schulform * Auf- gabenart	4.622	2	2.311	1.338	.265	.014
Fehler	319.431	186	1.727			
Gesamt	400.426	192	2.096			

Anmerkung:

*** Der Wert ist auf dem Niveau von 0.001 (2-seitig) signifikant

Die Unterschiede zwischen den verschiedenen Aufgabenarten werden mit Hilfe der Berechnung eines polynominalen Kontrasts untersucht, um einen möglichen linearen Trend des mittleren Itemparameters zu überprüfen. Dabei wird erwartet, dass die Wissensaufgaben einen geringeren Itemparameter aufweisen als die Unterrichtsaufgaben, da sich in der Analyse des Aufgabeneinsatzes im regulären Unterricht gezeigt hat, dass dieser sich zu einem Großteil aus Wissensaufgaben zusammensetzt, jedoch auch Verstehens- und Anwendungsaufgaben gestellt werden. Daher sollte der mittlere Itemparameter der Unterrichtsaufgaben höher sein als der der Wissensaufgaben, jedoch geringer als der mittlere Itemparameter der Verstehens- und Anwendungsaufgaben. Diese Hypothese wurde mit Hilfe des Kontrasts bestätigt, dessen Wert 1.051 ($p < .001$) beträgt und in einem Konfidenzintervall von .710 bis 1.393 liegt. Der lineare Trend der mittleren Itemschwierigkeit der Aufgabenart wird in Abbildung 12 verdeutlicht.

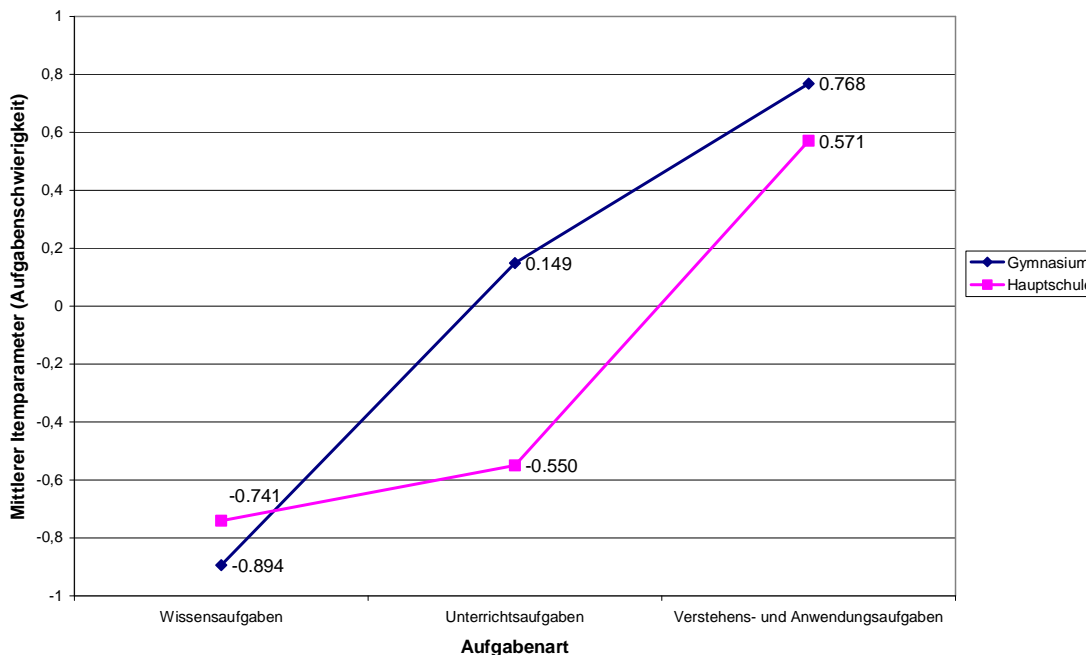


Abbildung 12: Mittlere Schwierigkeitsparameter der unterschiedlichen Aufgabenarten an den Schulformen

Insgesamt sind die Ergebnisse ein weiterer Beleg für die entwickelte Konstruktionsmethode von Aufgaben unterschiedlicher Schwierigkeit. Damit weisen die Ergebnisse empirische Evidenz für die Hypothese 4 nach. Aufgrund der Skalierungsergebnisse der Unterrichtsaufgaben bzgl. der Schwierigkeit, welche einen numerisch eher geringen Parameter aufweisen, kann auch Hypothese 2 bekräftigt werden.

Schulformspezifische Unterschiede der Unterrichtsaufgaben

Dem in der Abbildung 12 deskriptiven Unterschied zwischen den Schulformen bezüglich des mittleren Itemparameters der Unterrichtsaufgaben, welcher für die vorliegende Fragestellung besonders relevant erscheint, soll in der folgenden Auswertung nachgegangen werden. Dazu wurde ein t-Test für unabhängige Stichproben, also zwischen den Schulformen, bzgl. der mittleren Itemschwierigkeiten der - in dieser Auswertung - ausschließlich berücksichtigten Unterrichtsaufgaben berechnet. Die deskriptiven Ergebnisse sind in Tabelle 22 dargestellt.

Tabelle 22: Mittlere Schwierigkeitsparameter der Unterrichtsaufgaben an Gymnasien und Hauptschulen

Schultyp	Mittelwert	Standardabweichung	N
Gymnasium	.149	1.780	31
Hauptschule	-.549	1.107	30
Gesamt	-.169	1.539	61

Der mittlere Schwierigkeitsparameter der Unterrichtsaufgaben unterscheidet sich zwischen den Schulformen signifikant ($T_{(59)} = 1.739$; $p = .044$). Der Effekt der Schulform wurde einseitig getestet, da die gerichtete Hypothese besteht, dass die mittlere Aufgabenschwierigkeit an den Gymnasien höher ist als an den Hauptschulen. An den Hauptschulen sind die von den Lehrern eingesetzten Aufgaben im Mittel leichter als an den Gymnasien. Dort werden im Vergleich mit den Hauptschulen erwartungsgemäß schwierigere Aufgaben gestellt. Die schulformspezifische Hypothese bzgl. der Schwierigkeit der Unterrichtsaufgaben erhält somit aufgrund der Ergebnisse empirische Evidenz.

10.3.2 Schätzung der Personenparameter

Aufgrund der bereits berichteten PISA-Ergebnisse (Prenzel et al., 2001; Rost et al., 2004) wird erwartet, dass es bedeutsame Unterschiede im mittleren Fähigkeitsniveau zwischen den unterschiedlichen Schulformen gibt. Um diese Hypothese empirisch abzusichern, wird eine einfaktorielle Varianzanalyse mit der Schulform als unabhängige Variable und den Personen- bzw. Fähigkeitsparametern als abhängige Variable berechnet. Die deskriptiven Ergebnisse in Tabelle 23 weisen auf Unterschiede im mittleren Fähigkeitsparameter hin.

Tabelle 23: Mittlere Fähigkeitsparameter an Gymnasien und Hauptschulen

Schultyp	Mittelwert	Standardabweichung	N
Gymnasium	-.508	1.077	312
Hauptschule	-.802	1.044	175
Gesamt	-.614	1.073	487

Die ANOVA weist einen signifikanten Effekt der Schulform nach ($F_{(1, 485)} = 8.549$; $p = .004$, $\eta^2 = .017$), d.h. es besteht ein signifikanter Unterschied im mittlerem Fähigkeitsparameter zwischen den Schülern an Gymnasien und Hauptschulen (vgl. Tabelle 24).

Tabelle 24: Effekte der Schulform auf den mittleren Schwierigkeitsparameter

Quelle	Quadratsumme vom Typ III	df	Varianz	F	Signifikanz	Partielles Eta-Quadrat
Schulform	9.697	1	9.697	8.549**	.004	.017
Fehler	550.140	485	1.132			
Gesamt	743.392	487	1.526			

Anmerkung:

*** Der Wert ist auf dem Niveau von 0.01 (2-seitig) signifikant

Somit unterscheiden sich die Schüler in ihrer Fähigkeit an Gymnasien ($M = -.508$; $SD = 1.077$) und an den Hauptschulen ($M = -.802$; $SD = 1.043$) signifikant dahingehend voneinander, dass an den Gymnasien erwartungskonform der mittlere Fähigkeitsparameter einen höheren Wert aufweist.

10.3.3 Passung zwischen Aufgabenschwierigkeit und Schülerfähigkeit

Um die Frage zu untersuchen, inwieweit die Rasch-skalierte Schwierigkeit der Unterrichtsaufgaben an dem Fähigkeitsniveau der Klasse orientiert ist, wird der Zusammenhang zwischen mittlerer Unterrichtsaufgaben-Schwierigkeit und mittlerer Schülerfähigkeit in den skalierten Klassen bestimmt. In der vorliegenden Studie wird die Passung mit Hilfe der Aufgaben- (s) und Fähigkeitsparameter (f) der Rasch-Skalierung erfasst. Sind die Werte der beiden Parameter bekannt, so kann mit Hilfe der Formel $p = e^{(f-s)} / 1 + e^{(f-s)}$ die aufgrund der beiden Parameter zu erwartende Lösungswahrscheinlichkeit angegeben werden kann. Diese Lösungswahrscheinlichkeit kann pro Klasse berechnet werden, indem sowohl die Parameter s der in der Klasse eingesetzten Unterrichtsaufgaben als auch die Parameter f der in der Klasse befindlichen Schüler gemittelt werden. Bei einer mittleren Lösungswahrscheinlichkeit von 50% für die Unterrichtsaufgaben wird in der vorliegenden Arbeit von einer ausreichenden Passung gesprochen.

Das innerhalb der Klassen gemittelte Schwierigkeitsniveau der Unterrichtsaufgaben korreliert mit dem innerhalb der Klassen gemittelten Fähigkeitsniveau der Schüler über die

20 Klassen hinweg positiv mit $r = .29$ ($p = .001$). Entsprechend dieser vergleichsweise niedrigen Korrelation zeigen sich im Vergleich der einzelnen Klassen unterschiedlich große Diskrepanzen zwischen der Aufgabenanforderung (mittlere Rasch-Schwierigkeit der Unterrichtsaufgaben) und der mittleren Schülerfähigkeit. Dabei zeigt sich, dass besonders in Klassen, die sich im oberen Schwierigkeitsbereich der Unterrichtsaufgaben befinden, den Schülern sehr schwierige Aufgaben zur Bearbeitung vorgelegt werden. Die Differenz zwischen der mittleren Schülerfähigkeit und der mittleren Aufgabenschwierigkeit beträgt in diesen Klassen bis zu 2.4 Logits, so dass die mit Hilfe der Rasch-Formel berechnete durchschnittliche Lösungswahrscheinlichkeit der Unterrichtsaufgaben in diesen Klassen auf bis zu 8.3% sinkt. Die tatsächliche Lösungshäufigkeit der Unterrichtsaufgaben war in diesen Klassen etwas höher, sie betrug 13.2%. Umgerechnet auf die mittlere Schüleranzahl pro Klasse - die mittlere Klassengröße beträgt 24 Schüler - sind dies durchschnittlich 3 Schüler, welche die Unterrichtsaufgaben lösen können. Der Großteil der Schüler in diesen Klassen wird demnach mit überproportional schwierigen Aufgaben überfordert. Im Gegensatz dazu zeigt Abbildung 11 lediglich eine Klasse, welche oberhalb der Linie, die die 50%-Lösungswahrscheinlichkeit der Unterrichtsaufgaben markiert, liegt und in der somit die Unterrichtsaufgaben im Mittel mit einer Lösungswahrscheinlichkeit von etwa 70% von den Schülern richtig gelöst werden. Die tatsächliche Lösungshäufigkeit der wiederholten Unterrichtsaufgaben betrug in dieser Klasse 74%. Nur bei einigen der Klassen der Stichprobe konnte eine ausreichende Passung, also eine mittlere Lösungswahrscheinlichkeit der Unterrichtsaufgaben von 50% festgestellt werden, wobei sich diese Klassen eher im unteren Bereich der Aufgabenschwierigkeit (von -1.0 bis +0.5 Logits) befinden.

Abbildung 13 verdeutlicht den Zusammenhang zwischen der mittleren Schwierigkeit der vom Lehrer in seiner Klasse eingesetzten Unterrichtsaufgaben (x-Achse) und der innerhalb der jeweiligen Klasse gemittelten Schülerfähigkeit (y-Achse). Die Gymnasialklassen sind durch blaue Quadrate und die Hauptschulklassen durch gelbe gekennzeichnet. Die eingezeichnete Gerade der 50%-Lösungswahrscheinlichkeit verdeutlicht, dass die Passung im mittleren Bereich der Aufgabenschwierigkeit am ehesten gelingt. Die Verteilung der Aufgabenschwierigkeit nach den Schulformen macht deutlich, dass es keine strikte Trennung gibt, sondern im mittleren Schwierigkeitsbereich Klassen von Gymnasien und Hauptschulen vielmehr nahe beieinander liegen. Besonders schwierige Unterrichtsaufgaben, die eine große Diskrepanz zur mittleren Schülerfähigkeit aufweisen, werden besonders von Lehrern in Gymnasialklassen eingesetzt.

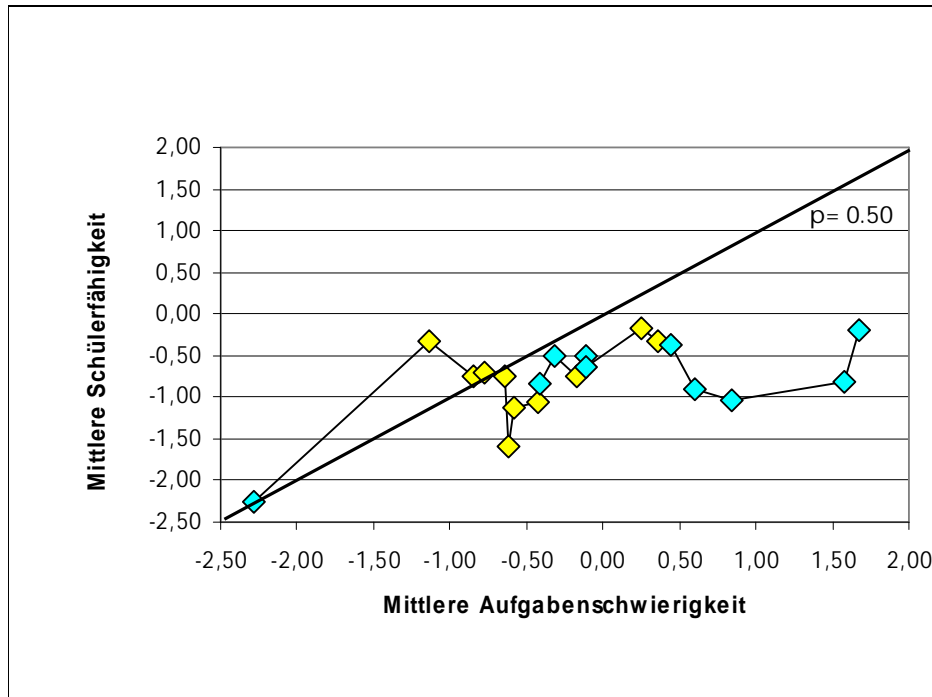


Abbildung 13: Zusammenhang zwischen mittlerer Aufgabenschwierigkeit der Unterrichtsaufgaben und mittlerem Personenparameter ($N = 20$ Klassen)

Ein Streudiagramm der Verteilung des Fähigkeitsparameters innerhalb der einzelnen Klassen, welche mit Hilfe unterschiedlicher Farben kenntlich gemacht wurden, auf der Y-Achse und der mittleren Schwierigkeit der skalierten Unterrichtsaufgaben auf der X-Achse verdeutlicht die große Variation des Fähigkeitsparameters innerhalb und zwischen den untersuchten Klassen (vgl. Abbildung 14). Die Werte der Personenparameter streuen im Bereich von +2 bis -4 Logits. Bei der Variation der mittleren Schwierigkeit der Unterrichtsaufgaben wird deutlich, dass ein Großteil der mittleren Aufgabenschwierigkeit im Bereich zwischen -1 und 0 Logits liegt. Einige wenige Klassen haben eine sehr geringe mittlere Aufgabenschwierigkeit im Bereich von -2 Logits. Die leichtesten Unterrichtsaufgaben einer Klasse haben einen mittleren Schwierigkeitsparameter von -2.3 Logits. In einigen Klassen wurden dagegen sehr schwierige Unterrichtsaufgaben im Bereich von 1 bis 2 Logits eingesetzt. Die Klasse mit den schwierigsten Unterrichtsaufgaben weist einen mittleren Schwierigkeitsparameter von 1.8 Logits auf.

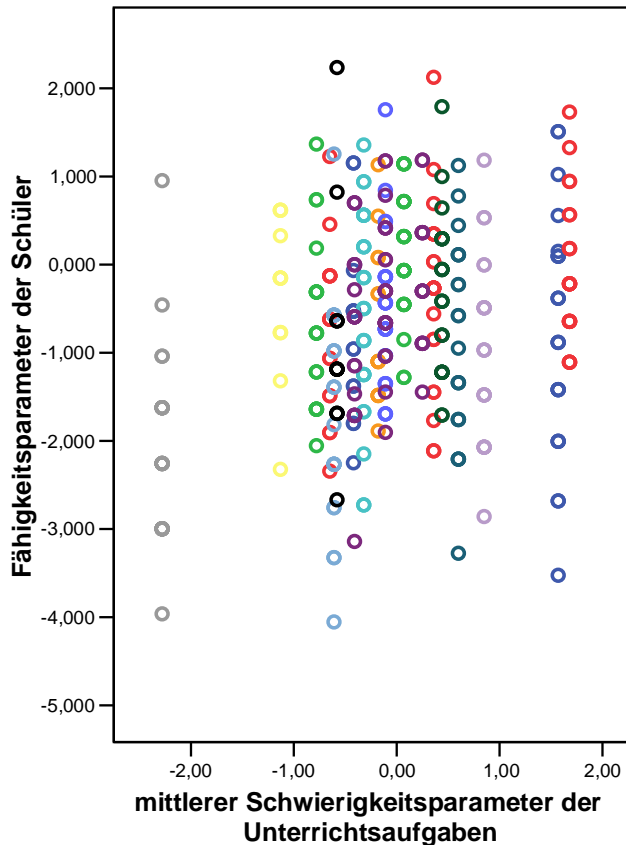


Abbildung 14: Verteilung der mittleren Schwierigkeitsparameter der Unterrichtsaufgaben und der Fähigkeitsparameter

Insgesamt zeigt sich aufgrund der Ergebnisse zur Passung zwischen der mittleren Schwierigkeit der Unterrichtsaufgaben und dem mittleren Fähigkeitsniveau der Klasse, dass keine ausreichende Passung über alle 20 untersuchten Schulklassen gegeben ist. Es gibt zwischen den untersuchten Klassen jedoch große Unterschiede. Werden besonders schwierige Unterrichtsaufgaben eingesetzt, ist keine Anpassung an das mittlere Fähigkeitsniveau gegeben. Insgesamt gelingt die Passung im mittleren Bereich der Aufgabenschwierigkeit (-1 bis 0 Logits) am ehesten. Auch in der Klasse mit den leichtesten Unterrichtsaufgaben und dem geringsten Fähigkeitsniveau ist die Passung gegeben. Lediglich in einer Hauptschulklasse beträgt die Lösungswahrscheinlichkeit mehr als 50%, die Unterrichtsaufgaben werden in dieser Klasse mit einer mittleren Lösungswahrscheinlichkeit von ca. 70% gelöst. Aufgrund dieser Ergebnisse kann die Hypothese 3 als zumindest teilweise evident angesehen werden, da für die Gesamtstichprobe der 20 Klassen keine ausreichende Passung gegeben ist.

11 Zusammenfassung und Diskussion der Ergebnisse aus Studie 1 und Ableitung der Fragestellung für Studie 2

In diesem Kapitel sollen zunächst die Ergebnisse der Studie 1 zusammengefasst und in Bezug auf die Hypothesen dargestellt werden, wobei die Befunde gleichzeitig kritisch betrachtet werden (Kapitel 11.1). Daran anschließend wird eine sich aus Studie 1 ergebende weiterführende Fragestellung für Studie 2 abgeleitet (Kapitel 11.2).

11.1 Zusammenfassung der Ergebnisse

Die globale Zielsetzung der Studie 1 lag in der Überprüfung des Zusammenhangs zwischen der Aufgabenschwierigkeit und der Schülerfähigkeit. Hierzu wurden spezifische Forschungsfragen formuliert. Die Teilnehmer an der Studie waren Hauptschüler und Gymnasiasten aus insgesamt 30 Klassen, die Analysedaten wurden im Fach Physik in der 9. und 10. Klassenstufe erfasst.

Fragestellung 1

Zunächst galt es zu ermitteln, welches kognitive Anspruchsniveau Lernaufgaben im Physikunterricht haben. Das kognitive Anspruchsniveau wurde anhand der zur Aufgabenbearbeitung erforderlichen kognitiven Prozesse Wissen, Verstehen und Anwenden klassifiziert (vgl. Anderson et al., 2001). Postuliert wurde in Hypothese 1, dass die im Unterricht von den Lehrern eingesetzten Lernaufgaben (Unterrichtsaufgaben) vorrangig der Wissenskategorie zuzuordnen sind, während Aufgaben zum Verstehen und Anwenden von Lehrern eher selten eingesetzt werden.

Diese Hypothese erhält durch die Ergebnisse der Studie empirische Evidenz: Die vom Lehrer in den aufgezeichneten Unterrichtsstunden eingesetzten Aufgaben wurden zum größten Teil auf dem Wissensniveau (62.4% aller Unterrichtsaufgaben) gestellt, gefolgt von den Verstehensaufgaben (32.3%). Mit Abstand am wenigsten Aufgaben wurden auf dem Niveau der Wissensanwendung (5.3%) gestellt. Diese Befunde stehen im Einklang mit den Ergebnissen einer Vielzahl anderer Studien, die gezeigt haben, dass Unterrichtsaufgaben zum großen Teil der Wissenskategorie zuzuordnen sind (meist als *low level*-Aufgaben bezeichnet; vgl. Kapitel 3.2.1) und dass es eine geringe Variation der kognitiven Prozesse gibt (vgl. Arnold et al., 1973; Au & Mason, 1981; Brophy & Good, 1986; Crooks, 1988; Doyle, 1988; Fleming & Chambers, 1983; Floyd, 1960; Gallagher, 1965; Goodlad, 1983; Graesser & Per-

son, 1994; Lindfors, 1980; Niegemann & Stadler, 2001; Renkl, 1991; Wilen, 1991; Wilson, 1969).

Des Weiteren wurde der Aufgabeneinsatz im Physikunterricht nach dem Anteil mündlicher und schriftlicher Aufgaben untersucht. Bei der Analyse dieses Aufgabenformats zeigte sich ein ebenfalls in Einklang mit der Literatur (vgl. Neubrand, 2002; Seidel, 2003) stehendes Ergebnis: Der weitaus größte Teil der Unterrichtsaufgaben wird in Form mündlicher Fragen während des Unterrichtsgesprächs gestellt (90.7% aller Unterrichtsaufgaben). In den untersuchten Stunden herrschte demnach der fragend-entwickelnde Unterrichtsstil vor. Der Anteil variierte jedoch abhängig von der Schulform: Hauptschülern wurden mehr mündliche Fragen gestellt als Gymnasiasten.

Fragestellung 2

In Weiterentwicklung der ersten Fragestellung ging es in der zweiten Fragestellung darum, welche Aufgabenschwierigkeiten sich für die Unterrichtsaufgaben mit Hilfe der Rasch-Skalierung schätzen lassen. Es wurde die Hypothese aufgestellt, dass die Unterrichtsaufgaben aufgrund des überwiegenden Einsatzes von Wissensaufgaben eher geringe bzw. numerisch negative Schwierigkeitsparameter aufweisen.

Der mittlere Rasch-Itemparameter der in den Leistungstests wiederholt zur Bearbeitung vorgelegten Unterrichtsaufgaben ist mit -0.17 Logits gering. In den schulformspezifischen Analysen zeigen sich bedeutsame Unterschiede dahingehend, dass der mittlere Itemparameter der Unterrichtsaufgaben an den Gymnasien signifikant höher ist als an den Hauptschulen. Dieses Ergebnis ist erwartungskonform mit der dieser Arbeit zugrundeliegenden schulformspezifischen Hypothese, dass die Schwierigkeit der Unterrichtsaufgaben an den Gymnasien signifikant höher sein sollte als an den Hauptschulen. Somit bestätigen die Ergebnisse die Hypothese zwei.

Fragestellung 3

Als zentrale Frage wurde untersucht, wie gut die Passung zwischen der Schwierigkeit der von Lehrern gestellten Unterrichtsaufgaben und der Schülerfähigkeit ist. Passung wird in der vorliegenden Arbeit mit Hilfe der Item- und Personenparameter definiert: Beträgt die mittlere Lösungswahrscheinlichkeit einer Klasse für die eingesetzten Unterrichtsaufgaben 50%, so wird von einer ausreichenden Passung gesprochen. Die Passung zwischen dem mittleren Schwierigkeitsniveau der Unterrichtsaufgaben und dem mittleren Fähigkeitsniveau der Schüler in den 20 Klassen ist zwar mit $r = .29$ moderat positiv, indiziert jedoch erhebliche Unterschiede zwischen den Klassen. Besonders im oberen Bereich der Aufgabenschwierig-

keit werden die meisten Schüler durch die eingesetzten Aufgaben überfordert. Die Differenz zwischen dem mittleren Fähigkeitsparameter und der mittleren Schwierigkeit der eingesetzten Unterrichtsaufgaben beträgt z.T. fast 2 Logits. Somit werden in den betreffenden Klassen Unterrichtsaufgaben eingesetzt, die eine gemittelte Lösungswahrscheinlichkeit von ca. 8% aufweisen. Die tatsächliche Lösungshäufigkeit ist mit ca. 13% nicht sehr viel höher. Rechnet man diese Lösungshäufigkeit auf eine mittlere Klassengröße der untersuchten Stichprobe von $M=24$ Schülern um, so bedeutet dieses Ergebnis, dass im Mittel nur 2-3 Schüler der gesamten Klasse die gestellten Aufgaben lösen konnten. Für die Mehrheit der Schüler waren die Unterrichtsaufgaben im Gegensatz dazu viel zu schwierig.

Für die untersuchte Stichprobe von 20 Physikklassen kann aufgrund der Ergebnisse zusammenfassend die Hypothese 3 als teilweise evident angesehen werden. Es zeigt sich zwar, besonders im unteren bis mittleren Bereich der Aufgabenschwierigkeit eine ausreichende Anpassung an die Schülerfähigkeit; im oberen Leistungsbereich jedoch wird deutlich, dass die dort beobachtbare Passung lediglich bei einigen wenigen leistungsfähigen Schülern gegeben ist. Auch wurden die Unterrichtsaufgaben nur in einer der zwanzig untersuchten Klassen mit einer Lösungshäufigkeit von 70% gelöst.

Inwieweit das in der vorliegenden Untersuchung angewendete Kriterium der Passung - 50% Lösungswahrscheinlichkeit - für alle bzw. die meisten im Unterricht vom Lehrer eingesetzten Aufgaben als ausreichend angesehen werden kann, ist fraglich. In internationalen Schulleistungsstudien ist es üblich, Testaufgaben einzusetzen, deren mittlere Lösungswahrscheinlichkeiten 62% beträgt, wenn der Personenparameter dem Schwierigkeitsparameter der Aufgabe entspricht (vgl. Köller, Watermann & Baumert, 2001), wodurch gewährleistet werden soll, dass die Aufgaben mit hinreichender Sicherheit gelöst werden. Würde eine höhere Lösungswahrscheinlichkeit in der vorliegenden Studie zugrunde gelegt, würde lediglich in der Klasse mit der Lösungswahrscheinlichkeit von 70 % eine ausreichende Passung gegeben sein.

Das vorliegende Ergebnis, dass die von den Lehrern eingesetzten Aufgaben z.T. viel zu schwierig für ihre Schüler sind, stimmt mit einem Befund, welcher im Rahmen der TIMS III-Studie (Baumert, Bos & Lehmann, 2000) nachgewiesen wurde, überein. Dabei zeigte sich, dass Experten der Fachdidaktiken die Lösungshäufigkeiten der TIMSS-Items signifikant überschätzten. Die Lösungshäufigkeiten wurden dabei sowohl in Grund- wie in Leistungskursen und jeweils in den Fächern Mathematik und Physik signifikant überschätzt. Es wurde nachgewiesen, dass, je höher das Anspruchsniveau der Aufgabe und je mehr fachliche Kenntnisse vorausgesetzt wurden, desto unrealistischer die Einschätzung ausfiel. Besonders im oberen Leistungsbereich neigen Experten also dazu, Kenntnisse und Fähigkeiten von

Schülern zu überschätzen (Klieme, 2000). Dieses Ergebnis entspricht dem vorliegenden Befund dahingehend, dass die Unterrichtsaufgaben im oberen Schwierigkeitsbereich die meisten Schüler überfordern. Aufgrund der Ergebnisse von Klieme (2000) kann angenommen werden, dass die in der vorliegenden Studie untersuchten Lehrer ebenfalls nicht die Aufgabenschwierigkeit anhand der Leistungsfähigkeit ihrer Schüler angemessen einschätzen können. Da die Unterrichtsaufgaben jedoch nicht von den Lehrern in ihrer Schwierigkeit explizit eingeschätzt wurden, kann aufgrund der Ergebnisse der vorliegenden Studie keine gesicherte Aussage zur diagnostischen Kompetenz, sowohl in Bezug auf die Einschätzung der Aufgabenschwierigkeit, als auch in Bezug auf die Einschätzung der Leistungsfähigkeit der Schüler für die untersuchte Lehrerstichprobe gemacht werden.

Das Ergebnis der vorliegenden Studie, dass die Aufgabenschwierigkeit z.T. sehr hoch ist, steht in Einklang mit dem Befund von Schnotz (1971), dass 70% der Aufgaben von Lehrern in ihrer Schwierigkeit überschätzt werden. Auch neigen Lehrer offensichtlich dazu, die Leistung ihrer Schüler zu überschätzen (vgl. Hosenfeld et al. 2002; Schrader, 1989; Schrader & Helmke, 1987). Als mögliche Gründe für die häufig unzureichende Passung kann angenommen werden, dass Lehrer den eigenen Unterrichtserfolg überschätzen und lediglich über unzureichende Kenntnis der Fähigkeiten und des Vorwissens ihrer Schüler verfügen. Womöglich orientieren sich Lehrer bei der Auswahl der Aufgaben und ihrer Schwierigkeit an den Leistungen, zu welchen ihre Schüler aufgrund des Unterrichts und des Lehrplans eigentlich in der Lage sein sollten, statt an der tatsächlich gezeigten Leistung (vgl. Kapitel 3.4). Diese möglichen Gründe für die z.T. nachgewiesene unzureichende Passung müssen noch empirisch überprüft werden.

Auch das Ergebnis der mit Hilfe der klassischen Testverfahren durchgeführten Auswertung der Leistungstests zeigt, dass die mittlere Lösungshäufigkeit der Unterrichtsaufgaben, welche eine Woche zuvor mit identischen Wortlaut Teil des Unterrichts waren, im Mittel lediglich 32.5% betrug und für die Schüler als eher schwierig zu bezeichnen sind. Dieses Ergebnis steht in Einklang mit Befunden, die gezeigt haben, dass Schüler schwierige Aufgaben im Unterricht gestellt bekommen (vgl. Fisher et al., 1980; Jorgenson, 1977).

Ein ausreichender theoretischer Erklärungsansatz für das Ergebnis der z.T. unzureichenden Passung zwischen Aufgabenschwierigkeit und Schülerfähigkeit liegt bislang nicht vor und muss daher noch entwickelt werden. Basierend auf den oben erwähnten Forschungsbefunden könnte in der mangelnden diagnostischen Kompetenz von Lehrern eine mögliche Ursache dafür gesehen werden.

Fragestellung 4

Ferner wurde geprüft, ob sich Aufgaben mit a priori angenommener unterschiedlicher Schwierigkeit auf Grundlage der Revision der Bloom'schen Taxonomie (Anderson et al., 2001) konstruieren lassen.

Die Aufgabenkonstruktion auf der theoretischen Grundlage der Revision der Bloom'schen Taxonomie (Anderson et al., 2001) in Kombination mit den Konstruktionsvorschriften für kriteriumsorientierte Tests von Klauer (1987) hat sich als Möglichkeit erwiesen, a priori Aufgaben unterschiedlicher Schwierigkeit konstruieren zu können. Anhand der Rasch-Skalierung konnten für die verschiedenen konstruierten Aufgabenarten signifikante Unterschiede in den mittleren Itemparametern nachgewiesen werden. Demnach weisen die Wissensaufgaben im Mittel einen signifikant geringeren Itemparameter auf als die zusammengefasst betrachteten Verstehens- und Anwendungsaufgaben. Auch zeigten sich bei den mit Hilfe der klassischen Testverfahren berechneten Lösungshäufigkeiten ebenfalls erwartungskonforme signifikante Unterschiede.

Die Besonderheit der gewählten Vorgehensweise liegt darin, dass das Schwierigkeitsniveau der Aufgaben, auf der theoretischen Grundlage der Taxonomie von Anderson et al. (2001), a priori festgelegt wurde und die tatsächliche Schwierigkeit der Aufgaben mit Hilfe der erhobenen Daten ermittelt wurde. Dabei konnte die Abstufung des Schwierigkeitsniveaus in Bezug auf die kognitiven Prozesse der Aufgabenbearbeitung weitgehend bestätigt werden. Aufgrund dieser Ergebnisse lassen sich Empfehlungen für die Konstruktion von Lernaufgaben, speziell für deren Schwierigkeit, im Unterricht ableiten.

Schulformspezifische Unterschiede

Insbesondere für die ersten drei Fragestellungen wurden schulformspezifische Unterschiede postuliert. Die Ergebnisse der Schulformvergleiche sind insgesamt erwartungskonform, jedoch nicht in dem erwarteten Ausmaß zu finden. Zwar unterscheiden sich sowohl die mittleren Fähigkeitsparameter zwischen Gymnasien und Hauptschulen als auch der mittlere Itemparameter der eingesetzten Unterrichtsaufgaben signifikant voneinander. Jedoch waren die insgesamt fast identischen Lösungshäufigkeiten aller Aufgabenarten zwischen den beiden Schulformen in dem vorliegenden Ausmaß nicht erwartet worden (vgl. Kapitel 10.2). Dieses Ergebnis war aufgrund der großen Leistungsunterschiede in den PISA-Erhebungen der naturwissenschaftlichen Kompetenz an Gymnasien und Hauptschulen nicht zugrunde gelegt worden (Prenzel et al., 2001, Rost et al., 2004). Besonders aufgrund des Ergebnisses, dass 52.4% der Hauptschüler der „Risikogruppe“ angehören - damit sind Schüler gemeint, welche lediglich die unterste Kompetenzstufe erreichen - waren größere Unterschiede zwi-

schen den beiden Schulformen als die hier gefundenen erwartet worden. Ungeachtet dessen muss bei der Interpretation der Ergebnisse berücksichtigt werden, dass im Gegensatz zu der PISA-Erhebung die Teilnahme an der vorliegenden Untersuchung für die Schulen und Lehrer freiwillig war. Es ist eine gewisse positive Selbstselektion der Stichprobe, besonders an den Hauptschulen, dahingehend zu vermuten, dass die hier untersuchte Lehrerstichprobe ihre Klassen als eher leistungsstark einschätzen würde. Dies wird möglicherweise durch die Videoaufzeichnung des Unterrichts noch verstärkt. Inwieweit die hier untersuchten Klassen tatsächlich leistungsstark sind, kann aufgrund des nicht möglichen Vergleichs mit einer realen Zufallsstichprobe aller (Haupt)Schulen nicht beantwortet werden. Das zunächst überraschende Ergebnis, dass eine der Gymnasialklassen sowohl die leichtesten Unterrichtsaufgaben als auch den niedrigsten mittleren Fähigkeitsparameter aufweist (vgl. Kapitel 10.3.3, Abbildung 13), ist jedoch dahingehend übereinstimmend, dass diese Klasse sowohl während der Unterrichtsbeobachtung bei der Videoaufzeichnung, als auch bei der Testdurchführung, einen sehr leistungsschwachen und unmotivierten Eindruck auf die Testleiterin gemacht hat. Somit kann ein Artefakt in diesem Falle relativ sicher ausgeschlossen werden.

Insgesamt sind die Ergebnisse in Bezug auf die eher mangelnde Passung von Aufgabenschwierigkeit und Schülerfähigkeit im Physikunterricht als erste Befunde anzusehen, da diese auf einer Stichprobe von nur 20 Klassen beruhen und somit eine begrenzte Aussagekraft haben. Um generalisierbare Aussagen treffen zu können, sollte die Untersuchung an einer größeren Stichprobe und womöglich einem anderen Fach repliziert werden.

11.2 Abgeleitete Fragestellung

Die Ergebnisse der Studie 1 haben gezeigt, dass die Lösungswahrscheinlichkeit der Unterrichtsaufgaben, nachdem sie bereits Teil des Physikunterrichts waren und eine Woche später erneut zur Bearbeitung vorgelegt wurden, im Mittel mit einer Lösungshäufigkeit von etwa 30% korrekt gelöst werden und dass für viele Schüler keine ausreichende Passung zwischen ihrer Fähigkeit und der Aufgabenschwierigkeit im Unterricht gegeben ist. Vor diesem Hintergrund wurde nach einer Möglichkeit gesucht, wie die Passung individuell erhöht werden kann, um auch die Schülerfähigkeit auf diese Weise zu erhöhen.

Dazu wurde die Idee des adaptiven Testens und des adaptiven Unterrichts aufgegriffen. Trotz z.T. inkonsistenter Ergebnisse konnte in der diesbezüglichen Literatur gezeigt werden, dass, wenn spezifische Charakteristika der Schüler und des Unterrichts berücksichtigt werden, adaptiver Unterricht eine Möglichkeit zur Leistungsverbesserung darstellen kann. Da mit Hilfe identischer Aufgaben für alle Schüler einer Klasse in Studie 1 keine aus-

reichende Passung mit der mittleren Schülerfähigkeit erreicht wurde, erfolgt in Studie 2 eine individuelle Anpassung der Aufgabenschwierigkeit an den Fähigkeitsparameter des einzelnen Schülers (vgl. Corno & Snow, 1986; Kulik & Kulik, 1991; Snow, 1989).

Es wird somit in der im folgenden beschriebenen Studie 2 versucht, durch an die individuellen Schülerfähigkeiten angepassten Schwierigkeiten der zu bearbeitenden Aufgaben die Lösungshäufigkeit und damit auch den Fähigkeitsparameter der Schüler zu erhöhen. Ob schon die Ergebnisse der ATI-Forschung den Zusammenhang zwischen kognitivem Anspruchsniveau des Unterrichts und den Schülerfähigkeiten belegt haben, gibt es bislang kaum empirische Untersuchungen zur Beziehung zwischen dem adaptivem Aufgabeneinsatz bzw. der Variation der Aufgabenschwierigkeit im Unterricht in Abhängigkeit der individuellen Schülerfähigkeiten. Daher soll der Zusammenhang zwischen individueller Schülerfähigkeit und Aufgabenschwierigkeit, welcher möglicherweise einen Einfluss auf die Leistungsfähigkeit bzw. die Veränderung des Fähigkeitsparameters der Schüler hat, in Studie 2 systematisch untersucht werden.

Bevor auf die Studie 2 im Einzelnen eingegangen wird, soll im folgenden Kapitel 12 kurz auf die theoretischen Überlegungen und bisherige Ergebnisse zum adaptiven Testen und Unterricht eingegangen werden.

12 Studie 2: Adaptives Testen und Unterrichten

Für die im Folgenden dargestellte Studie 2 ist neben dem theoretischen Hintergrund, welcher in den Kapiteln 2 bis 6 dargelegt wurde, die Idee des adaptiven Testens bzw. Unterrichts von besonderer Relevanz, da in dieser Studie der Zusammenhang zwischen Schülerfähigkeit und Aufgabenschwierigkeit mit Hilfe einer adaptiven Anpassung experimentell untersucht wird. Daher wird in dem Kapitel zunächst die Bedeutung der Adaptation für den Unterricht dargestellt (Kapitel 12.1) sowie wichtige Entwicklungen des Konstrukts skizziert (12.2), bevor in Kapitel 12.3 auf den Stand der Forschung zum adaptiven Testen bzw. Unterricht eingegangen wird. Im Anschluss werden Schlussfolgerungen für die Fragestellung der Studie 2 erläutert.

12.1 *Bedeutung der Adaptation für den Unterricht*

Eines der schwierigsten Probleme der Unterrichtsgestaltung stellt nach Slavin (1997) der Umgang mit den unterschiedlichen Wissens- und Fähigkeitsniveaus der einzelnen Schüler dar. Unterrichtet der Lehrer die gesamte Klasse nach einer Methode, ist es selbstverständlich, dass die Schüler unterschiedlich schnell und unterschiedlich viel in diesem Unterricht lernen. Verschiedene Autoren (z.B. Cunningham, 1987; Seel, 1981) haben darauf hingewiesen, dass es keinen Unterricht gibt, der allen Schülern gleichermaßen optimales Lernen ermöglicht. Seel (1981) weist darauf hin, dass der identische Einsatz von Aufgaben für alle Schüler erhebliche Nachteile mit sich bringen kann. Zum einen können die Anforderungen für einen Teil der Schüler zu hoch angesetzt sein, so dass die Aufgabenschwierigkeit über alle Maßen gesteigert wird und zu Lernunwirksamkeit führt. Zum anderen können sich die aufgrund der durchschnittlichen Lernvoraussetzungen ausgewählten Lernaufgaben zumindest für einen Teil der Schüler als zu einfach in ihren Anforderungen erweisen, so dass sie deaktivierend und unterfordernd wirken. Daher hebt der Autor die Bedeutsamkeit der adaptiven Anpassung der Aufgabenkomplexität – und somit der Aufgabenschwierigkeit – an die individuellen Voraussetzungen der Schüler hervor.

Der Grad der Interaktion zwischen dem Schüler und der Lernaufgabe ist in erster Linie von den individuellen kognitiven Voraussetzungen des Schülers abhängig (Reinmann-Rothmeier & Mandl, 1998). Daher wird mit Hilfe verschiedener Programme und Methoden versucht, den Unterricht zu individualisieren, indem dieser an die Bedürfnisse und Fähigkeiten der einzelnen Schüler adaptiert wird. Kounin (1976) hebt die Anpassung der Anforder-

rungen an die Schüler und die Bedeutung eines angemessenen Anspruchsniveaus hervor. Für Unterrichtsansätze, die sich mit der Gestaltung solcher Anpassungsvorgänge befassen, wurde von Schwarzer und Steinhagen (1975) die Bezeichnung „adaptiver Unterricht“ eingeführt. Insgesamt geht das Konzept des adaptiven Unterrichts, ähnlich wie das Konzept des *mastery learnings* (vgl. Kapitel 4.7), von der Annahme aus, dass fast alle Schüler ein bestimmtes Lernziel erreichen können, wenn der Unterricht an ihrem jeweiligen Kenntnisstand ansetzt, d.h. die Lernbedingungen so verändert werden, dass jeder Schüler optimale Lerngelegenheiten erhält (Gruehn, 2000).

Generell gibt es eine Vielfalt von Möglichkeiten, die Schülermerkmale und die optimale Lernumgebung individuell aneinander anzupassen. Hierzu müssen jedoch sowohl die Merkmale der Schüler als auch die Merkmale der Lernumgebung identifiziert werden. Demnach sind die Lehrer, im Bezug auf die Regulierung des Unterrichtstempos und die Gestaltung der Aufgabenschwierigkeit, auf Diagnose- und Steuerungskompetenzen angewiesen (vgl. Kapitel 3.4).

Die geringe Anpassung des Unterrichts an die Lern- und Leistungsvoraussetzungen der Schüler wird von Schrader (1989) als ein Kernproblem effektiven Unterrichts angesehen. Das Problem, trotz unterschiedlicher Leistungsvoraussetzungen für alle Lernenden optimale Lernergebnisse bzw. zumindest gewisse Minimalziele zu erreichen, erfordert Anpassungsleistungen verschiedener Art, die sowohl in der Anpassung der Lernenden an verschiedene Unterrichtsformen als auch in der Anpassung des Unterrichts an den Lernenden liegen können.

12.2 Entwicklung des Konstrukts adaptiver Unterricht

Aufbauend auf der Erkenntnis, dass bestimmte Instruktionsformen in Abhängigkeit von einzelnen Schülermerkmalen unterschiedlich wirken, wird mit dem Konstrukt des adaptiven Unterrichts versucht, Instruktionsprozesse so zu gestalten, dass interindividuelle Unterschiede zwischen Schülern einer Klasse durch den Einsatz verschiedener Unterrichtsmethoden berücksichtigt und damit Lernprozesse optimiert werden. Die Idee des adaptiven Unterrichts geht auch auf Befunde der *Aptitude-Treatment-Interaction-Forschung* (ATI-Forschung) zurück. Dabei steht der Zusammenhang zwischen spezifischen instruktionalen Prozeduren und Strategien und spezifischen Schülermerkmalen - meist kognitive Fähigkeiten und affektive bzw. motivationale Variablen - im Mittelpunkt des Forschungsinteresses (vgl. Cronbach & Snow, 1977; Park & Lee, 2003). In der ATI-Forschung werden also die Wechselwirkungen zwischen alternativen „Behandlungen“ und unterschiedlich ausgeprägten

Schülermerkmalen erforscht. Es werden Möglichkeiten einer optimalen Passung von verschiedenen Unterrichtsmethoden und individuellen Merkmalsausprägungen untersucht.

Die Adaptivität eines Unterrichtssystems bemisst sich nach dem Umfang, in dem unterschiedliche Ziele und alternative unterrichtliche Maßnahmen zur Erreichung der Ziele berücksichtigt werden, und dem Ausmaß, in dem die Maßnahmen auf die bei den Schülern diagnostizierten Lernvoraussetzungen abgestimmt werden (Corno & Snow, 1986).

Corno und Snow (1986) unterscheiden zwei Ebenen, auf denen die Anpassung – man spricht in diesem Zusammenhang von der Adaptionsrate – stattfinden: (1) auf der Ebene mittel- bis langfristiger Entscheidungen und Handlungsstrategien (*Makroadaptation*), z.B. Anpassung in Form äußerer Differenzierung durch die Bildung leistungshomogener Klassen, und (2) auf der Ebene kurzfristiger „momentaner“ Anpassungsprozesse an die Lernbedürfnisse der Schüler im Unterricht (*Mikroadaptation*; vgl. Corno & Snow, 1986). Der in Kapitel 4 skizzierte *Programmierte Unterricht* ist eine Form der Mikroadaptation. Die Idee der makro-adaptiven Gestaltung von Lehr- und Lernsituationen ist vor allem durch die ATIForschung bekannt geworden. Grundgedanke ist, dass Merkmale der Lehr-Lernsituation und Personenmerkmale des Lernenden in komplexer Weise interagieren und zusammen den Erfolg einer bestimmten Unterrichtsmaßnahme ausmachen (Cronbach & Snow, 1977).

Eine weitere Möglichkeit der Unterscheidung bezieht sich auf den Adaptionszweck. Salomon (1972) unterscheidet drei Modelle: Förder-, Kompensations- und Präferenzmodell, nach denen Adaptionsmaßnahmen gegliedert werden können. Diese Modelle zielen auf die Anpassung von Lehrzielen bzw. Lernfunktionen, Lehrzeit und Lehrmethode. Mit Fördermodellen sind Adaptionsmaßnahmen gemeint, mit denen versucht wird, individuelle Defizite bei der Lehrzielerreichung durch zusätzlichen Unterricht zu beseitigen. Durch geeignete Lernhilfen und Maßnahmen wird im Rahmen des Kompensationsmodells versucht, allgemeine Lernvoraussetzungsdefizite auszugleichen. Von den in diesen Modellen eingesetzten Maßnahmen profitieren lediglich Schüler mit Defiziten, denn Schüler ohne erfassbare und lehrzielbezogene Defizite oder Schwächen haben keinen direkten Vorteil und werden im schlimmsten Fall sogar unterfordert und demotiviert (vgl. Leutner, 1992; Nix, 1990; Snow, 1977). Im Gegensatz dazu profitieren von dem Präferenzmodell alle Schüler, da hierbei die individuellen Stärken der Lernvoraussetzungen besonders hervorgehoben werden. Für eine genauere Beschreibung der verschiedenen Adaptionsmodelle sei auf Salomon (1972) verwiesen.

Adaptiver Unterricht kann des Weiteren nach der Art der Adaptionsmaßnahme unterschieden werden. So können das Lehrziel, die Lehrmethode oder die Lernzeit individuell an den einzelnen Schüler angepasst werden. Leutner (1992) fasst die genannten Facetten der

Adaptationsmaßnahmen zusammen und differenziert in Anlehnung an die Modelle von Cronbach (1967) und Glaser (1977) acht Maßnahmen die in Tabelle 25 dargestellten werden.

Tabelle 25: Facetten der Adaption nach Leutner (1992) in Anlehnung an die Modelle von Cronbach (1967) und Glaser (1977)

	Lehrmethode	Lernzeit	
		fixiert	variabel
Lehrziele fixiert	fixiert	1. Fortschreitende Auslese	5. Mastery learning bzgl. Lernziel
	variabel	2. Innere Differenzierung	6. Förder-, Nachhilfeunterricht
Lehrziele variabel	fixiert	3. Notengebung im herkömmlichen Unterricht	7. Mastery learning bzgl. Notengebung
	variabel	4. Äußere (Schul-) Differenzierung	8. Gegliedertes Schulsystem

Sind, wie in Maßnahme 1, sowohl Lehrziel, Lernmethode als auch Lernzeit fixiert, ist die einzige Möglichkeit der Adaptation, nur fähige Schüler zum Unterricht zuzulassen und während des Unterrichts die Schüler auszuschließen, die das fixierte Ziel nicht erreichen (fortschreitende Auslese). Ist die Lernzeit fixiert, die Lernmethode jedoch variabel, kann die Möglichkeit der inneren Differenzierung angewendet werden (Maßnahme 2). Steht nur eine Lernmethode zur Verfügung, die Lernzeit ist fixiert, das Lernziel jedoch variabel, so kann dieses dem Schüler angepasst werden, wie es beispielsweise bei der traditionellen Notengebung der Fall ist (Maßnahme 3). Im Gegensatz dazu ergibt sich eine äußere Differenzierung (Maßnahme 4), wenn die Lehrmethode und das Lehrziel variabel sind, doch die Lernzeit fixiert ist, wie z.B. bei der Einteilung der Schüler in Grund- und Leistungskurse.

Im Unterschied dazu ist bei den folgenden Maßnahmen die Lernzeit variabel: Sind Lehrziel und –methode festgelegt, kann im Sinne des *mastery learnings* (vgl. Kapitel 4.6) jeder Schüler so lange unterrichtet werden, bis er den Inhaltsbereich verstanden hat (Maßnahme 5). Bei einem fixierten Lehrziel, welches mit variablen Methoden und variabler Lernzeit erreicht werden kann, bieten sich die Möglichkeiten des Förderunterrichts an, um die individuellen Defizite auszugleichen (Maßnahme 6). Die Möglichkeiten des *mastery learnings* bzgl. der Notengebung können genutzt werden, wenn das Lehrziel und die Lernzeit variabel sind, die Lehrmethode jedoch fixiert ist (Maßnahme 7). Die achte Möglichkeit der Adaptation besteht bei Variabilität aller drei Facetten in der Differenzierung nach einem ge-

gliederten Schulsystem, indem unterschiedliche Lehrziele mit verschiedenen Methoden und Lernzeiten erreicht werden.

Als Beispiel eines förderungsorientierten Unterrichts gilt das Modell des zielerreichenden Lernens (Bloom, 1976, siehe Kapitel 4.6), wonach sich die Schüler in der für die Aneignung des Lernstoffs benötigten Lernzeit unterscheiden. Damit wird jedoch lediglich ein Teilaspekt des adaptiven Unterrichts berücksichtigt. Für andere Aspekte, wie z.B. die Wahl des Unterrichtstempos und der Aufgabenschwierigkeit, existieren nach Schrader (1989) keine theoretischen und empirisch ähnlich gut fundierten Ansätze wie das Modell des zielerreichenden Lernens.

In den 1970er Jahren wurde eine Reihe von Versuchen unternommen, die verschiedenen Methoden umzusetzen. Beispielsweise erfordert die Umsetzung der von Wygotsky (1978) eingeführte Zone der proximalen Entwicklung (vgl. Kapitel 4.3) im Unterricht vom Lehrer eine Differenzierung und Individualisierung des Unterrichts durch die Anpassung der Schwierigkeit des Unterrichtsstoffs und –tempos und somit den Einsatz unterschiedlicher Aufgaben für die einzelnen Schüler. In diesem Fall müssen sowohl Lernzeit als auch Lernmethode – und somit Aufgaben als Teil dessen - variabel an die individuellen Voraussetzungen der Schüler adaptiert werden. Die Auffassung wurde von Helmke (2004) auf die im Unterricht eingesetzten Aufgaben übertragen, indem er annahm, dass diese hinsichtlich ihrer Schwierigkeit an der Zone der proximalen Entwicklung der Schüler ausgewählt werden sollten. Das bedeutet, dass sie in ihrer Schwierigkeit etwas oberhalb des aktuellen Fähigkeits- und Wissensniveaus der Schüler liegen sollten. Diese Hypothese wurde bislang jedoch noch nicht einer empirischen Überprüfung unterzogen, was in der in den Kapitel 12-16 beschriebenen Studie 2 erfolgen soll. Dabei werden die von den Schülern zu bearbeitenden Aufgaben in ihrer Schwierigkeit individuell an die Schülerfähigkeit adaptiert.

Weitere adaptive Programme sind die *Individually Guided Education* (IGE: (Klausmeier, 1975), *PLAN* (*Program for Learning in Accordance with Needs*; Flanagan, Shanner, Bruder & Marker, 1975) sowie *Learning for Mastery* (Block, 1971; Bloom, 1976) und *Personalized System of Instruction* (PSI oder Keller Plan; Block & Burns, 1976; Keller & Sherman, 1974). Auf diese Programme kann im Rahmen dieser Arbeit nicht näher eingegangen werden. Es sei lediglich angemerkt, dass all diesen Programmen gemeinsam ist, dass sie versucht haben, den Unterricht zu individualisieren, indem sie primär auf die individuellen Wissensunterschiede der Schüler zu Beginn des Unterricht fokussiert haben (Shull, 1980).

Auch wurde bereits früh versucht, Adaptivität durch den Einsatz von Computern zu erreichen. Erste Ansätze in diese Richtung erfolgten mit der programmierten Instruktion, bei der sich die Individualisierung allerdings auf sehr einfache Faktoren wie Lernzeit und Auf-

gabensequenzierung beschränkte (Fletcher, 1992). Adaptive *Computer-Assisted-Instruction-Systeme* stellen einen weiteren Versuch dar, die Möglichkeiten von Computern für eine adaptive Steuerung von Lernprozessen zu nutzen (vgl. Leutner, 1992).

Die Idee des adaptiven Unterrichts wurde in jüngerer Zeit von Weinert (1997) wieder aufgegriffen und hat somit auch heute nicht ihre Aktualität eingebüßt (vgl. auch Leutner, 2004). Helmke (2004) spricht in diesem Zusammenhang aufgrund der Heterogenität der Schüler bzgl. ihrer Fähigkeiten und ihres Vorwissens von der „Notwendigkeit der Individualisierung“ (S. 72) der unterschiedlichen Methoden, Lernmaterialien, Lerninhalte, Lernzielniveaus und Motivierungstechniken.

Wie im vorangegangenen Abschnitt verdeutlicht wurde, stellt die Anpassung der Aufgabenschwierigkeit an die individuellen Schülerfähigkeiten eine bedeutsame Variable im adaptiven Unterricht dar. Ebenso wie der adaptive Unterricht ist auch das adaptive Testen verstärkt in das Forschungsinteresse gerückt. So wurde an konventionellen Tests kritisiert, dass bei dieser Testform allen Probanden alle Aufgaben vorgelegt werden. Dieses Vorgehen ist insbesondere bei Probanden mit extremen Merkmalsausprägungen unökonomisch. Da viele Tests als *power*-Tests konzipiert werden, d.h. dass die Aufgabenschwierigkeit im Verlauf des Tests ansteigt, sind solche Tests für leistungsschwächere Probanden im schwierigen Testteil oft frustrierend. Im Gegensatz dazu sind solche Tests für leistungsstärkere Probanden aufgrund der für sie sehr leichten Aufgaben zu Beginn des Tests z.T. langweilig und unterfordernd (Lienert & Raatz, 1998). Dieses Problem versucht man durch das adaptive Testen zu lösen, indem jeder Proband entsprechend seinem Leistungsvermögen angepasst andere Items erhält (z.B. Hornke, 1977). Bei dem Vorgehen des *Computer-Adaptiven Testens* (CAT) bekommt der Proband zu Beginn ein Item mittlerer Schwierigkeit zur Bearbeitung vorgelegt. Mittels der Antwort wird die Fähigkeit des Probanden geschätzt und dann ein weiteres an den Fähigkeitsparameter in der Schwierigkeit adaptiertes Item mit einer erwarteten Lösungswahrscheinlichkeit von 50% vorgelegt. Ist die Antwort korrekt, ist das nachfolgende Item schwieriger. Ist die Antwort falsch, hat das anknüpfende Item eine geringere Schwierigkeit. Items, die zu leicht oder zu schwer für den Probanden sind, stellen wenig Information über die wahre Fähigkeit zur Verfügung. Die verbreitetste Methode zur Eichung der Itemschwierigkeiten und Schätzung der Fähigkeiten beim computer-adaptiven Testen erfolgt auf der Grundlage der Überlegungen der *Item-Response*-Theorie (IRT) (Lienert & Raatz, 1998). Ist dabei ein Item zu schwer oder zu leicht für den Probanden, trägt dieses Items kaum zur Fähigkeitsschätzung der Person bei (Green, Bock, Humphreys, Linn & Reckase, 1984). Computer-adaptives Testen passt die Testschwierigkeit der gegenwärtig geschätzten Fähigkeit der zu testenden Person an, so dass leistungsstärkere Personen schwieri-

gere Items bearbeiten und Leistungsschwächere leichtere Items. Dadurch wird der Standardfehler der Messung minimiert. Somit kann auch die Testlänge gekürzt werden, ohne dass die Genauigkeit der Testung beeinträchtigt ist (z.B. Bergstrom, Lunz & Gershon, 1992; McKinley & Reckase, 1980; Weiss & Kingsbury, 1984).

12.3 Forschungsergebnisse zum adaptiven Testen und Unterricht

Im Folgenden soll kurz auf die bislang vorliegenden Ergebnisse zum adaptiven Testen und zum adaptiven Unterricht eingegangen werden.

Bloom (1976) hat zwar gezeigt, dass Lehrer über eine Reihe von Lernmethoden verfügen und diese modifizieren und an den Schüler anpassen. Andere Studien zeigen jedoch (Dahllöf, 1971; Good & Brophy, 1971; Lundgren, 1972), dass sich die Lehrer bei der Unterrichtsplanung gewöhnlich nur auf einen kleinen Teil der Schüler ihrer Klasse, deren Leistung etwa im Bereich zwischen dem 10. und 25. Perzentil der Leistungsverteilung der Klasse zu lokalisieren ist, beschränken. Somit gibt es in den meisten Klassen kaum eine Binnendifferenzierung der Aufgabenschwierigkeit an die Schülerfähigkeiten. Die kleine Gruppe der besseren Schüler dient dem Lehrer als Referenzgruppe zur Auswahl der Aufgaben. Erwartungsgemäß passen die Aufgaben nicht mit der Leistungsfähigkeit der anderen bei der Auswahl unberücksichtigten Schüler überein. Im Gegensatz dazu konnten Baumert, Schmitz, Sang und Roeder, (1987) dieses Konzept der Referenzgruppe, welche sich aus den leistungsschwachen Schülern zusammensetzte, für die Setzung des Anspruchsniveaus des Unterrichts nicht bestätigen. Baumert et al. (1987) nehmen dennoch an, dass bestimmte Verhaltensweisen der Schüler (sog. *cues*), wie z.B. nachlassende Konzentration, dem Lehrer wichtige Hinweise zur Anpassung der Aufgabenschwierigkeit geben können. Wann und wie viel neuer Unterrichtsstoff eingeführt wird, ob die Aufgabenschwierigkeit erhöht werden soll oder welches Maß an Wiederholung und Übung erforderlich ist, darüber müsse der Lehrer – so die Autoren - im einzelnen in der konkreten Unterrichtssituation entscheiden. Die dabei eingesetzten Entscheidungskriterien der Lehrer sind bislang weitgehend unbekannt und unerforscht. Auch eine Identifikation der oben genannten Schülergruppe, anhand deren die Unterrichtsplanung erfolgt, ist mit Hilfe von Unterrichtsbeobachtung und Lehrerbefragung bislang nicht gelungen, wodurch keine weiteren, gesicherten Aussagen über diese mögliche Referenzgruppe gemacht werden können (vgl. Baumert et al., 1987).

In einer Metaanalyse über 254 Studien, welche die Effektivität von computeradaptivem Unterricht mit regulärem Unterricht in Bezug auf die Schülerleistung verglichen, konnten Kulik und Kulik (1991) zeigen, dass die Effektstärke $d = 0.30$ betrug, d.h. die Leis-

tung der Schüler, welche mit Hilfe von computer-adaptivem Unterricht gelernt haben, liegt um 0.30 Standardabweichungen über der Leistung der Schüler, die regulären Unterricht erhalten haben. Auch wenn der über alle berücksichtigten Studien gemittelte Effekt als moderat zu bezeichnen ist, so zeigen sich große Unterschiede zwischen den berücksichtigten Studien. Die größte positive Effektstärke wurde in der Studie von Cartwright, Cartwright und Robine (1972) mit 2.17 Standardabweichungen nachgewiesen. Der größte negative Effekt betrug im Gegensatz dazu -1.20 Standardabweichungen (Diem, 1982). Kulik, Kulik und Cohen (1979) konnten nachweisen, dass makroadaptive Lernprogramme, wie der *Keller-Plan* (Keller & Sherman, 1974), im Vergleich zum herkömmlichen Unterricht zu besseren Lernergebnissen (Effektstärke $d = 0.50$) führen. Die Autoren weisen im Zusammenhang mit diesem Befund jedoch darauf hin, dass mit Hilfe solcher makroadaptiver Lernprogramme bislang die Fähigkeitsunterschiede zwischen den Schülern nicht beseitigt werden konnten, sondern auch nach der Bearbeitung solcher adaptiver Programmen bestehen bleiben. Niemiec und Walberg (1987) nehmen an, dass die Effektivität von computer-adaptivem Unterricht von dem Bildungsniveau abhängt. Die Autoren konnten zeigen, dass die Effektstärke des computer-adaptivem Unterrichts am College 0.26; an weiterführenden Schulen (*secondary schools*) 0.32; an Grundschulen 0.46 und an Sonderschulen 0.56 betrug.

In der *ATI*-Forschung können trotz zahlreicher Studien mit verschiedenen Personen- und Unterrichtsvariablen nicht zuletzt aufgrund methodischer Schwierigkeiten (vgl. Cronbach & Snow, 1981) nur wenige der *aptitude treatment*-Interaktionen als empirisch gesichert gelten: Viele der gefundenen Interaktionen konnten nur unzureichend repliziert werden. Dies wird häufig mit der Eigenschaft begründet, dass die in diesem Forschungsbereich untersuchten Interaktionen von Effekten generell schwieriger nachzuweisen bzw. zu replizieren sind als Haupteffekte (vgl. Stevens, 1992). Auch ist nicht zu erwarten, dass die in experimentellen Studien nachgewiesenen ATI-Effekte in naturalistischen Untersuchungen replizierbar sind. Corno und Snow (1986) kritisieren an der Forschung zum adaptiven Unterricht, dass in den meisten Untersuchungen einzelne Programme vorgestellt werden, als dass die Effekte dieser im Vergleich zum regulären Unterricht analysiert werden.

Zu den als gesichert geltenden Interaktionen zählt allerdings die Wechselwirkung zwischen kognitiven Fähigkeiten und dem kognitiven Anspruchsniveau des Unterrichts: Relativ wenig strukturierte und damit komplexe Unterrichtsarten kommen vor allem fähigen Schülern zugute, die ihre Kompetenzen flexibel zur Lösung der Probleme einsetzen können, während schwächere Schüler unter solchen Bedingungen zu scheitern drohen. Diese Schüler profitieren eher von einem Unterricht, der darauf abzielt, ihre Wissenslücken zu schließen (Treiber, 1982; Corno & Snow, 1986), der hochstrukturiert und durch klar definierte Aufga-

benstellungen des Lehrers gekennzeichnet ist (vgl. Snow, 1989). Wilen (1991) konnte die Bedeutung der Adaptivität des Schwierigkeitsniveaus der Aufgaben an die Fähigkeit der Schüler für deren Leistung nachweisen. Levin und Long (1981) zeigten zudem, dass sich Schüler nicht lediglich durch ihr Vorwissen und ihre individuellen Fähigkeiten unterscheiden, sondern dass auch die Bearbeitungsgeschwindigkeit und -dauer zwischen Schülern variiert. Diese werden als wichtige Einflussfaktoren für die Lernbeteiligung betrachtet. Die Autoren schlussfolgern, dass vom Lehrer die jeweils angemessenen Lernbedingungen so adaptiert werden sollten, dass sie optimal für die Lerneffektivität jedes einzelnen Schülers wirken können.

In neuerer Zeit wurde im Rahmen der PISA-Erhebung gezeigt, dass es deutschen Lehrern offensichtlich schwer fällt, Schüler individuell zu unterrichten und zu fördern (Rost et al., 2004). Eine besondere Problematik des in Deutschland vorherrschenden fragend-entwickelnden Unterrichts ist (Neubrand, 2002; vgl. Seidel, 2003, für den Physikunterricht), dass auf die individuellen Unterschiede der Schüler bzgl. der für sie angemessenen Aufgabenschwierigkeit in dieser Unterrichtsform nicht bzw. nur unzureichend eingegangen werden kann. Im Unterrichtsgespräch wird vielmehr zu einem Zeitpunkt lediglich eine Aufgabe gestellt, welche für alle Schüler gleich ist, und diese Aufgabe kann somit in ihrer Schwierigkeit nicht individuell an die jeweilige Schülerfähigkeit angepasst werden. Über das Zustandekommen der Entscheidung des Lehrers, welche Aufgabe mit welcher Schwierigkeit im Unterricht eingesetzt werden soll, liegen bislang keine empirischen Befunde vor. Auch die dieser Entscheidung zugrunde liegenden Mechanismen und Prozesse sind bislang kaum bekannt.

Mit Hilfe verschiedener Reviews zur Effektivität des Computer-Adaptiven Testens (CAT) konnte gezeigt werden, dass dies meist eine effektive Methode darstellt, um die individuellen Unterschiede in den Lernvoraussetzungen zu verringern (Corno & Snow, 1986; Snow, 1989; Fletcher, 1992). In Bezug auf die Kompensation individueller Unterschiede ist CAT oft erfolgreich, wenn es zusätzlich zum regulären Unterricht eingesetzt wird. Insgesamt kann festgehalten werden, dass die Effektivität des CAT von der Art des Einsatzes abhängt (Slavin, 1997). Bezüglich der motivationalen Aspekte beim Lernen mit Hilfe des computeradaptierten Testens konnten Betz und Weiss (1976) zeigen, dass durch CAT nicht nur leistungsstarke Personen motiviert werden, eine gute Testleistung zu zeigen, sondern auch leistungsschwächere Schüler.

12.4 Schlussfolgerung für Fragestellung der Studie 2

Trotz der z.T. unbefriedigenden Forschungsergebnisse kann zusammenfassend festgehalten werden, dass, wenn spezifische Charakteristika der Schüler und des Unterrichts berücksichtigt werden, adaptives Unterrichten und Testen eine Möglichkeit zur Leistungsverbesserung darstellen kann. Da sich identische Aufgaben für alle Schüler nicht als lernförderlich herausgestellt haben, erfolgt in Studie 2 eine individuelle Anpassung der Aufgabenschwierigkeit an den Fähigkeitsparameter nach der Idee des adaptiven Testens bzw. Unterrichtens (Corno & Snow, 1986; Kulik & Kulik, 1991; Snow, 1989). Obschon aufgrund der theoretischen Überlegungen zum adaptiven Unterricht angenommen werden kann, dass ein Zusammenhang zwischen dem kognitivem Anspruchsniveau des Unterrichts – erfasst über die Aufgabenschwierigkeit - und den Schülerfähigkeiten besteht, gibt es bislang kaum empirischen Untersuchungen, welche diesen Zusammenhang systematisch untersucht haben. Daher soll dieser vermutete, für die Leistungsfähigkeit der Schüler relevante Zusammenhang in der im Folgenden dargelegten Studie 2 systematisch untersucht werden. Dabei erfolgt die Adaptivität an die lernrelevanten Unterschiede zwischen den Schülern, indem die Schüler ihrem Fähigkeitsniveau entsprechende Aufgaben zur Bearbeitung erhalten. Aufgrund der Ergebnisse von Redfield und Rousseau (1981) sowie Crooks (1988) zum Zusammenhang der Dauer der Experimente und ihrer Effektstärke bekommen die Schüler in Studie 2 über einen längeren Zeitraum verteilt Übungsaufgaben vorgelegt, um den möglichen Effekt individuell angepasster unterschiedlicher Schwierigkeitsniveaus auf die Schülerleistung empirisch untersuchen zu können. Aus den inkonsistenten Ergebnissen der Forschung zur Schwierigkeit von Lernaufgaben (vgl. Huber, 1966; Nussbaum & Leutner, 1986; Sweller, 1976) kann bislang allerdings keine präzise Erwartung hinsichtlich der optimalen Aufgabenschwierigkeit getroffen werden. Daher soll in Studie 2 dieser Frage zunächst mit Hilfe dreifach gestufter Adaptationsbedingungen (individuell adaptierte Lösungswahrscheinlichkeit 30%, 50% und 80%) nachgegangen werden.

13 Fragestellung

Aufgrund der Ergebnisse der Studie 1 wurde in der anschließenden Studie 2 die Möglichkeit einer Leistungssteigerung der Schüler durch besser an ihre Fähigkeitsparameter angepasste Lernaufgaben über einen längeren Zeitraum hinweg, d.h. über eine Unterrichtsreihe von drei Monaten, untersucht. Durch eine „geschickte“ Auswahl von Lernaufgaben mit einem bestimmten Schwierigkeitsparameter sollte sich mit Hilfe der probabilistischen Testtheorie eine optimale Passung mit dem jeweiligen Fähigkeitsparameter der Schüler erreichen lassen.

In Studie 2 wurden die folgenden spezifischen Fragestellungen untersucht:

1. Können Schüler ihre Leistungen verbessern, wenn sie gezielt Aufgaben bearbeiten, die in Anlehnung an die Idee des adaptiven Testens an ihre Fähigkeit angepasst werden?
2. Lassen sich in Abhängigkeit von der Schwierigkeit der eingesetzten Übungsaufgaben die Schülerfähigkeiten unterschiedlich stark verändern?

Um die Fragestellungen der Studie 2 beantworten zu können, wurde eine Teilmenge der in Studie 1 konstruierten und skalierten Aufgaben, in einer Unterrichtsreihe in zwei Gymnasialklassen zu zwei Übungszeitpunkten eingesetzt. Während der Übungsphasen wurde in Anlehnung an Nussbaum und Leutner (1986) die Schwierigkeit der zu bearbeitenden Übungsaufgaben dreistufig - mit Lösungswahrscheinlichkeiten von 30%, 50% und 80% - experimentell variiert. Innerhalb der Untersuchungsgruppen wurde die Aufgabenschwierigkeit an die jeweilige Schülerfähigkeit individuell adaptiert, so dass sich die geforderten Lösungswahrscheinlichkeiten bei jedem einzelnen Schüler ergeben konnten. Somit wurden in Studie 2 drei Gruppen, welche sich in ihrer mittleren Lösungswahrscheinlichkeit (30%, 50% und 80%) unterschieden, untersucht.

Im Folgenden werden die Hypothesen zu den einzelnen Fragestellungen formuliert.

Hypothese 1

Es wird aufgrund der Effekte des adaptiven Unterrichts auf die Schülerleistung (vgl. Corno & Snow, 1986; Kulik & Kulik, 1991; Snow, 1989; Treiber, 1981) erwartet, dass sich der Fähigkeitsparameter der Schüler erhöhen lässt, wenn sie an ihre Fähigkeit angepasste Aufgaben bearbeiten. Die Aufgabenschwierigkeit muss auf die Lernvoraussetzungen der

Schüler abgestimmt sein, um eine lernförderliche Wirkung mit Hilfe der Aufgaben erzielen zu können (Brophy & Good, 1986; Schrader, 1989). Brophy (2000) schließt aus den bisherigen Forschungsergebnissen zur Optimierung des Unterrichts, dass Schüler ein hohes Niveau an Wissen erreichen können, wenn die Lernaktivitäten so schwierig sind, dass sie für die Schüler eine Herausforderung darstellen und ihr Wissen erweitern. Jedoch sollten die Aufgaben nicht zu schwierig sein, so dass die Schüler frustriert und verwirrt werden. Daher sollte der Unterricht nahe an Wygotskys (1978) Zone der proximalen Entwicklung erfolgen, welche in Kapitel 4.4 thematisiert wurde. In der vorliegenden Untersuchung ist mit dieser Zone der Wissens- und Fertigkeitsbereich gemeint, welchen die Schüler noch nicht selbstständig, jedoch mit Hilfe ihres Lehrers erreichen können (Brophy, 2000).

Hypothese 2

In Bezug auf die Veränderbarkeit des Fähigkeitsparameters, welche mit Hilfe individuell angepasster Aufgabenschwierigkeit erreicht werden soll, können in Bezug auf den Einfluss der unterschiedlichen Schwierigkeit der eingesetzten Aufgaben verschiedene Überlegungen angestellt werden.

Es ist zu vermuten, dass bei Schülern, die mäßig schwierige Aufgaben mit einer individuell angepassten Lösungswahrscheinlichkeit von 30% bearbeiten, eine Erhöhung des Personenparameters möglich sein wird. Nach Helmke (2004) sollte die Schwierigkeit der unterrichtlichen Anforderungen in einer Schwierigkeitszone liegen, die, in Anlehnung an die Zone der proximalen Entwicklung von Wygotsky (1978), oberhalb des aktuellen Wissensstands der Schüler liegt, aber nicht zu weit davon entfernt ist. Nach Wygotskys Terminologie ist der Entwicklungsstand der Zone der proximalen Entwicklung nicht für jeden Schüler identisch, sondern individuell unterschiedlich. Somit muss die Aufgabenschwierigkeit adaptiv an die Schülerfähigkeit angepasst werden, d.h. für verschiedene Schüler ist die gleiche Aufgabe unterschiedlich schwer, und so muss, um allen Schülern mäßig schwierige Aufgaben zur Bearbeitung aufzugeben, die Aufgabenschwierigkeit individuell, in Abhängigkeit von der jeweiligen Schülerfähigkeit, angepasst werden. Auch Doyle (1988) betrachtet schwierige Aufgaben als geeignetes Mittel zur Förderung von Verständniswissen und Problemlösekompetenzen.

Als weiteren Beleg für die Annahme einer möglichen Effektivität mäßig schwieriger Aufgaben können die Ergebnisse zu Leistungsunterschieden in Bezug auf *low level*- und *high level*-Aufgaben herangezogen werden (vgl. Hamaker, 1986; Redfield & Rousseau, 1981; Rosenshine, 1971; Samson et al., 1987; Winne, 1979). Levin (2005) fasst die Forschungsergebnisse zum Zusammenhang zwischen Aufgabenniveau und Schülerleistung da-

hingehend zusammen, dass insgesamt Aufgaben auf einem höheren kognitiven Niveau zu besseren Lernergebnissen führen als Aufgaben auf kognitiv niedrigerem Niveau. Aufgrund der Forschungsergebnisse ist anzunehmen, dass die Aufgaben, welche überwiegend in der 30%-Gruppe eingesetzt werden, auf dem *high level*-Niveau zu klassifizieren sind und somit zu besseren Lernergebnissen führen. Im Gegensatz dazu sollten die *low level*-Aufgaben, welche vermehrt in der 80%-Gruppe eingesetzt werden und damit als überwiegend leicht zu klassifizieren sind, einen weniger leistungsförderlichen Effekt haben.

Im Gegensatz dazu legen andere Forschungsergebnisse zur optimal lernförderlichen Aufgabenschwierigkeit, wie z.B. die Untersuchung von Nussbaum und Leutner (1986), die Vermutung nahe, dass die Schülerfähigkeit besonders gut mit Hilfe leichter Aufgaben erhöht werden kann (vgl. Kapitel 3.3). Diese Annahme steht im Einklang mit Ergebnissen lerntheoretischer Untersuchungen (vgl. Fürntratt, 1978; Huber, 1966; Kloep & Weimann, 1982; Sweller, 1976). Auch aufgrund der Ergebnisse der hier dargestellten Studie 1, welche gezeigt hat, dass im Unterricht schwierige und z.T. sehr schwierige Aufgaben eingesetzt werden, welche im Mittel nur von einigen Schülern einer Klasse gelöst werden, ist zu erwarten, dass Schüler mit Hilfe leichter bis mittelschwieriger Aufgaben (80% bzw. 50% Lösungswahrscheinlichkeit) bessere Leistungen erbringen und somit eine Erhöhung des Fähigkeitsparameters möglich sein sollte.

Aufgrund dieser insgesamt widersprüchlichen Ergebnisse wird die Fragestellung bzgl. der Effektivität der verschiedenen Aufgabenschwierigkeiten im Hinblick auf die Veränderbarkeit der Schülerfähigkeit als offene Untersuchungsfrage formuliert. Es können keine gerichteten Hypothesen in Bezug auf mögliche Unterschiede in den drei Untersuchungsgruppen aufgestellt werden. Obschon der Zusammenhang zwischen kognitivem Anspruchsniveau des Unterrichts und den Schülerfähigkeiten als belegt angesehen werden kann (vgl. Kunter et al., 2006; Slavin, 1997), gibt es bislang wenige empirische Untersuchungen zur Beziehung zwischen adaptivem Aufgabeneinsatz bzw. der Variation der Aufgabenschwierigkeit im Unterricht in Abhängigkeit von der individuellen Schülerfähigkeit (vgl. Nussbaum & Leutner, 1986).

14 Methode

Die Beschreibung der Methode erfolgt in den folgenden Unterkapiteln. Zunächst wird die Stichprobe beschrieben (Kapitel 14.1), dann das Design der Studie mit der Durchführung (Kapitel 14.2). Im Anschluss wird die adaptive Auswahl der Aufgabenschwierigkeit anhand des individuellen Fähigkeitsparameters als zentrales Merkmal dieser Studie beispielhaft dargelegt (Kapitel 14.3), bevor im letzten Unterkapitel (Kapitel 14.4) auf die Datenauswertung eingegangen wird.

14.1 Stichprobe

Die Untersuchungsstichprobe bestand aus zwei Schulklassen des 10. Schuljahrs zweier Gymnasien. Es wurden Gymnasialklassen ausgewählt, da sich in Studie 1 bei diesen Klassen gezeigt hatte, dass die Passung insbesondere im oberen Bereich der Aufgabenschwierigkeit gering ist. Die beiden Lehrer der zwei Klassen hatten bereits an der Erhebung 1 des Projekts teilgenommen und erklärten sich bereit, mit einer weiteren Klasse an dieser Untersuchung teilzunehmen. Die Datenerhebung fand im 2. Halbjahr des Schuljahrs 2005/2006 statt.

In der Datenauswertung wurden nur vollständige Datensätze berücksichtigt, so dass die Daten von 43 Schülern in die Auswertung eingingen. Die Stichprobe setzt sich aus 20 Mädchen und 23 Jungen zusammen. Da keine demographischen Daten der Schüler erhoben wurden, kann keine Aussage zum durchschnittlichen Alter der Schüler gemacht werden.

Die Untersuchungsgruppen unterscheiden sich in der individuell angepassten mittleren Lösungswahrscheinlichkeit der in den Übungsphasen zu bearbeitenden Aufgaben, diese betrug 30%, 50% oder 80%. Auf eine Kontrollgruppe, die keine Lernaufgaben zu bearbeiten hatte, wurde verzichtet, da es aufgrund zahlreicher empirischer Befunde hinreichend gesichert ist, dass eine Beantwortung von Lernaufgaben, im Vergleich zu keiner Aufgabenbearbeitung, einen positiven Effekt auf die Leistung von Lernenden ausübt (Duchastel, 1981; Hamaker, 1986; Häfele, 1995).

Jeder Schüler bekam nach Abschluss der gesamten Datenerhebung 10 € für die Teilnahme an der Studie ausgehändigt. Diese Probandenvergütung war zuvor nicht angekündigt worden, wodurch mögliche Motivationseffekte aufgrund des Geldes ausgeschlossen werden können.

14.2 Design und Durchführung

Die Studie 2 wurde im Sinne eines *Mehr-Gruppen-Pretest-Posttest-Designs* konzipiert (vgl. Cook & Campbell, 1979). Der Versuchsaufbau ist in Abbildung 15 zur Verdeutlichung graphisch dargestellt.

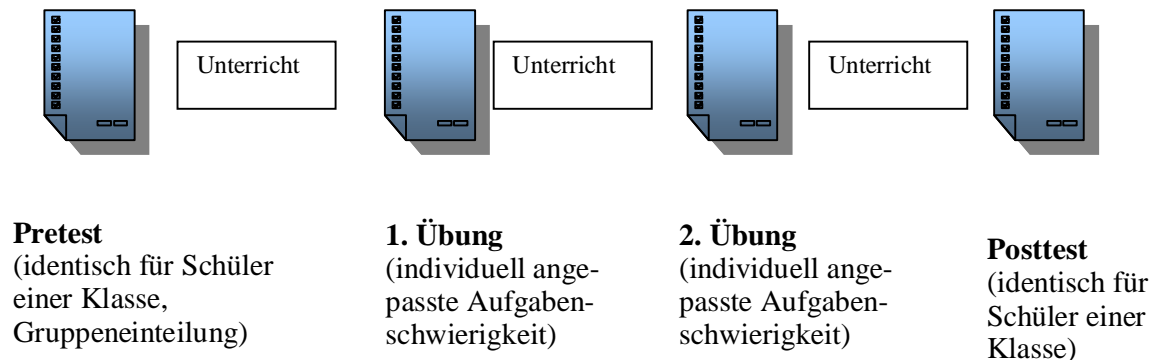


Abbildung 15: Versuchsaufbau der Studie 2

Die experimentell variierte Variable war die individuell angepasste Lösungswahrscheinlichkeit der zu bearbeitenden Übungsaufgaben, die 30%, 50% und 80% betrug. Die Aufgaben mit einer Lösungswahrscheinlichkeit von 30% werden als mäßig schwierig definiert, die Aufgaben mit einer Lösungswahrscheinlichkeit von 50% werden als mittelschwierig angesehen, und die Aufgaben mit einer Lösungswahrscheinlichkeit von 80% werden als leicht definiert. Die Aufgaben sind z.T. dem in der Studie 1 IRT-skalierten Aufgabenpool entnommen. Zudem wurden gleichermaßen neue Aufgaben konstruiert, um eine genügend große Anzahl von Aufgaben zu erhalten, welche den unterrichteten Inhaltsbereich thematisieren. Aufgrund der Ergebnisse zur Konstruktion von Aufgaben unterschiedlicher Schwierigkeit in Studie 1 (vgl. Kapitel 10.3.1) wurde in Studie 2 dieselbe Konstruktionsmethode erneut eingesetzt, um weitere inhaltsbezogene Aufgaben entwickeln zu können. Da die Schätzung der Itemparameter erst nach der Bearbeitung durch die Schüler erfolgen konnte, wurde das kognitive Anspruchsniveau als Richtlinie zum adaptiven Einsatz dieser neu konstruierten Aufgaben genutzt. Demnach wurden „Anwendungsaufgaben“ in der 30%-Gruppe besonders bei Schülern mit einem hohen Fähigkeitsparameter eingesetzt. Im Gegensatz dazu wurden beispielsweise „Verstehensaufgaben“ in der 30%-Gruppe eher Schülern mit einem geringeren Fähigkeitsparameter zugeordnet. In der 50%-Gruppe wiederum wurden einem Schüler mit einem hohen Fähigkeitsparameter hauptsächlich Verstehensaufgaben vorgelegt,

einem Schüler mit einem geringeren Fähigkeitsparameter sowohl Wissens- als auch Verstehensaufgaben. In der Untersuchungsgruppe mit der mittleren Lösungswahrscheinlichkeit von 80% erhielt ein Schüler mit einem hohen Fähigkeitsparameter wiederum sowohl Wissens- als auch Verstehensaufgaben, wohingegen ein Schüler mit einem geringen Fähigkeitsparameter fast ausschließlich Wissensaufgaben zur Bearbeitung vorgelegt bekam. Im Anschluss an die IRT-Skalierung dieser neuen Aufgaben erfolgte ein Treatment-Check, indem die aufgrund der Fähigkeits- und Itemparameter theoretisch erwartbaren Lösungswahrscheinlichkeiten berechnet und mit den beobachteten relativen Lösungshäufigkeiten verglichen wurden (vgl. Kapitel 15.1). Eine Auflistung der Aufgaben, die in den einzelnen Tests enthalten sind (Pretest, Posttest und Übungsaufgaben zu den Zeitpunkten 1 und 2) befindet sich im Anhang G, welcher sich auf der beiliegenden CD befindet.

Die Pretestleistung wurde über den zuvor unterrichteten Inhaltsbereich nach Abschluss der vorangegangenen Unterrichtsreihe zu Beginn des 2. Halbjahrs des 10. Schuljahrs mittels eines Leistungstests erhoben. Dieser Pretest setzt sich ausschließlich aus Aufgaben zusammen, welche in Studie 1 konstruiert und skaliert worden waren. Basierend auf den Testergebnissen wurde der Fähigkeitsparameter jedes Schülers auf der Grundlage der *Item-Response*-Theorie skaliert. Im Anschluss daran wurden die Schüler aufgrund ihres Fähigkeitsparameters den drei Untersuchungsgruppen (30% ,50% und 80% Lösungswahrscheinlichkeit) so zugeordnet, dass sich die Untersuchungsgruppen im Pretest nicht im mittleren Fähigkeitsparameter unterscheiden. Diese Zuordnung wurde mit Hilfe einer einfaktoriellen Varianzanalyse überprüft, wobei der Fähigkeitsparameter als abhängige Variable diente und die Gruppenzugehörigkeit als unabhängige Variable; ($F_{(2, 41)} < 1$). Es besteht somit kein signifikanter Unterschied im mittleren Pretest-Fähigkeitsparameter zwischen den Untersuchungsgruppen. In den drei Gruppen befanden sich etwa gleich viele Schüler mit $N_{30\%} = 16$ Schülern, $N_{50\%} = 12$ Schülern und $N_{80\%} = 15$ Schülern.

Die Tatsache, dass die eingesetzte Aufgabenmenge Rasch-skaliert ist, macht sie für die Beantwortung der Fragestellung besonders geeignet: Ist auch der Fähigkeitsparameter einer Person bekannt, kann für eine zu bearbeitende Aufgabenmenge die Lösungswahrscheinlichkeit berechnet werden. Auf diese Weise lässt sich die Schwierigkeit der Aufgaben gezielt auf individuellem Niveau experimentell variieren. Mit der Formel $p = e^{(f-s)} / (1 + e^{(f-s)})$ lässt sich die Lösungswahrscheinlichkeit durch Einsetzen der Werte für den Fähigkeitsparameter und den Itemparameter berechnen. Anhand dieses Vorgehens wurden jeweils die bereits in Studie 1 skalierten Aufgaben, welche von den einzelnen Schülern in den Übungsphasen bearbeitet werden sollten, ausgewählt.

Zwischen dem Pretest und der ersten Übungsphase fand regulärer Unterricht zu einem neuen Thema statt. Dieser physikalische Inhaltsbereich unterschied sich in den beiden untersuchten Klassen. In der einen Klasse wurde vom Lehrer die Atomphysik, in der anderen Klasse die Halbleiter-Elektronik unterrichtet. Die Testleiterin nahm als teilnehmende Beobachterin an diesen Unterrichtsstunden teil. Den beiden Lehrern wurden keine Vorgaben zu den von ihnen gehaltenen Unterrichtsstunden gemacht. Zwischen den einzelnen Testungen bestand ein Zeitraum von vier Wochen mit drei Unterrichtsstunden pro Woche. Dies entspricht 12 Physikstunden zwischen den einzelnen Übungsphasen bzw. Erhebungen.

Die Schüler bearbeiteten eigenständig und schriftlich in der zuvor nicht angekündigten Übungsphase Aufgaben zu dem jeweils unterrichteten Inhaltsbereich, deren Aufgabenschwierigkeit individuell an die Fähigkeitsparameter angepasst wurde. Die Übungsphasen umfassten jeweils eine 45 minütige Unterrichtsstunde. Zwischen der 1. und 2. Übungsphase erfolgten ebenfalls vier Wochen bzw. 12 Stunden regulärer Physikunterricht des Lehrers zum jeweiligen Inhaltsbereich. Das für die erste Übungsphase beschriebene Vorgehen wurde auch in der 2. Übungsphase angewendet. Zwischen der 2. Übungsphase und der Posttest-Erhebung wurden wiederum 12 Unterrichtsstunden vom Lehrer gehalten. Nach Abschluss der jeweiligen Unterrichtsreihe bearbeiteten die Schüler einer Klasse einen identischen Posttest, der thematisch den gesamten unterrichteten Inhaltsbereich abdeckte.

14.3 Instrumente – Adaptive Anpassung der Aufgabenschwierigkeit an die individuelle Schülerfähigkeit

Da die Adaptation der Aufgabenschwierigkeit in den beiden Übungsphasen für jeden Schüler individuell erfolgte und das zentrale Merkmal der Studie darstellt, soll dies beispielhaft für zwei Schüler, welche der 50%-Gruppe zugeordnet wurden, in Tabelle 26 dargestellt werden. Die Zuordnung der jeweiligen Aufgaben in den Übungsphasen für alle weiteren Schüler ist in Anhang D zu finden, welcher sich auf der beiliegenden CD befindet.

Tabelle 26: Aufgabenauswahl für die Übungsphasen für zwei Schüler der 50%-Gruppe

Personen- Parameter (f)	Gruppe	1.Übung		2.Übung	
		Items	Item- Parameter (s)	Items	Item- Parameter (s)
-0.856	5	L313	-2.014	L51a	-1.055
		L513	-1.719	L521b	-1.419
		L81	-1.910	L524a	-0.132
		L83	0.604	L525a	-2.075
		L84	-0.990	L526	-1.417
		L85	1.041	L527	-0.507
		L86a	-0.991	L821	-1.856
		L86b	-0.788	L822	-1.012
		L86c	0.784	L827	-1.861
		L87	-0.992	L828	-1.862
		L88	-1.915	L829	-0.345
		L89	-1.915	L8210	-0.345
		L810	0.603	L8211	-1.019
L811	-2.552				
Mittlerer					
Itemparameter			-0.911		-1.197
Lösungswahrscheinlichkeit			51.4%		58.4%
1.334	5	L32	3.856	L323	0.848
		L33	0.562	L324	2.926
		L312	0.796	L326b	2.358
		L52	0.354	L327	0.985
		L56a	0.669	L328	2.064
		L57	1.302	L329	2.402
		L58	2.047	L3212	2.190
		L59	2.046	L522	0.816
		L511	3.239	L524b	1.127
		L82	2.654	L5212	0.776
		L33	0.604	L822	-1.012
		L85	1.041	L824	0.252
		L86c	0.784	L829	-0.345

<i>Personen- Parameter (f)</i>	<i>Gruppe</i>	<i>1.Übung</i>		<i>2.Übung</i>	
		<i>Items</i>	<i>Item- Parameter (s)</i>	<i>Items</i>	<i>Item- Parameter (s)</i>
		L810	0.603	L8211	-1.019
Mittlerer Itemparameter			1.468		1.260
Lösungswahrscheinlichkeit			46.7%		51.8%

Anmerkung:

^a berechnet mit Hilfe der Formel: $p = e^{(f-s)} / (1 + e^{(f-s)})$

Tabelle 26 verdeutlicht, dass Schüler mit unterschiedlichen Personenparametern (-0.856 bzw. 1.334), die sich in derselben Untersuchungsgruppe (50% Lösungswahrscheinlichkeit) befanden, unterschiedlich schwierige Aufgaben in den Übungsphasen zur Bearbeitung vorgelegt bekamen. Damit konnte gewährleistet werden, dass sie die Aufgaben der jeweiligen Übungsphase mit einer mittleren Lösungswahrscheinlichkeit von etwa 50% richtig lösten. Dabei variierte der Itemparameter zwischen den zu bearbeitenden Aufgaben, um mögliche Frustrations- bzw. Misserfolgseffekte, besonders in der 30%-Gruppe, zu vermeiden.

14.4 Auswertung

Die Aufgaben der Leistungstests wurden, wie in Studie 1, mit Hilfe von Lösungsbeispielen ausgewertet, um so die Auswertungsobjektivität des offenen Antwortformats zu erhöhen (siehe Lienert & Raatz, 1998). Dabei wurde der Wert „null“ bzw. kein Punkt für eine falsche Antwort und für eine nicht-bearbeitete Aufgabe vergeben. Der Wert „eins“ bzw. ein Punkt wurde für eine richtige bzw. ausreichende Antwort im Sinne der Beispiellösungen vergeben. Aus allen im Leistungstest erhaltenen Punkten wurde der Summenscore gebildet, der das Gesamtergebnis des Tests darstellt. Die beobachteten relativen Lösungshäufigkeiten der Aufgaben wurde mit Hilfe der klassischen Testverfahren berechnet; sie sind als Treatment-Check für die Adaptation der Itemschwierigkeit an die individuellen Fähigkeitsparameter relevant.

Die Personenparameter wurden nach der 1. Übungsphase nochmals geschätzt, um mögliche Veränderungen zwischen dem Pretest und der 1. Übungsphase identifizieren und diese mögliche Veränderung des Personenparameters für die 2. Übungsphase zur Anpassung

der Aufgabenschwierigkeit berücksichtigen zu können. In Tabelle 27 sind die deskriptiven Ergebnisse der Personenparameter-Schätzungen zu den Testzeitpunkten Pretest, 1. und 2. Übungsphase angegeben.

Tabelle 27: Personenparameter bei Pretest sowie 1. und 2. Übungsphase

	<i>Gruppenzugehörigkeit</i>	<i>Mittelwert</i>	<i>Standard- abweichung</i>	<i>N</i>
Pretest- Personenparameter	30% Lösungswahrscheinlichkeit	.301	1.003	16
	50% Lösungswahrscheinlichkeit	.122	.804	12
	80% Lösungswahrscheinlichkeit	.090	.802	15
1.Übung- Personenparameter	30% Lösungswahrscheinlichkeit	.394	.859	16
	50% Lösungswahrscheinlichkeit	.421	.907	12
	80% Lösungswahrscheinlichkeit	.221	.846	15
2.Übung- Personenparameter	30% Lösungswahrscheinlichkeit	.291	.780	16
	50% Lösungswahrscheinlichkeit	.156	.652	12
	80% Lösungswahrscheinlichkeit	.100	.664	15

Im Vergleich zwischen der Pretest- und 1. Übungs-Schätzung der Personenparameter zeigen sich keine signifikanten Unterschiede ($F_{(1, 40)} = 2.47$; $p = .124$; $\eta^2 = .058$), so dass die Aufgabenauswahl für die 2. Übung ebenfalls mit Hilfe der Pretest-Personenparameter erfolgen konnte. Eine 3. Schätzung des Personenparameters anhand der Testergebnisse der 2. Übungsphase ist nicht notwendig, da die Aufgaben des Posttests nicht individuell angepasst werden, sondern - wie beim Pretest - die Aufgaben für alle Schüler einer Klasse identisch sind. Zur Absicherung der Ergebnisse wird jedoch auch diese 3. Schätzung der Personenparameter durchgeführt. Zwischen den Erhebungen zur 1. und 2. Übungsphase zeigt sich keine bedeutsame Veränderung im Personenparameter: $F_{(1, 40)} = 3.96$; $p = .053$; $\eta^2 = .090$.

Insgesamt wurden die in den Übungsphasen adaptierten Aufgaben basierend auf dem im Pretest geschätzten Personenparameter individuell angepasst. Es erfolgte somit, in Anlehnung an Nussbaum und Leutner (1986), eine Adaptation anhand des initialen Personenparameters.

15 Ergebnisse

In der Darstellung der Ergebnisse wird zunächst überprüft, inwieweit die theoretisch berechneten Lösungswahrscheinlichkeiten mit den tatsächlichen Lösungshäufigkeiten in den verschiedenen Untersuchungsgruppen während der beiden Übungsphasen übereinstimmen (Kapitel 15.1). Im Anschluss daran wird die Rasch-Skalierung der Aufgaben (Kapitel 15.2.1) und der Personenparameter in den verschiedenen Gruppen (Kapitel 15.2.2) dargestellt.

15.1 *Treatment-Check*

Es wird überprüft, inwieweit die in den drei experimentellen Gruppen beobachtete durchschnittliche relative Lösungshäufigkeit der Übungsaufgaben mit der im experimentellen Design beabsichtigten, d.h. der erwarteten durchschnittlichen Lösungswahrscheinlichkeit übereinstimmt. Dazu werden, mit Hilfe der klassischen Testtheorie, die relativen Lösungshäufigkeiten innerhalb der verschiedenen Gruppen für die jeweils bearbeiteten Aufgaben berechnet und gemittelt. Die Analyseeinheit war die einzelne Aufgabe bzw. deren relative Lösungshäufigkeit. Die Ergebnisse für die drei Gruppen zu beiden Übungsphasen sind in Abbildung 16 mit Mittelwert und Standardabweichung dargestellt. Dabei zeigte sich eine zufriedenstellende bis gute Übereinstimmung zwischen den theoretisch erwarteten Lösungswahrscheinlichkeiten und den tatsächlich beobachteten relativen Lösungshäufigkeiten.

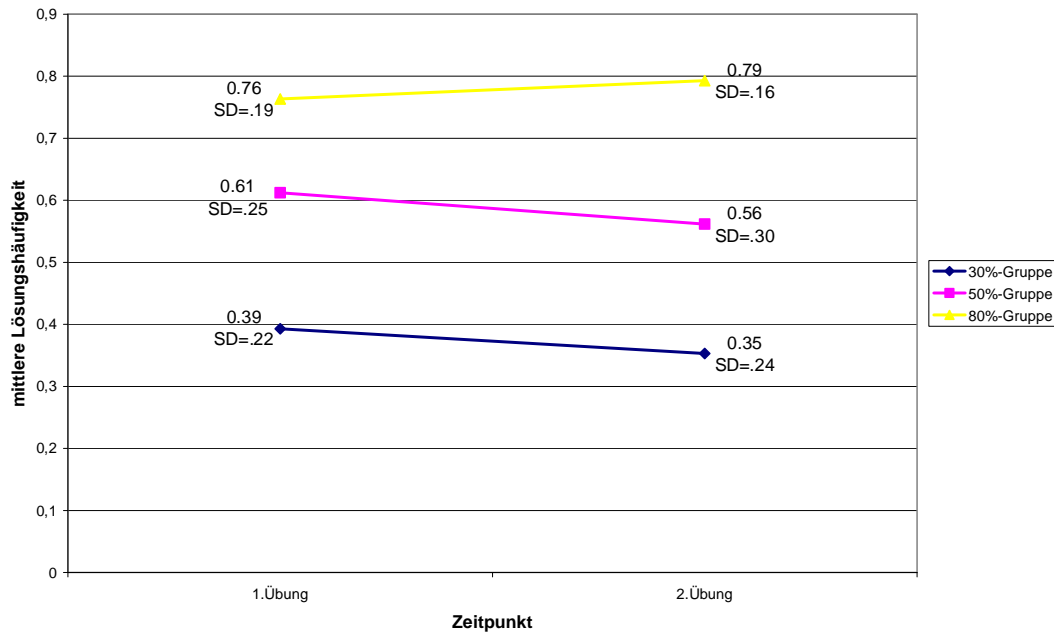


Abbildung 16: Mittlere relative Lösungshäufigkeiten in den Untersuchungsgruppen zu den Übungszeitpunkten

Die zugehörige zweifaktorielle Varianzanalyse mit der Gruppenzugehörigkeit (30%-, 50%-, 80%-Lösungswahrscheinlichkeit) und dem Übungszeitpunkt (1. und 2. Übung) als unabhängige Variablen und der tatsächlichen, relativen Lösungshäufigkeit der Übungsaufgaben als abhängige Variable bestätigt die deskriptiven Ergebnisse. Die in der Tabelle 28 dargestellten Lösungshäufigkeiten entsprechen den in Abbildung 16 dargestellten. Die in den Übungsphasen eingesetzten Aufgaben bzw. deren relative, klassisch berechnete Lösungshäufigkeit stellen dabei die Analyseeinheit dar. Es zeigt sich ein Effekt der Gruppenzugehörigkeit: $F(2, 141) = 37.133$, $p = .001$, $\eta^2 = .345$. Im Gegensatz dazu wird für den Übungszeitpunkt kein signifikanter Effekt festgestellt ($F(1, 141) < 1$). Somit unterscheiden sich die mittleren Lösungshäufigkeiten innerhalb der Gruppen zwischen den beiden Übungszeitpunkten nicht signifikant. Auch der Interaktionsterm von Gruppe und Übungszeitpunkt ist nicht signifikant ($F(2, 141) < 1$). Die Ergebnisse der Varianzanalyse sind in Tabelle 29 dargestellt.

Tabelle 28: Mittlere relative Lösungshäufigkeiten der Aufgaben in den Untersuchungsgruppen zu den Übungszeitpunkten

Gruppe	Übungszeitpunkt	Mittelwert	Standardabweichung	N
30% Lösungswahrscheinlichkeit	1.Übung	.393	.219	30
	2.Übung	.353	.244	24
50% Lösungswahrscheinlichkeit	1.Übung	.612	.248	24
	2.Übung	.562	.303	24
80% Lösungswahrscheinlichkeit	1.Übung	.763	.191	24
	2.Übung	.793	.159	21

Tabelle 29: Effekte der Gruppenzugehörigkeit und des Übungszeitpunktes auf die relative Lösungshäufigkeit der Aufgaben

Quelle	Quadratsumme vom Typ III	df	Varianz	F	Signifikanz	Partielles Eta-Quadrat
Gruppe	4.023	2	2.012	37.133***	.001	.345
Übungszeitpunkt	.015	1	.015	.277	.599	.002
Übungszeitpunkt * Gruppe	.044	2	.022	.404	.668	.006
Fehler	7.639	141	.054			
Gesamt	59.010	147	.401			

Anmerkung:

*** Der Wert ist auf dem Niveau von 0.001 (2-seitig) signifikant

Zur weiteren Überprüfung des Treatments wurde eine Varianzanalyse mit Messwiederholung durchgeführt (vgl. Tabellen 30 und 31), wobei die relative Lösungshäufigkeit pro Schüler als abhängige Variable diente. Demnach stellt bei dieser Berechnung der einzelne Schüler bzw. seine durchschnittliche Aufgabenlösung die Analyseeinheit dar, im Gegensatz dazu war bei der oben dargestellten zweifaktoriellen ANOVA die Aufgabe die Analyseeinheit. Die Lösungshäufigkeit wurde definiert als die Summe der richtigen Lösungen dividiert durch die Anzahl der zu bearbeiteten Aufgaben und für beide Übungszeitpunkte pro Schüler bestimmt. Der Übungszeitpunkt stellt den Innersubjektfaktor und die Gruppenzugehörigkeit den Zwischensubjektfaktor dar.

Tabelle 30: Mittlere relative Lösungshäufigkeiten der Schüler in den Untersuchungsgruppen zu den Übungszeitpunkten

Gruppe	Gruppe	Mittelwert	Standardabweichung	N
1.Übungsphase	30% Lösungswahrscheinlichkeit	.323	.074	16
	50% Lösungswahrscheinlichkeit	.542	.143	12
	80% Lösungswahrscheinlichkeit	.771	.117	15
2.Übungsphase	30% Lösungswahrscheinlichkeit	.312	.071	16
	50% Lösungswahrscheinlichkeit	.567	.082	12
	80% Lösungswahrscheinlichkeit	.793	.090	15

Tabelle 31: Effekte der Übungszeitpunkte und der Gruppenzugehörigkeit auf die relative Lösungshäufigkeit der Schüler

Quelle	Quadratsumme vom Typ III	df	Varianz	F	Signifikanz	Partielles Eta-Quadrat
Zwischensubjekt-effekt						
Gruppe	3.343	2	1.672	148.566***	.001	.889
Fehler	.450	40	.011			
Innersubjekteffekt						
Übungszeitpunkt	.003	1	.003	.393	.534	.010
Übungszeitpunkt * Gruppe	.006	2	.003	.368	.695	.018
Fehler	.317	40	.008			
Gesamt	.326	43	.008			

Anmerkung:

*** Der Wert ist auf dem Niveau von 0.001 (2-seitig) signifikant

Erwartungskonform zeigt sich ein signifikanter Zwischensubjekteffekt für die Variable Gruppenzugehörigkeit: $F(2, 40) = 148.566$; $p < .001$; $\eta^2 = .889$. Das bedeutet, dass die jeweils an den individuellen Fähigkeitsparametern adaptierte Aufgabenschwierigkeit zu signifikant unterschiedlichen Lösungshäufigkeiten zwischen den Gruppen führt. Konform zur Zielsetzung des Treatments ergibt sich kein signifikanter Innersubjekteffekt für den Übungszeitpunkt ($F(2, 40) < 1$), somit unterscheidet sich die mittlere Lösungshäufigkeit eines Schülers zwischen den beiden Übungszeitpunkten nicht signifikant. Auch die Interaktion von Übungszeitpunkt und Gruppe ist nicht signifikant: $F(1, 40) < 1$.

Insgesamt kann das Ergebnis als Beleg für eine „erfolgreiche“ Adaptation der Übungsaufgaben betrachtet werden, da die Schüler der verschiedenen Untersuchungsgruppen die ihnen vorgelegten Aufgaben mit einer signifikant unterschiedlichen mittleren Lösungshäufigkeit lösen. Als Fazit dieses Auswertungsteils ist festzuhalten, dass mit Hilfe der eingesetzten Methode zur individuellen Adaptation der Aufgabenschwierigkeit erreicht wurde, dass die theoretisch berechneten Lösungswahrscheinlichkeiten in den verschiedenen Untersuchungsgruppen mit den tatsächlichen Lösungshäufigkeiten zu beiden Übungszeitpunkten gut übereinstimmen.

15.2 Raschskalierung

In den folgenden Unterkapiteln werden die Schätzung der Item- und Personenparameter dargestellt, wobei auf den Vergleich bzw. die Veränderungen zwischen Pretest und Posttest eingegangen wird.

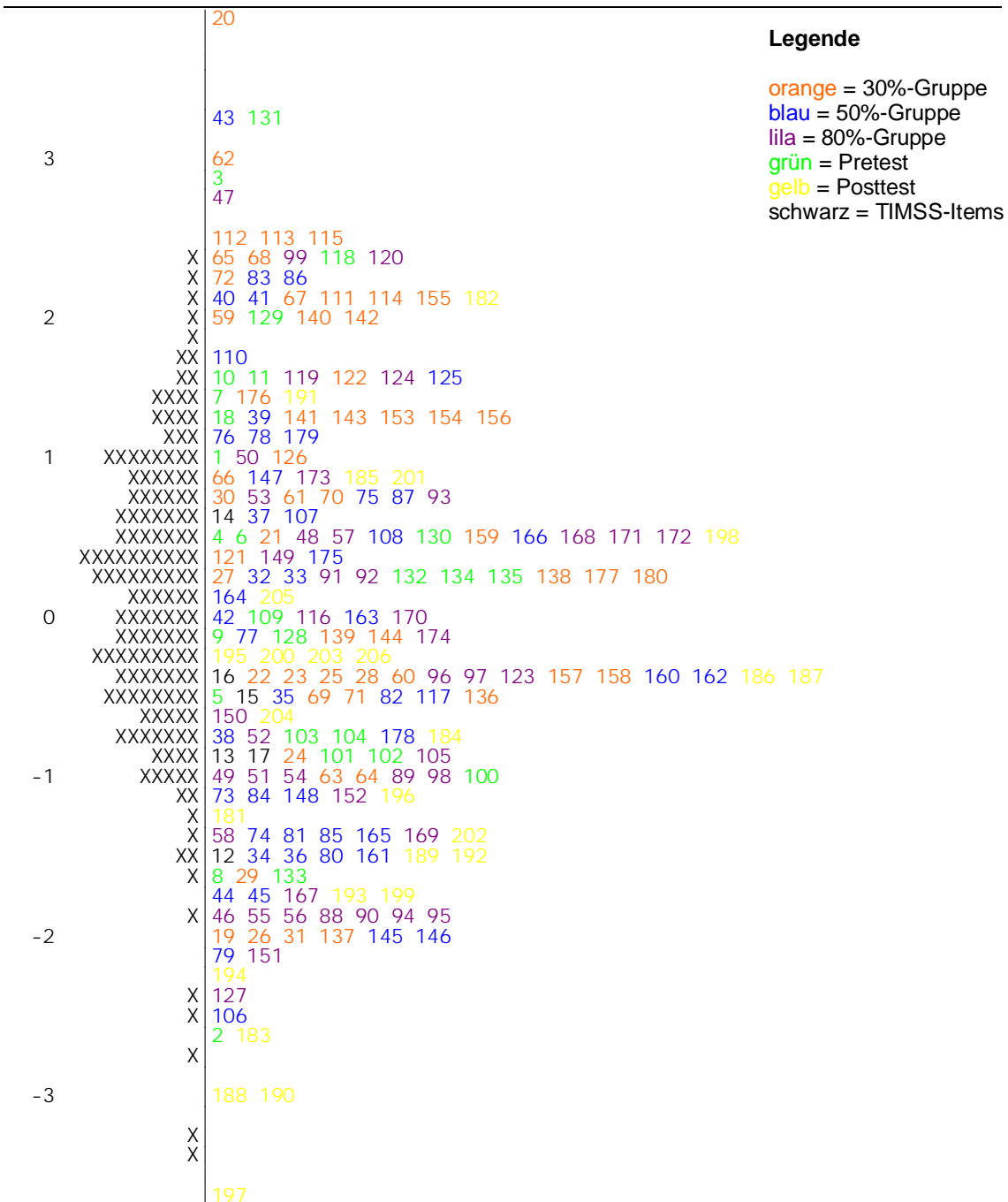
15.2.1 Schätzung der Itemparameter

Da die Anzahl der vorhandenen inhaltsbezogenen Aufgaben aus Studie 1 nicht ausreichend war, mussten weitere, unterschiedlich schwierige Aufgaben konstruiert werden. Ihre Itemparameter wurden gemeinsam mit den Items, deren Schwierigkeit bereits im Rahmen der Studie 1 skaliert wurde, geschätzt. Dabei wurden die Itemparameter der in Studie 1 skalierten Aufgaben in ihren Schwierigkeitsparametern fixiert und die zusätzlich für die Studie 2 konstruierten Aufgaben frei geschätzt (vgl. Rost, 2004).

Insgesamt erfolgte die Skalierung ähnlich wie in Studie 1 (vgl. Kapitel 9.5 und 10.3.1). Daher wird im Folgenden lediglich auf Besonderheiten bzw. Unterschiede im Vorgehen eingegangen. Zunächst erfolgte die Schätzung der Aufgabenschwierigkeiten für die neu konstruierten Aufgaben. Dieses iterative Verfahren wurde mit jeder weiteren Datenerhe-

bung (1.Übungsphase, 2.Übungsphase und Posttest) wiederholt, d.h. zunächst wurden die aus Studie 1 übernommenen Items auf ihre Schwierigkeitsparameter fixiert und die Fähigkeitsparameter des Pretests geschätzt. Im nächsten Schritt wurden die aus Studie 1 übernommenen und in den Übungsphasen eingesetzten Items fixiert, und die z.T. neu konstruierten Aufgaben der 1. Übungsphase wurden frei geschätzt. Dieses Vorgehen wurde so lange wiederholt, bis alle Aufgaben in ihrer Schwierigkeit skaliert waren.

Zur Kontrolle bzw. Aufdeckung möglicher Veränderungen der Personenparameter wurden diese bei jeder Erhebung ebenfalls neu geschätzt und mit der vorherigen Schätzung verglichen, so dass insgesamt vier Schätzungen der Personenparameter vorliegen. Es zeigten sich während der Übungsphasen keine signifikanten Veränderungen im Personenparameter, die Ergebnisse der dazu durchgeführten Varianzanalysen wurden bereits in Kapitel 14.4 dargestellt. Die Skalierung der in Studie 2 eingesetzten Items und Personen zeigt die in Abbildung 17 dargestellte Verteilung.

**Anmerkung:**

Jedes X repräsentiert 1.1 Fälle.

Abbildung 17: Skalierung der Aufgaben- und Personenstichprobe

Die Itemparameter der eingesetzten Aufgaben streuen im Bereich zwischen 4 und -3.5 Logits mit einem fixierten Mittelwert von 0.000 Logits und einer Standardabweichung von 1.554. Die in den verschiedenen Untersuchungsgruppen in den beiden Übungsphasen adaptiv eingesetzten Aufgaben sind in Abbildung 17 durch verschiedene Farben markiert. Dabei zeigt sich erwartungsgemäß keine eindeutige Abgrenzung der eingesetzten Aufgaben zwischen den Untersuchungsgruppen, denn eine Person in der 30%-Gruppe mit einem geringen Personenparameter bekam leichtere Aufgaben in der Übungsphase vorgelegt als eine Person in derselben 30%-Gruppe mit einem höheren Personenparameter. Daher variiert die Aufgabenschwierigkeit auch innerhalb der drei Untersuchungsgruppen.

In allen vier Erhebungen wurden die gleichen TIMSS-Items eingesetzt. Damit war es möglich, die Aufgaben und Personen trotz der unterschiedlichen Inhaltsbereiche und der verschiedenen, von den Schülern bearbeiteten Aufgaben zu vergleichen. Dieses Vorgehen hat sich in Studie 1 bewährt. Aufgrund der wiederholten Darbietung der TIMSS-Items müssen Lerneffekte angenommen werden, diese würden sich in veränderten Itemparametern der TIMSS-Items widerspiegeln. Daher wurden deren Parameter bei den Skalierungsläufen jeweils neu geschätzt und nicht auf die Werte, der vorherigen Erhebungen fixiert, um so mögliche Veränderungen der Itemparameter identifizieren zu können. Die in Abbildung 17 dargestellten TIMSS-Itemparameter sind die der Posttestskalierung.

Die Personenparameter der untersuchten Stichprobe variieren zwischen 2.4 und -3.4 Logits ebenfalls deutlich, der Mittelwert beträgt 0.130 und die Standardabweichung 0.750. Die *final deviance*, welche als Maß für die Modellgüte gilt (vgl. Kapitel 10.3.1), hat den Wert 4002.744 und ist im Vergleich zu den *final deviance*-Werten der vorangegangenen Modellschätzungen im iterativen Skalierungsprozess, welche die Werte 4189.256 und 4332.333 aufweisen, der beste Modellfit.

Die Fit-Statistiken auf Aufgabenebene (*infit*) weisen für die die vorliegende Modellschätzung aus, dass 71.4% der Aufgaben in dem Bereich zwischen 1.20 und 0.80 liegen (vgl. Kapitel 10.3.1). Folgt man Adams und Khoo (1996), so liegt der akzeptable Bereich der Fitstatistiken zwischen 0.60 und 1.60. In diesen Bereich fallen insgesamt 98.1% der Aufgaben, so dass nur 1.9% der Aufgaben nach diesen Schwellenwerten als problematisch betrachtet werden können. Die Fit-Indizes der Skalierung auf Aufgabenebene werden - wie für Studie 1 - in Anhang E detailliert dargestellt, welcher sich auf der beiliegenden CD befindet.

Insgesamt ist festzustellen, dass eine IRT-Skalierung der Aufgaben möglich ist und dass die Ergebnisse der Skalierung ohne größere Einschränkungen als Grundlage der weiteren Datenanalysen genutzt werden können.

Die Unterschiedlichkeit der Aufgabenschwierigkeit zwischen den Untersuchungsgruppen wurde mit Hilfe einer Varianzanalyse mit den drei Experimentalgruppen als unabhängige Variable und den skalierten Itemparametern aller Übungsaufgaben als abhängige Variable untersucht. Die deskriptiven Ergebnisse sind in Tabelle 32 dargestellt. Die Analyse zeigt einen signifikanten Effekt der Gruppe ($F_{(2, 144)} = 17.692$, $p < .001$, $\eta^2 = .197$), wie Tabelle 33 deutlich macht. Inhaltlich bedeutet das Ergebnis, dass der mittlere Itemparameter, d.h. die Aufgabenschwierigkeit, zwischen den drei Untersuchungsgruppen signifikant unterschiedlich ist.

Tabelle 32: Mittlerer IRT-Schwierigkeitsparameter in den drei Untersuchungsgruppen

Gruppe	Mittelwert	Standardabweichung	N
30% Lösungswahrscheinlichkeit	.883	1.250	54
50% Lösungswahrscheinlichkeit	-.042	1.073	48
80% Lösungswahrscheinlichkeit	-.592	1.421	45

Tabelle 33: Effekt der Untersuchungsgruppe auf den mittleren IRT- Schwierigkeitsparameter in den Übungsphasen

Quelle	Quadratsumme vom Typ III	df	Varianz	F	Signifikanz	Partielles Eta-Quadrat
Gruppe	55.451	2	27.726	17.692***	.001	.197
Fehler	225.661	144	1.567			
Gesamt	283.576	147	1.939			

Anmerkung:

*** Der Wert ist auf dem Niveau von 0.001 (2-seitig) signifikant

Dieser Effekt spiegelt sich in einer polynomialen Kontrastberechnung wider: Es zeigt sich ein linearer Trend der mittleren Itemschwierigkeit über die Untersuchungsgruppen hinweg. Der Kontrastparameter beträgt -1.043 ($p < .001$) mit einem Konfidenzintervall von -1.396 bis -0.689 .

Auch die Korrelationen zwischen den beobachteten relativen Lösungshäufigkeiten und den skalierten Itemparametern, welche als Hinweis auf die Modellgültigkeit der Itemschätzung betrachtet werden können, zeigen, dass es jeweils einen signifikanten Zusammenhang zwischen den beiden Schwierigkeitsindizes gibt (vgl. Tabelle 34). Die negativen Korre-

lationen kommen dadurch zustande, dass ein negativer Itemparameter in der IRT-Skalierung eine geringe Schwierigkeit bedeutet. Im Gegensatz dazu ist bei der Berechnung der Schwierigkeit mit Hilfe der klassischen Testverfahren eine geringe Aufgabenschwierigkeit durch einen hohen Wert bei der relativen Lösungshäufigkeit gekennzeichnet.

Tabelle 34: Korrelationen zwischen den mittleren IRT-Schwierigkeitsparametern und den klassisch berechneten relativen Lösungshäufigkeiten

Experimentalgruppe	r	p
Gesamtstichprobe	-.444**	<.001
30%	-.324*	.017
50%	-.502**	.001
80%	-.734**	<.001

Anmerkungen:

* Die Korrelation ist auf dem Niveau von 0.05 (2-seitig) signifikant.

** Die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant.

Insgesamt kann aufgrund der vorliegenden Ergebnisse davon ausgegangen werden, dass sich die mittlere Aufgabenschwierigkeit zwischen den drei Experimentalgruppen signifikant voneinander unterscheidet und die Schüler in den drei Gruppen Aufgaben bearbeiten, die sich in ihrer Schwierigkeit bedeutsam voneinander unterscheiden.

15.2.2 Schätzung der Personenparameter

Im Rahmen der Zuordnung der Personen zu den drei Experimentalgruppen wurde darauf geachtet, dass sich die Gruppen im mittleren Pretest-Personenparameter nicht bedeutsam unterscheiden (vgl. Kapitel 14.4). Eine Varianzanalyse mit den Pretest-Personenparametern als abhängige Variable und der Gruppenzugehörigkeit als unabhängige Variable zeigte erwartungskonform keine signifikanten Unterschiede zwischen den Untersuchungsgruppen in Bezug auf die geschätzten Personenparameter, $F_{(2, 40)} < 1$.

Um die Frage zu beantworten, ob es eine signifikante Veränderung des Personenparameters im Vergleich der Pretest-Werte und der Posttest-Werte sowohl innerhalb der experimentellen Gruppen als auch Unterschiede in der Veränderung zwischen den Gruppen gibt, wurde eine ANOVA mit Messwiederholung gerechnet. Die deskriptiven Ergebnisse sind in Tabelle 35 dargestellt.

Tabelle 35: Mittlerer Personenparameter in den Untersuchungsgruppen bei Pre- und Posttest-Messung

	Gruppe	Mittelwert	Standard- abweichung	N
Pretest- Personenparameter	30% Lösungswahrscheinlichkeit	.301	1.003	16
	50% Lösungswahrscheinlichkeit	.122	.804	12
	80% Lösungswahrscheinlichkeit	.091	.802	15
Posttest- Personenparameter	30% Lösungswahrscheinlichkeit	.527	.772	16
	50% Lösungswahrscheinlichkeit	-.022	.618	12
	80% Lösungswahrscheinlichkeit	-.172	.671	15

Die Gruppenzugehörigkeit stellt bei der ANOVA den Zwischenssubjektfaktor und der Messzeitpunkt den Innersubjektfaktor dar. Es zeigt sich kein signifikanter Effekt des Messzeitpunktes, $F_{(1, 40)} < 1$, d.h., dass es über die drei Gruppen hinweg zu keiner signifikanten Veränderung des Personenparameters zwischen Pre- und Posttest kam. Auch der Gruppeneffekt ist nicht signifikant, $F_{(2, 40)} = 1.587$, $p = .217$, $\eta^2 = .074$. Jedoch ist die Interaktion zwischen Messzeitpunkt und Untersuchungsgruppe statistisch bedeutsam, $F_{(2, 40)} = 3.518$, $p = .039$, $\eta^2 = .150$ (vgl. Tabelle 36).

Tabelle 36: Effekt des Messzeitpunkts und der Gruppenzugehörigkeit auf den Personenparameter

Quelle	Quadrat- summe vom Typ III	df	Varianz	F	Signifi- kanz	Partielles Eta-Quadrat
Zwischensubjekt- effekt						
Gruppe	3.566	2	1.783	1.587	.217	.074
Fehler	44.941	40	1.124			
Innersubjekteffekt						
Messzeitpunkt	.076	1	.076	.534	.469	.013
Messzeitpunkt * Gruppe	1.005	2	.503	3.518*	.039	.150
Fehler	5.715	40	.143			
Gesamt	6.796	43	.158			

Anmerkung:

* Der Wert ist auf dem Niveau von 0.05 (2-seitig) signifikant

Die Veränderung des Personenparameters von der Pretest-Erhebung zur Posttest-Erhebung für die drei Untersuchungsgruppen ist in Abbildung 18 dargestellt.

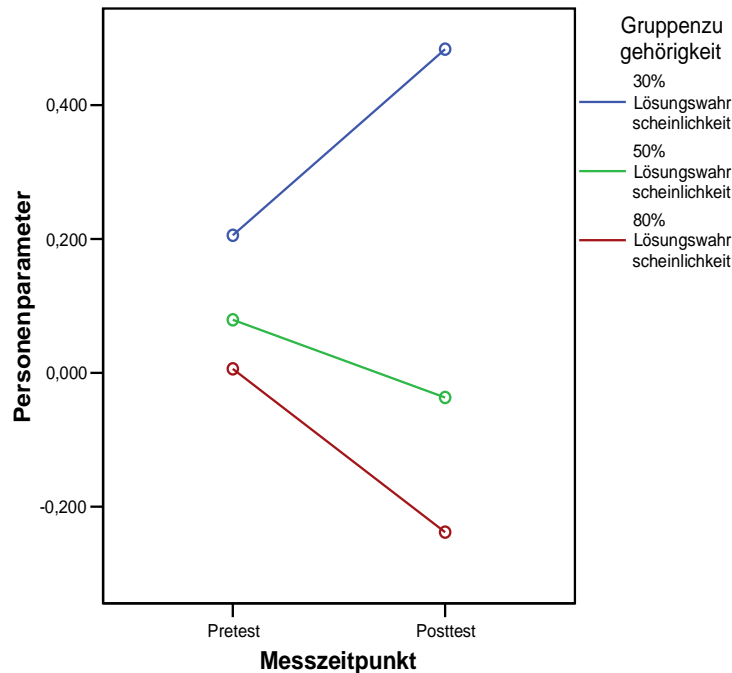


Abbildung 18: Veränderung des Personenparameters zwischen Pretest- und Posttest-Erhebung

Zur Berechnung der Post-Hoc-Vergleiche innerhalb der Untersuchungsgruppen wurden geplante Kontraste berechnet. Dabei zeigte sich lediglich bei der 30%-Gruppe ein signifikanter Zuwachs im mittleren Personenparameter: $t_{(15)} = 2.43$; $p = .028$. Bei der 50%- und der 80%-Gruppe unterscheiden sich Pre- und Posttestwerte nicht: 50%-Gruppe: $t_{(11)} = <1$; 80%-Gruppe: $t_{(14)} = -1.64$; $p = .124$. Somit sind die in Abbildung 18 augenscheinlichen Veränderungen zwischen Pre- und Posttest-Werten lediglich in der 30%-Gruppe statistisch bedeutsam. Die in der Abbildung (grüne bzw. rote Linie) deskriptiv sichtbare Reduktion in der 50%- und der 80%-Gruppe bedeutet keine signifikante Veränderung im mittleren Personenparameter zwischen Pre- und Posttest innerhalb dieser Gruppen.

Die Berechnung der Effektstärken erfolgt auf Grundlage der Formel: $\varepsilon = \mu_1 - \mu_2 / \sigma_1$ und weist die Posttest-Unterschiede zwischen den experimentellen Gruppen als praktisch bedeutsam aus. Bei Kontrolle des Pretests schneidet die 30%-Gruppe im Posttest mit einer Effektstärke von $d = 0.79$ deutlich besser ab als die 50%-Gruppe. Der Vergleich zwischen der

30%- und 80%-Gruppe weist eine Effektstärke von $d = 0.97$ aus. Beide Effekte können nach Bortz und Doring (2002) als große Effekte ($d \geq 0.8$) angesehen werden. Inhaltlich bedeutet eine Effektstärke von $d = 0.79$ für den Unterschied zwischen der 30%- und der 50%-Gruppe bzw. von $d = 0.97$ für den Unterschied zwischen der 30%- und der 80%-Gruppe, dass mit den 30%-Lösungswahrscheinlichkeit-Aufgaben die Position des Mittelwertes dieser Gruppe um 0.79 im Vergleich zu der 50%-Gruppe, bzw. um 0.97 Standardabweichungen im Vergleich zum Mittelwert der 80%-Gruppe verschoben ist.

In der Abbildung 19 der Regression der Posttest-Personenparameter auf die Pretest-Personenparameter ist zu erkennen, dass nicht nur die leistungsfähigen Schüler der 30%-Gruppe von den für sie mäßig schwierigen Aufgaben profitieren. Auch für die leistungsschwächeren Schüler der 30%-Gruppe sind die für sie individuell mäßig schwierigen Aufgaben ebenfalls von Nutzen. Dies wird dadurch deutlich, dass auch im unteren Bereich der Abszisse die Schüler der 30%-Gruppe oberhalb der fett eingetragenen (schwarzen) Regressionsgerade liegen. Dieses Ergebnis wird in Abbildung 20, in der anstatt der jeweiligen Regressionsgerade die Winkelhalbierende eingezeichnet wurde ebenfalls dadurch verdeutlicht, dass der überwiegende Teil der Schüler der 30%-Gruppe oberhalb der Winkelhalbierenden liegt, auch die Schüler mit einem niedrigen Fähigkeitsparameter. Im Gegensatz dazu liegen nur einige wenige Schüler der 50%- und der 80%-Gruppe oberhalb der Winkelhalbierenden.

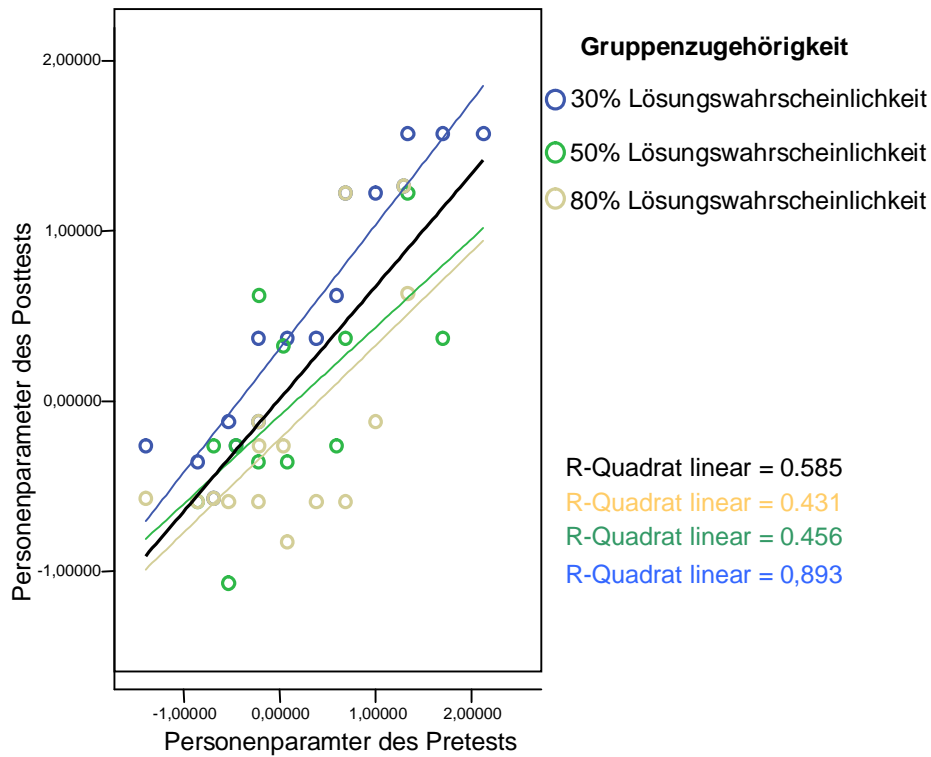


Abbildung 19: Regression der Posttest- auf die Pretestpersonenparameter

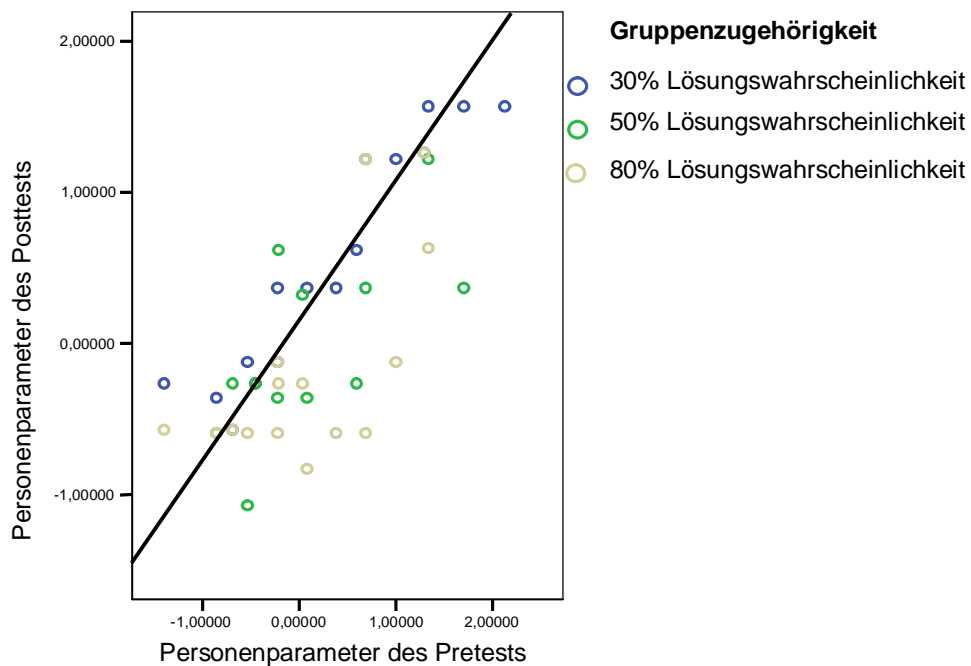


Abbildung 20: Darstellung der Personenparameter beider Messungen mit eingezeichneten Winkelhalbierenden

Um zu überprüfen, ob die in Abbildung 19 augenscheinlichen Unterschiede in den jeweiligen Steigungen der Regressionsgeraden der einzelnen Untersuchungsgruppen statistisch signifikant unterschiedlich sind, wurde per Varianzanalyse (UV: Gruppenzugehörigkeit und Pretest-Personenparameter, AV: Posttest; sequentielle Varianzzerlegung) ein Test auf Interaktion zwischen Gruppenzugehörigkeit und Pretest-Personenparameter durchgeführt. Es zeigten sich keine bedeutsamen Unterschiede zwischen den Steigungen ($F_{(2, 40)} < 1$).

Um bei der Beurteilung des Leistungszuwachses das Vorwissen bzw. die Leistung des Pretests kontrollieren zu können, werden residualisierte Zuwachswerte verwendet, d.h. aus der Posttestleistung wurde die korrespondierende Pretestleistung auspartialisiert.⁷ Die dazugehörigen deskriptiven Daten sind in Tabelle 37 dargestellt. Auch die Berechnung einer einfaktoriellen Varianzanalyse mit den Posttest-Residuen nach Auspartialisierung des Pretests, d.h. mit dem „residualer Lerngewinn“ als abhängige Variable weist einen signifikanten Effekt der Gruppenzugehörigkeit aus, $F_{(2, 40)} = 7.426$, $p = .002$, $\eta^2 = .271$ (vgl. Tabelle 38).

Tabelle 37: Mittlere Posttest-Residuen der Untersuchungsgruppen

Gruppe	Mittelwert	Standardabweichung	N
30% Lösungswahrscheinlichkeit	.316	.261	16
50% Lösungswahrscheinlichkeit	-.115	.470	12
80% Lösungswahrscheinlichkeit	-.245	.514	15

Tabelle 38: Effekt der Gruppenzugehörigkeit auf die Posttest-Residuen

Quelle	Quadratsumme vom Typ III	df	Varianz	F	Signifikanz	Partielles Eta-Quadrat
Gruppe	2.657	2	1.328	7.426**	.002	.271
Fehler	7.155	40	.179			
Gesamt	9.812	43	.228			

Anmerkung:

**Der Wert ist auf dem Niveau von 0.01 (2-seitig) signifikant

⁷ Es soll nicht näher auf die Vor- und Nachteile verschiedener Maße der Veränderung eingegangen werden (siehe z.B. Cronbach & Furby, 1970). Es wird sich dem in der Unterrichtsforschung üblichen Vorgehen, residualisierte Zuwachsmäße zu verwenden, angeschlossen. Diese Zuwachsmäße können am ehesten als Unterrichtseffekte gedeutet werden.

Mit Hilfe einer polynominalen Kontrastberechnung wurde der linearer Trend der mittleren Posttest-Residuen über die Untersuchungsgruppen hinweg bestätigt: Der berechnete Kontrastwert beträgt $-.397$ ($p = .001$) und liegt in einem Konfidenzintervall von $-.614$ bis $-.179$. Dieses Ergebnis kann als Bestätigung der Ergebnisse der oben berichteten ANOVA mit Messwiederholung im Pretest und Posttest angesehen werden: Dort ergab sich lediglich für den mittleren Personenparameter der 30%-Gruppe ein bedeutsamer Zuwachs, d.h. ein Lerngewinn vom Pretest zum Posttest.

16 Zusammenfassung und Diskussion der Ergebnisse der Studie 2

In diesem Kapitel sollen zunächst die Ergebnisse zusammengefasst und in Bezug zu den Hypothesen dargestellt werden (Kapitel 16.1), daran anschließend folgt die Diskussion der Ergebnisse (Kapitel 16.2).

16.1 Zusammenfassung der Ergebnisse

In der Studie 2 wurde der Frage nachgegangen, ob sich Schülerleistungen verbessern lassen, wenn Schüler Lernaufgaben bearbeiten, die in Anlehnung an die Idee des adaptiven Testens in ihrer Schwierigkeit individuell an die jeweilige Fähigkeit des Schülers angepasst werden. Ferner wurde untersucht, ob sich die Schülerfähigkeit in Abhängigkeit von der Aufgabenschwierigkeit der eingesetzten Lernaufgaben unterschiedlich stark verändern lässt.

Es zeigte sich, dass sich der Fähigkeitsparameter erhöhen lässt, wenn Schüler für sie individuell mäßig schwierige Lernaufgaben, mit einer Lösungswahrscheinlichkeit von 30%, während der Übungsphasen bearbeiteten. Wenn die Schüler hingegen mittelschwere (50%) oder leichte (80%) Aufgaben bearbeiteten, führte dies zu keinen signifikanten Veränderungen ihres Personenparameters. Insgesamt lässt sich festhalten, dass nur die Bearbeitung von individuell mäßig schwierigen Aufgaben zu einer Leistungssteigerung bzw. einer Erhöhung des Personenparameters führt. Dieses Ergebnis ist konform mit der Annahme von Helmke (2004), dass das Ausmaß der unterrichtlichen Anforderungen in einer Schwierigkeitszone liegen sollte, die oberhalb des aktuellen Wissenstands der Schüler liegt, aber, in Anlehnung an die Zone der proximalen Entwicklung von Wygotsky (1978), nicht zu weit davon entfernt ist. Als weitere Belege für die Plausibilität der Ergebnisse können die in der Literatur dokumentierten lernförderlichen Effekte von *high level*-Aufgaben herangezogen werden, da die Aufgaben, welche überwiegend in der 30%igen-Gruppe eingesetzt wurden, auf dem *high level*-Niveau zu klassifizieren sind (vgl. Hamaker, 1986; Redfield & Rousseau, 1981; Roushine, 1971; Samson et al., 1987; Winne, 1979).

Insgesamt konnte gezeigt werden, dass mäßig schwierige Aufgaben, welche individuell an die jeweilige Schülerfähigkeit adaptiert wurden, zu einer signifikanten Erhöhung des Fähigkeitsparameters führen.

16.2 *Diskussion der Ergebnisse*

Die Ergebnisse der Studie 2 werden mit den bislang vorliegenden Ergebnissen zum adaptiven Testen und dem adaptiven Aufgabeneinsatz in Beziehung gesetzt.

Der Einfluss des Anspruchsniveaus des Unterrichts, welcher z.B. von Cronbach und Snow (1981) nachgewiesen wurde, zeigt sich auch in der hier vorgestellten Studie 2. Wird das Anspruchsniveau des Unterrichts, operationalisiert über die Aufgabenschwierigkeit (IRT-Itemparameter) in den Übungsphasen an den IRT-Fähigkeitsparameter adaptiert, so lässt sich dieser Fähigkeitsparameter unter der 30%-Lösungswahrscheinlichkeits-Bedingung erhöhen. Somit kann bestätigt werden, dass ein auf diese Weise adaptierter Unterricht eine Möglichkeit zur Leistungsverbesserung darstellen kann, was im Einklang mit verschiedenen Metaanalysen steht (Corno & Snow, 1986; Kulik & Kulik, 1991; Snow, 1989).

Bei der Interpretation der Ergebnisse der Studie 2 muss berücksichtigt werden, dass der Nutzen der adaptiven Anpassung der Aufgabenschwierigkeit lediglich für diejenigen Schüler nachgewiesen werden könnte, die mäßig schwierige Aufgaben (30% Lösungswahrscheinlichkeit) bearbeitet hatten, jedoch nicht für Schüler, die mittelschwere (50% Lösungswahrscheinlichkeit) oder leichte (80% Lösungswahrscheinlichkeit) Aufgaben bekamen. Dieser Zusammenhang zwischen der Variation der Aufgabenschwierigkeit in Verbindung mit der adaptiven Anpassung an die individuelle Schülerfähigkeit wurde bislang im regulären Unterricht, d.h. in einer feldexperimentellen Untersuchung noch nicht empirisch untersucht. Ein Vergleich mit anderen Studien zum Effekt der Aufgabenschwierigkeit zeigt im Übrigen, dass in diesen Studien üblicherweise keine individuelle Anpassung an die Schülerfähigkeit vorgenommen wird, sondern alle Schüler einer Untersuchungsgruppe identische Aufgaben vorgelegt bekommen und sich somit die mit Hilfe der klassischen Testtheorie berechnete Schwierigkeit lediglich zwischen den Untersuchungsgruppen unterscheidet. Somit ist keine individuelle Anpassung der Aufgabenschwierigkeit innerhalb der Untersuchungsgruppen gegeben, wie es in der vorgestellten Studie 2 erfolgte. So konnten verschiedene Autoren nachweisen, dass Schüler bessere Leistungen zeigen, wenn sie eine Serie von Lernaufgaben mit leichten Aufgaben beginnen. Im Gegensatz dazu konnten Personen, die eine Serie von Lernaufgaben mit unlösbaren (Sweller, 1976) oder mittelschwierigen (Fürntratt, 1978; Kloep & Weimann, 1982) Aufgaben begannen, die nachfolgenden leichten Aufgaben überhaupt nicht bzw. schlechter lösen. Diese Studien treffen jedoch keine Aussagen über die individuellen Leistungen der Schüler innerhalb der einzelnen Gruppen, so dass die Ergebnisse mit den Befunden der Studie 2 nur bedingt vergleichbar sind. Aufgrund der vorliegenden Ergebnisse ist zu vermuten, dass leistungsfähigere Schüler in der Gruppe mit den mittelschwieri-

gen Aufgaben - etwa bei der Untersuchung von Fürntratt (1978) - diese Aufgaben besser lösen können als leistungsschwächere Schüler, da die Aufgabenschwierigkeit ihrem Leistungsniveau eher entspricht.

Im Unterschied zu den bislang erwähnten Studien wurde in einer Untersuchung von Nussbaum und Leutner (1986) die Aufgabenschwierigkeit individuell an den jeweiligen Fähigkeitsparameter der Versuchspersonen adaptiert. Es wurde das entdeckende Lernen (nach Bruner, 1961) von Lösungsregeln aus figuralen Aufgabenmerkmalen (sog. Matrizenaufgaben) untersucht, wobei die Aufgabenschwierigkeit systematisch variiert wurde. Die Ergebnisse lassen sich dahingehend zusammenfassen, dass die besten Leistungen bei geringer Aufgabenschwierigkeit (95% Lösungswahrscheinlichkeit) erzielt wurden. Ein Vergleich dieser Untersuchung mit der vorliegenden Studie zeigt folgende Differenz: In der Untersuchung von Nussbaum und Leutner (1986) wurde entdeckendes Lernen bei Matrizenaufgaben mit unterschiedlicher Aufgabenschwierigkeit variiert. Im Gegensatz dazu war der Inhaltsbereich der vorliegenden Untersuchung weniger gut strukturiert und bezog sich auf die fachlichen Inhalte zur Atomphysik bzw. zur Halbleiter-Elektronik. Möglicherweise bestehen auch Unterschiede in Bezug auf die von den Versuchspersonen zu bearbeitenden Aufgaben: In der Studie von Nussbaum und Leutner (1986) mussten die Versuchspersonen verschiedene figurale Muster bzw. Vorlagen miteinander vergleichen, um Gesetzmäßigkeiten zu finden und Lösungsregeln abzuleiten. Im Gegensatz dazu mussten die Schüler in der vorliegenden Studie 2 schulische Inhalte erinnern, verstehen und anwenden. Ein weiterer Unterschied zwischen den beiden Untersuchungen besteht im Vorwissen: Bei Nussbaum und Leutner (1986) verfügen die Versuchspersonen über keinerlei Erfahrungen mit der untersuchten Aufgabenart und können somit kein Vorwissen bei der Bearbeitung der Matrizenaufgaben einsetzen. In der Studie 2 verfügen die Schüler über physikalisches Vorwissen aufgrund des vorangegangenen Unterrichts zu dem Inhaltsbereich, welcher in den Übungsaufgaben thematisiert wurde. Daher kann der lernförderliche Effekt der mäßig schwierigen Aufgaben (30% Lösungswahrscheinlichkeit) dahingehend erklärt werden, dass diese bei den Schülern zu einer Aktivierung des Vorwissens führen, welches sie dann bei der Aufgabenbearbeitung einsetzen. Eine empirische Überprüfung dieses Erklärungsansatzes für die Effektivität der mäßig schwierigen Lernaufgaben steht noch aus.

Insgesamt scheinen die Befunde der hier vorgestellten Untersuchung im Vergleich zu den bisherigen Untersuchungen zum Einfluss der Aufgabenschwierigkeit auf die Lernleistungen eher erwartungskonträr zu sein. Zieht man jedoch Wygotskys (1978) Konstrukt der Zone der proximalen Entwicklung (vgl. Kapitel 4.3) als Erklärungsansatz heran, so erscheint der erhöhte Fähigkeitsparameter derjenigen Schüler, welche mäßig schwierige Aufgaben be-

arbeitet haben, damit erklärbar, dass diese Aufgaben sich am oberen Rand der jeweiligen individuellen Zone befinden und von den Schülern gerade noch gelöst werden können. Diese Auffassung wurde von Helmke (2004) auf die im Unterricht eingesetzten Aufgaben übertragen, indem er annahm, dass diese hinsichtlich ihrer Schwierigkeit an der Zone der proximalen Entwicklung der Schüler ausgewählt werden sollten (vgl. Brophy, 2000). Im vorliegenden Fall befinden sich vermutlich die mäßig schwierigen Aufgaben mit einer Lösungswahrscheinlichkeit von 30% in diesem Bereich und haben sich als lernförderlich herausgestellt.

Da sich nicht alle Schüler einer Klasse auf demselben Entwicklungsniveau befinden, müssen die Aufgabenschwierigkeiten individuell an den jeweiligen Schüler adaptiert werden. Somit lässt sich erklären, dass auch die leistungsschwächeren Schüler der 30%-Gruppe von für sie mäßig schwierigen Aufgaben profitiert haben, mit der Idee des adaptiven Unterrichts wurde die Hoffnung verbunden, dass seine leistungsförderliche Wirkung vor allem leistungsschwächeren Schülern zugute kommt (Schrader, 1989). In der dargestellten Studie 2 haben alle Schüler, auch die leistungsschwächeren, von für sie mäßig schwierigen Aufgaben profitiert, was also für deren leistungsförderliche Wirkung spricht.

Ein weiterer Erklärungsansatz der leistungsförderlichen Wirkung mäßig schwieriger Lernaufgaben kann in den Ergebnissen zur Effektivität sogenannter *high level*-Aufgaben gesehen werden. Diese haben sich in den meisten Analysen im Vergleich mit *low level*-Aufgaben als lernwirksamer herausgestellt (vgl. Kapitel 3.2.1; Hamaker, 1986; Redfield & Rousseau, 1981; Rosenshine, 1971; Samson et al., 1987). Dabei werden *high level*-Aufgaben als kognitiv anspruchsvollere Aufgaben betrachtet (Hamaker, 1986; Hunkins, 1976). Da ein Großteil der in der 30%-Gruppe gestellten Lernaufgaben auf diesem *high level*-Niveau zu klassifizieren ist und meist Verstehens- bzw. Anwendungsprozesse erfordern, ist anzunehmen, dass diese zu einer tieferen Verarbeitung des Wissensinhalts und dadurch zu einer besseren Lernleistung führen. Im Vergleich dazu ist ein Großteil der Aufgaben, welche in der 80%-Gruppe bearbeitet wurden, auf dem Wissensniveau (*low level*-Aufgaben) zu klassifizieren. Mit dem Befund, dass sich diese Aufgabenart in den meisten Untersuchungen, verglichen mit *high level*-Aufgaben, als wenig lernförderlich herausgestellt hat, ist das Ergebnis der 80%-Gruppe, in der sich keine Veränderung des Personenparameters im Pre-Posttest-Vergleich zeigte, erklärbar.

Aufgrund der vorliegenden Ergebnisse sowie dem Befund von Redfield und Rousseau (1981), dass Aufgaben auf kognitiv höherem Niveau bzw. mäßig schwierige Aufgaben dann zu besseren Leistungen führen, wenn sie regelmäßiger Bestandteil des Unterrichts sind, ist die Empfehlung abzuleiten, dass diese Aufgaben vermehrt in der Übungsphase des regulären Unterrichts Einsatz finden sollten. Da in der vorliegenden Studie lediglich zwei Ü-

bungsphasen stattfanden, sollten die Ergebnisse in weiteren Untersuchungen mit mehreren Übungsphasen, in denen die Aufgaben von den Schülern bearbeitet werden, repliziert werden. Damit könnten mögliche Übungseffekte im zeitlichen Verlauf besser abgebildet werden. Diese empirische Prüfung steht noch aus.

Insgesamt können die entwickelte Methode der Aufgabenkonstruktion und die Rasch-Skalierung als Möglichkeit der Adaptation der Aufgabenschwierigkeit an die Schülerfähigkeit betrachtet werden, um den Fähigkeitsparameter in bedeutsamer Weise zu erhöhen. Mit Hilfe dieser Methode konnte in der vorliegenden Studie 2 eine angemessene und individuelle Auswahl von Übungsaufgaben erfolgen.

Bei der Ergebnisinterpretation der vorliegenden Studie 2 muss die insgesamt kleine Stichprobe von 43 Schülern, welche nochmals auf drei Untersuchungsgruppen aufgeteilt wurden, einschränkend berücksichtigt werden. Somit sollten die Untersuchungsergebnisse zunächst an einer größeren Stichprobe und anderen Unterrichtsfächern repliziert werden, bevor gesicherte Empfehlungen für die Unterrichtsgestaltung ausgesprochen werden. Auf die Schwächen der vorliegenden Untersuchung, sowie offene Fragen für anschließende Forschung wird in Kapitel 17 näher eingegangen.

Aus den vorliegenden Ergebnissen lässt sich, mit den oben erwähnten Einschränkungen, zusammenfassend bereits zum gegenwärtigen Zeitpunkt die Empfehlung für den Einsatz von Lernaufgaben dahingehend ableiten, dass, um Lernaufgaben möglichst effektiv im Unterricht einsetzen zu können, diese in ihrer Schwierigkeit individuell an die jeweilige Schülerfähigkeit angepasst werden sollten. Dabei sollten, den vorliegenden Befunden entsprechend, die Lernaufgaben für den Schüler eine mittlere Lösungswahrscheinlichkeit von etwa 30% aufweisen, um eine signifikante Erhöhung des individuellen Fähigkeitsparameters zu erzielen.

17 Zusammenfassende Diskussion der Ergebnisse beider Studien und Forschungsausblick

In diesem Kapitel werden abschließend die Ergebnisse der beiden vorliegenden Studien gemeinsam betrachtet. Dazu werden zunächst die jeweiligen Ergebnisse zusammenfassend dargestellt, um sie gemeinsam diskutieren zu können (Kapitel 17.1). Im Anschluss daran wird ein Ausblick auf weitere, noch offene Forschungsfragen gegeben (Kapitel 17.2), bevor zum Abschluss die Grenzen bzw. Schwächen der beiden Studien erörtert werden (Kapitel 17.3).

17.1 Zusammenfassende Diskussion der Ergebnisse beider Studien

In Studie 1 hat sich in Bezug auf den Aufgabeneinsatz im Physikunterricht gezeigt, dass die Lehrer in den aufgezeichneten Unterrichtsstunden zum größten Teil Wissensaufgaben (62.4% aller Unterrichtsaufgaben) gestellt haben, gefolgt von Verstehensaufgaben (32.3%). Mit Abstand am wenigsten Aufgaben wurden auf dem Niveau der Wissensanwendung (5.3%) gestellt. Dieser Befund steht im Einklang mit den Ergebnissen einer Vielzahl anderer Studien, die gezeigt haben, dass Unterrichtsaufgaben zum großen Teil der Wissenskategorie zuzuordnen sind, welche meist als *low level*-Aufgaben bezeichnet werden und dass es eine geringe Variation der kognitiven Prozesse gibt (vgl. Kapitel 3.2.1).

Des Weiteren wurde in Studie 1 der Zusammenhang zwischen der Schwierigkeit der vom Lehrer eingesetzten Unterrichtsaufgaben, die den Schülern eine Woche nach dem Unterricht erneut zur Bearbeitung vorgelegt wurden, und der mittleren Schülerfähigkeit in der jeweiligen Klasse untersucht. Dabei zeigte sich insgesamt ein signifikanter, moderat positiver Zusammenhang ($r = .29$). Allerdings wurden große Unterschiede zwischen den untersuchten Physikklassen bezüglich der Passung zwischen Aufgabenschwierigkeit und Schülerfähigkeit deutlich. Besonders in Klassen, in denen die mittlere Schwierigkeit der Unterrichtsaufgaben hoch war, war die Passung – definiert als eine mittlere Lösungswahrscheinlichkeit der Unterrichtsaufgaben von 50% - nicht mehr gegeben. Die mittlere Lösungswahrscheinlichkeit in diesen Klassen betrug z.T. nur 8%, so dass diese Unterrichtsaufgaben für den überwiegenden Teil der Klasse zu schwierig waren. Die Schüler konnten die Aufgaben nicht erfolgreich beantworten. Insgesamt kann aufgrund der vorliegenden Ergebnisse für die untersuchte Stichprobe abgeleitet werden, dass eine Passung zwischen der Schwierigkeit der von den Lehrern im Unterricht eingesetzten Aufgaben und der mittleren Schülerfähigkeit der

Klasse nicht in ausreichendem Maße gegeben ist. Ein möglicher Grund in der teilweise geringen Passung kann in der fehlenden bzw. unzureichenden diagnostischen Kompetenz der Lehrer gesehen werden, welche in Kapitel 3.4 dargestellt wurde (vgl. Hosenfeld et al., 2002; Schnotz, 1971; Schrader, 1989; Schrader & Helmke, 1987).

Die Möglichkeiten der Aufgabenkonstruktion auf der theoretischen Grundlage der Revision der Bloom'schen Taxonomie (Anderson et al., 2001) in Kombination mit den Konstruktionsvorschriften von Klauer (1987) wurden ebenfalls in Studie 1 untersucht. Diese Art der Aufgabenkonstruktion hat sich als Möglichkeit bewährt, Aufgaben unterschiedlicher Schwierigkeit zu gestalten. Die statistischen Auswertungen der drei konstruierten Aufgabenarten - Wissens-, Verstehens- und Anwendungsaufgaben - haben gezeigt, dass sie sich in bedeutsamer Weise in ihrem mittleren Schwierigkeitsparameter voneinander unterscheiden.

Die vorliegende Studie 1 hat demzufolge erwiesen, dass Aufgaben im schulischen Bereich in Bezug auf ihre Schwierigkeit in den meisten der untersuchten Klassen nicht an die mittlere Schülerfähigkeit angepasst werden. Ferner hat sich gezeigt, dass mit Hilfe der hier eingesetzten Konstruktionsmethode Aufgaben unterschiedlicher Schwierigkeit entwickelt werden können.

Aufgrund der in Studie 1 teilweise deutlich gewordenen fehlenden Passung zwischen Aufgabenschwierigkeit und Schülerfähigkeit im Unterricht sowie aufgrund der insgesamt unbefriedigenden Ergebnisse der Leistungstests bzgl. der relativen Lösungshäufigkeiten der verschiedenen Aufgabenarten wurde in Studie 2 die Effektivität von Lernaufgaben, die in ihrer Schwierigkeit an die jeweilige Fähigkeit der Schüler adaptiert wurden, untersucht. Dazu wurde die Lösungswahrscheinlichkeit der Lernaufgaben experimentell in drei Gruppen (30%, 50%, und 80% Lösungswahrscheinlichkeit) variiert. In zwei Übungsphasen bearbeiteten die Schüler Aufgaben, welche aufgrund der zuvor geschätzten Item- und Fähigkeitsparameter bezüglich der Lösungswahrscheinlichkeit individuell angepasst waren. Die Ergebnisse zeigen, dass mit Hilfe individuell adaptierter, mäßig schwieriger Aufgaben (30% Lösungswahrscheinlichkeit) der Fähigkeitsparameter signifikant erhöht werden kann.

Die Ergebnisse der beiden vorliegenden Untersuchungen wirken auf der ersten Blick widersprüchlich, denn in Studie 1 wurde gezeigt, dass Schüler im Unterricht z.T. durch zu schwierige Aufgaben überfordert werden. Im Gegensatz dazu erwiesen sich in Studie 2 die mäßig schwierigen Aufgaben (30% Lösungswahrscheinlichkeit) als lernförderlich, mit Hilfe dieser Aufgaben konnte der Fähigkeitsparameter in bedeutsamer Weise erhöht werden. Ein genauerer Vergleich der beiden Untersuchungen zeigt folgende Unterschiede, die als mögliche Erklärungsansätze herangezogen werden können. In Studie 1 wurden von allen Schülern einer Klasse dieselben Aufgaben bearbeitet. Diese erwiesen sich bei einigen Klassen, beson-

ders im oberen Schwierigkeitsbereich, als im Mittel sehr schwierig, so dass lediglich ein bis drei Schüler die Unterrichtsaufgaben richtig beantworten konnten. Die Lehrer orientieren sich bei der Auswahl der Aufgabenschwierigkeit augenscheinlich an den leistungsfähigsten Schülern ihrer Klasse. Dieses Ergebnis steht im Einklang mit Untersuchungen, die gezeigt haben, dass Schüler im Unterricht häufig sehr schwierige Aufgaben zur Bearbeitung vorgelegt bekommen (vgl. Fisher et al., 1980; Jorgenson, 1977). Dieses Ergebnis hinsichtlich der schwierigen Unterrichtsaufgaben steht jedoch im Widerspruch mit dem Befund von Renkl (1991), der zeigen konnte, dass die mittlere relative Lösungshäufigkeit von Unterrichtsaufgaben 83.3% beträgt. Im Unterschied zu der vorliegenden Studie 1 untersuchte Renkl (1991) jedoch den Aufgabeneinsatz im Mathematikunterricht an Grundschulen. Auch kann die Untersuchung von Schnotz (1971) herangezogen werden, in der gezeigt wurde, dass Lehrer die Aufgabenschwierigkeit bei 70% der einzuschätzenden Aufgaben überschätzten. Dieses Ergebnis steht in Einklang mit dem Befund von Klieme (2000), dass auch Fachdidaktiker die relative Lösungshäufigkeit der TIMSS-Items, besonders im oberen Leistungsbereich (Leistungskurse), überschätzen. In der vorliegenden Studie 1 war speziell im oberen Schwierigkeitsbereich die Passung der Unterrichtsaufgaben mit der mittleren Schülerfähigkeit kaum gegeben.

Im Unterschied dazu wurde in Studie 2 die Aufgabenschwierigkeit individuell für jeden Schüler an seine Fähigkeit angepasst, so dass Schüler, welche derselben Untersuchungsgruppe angehörten, je nach ihrer Fähigkeit unterschiedliche Aufgaben bearbeitet haben. Der Befund der Studie 2 steht im Einklang mit Ergebnissen zum adaptiven Testen und Unterrichten, welche zeigen konnten, dass die Anpassung des kognitiven Anspruchsniveaus des Unterrichts an die kognitiven Fähigkeiten der Schüler eine effektive Methode darstellt, um die Schülerleistungen zu erhöhen (vgl. Cronbach & Snow, 1981; Slavin, 1997; Wilen, 1991).

Somit sind die unterschiedlichen Ergebnisse der beiden Studien womöglich dadurch erklärbar, dass die Aufgabenschwierigkeit in Studie 1 für die meisten Schüler einer Klasse viel zu hoch war, denn die z.T. gefundene Lösungswahrscheinlichkeit von 8% ist über den in einer Klasse vorfindbaren gesamten Fähigkeitsbereich gemittelt worden, so dass die eingesetzten Unterrichtsaufgaben für einen Großteil der Schüler unlösbar schwierig waren. Im Gegensatz dazu wurde die Aufgabenschwierigkeit in Studie 2 jeweils individuell an den Fähigkeitsparameter des einzelnen Schülers angepasst. Aufgrund der vorliegenden Ergebnisse erscheint diese individuelle Adaptation der Aufgabenschwierigkeit als ausschlaggebend für die Erhöhung des Fähigkeitsparameters.

Gleichfalls stellt die jeweilige Aufgabenschwierigkeit einen relevanten Unterschied zwischen den beiden vorliegenden Studien dar und kann möglicherweise ebenfalls zur Erklä-

rung der unterschiedlichen Ergebnisse herangezogen werden. Die von den Lehrern im Unterricht eingesetzten Aufgaben in Studie 1, mit einer mittleren Lösungshäufigkeit von z.T. unter 10%, sind als sehr schwierig zu bezeichnen und können für die meisten Schüler der Klasse als überfordernd angesehen werden, da diese Aufgaben im Mittel von lediglich zwei bis drei Schülern korrekt beantwortet werden können. Im Unterschied dazu ist die Lösungswahrscheinlichkeit von 30% in Studie 2 als mäßig schwierig einzustufen und scheint eine herausfordernde Lernwirkung zu haben. Im Sinne der Zone der proximalen Entwicklung (Wygotsky, 1978) befinden sich mäßig schwierige Aufgaben in einem für den Schüler aufgrund seines Wissens und seiner Fähigkeiten noch zu lösenden Bereich. Dieses Ergebnis kann als Bestätigung der Annahme von Helmke (2004) betrachtet werden, dass sich die im Unterricht eingesetzten Aufgaben, in ihrer Schwierigkeit in der Zone der proximalen Entwicklung der Schüler befinden sollten. Das bedeutet, dass sie in ihrer Schwierigkeit etwas oberhalb des aktuellen Fähigkeits- und Wissensniveaus der Schüler liegen sollten. Die Grenzen zwischen mäßig schwierigen und herausfordernden Aufgaben (30% Lösungswahrscheinlichkeit) und sehr schwierigen, z.T. überfordernden Aufgaben (10% Lösungswahrscheinlichkeit) lassen sich aufgrund der vorliegenden Ergebnisse nicht bemessen und stehen somit noch zur empirischen Überprüfung aus. In einer möglichen Untersuchung sollte dieser obere Bereich der Aufgabenschwierigkeit ausdifferenziert werden, indem beispielsweise Gruppen mit 10%-, 20%- und 30%-Lösungswahrscheinlichkeit verglichen werden. Aufgrund der bisherigen Ergebnisse ist zu erwarten, dass es mit immer schwierigeren Aufgaben zu einer geringeren bzw. keiner Veränderung des Fähigkeitsparameters kommt. Demnach sollte der lernförderliche Effekt der mäßig schwierigen Aufgaben replizierbar sein, sowie denkbare Unterschiede zwischen den Aufgaben mit 20% und 10% Lösungswahrscheinlichkeit sollten aufgezeigt werden können.

Die z.T. erwartungskonträren Ergebnisse der vorliegenden Untersuchungen zeigen verschiedene theoretische Implikationen auf. Die Ergebnisse lassen sich mit Hilfe der vorliegenden theoretischen Modelle zur Aufgabenbearbeitung, insbesondere in Bezug auf die Aufgabenschwierigkeit, nicht ausreichend erklären. Zwar konnten die auf der Grundlage der Zone der proximalen Entwicklung (Wygotsky, 1978) formulierten Annahmen (vgl. Helmke, 2004) empirisch bestätigt werden. Das Konzept von Wygotsky (1978) scheint eine bedeutende Rolle zu spielen, da die individuellen Unterschiede der einzelnen Lerner betont werden und die Idee der adaptiven Anpassung der Aufgabenschwierigkeit aufgegriffen wird. Die immense Bedeutung von Theorien über die Art der mentalen Repräsentation von Wissen und des Wissenserwerbs für das Verständnis der Aufgabenbearbeitung bzw. der dazu notwendigen kognitiven Prozesse sollte berücksichtigt werden (Gardner, 1989). Erst wenn die Funkti-

on der Aufgabenbearbeitung für den Wissenserwerb erklärbar ist, können gezielt Verbesserung des Aufgabeneinsatzes in Bezug auf den Wissenserwerb untersucht werden. Es bedarf folglich weiterer theoriegeleiteter Forschung bzgl. der Wirkmechanismen von Aufgaben auf die Lernleistung der Schüler. Doch, wie bereits in Kapitel 4 verdeutlicht wurde, reichen die bisherigen theoretischen Überlegungen zur vollständigen Erklärung der gefundenen Ergebnismuster nicht aus. Besonders wichtig erscheint die theoriegeleitete Unterscheidung der verschiedenen kognitiven Prozesse der Aufgabenbearbeitung, denn nur so können differenzierte Aussagen über den Einfluss des Einsatzes verschiedener Aufgabenarten auf die Leistungen der Schüler gemacht werden (vgl. Renkl & Helmke, 1992). Daher bedarf es insgesamt differenzierterer theoretischer Modelle zur Wissensverarbeitung und zum Wissensaufbau durch die Bearbeitung von Lernaufgaben. Ebenfalls fehlen für die Erklärung der lernförderlichen Effekte der mäßig schwierigen Lernaufgaben bislang umfassende theoretische Erklärungsmodelle, wobei auf den Ergebnissen und Überlegungen zum adaptiven Unterricht aufgebaut werden kann (vgl. Kapitel 12).

Aus den vorliegenden Ergebnissen lassen sich zumindest teilweise anwendungsorientierte Implikationen für den Unterricht in Bezug auf den Einsatz von Lernaufgaben, speziell für deren Schwierigkeit, ableiten. Um Lernaufgaben möglichst effektiv im Unterricht einsetzen zu können, sollten diese in ihrer Schwierigkeit individuell an die jeweilige Schülerfähigkeit angepasst werden. Dabei sollten, den vorliegenden Befunden entsprechend, die Lernaufgaben für den Schüler mäßig schwierig sein, d.h. eine Lösungswahrscheinlichkeit von etwa 30% aufweisen. Des Weiteren sollte ein angemessenes „Mischungsverhältnis“ der unterschiedlichen Aufgabenarten bzgl. der kognitiven Prozesse der Aufgabenbearbeitung im Unterricht eingesetzt werden. Der Befund der vorliegenden Studie 1, welcher mit zahlreichen Ergebnissen aus der Literatur im Einklang steht, dass der überwiegende Teil der vom Lehrer im Unterricht gestellten Aufgaben dem Wissensniveau zuzuordnen ist, sollte dahingehend zu Veränderungen des Aufgabeneinsatzes im Unterricht führen, dass eine ausgewogene Mischung der verschiedenen Aufgabenarten zum Einsatz kommt, denn allen Aufgabenarten kommt eine lernförderliche Bedeutung zu. Erst ein gesichertes (Fakten-)Wissen ermöglicht es den Schülern, vertiefende Wissensstrukturen und ein Verständnis des Inhaltsbereichs aufzubauen, um im Anschluss das Wissen erfolgreich auf andere Aufgaben bzw. Inhaltsbereiche anwenden zu können (vgl. Levin, 2005). Somit sollte nicht der einseitige Einsatz einer Aufgabenart im Unterricht erfolgen bzw. eine Aufgabenart explizit und schwerpunktmäßig eingesetzt werden. Vielmehr erscheint ein Mischungsverhältnis der verschiedenen Aufgabenarten angebracht, wie es in der Literatur bereits von verschiedenen Au-

toren vorgeschlagen wurde (vgl. Brophy & Good, 1986; Burnstein, 1980; Gall et al., 1978; Renkl, 1991; Renkl & Helmke, 1992; Resnick & Ford, 1981; Tobin & Capie, 1982).

17.2 Ausblick und offene Fragen

Die vorliegende Arbeit liefert selbstverständlich nicht nur Antworten auf Forschungsfragen, sondern zeigt ebenfalls offene Fragen auf, welche in zukünftigen Studien untersucht werden sollten. Wie bereits in Kapitel 17.1 beschrieben, erscheint die Ausdifferenzierung des oberen Schwierigkeitsbereichs besonders relevant, um mögliche unterschiedliche Effekte mäßig schwieriger (30% Lösungswahrscheinlichkeit) und sehr schwieriger (weniger als 10% Lösungswahrscheinlichkeit) Lernaufgaben differenzieren und erklären zu können. Auch sollte in künftigen Studien gewährleistet sein, dass eine im Vergleich zu den kleinen Stichproben innerhalb der einzelnen Untersuchungsgruppen in Studie 2 größere Schülerstichproben zur Verfügung stehen.

Des Weiteren sollten die hier berichteten Ergebnisse, aufgrund der nachgewiesenen Unterschiede im mittleren Fähigkeitsparameter zwischen Gymnasium und Hauptschule, an einer anderen Schulform repliziert werden. Da in der 30%-Gruppe auch die leistungsschwächeren Schüler von der Bearbeitung mäßig schwieriger Aufgaben profitieren konnten, ist anzunehmen, dass dieser lernförderliche Effekt bzw. die Erhöhung des Fähigkeitsparameters ebenfalls an der Hauptschule bzw. an den anderen Schulformen des deutschen Bildungssystems nachgewiesen werden kann.

Da in beiden Studien ausschließlich physikalische Inhalte thematisiert wurden, sollten Untersuchungen sowohl in anderen Inhaltsbereichen als auch in anderen Fächern durchgeführt werden, um die Gültigkeit der vorliegenden Befunde auf andere Bereiche übertragen zu können.

Bei der entwickelten Methode der Aufgabenkonstruktion wurden den verschiedenen Aufgabenarten a priori unterschiedliche Schwierigkeiten zugewiesen, welche in der statistischen Überprüfung sowohl bei der IRT-Skalierung der Itemparameter als auch der Berechnung der Lösungshäufigkeiten, auf Grundlage klassischer Testverfahren bestätigt werden konnten. Dennoch sollte diese Möglichkeit zur Konstruktion von Aufgaben unterschiedlicher Schwierigkeit ebenfalls anhand weiterer Analysen repliziert werden, da in den vorliegenden Studien die Stichprobengrößen, auf deren Grundlage die Skalierung der Itemparameter erfolgte, vergleichsweise klein gewesen sind und die Ergebnisse der Skalierung und damit auch die geschätzten Itemparameter mit Vorsicht betrachtet werden sollten (vgl. Kapitel 17.3). Dies gilt auch, obschon der Modellfit der IRT-Skalierung auf Aufgabenebene und der

Treatment-Check, welcher mit Hilfe der tatsächlichen Lösungshäufigkeiten der konstruierten Aufgaben erfolgte, zufriedenstellende bzw. gute Ergebnisse in Bezug auf die unterschiedliche Schwierigkeit der konstruierten Aufgabenarten liefern. Nach den vorliegenden Befunden unterscheiden sich die mittleren Itemparameter und Lösungshäufigkeiten der konstruierten Aufgabenarten signifikant voneinander.

17.3 Grenzen der Studien

Nachdem in den vorangegangenen Kapiteln die Ergebnisse der vorliegenden Untersuchungen im Zusammenhang mit bisherigen Forschungsergebnissen diskutiert und offene Forschungsfragen aufgezeigt wurden, soll in diesem Kapitel auf die Grenzen bzw. Einschränkungen der Studien eingegangen werden.

In Bezug auf die IRT-Skalierung der Aufgabenschwierigkeit und Schülerfähigkeit sollte bei der Interpretation der Ergebnisse berücksichtigt werden, dass die Anzahl der Personen, welche die Aufgaben bearbeitet haben, klein war, z.T. weniger als 20 Schüler. Somit sollten die zugewiesenen Item- und Personenparameter mit Vorsicht interpretiert werden. Die Literatur zur Parameterschätzung gibt an, dass große Stichprobengrößen zur Parameterschätzung notwendig sind. Eine Stichprobe von 1000 Fällen wird mitunter als minimal und 2000 als adäquat empfohlen (Green et al., 1984). Lord (1980) konnte im Gegensatz dazu jedoch zeigen, dass bereits kleinere Stichproben ($N= 150-200$) hinreichend für die Schätzung eines eindimensionalen Rasch-Modells sind. In den vorliegenden Studien sind jedoch auch diese Stichprobengrößen nicht realisierbar gewesen, da die Lehrer sehr unterschiedliche Inhaltsbereiche in den Unterrichtsaufgaben thematisierten. Dieser jeweils in den Unterrichtsaufgaben behandelte Inhaltsbereich wurde in den für die Leistungstests neu konstruierten Aufgaben aufgegriffen, so dass auch diese neu konstruierten Aufgaben meistens nur in einer Klasse und somit in einer kleinen Stichprobe eingesetzt werden konnten. Aufgrund dessen sollten die Ergebnisse, wie bereits in Kapitel 17.2 diskutiert, mit Hilfe einer größeren Schülerstichprobe und einem konstant gehaltenen Inhaltsbereich repliziert werden.

Bezüglich des Designs beider Untersuchungen ist zu kritisieren, dass keine Kontrollgruppe berücksichtigt wurde. Daher kann nicht vollständig ausgeschlossen werden, dass andere, nicht auf den Aufgabeneinsatz zurückführbare Variablen oder Ereignisse ebenfalls einen Einfluss auf die Befunde gehabt haben könnten. Bei feldexperimentellen Untersuchungen, besonders im vorliegenden Fall im schulischen Bereich, in denen die Untersuchungen als Teil des regulären Unterrichts stattgefunden haben, ist die Berücksichtigung einer Kontrollgruppe kaum zu realisieren. Da die Lerneffektivität des generellen Aufgabeneinsatzes im

Vergleich zur Bedingung, in der keine Lernaufgaben eingesetzt wurden, hinreichend bestätigt wurde (vgl. Hamaker, 1986), ist die fehlende Kontrollgruppe in den vorliegenden Studien als nicht sehr gravierend anzusehen.

Dass keine weiteren Unterrichtsmerkmale, außer den eingesetzten Lernaufgaben, in den Studien berücksichtigt wurden, erscheint problematisch, da die Lernprozesse im Schulkontext durch eine Vielzahl von Faktoren und deren Wechselwirkungen bestimmt werden. Neben der Aufgabenschwierigkeit spielen weitere Merkmale eine Rolle, wie z.B. die Sequenzierung von Aufgabenfolgen und die Wartezeit nach dem Stellen der Aufgabe durch den Lehrer, bevor er einen Schüler zur Beantwortung aufruft (vgl. Brophy & Good, 1986). Die möglichen Effekte dieser Merkmale auf die Leistung wurden in den vorliegenden Studien nicht miterfasst.

Eine weitere Schwäche kann in der ausschließlichen Betrachtung von Leistungsvariablen gesehen werden, da Lernprozesse und Lernleistungen auch von motivationalen Faktoren abhängig sind. Hierzu zählen etwa die überdauernde Einstellung zu bestimmten Aufgabenarten, die Anstrengungsbereitschaft und das Lerner-Selbstkonzept. Ferner sind als mögliche Einflussvariablen, welche die individuelle Leistung beeinflussen können, die Leistungszielorientierung und die Aufgabenorientierung zu nennen (vgl. Heckhausen & Kuhl, 1985) ebenso wie die Konzentrationsfähigkeit eines Schülers (Lienert & Raatz, 1998), welche ebenfalls in den vorliegenden Studien nicht erfasst wurden.

Aufgrund der dargestellten Schwächen bzw. Grenzen der vorliegenden Studien erscheinen weitergehende Untersuchungen bzgl. der Schwierigkeit von Lernaufgaben notwendig. Insgesamt sollte somit, aufgrund der Bedeutung von Lernaufgaben für das schulische Lernen, aufbauend auf den vorliegenden Befunden weitere Forschung zu Lernaufgaben und der Passung von Aufgabenschwierigkeit und Schülerfähigkeit angestoßen werden.

Literaturverzeichnis

- Adams, R. J. & Wu, M. (2002). *PISA 2000-Technical report*. Paris: OECD.
- Adams, R. J. (2002). Scaling PISA cognitive data. In R. Adams & M. Wu (Eds.), *PISA 2000 technical report* (pp. 99-108). Paris: OECD.
- Adams, R.J. & Khoo, S.-T. (1996). *Quest: The interactive test analysis system*. Hawthorn: ACER.
- Aebli, H. (2006). *Zwölf Grundformen des Lehrens*. (13.Auflage). Stuttgart: Klett-Cotta.
- Anderson, J. R., Reder, L. M. & Simon, H. A. (1996). Situated learning and education. *Educational Researcher*, 25, 5-11.
- Anderson, L. W. & Sosniak, L. A. (Eds.). (1994). *Bloom's taxonomy: a forty-year retrospective. Ninety-third yearbook of the national society for the study of education*. Chicago: University of Chicago Press.
- Anderson, L. W. (1973). *Time and school learning*. Chicago: University.
- Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J. & Wittrock, M.C. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Anderson, R. C. & Biddle, W. (1975). On asking people questions about what they are reading. In G.H. Bower (Ed.), *The psychology of learning and motivation*. Vol. 9 (pp. 0-132). San Diego, CA: Academic Press.
- Anderson, R. C. & Faust, G. W. (1973). *Educational psychology*. New York: Dodd Mead.
- Anderson, R. C. (1970). Control of student mediating processes during verbal learning and instruction. *Review of Educational Research*, 40, 349-369.
- Anderson, R. C. (1972). How to construct achievement tests to assess comprehension. *Review of Educational Research*, 42 (2), 145-170.
-

- Andre, T. (1979). Does answering higher level questions while reading facilitate roductive learning. *Review of Educational Research*, 49, 280-318.
- Andrich, D. (1988). *Rasch models for measurement*. Newbury Park: Sage.
- Arlin, M. & Webster, J. (1983). Time cost of mastery learning. *Journal of Educational Psychology*, 75 (2), 187-195.
- Arlin, M. (1984). Time, equality, and mastery learning. *Review of Educational Research*, 54 (1), 65-86.
- Arnold, D. S., Atwood, R. K., & Rogers, V. M.(1973). An investigation of the relationship among question level, response level, and lapse time. *School Science and Mathematics*, 73, 591-594.
- Arnold, K.-H. & Neber, H. (2004). Lernen und Fragen- Eine Einführung. *Unterrichtswissenschaft*, 32 (4), 290-294.
- Aschner, M. (1961). Asking questions to trigger thinking. *NEA Journal*, 50, 44-46.
- Atkinson, J. W. (1957). Motivational determinants of risk-taking behaviour. *Psychological Review*, 64, 359-372.
- Atkinson, R. K., Derry, S. J., Renkl, A. & Wortham, D. W. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, 70, 181-214.
- Au, K. H., & Mason, J. M. (1981). Social organization factors in learning to read: The balance of rights hypothesis. *Reading Research Quarterly*, 17, 115-152.
- Ausubel, D. P. & Robinson, F. G. (1969). *School learning: An introduction to educational psychology*. New York: Holt, Rinehart and Winston.
- Ausubel, D. P. (1974). *Psychologie des Unterrichts*. Weinheim: Beltz.
- Ballstaedt, S. P., Mandl, H., Schnotz, W. & Tergan, S. O. (1981). *Texte verstehen, Texte gestalten*. München: Urban & Schwarzenberg.
- Baumert, J. (1998). Fachbezogenes-fachübergreifendes Lernen/ Erweiterte Lern- und Denkstrategien. In Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und
-

- Kunst (Hrsg.), *Wissen und Werte für die Welt von morgen. Dokumentation zum Bildungskongress* (S. 213-231). München: Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und Kunst.
- Baumert, J., & Köller, O. (2000). Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe. In: Baumert, J., Bos, W. & Lehmann, R. (Hrsg.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn, Bd. 2 Mathematische und naturwissenschaftliche Grundbildung am Ende der gymnasialen Oberstufe* (S. 271-315). Opladen: Leske + Budrich.
- Baumert, J., Bos, W. & Watermann, R. (1998). TIMSS/III-Schülerleistungen in Mathematik und Naturwissenschaften am Ende der Sekundarstufe II im internationalen Vergleich. Opladen: Leske + Budrich.
- Baumert, J., Bos, W., & Lehmann, R. (Hrsg.) (2000). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn, Bd. 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit*. Opladen: Leske + Budrich.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K. J. & Weiß, M. (Hrsg.). (2001). *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J., Kunter, M., Brunner, M., Krauss, S., Blum, W. & Neubrand, M. (2004). Mathematikunterricht aus Sicht der PISA-Schülerinnen und –Schüler und ihrer Lehrkräfte. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland - Ergebnisse des zweiten internationalen Vergleichs* (S. 314-254). Münster: Waxmann.
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B, Clausen, M., Hosenfeld, I., Köller, O. & Neubrand, J. (1997). *TIMSS-Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich: Deskriptive Befunde*. Opladen: Leske + Budrich.
-

- Baumert, J., Schmitz, B., Sang, F. & Roeder, P. M. (1987). Zur Kompatibilität von Leistungsförderung und Divergenzminderung in Schulklassen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 19, 249-265.
- Bean, J. P. (1985). Interaction effects based on class level in an explanatory model of college student dropout syndrome. *American Educational Research Journal*, 22 (1), 35-64.
- Bednar, A. K., Cunningham, D., Duffy, T. M. & Perry, J. D. (1992). Theory into practice: How do we link? In T. M. Duffy & D. H. Jonassen (Eds.), *Constructivism and the technology of instruction* (pp.17-34). Hillsdale: Erlbaum.
- Bereiter, C. & Scardamalia, M. (1998). Beyond Bloom's taxonomy: rethinking knowledge for the knowledge age. In A. Hargraves, A.Liderman, M. Fullan & D. Hopkins (Eds.), *International handbook of educational change* (pp. 675-692). London: Kluwer Academic Publishers.
- Bergstrom, B. A., Lunz, M. E. & Gershon, R. C. (1992). Altering the level of difficulty in computer adaptive testing. *Applied Measurement in Education*, 5 (2), 137-149.
- Berliner, D. C. & Fisher, C. W. (1985). *Perspectives on instructional time*. New York: Longman.
- Betz, N. E. & Weiss, D. J. (1976). *Psychological effect of immediate knowledge of results and adaptive ability testing* (Research Rep. 76-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program.
- Biggs, J. & Collins, K. (1982) *Evaluating the quality of learning: the SOLO taxonomy* New York: Academic Press.
- BLK-Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (Hrsg.). (1997). *Gutachten zur Vorbereitung des Programms „Steigerung der Effizienz des mathematisch- naturwissenschaftlichen Unterrichts“* (= BLK-Materialien zur Bildungsplanung und Forschungsförderung, Heft 60). Bonn: BLK.
- Block, J. H. & Burns, R. B. (1976). Mastery learning. In L. S. Shulman (Ed.), *Review of research in education* (Vol.4) (pp. 3-49). Itasca, Ill: Peacock.
- Block, J. H. (1971). (Ed.). *Mastery learning: theory and practice*. New York: Holt, Rinehart & Winston.
-

- Bloom, B.S. (1976). *Human characteristics and school learning*. New York: Wiley.
- Bloom, B.S., Engelhart, M.D., Frust, E.J., Hill, W.H. & Krathwohl, D.R. (1956). *Taxonomy of educational objectives: Handbook I: Cognitive domain*. New York: McKay.
- Blosser, P.E. (1979). *Review of research: teacher questioning behavior in science classrooms*. Columbus, Ohio: EPIC Clearinghouse for Science, Mathematics, and Environmental Education, 1979. ERIC Document ED 184 818.
- Bortz, J. & Döring, N. (2002). *Forschungsmethoden und Evaluation*. Berlin: Springer.
- Bransford, J. D., Brown, A. L. & Cocking, R. R. (1999). *How people learn: Brain, mind, experience and school*. Washington, D.C: National Academy Press.
- Bratman, M.E. (1987). *Intentions, plans, and practical reasons*. Cambridge, MA: Harvard University Press.
- Briggs, L. J. (1977). *Instructional design: principles and applications*. Englewood Cliffs, NJ: Educational Technology Publications.
- Bromme, R., Seeger, F. & Steinbring, H. (1990). *Aufgaben als Anforderungen an Lehrer und Schüler* (= IDM-Untersuchungen zum Mathematikunterricht, Bd. 14). Köln: Aulis.
- Brophy, J. E. & Evertson, C. M. (1976). *Learning from teaching: a developmental perspective*. Boston: Allyn and Bacon, Inc.
- Brophy, J. E. & Good, T. L. (1986). Teacher behaviour and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd Ed.) (pp. 328-375). New York: Macmillian.
- Brophy, J. E. (1983). Classroom organization and management. *The Elementary School Journal*, 83, 265-285.
- Brophy, J. E. (2000). *Teaching*. (Educational Practices Series, Vol.1). Brüssel: International Academy of Education and International Bureau of Education.
-

- Brown, A. L. (1989). Analogical learning and transfer: What develops? In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 369-412). Cambridge, England: Cambridge University Press.
- Brualdi, A. C. (1998). Classroom questions. *Practical Assessment, Research and Evaluation*, 6, (6), ERIC Document reproduction No: ED 422407.
- Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, 31, 21-32.
- Burns, R. (1984). The process and context of teaching: A conceptual framework. *Evaluation in Education*, 8, 95-112.
- Burnstein, L. (1980). Issues in the aggregation of data. *Review of Research in Education*, 8, 158-233.
- Bynes, J. P. (1996). *Cognitive development and learning in instructional contexts*. Boston: Allyn & Bacon.
- Carroll, J.B. (1963). A model of school learning. *Teachers College Report*, 64, 723-733.
- Cartwright, C.A., Cartwright, G.P. & Robine, G.E. (1972). CAI Course in the early identification of handicapped children. *Exceptional Children*, 38 (6), 453-459.
- Case, R. (1998). The development of conceptual structures. In W. Damon (Series Ed.), D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol.2. Cognition, perception, and language* (5th ed.) (pp. 745-800). New York: Wiley.
- Christiansen, B. & Walther, G. (1986). Task and activity. In B. Christiansen, A.G. Howsen & M. Otte (Eds.), *Perspectives on mathematics education* (pp.243-307). Dordrecht: Reidel Publishing Company.
- Clegg, A. A. (1987). Why questions?. In W. W. Wilen (Ed.), *Questions, questioning techniques, and effective teaching* (pp. 11-22). Washington, D.C.: National Education Association.
- Cobb, P., Perlwitz, M. & Underwood, D. (1996). Constructivism and activity theory: a consideration of their similarities and differences as they relate to mathematics education. In H. Mansfeld, N. A. Pateman & N. Bednarz (Eds.), *Mathematics for tomorrow's young children* (pp. 10-59). London: Kluwer Academic Publishers.
-

- Cohen, I. S. (1963). Programmed learning and the socratic dialogue. *American Psychologist*, 17, 772-775.
- Cohen, J. (1977). *Statistical power analysis for the behavioral science*. New York: Academic Press.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation. Design and analysis for field study*. Chicago: Rand McNally.
- Corno, L. & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 605-629). New York: Macmillan.
- Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.
- Cronbach, L. J. & Furby, L. (1970). How we should measure „change“-or should we? *Psychological Bulletin*, 74, 68-80.
- Cronbach, L. J. & Snow, R. E. (1977). *Aptitudes and instructional methods*. New York: Irvington Press.
- Cronbach, L. J. & Snow, R. E. (1981). *Aptitudes and instructional methods: A handbook for research in interactions*. New York: Irvington.
- Cronbach, L. J. (1967). How can construction be adapted to individual differences? In R. M. Gagné (Ed.), *Learning and individual differences* (pp. 23-39). Columbus, Ohio: Merrill.
- Crooks, T.J. (1988). The impact of classroom evaluation practice on students. *Review of Educational Research*, 58 (4), 438-481.
- Cunningham, R. T. (1987). What kind of question is that? In W. W. Wilen (Ed.) *Questions, questioning techniques, and effective teaching* (pp.67-94). Washington, D.C: National Education Association.
- Dahllöf, U. (1971). *Ability grouping, content validity, and curriculum process analysis*. NY: Teachers College Press.
-

- DeLandsheere, V. (1977). On defining educational objectives. *Evaluation in Education: International Review Series*, 1, 73-190.
- Detterman, D. K. & Sternberg, R. J. (1993). *Transfer on trail: intelligence, cognition, and instruction*. Norwood, NJ: Ablex.
- Dick, W. & Carey, L. (1996). *The systematic design of instruction* (4th ed.). New York: Harper Collins.
- Diem, D. C. (1982). *The effectiveness of computer assisted instruction in College Algebra*. Florida: Atlantic University.
- Dillon, J. T. (1982a). The multidisciplinary study of questioning. *Journal of Educational Psychology*, 74, 147-165.
- Dillon, J. T. (1982b). Cognitive correspondence between question/statement and response. *American Educational Research Journal*, 19, 540-552.
- Dillon, J. T. (1994). *Using discussion in classrooms*. Milton Keynes, UK: Open University Press.
- Ditton, H. (2006). Unterrichtsqualität. In K.-H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (S.235-243). Bad Heilbrunn: Klinkhardt.
- Dörner, D. (1976). *Problemlösen als Informationsverarbeitung*. (2. Aufl.) Stuttgart: Kohlhammer.
- Doyle, W. (1983). Academic work. *Review of Educational Research*, 53, 159-199.
- Doyle, W. (1986). Classroom organization and management. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 392-431). London: Macmillan.
- Doyle, W. (1988). Work in mathematics classes: the context of students' thinking during instruction. *Educational Psychologist*, 23, 167-180.
- Dubs, R. (1978). *Aspekte des Lehrerverhaltens*. Aarau: Sauerländer.
- Duchastel, P. C. (1981). Retention of prose following testing with different types of tests. *Contemporary Educational Psychology*, 6, 217-226.
-

- Ellis, K. (1993). *Teacher questioning behavior and student learning: What research says to teachers*. Columbus, Ohio: EPIC Clearinghouse for Science, Mathematics, and Environmental Education, 1979. ERIC Document ED 359 572.
- Feger, B. (1984). Die Generierung von Testitems zu Lehrtexten. *Diagnostica*, 30 (1), 24-46.
- Fischer, G. H. (1968). *Psychologische Testtheorie*. Bern: Huber. (Psychologisches Kolloquium; Bd. 5).
- Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests: Grundlagen und Anwendungen*. Bern: Huber.
- Fisher, C. W., Berliner, D. C., Filby, N. N., Marliave, R., Cahen, L. S. & Sishaw, M. M. (1980). Teaching behaviors, academic learning time, and student achievement: An overview. In C. Denham & A. Leiberman (Eds.), *Time to learn* (pp. 7-323). Washington, DC: Department of Health, Education, and Welfare, National Institute of Education.
- Flanagan, J. C., Shanner, W. M., Brudner, H. J. & Marker, R. W. (1975). An individualized instructional system: PLAN. In H. Talmadge (Ed.), *Systems of individualized education* (pp. 136-167). Berkeley: McCutchan.
- Fleischer, J., Wirth, J. & Leutner, D. (in Druck). Testmethodische Grundlagen der Lernstandserhebungen NRW: Erfassung von Schülerkompetenzen für Vergleiche mit kriterialen und sozialen Bezugsnormen. In Landesinstitut für Schule/Qualitätsagentur (Hrsg.), *Umgang mit Ergebnissen von Lernstandserhebungen im Fach Mathematik – das Beispiel NRW*. Stuttgart: Klett.
- Fleming, M. & Chambers, B. (1983). Teacher-made tests: windows to the classroom. In W. E. Hathaway (Ed.), *Testing in the schools* (pp. 29-47). San Francisco: Jossey-Bass.
- Fletcher, J. D. (1992). *Individualized systems of instruction*. Educational Resources Information Center, ERIC 355 917.
- Floyd, W. D. (1960). *An analysis of the oral questioning activities in selected colorado primary classroom*. Colorado: State College.
- Foster, P. (1983). Verbal participation and outcomes. In L. Ellner & C. Barnes (Eds.), *Studies of college teaching* (pp. 117-160). Toronto, ON: Heath.
-

- Frase, L. T. (1968). Some unpredicted effects of different questions upon learning from connected discourse. *Journal of Educational Psychology*, 59, 197-200.
- Fricke, R. (1974). *Kriteriumsorientierte Leistungsmessung*. Stuttgart: Kohlhammer.
- Fricke, R. & Lühmann, R. (1983). Kriteriumsorientierte Tests-Theorie und Praxis. *Psychologie in Erziehung und Wissenschaft*, 30 (3), 173-182.
- Fürntratt, E. (1978). Aufgabenschwierigkeit, Übungsfortschritt und Arbeitsmotivation. *Psychologie in Erziehung und Unterricht*, 25, 221-230.
- Gage, N. L. & Berliner, D. C. (1996). *Pädagogische Psychologie* (6. Aufl.). Weinheim: Beltz.
- Gagné, R. M. (1962). The acquisition of knowledge. *Psychological Review*, 69, 355-365.
- Gagné, R. M. (1965). *The conditions of learning*. New York: Holt, Rinehart & Winston.
- Gagné, R. M. (1968). Learning hierarchies. *Educational Psychologist*, 6, 3-6.
- Gagné, R. M. (1972). Domains of learning. *Interchange*, 3 (1), 1-8.
- Galambos, J. A. & Black, J. B. (1985) Using knowledge of activities to understand and answer questions. In A. C. Graesser & J. B. Black (Eds.), *The psychology of questions* (pp. 219–226). Hillsdale, NJ: Erlbaum.
- Gall, M. D. & Artero-Boname, M. T. (1995). Questioning. In L. W. Anderson (Ed.), *International encyclopedea of teaching and teacher education* (2nd ed.) (pp. 242-248). Oxford: Pergamon.
- Gall, M. D. & Rhody, T. (1987). Review of research on questioning techniques. In: W. W. Wilen (Ed.), *Questions, questioning techniques, and effective teaching*. (23-48) Washington, D.C: National Education Association.
- Gall, M. D. (1970). The use of questions in teaching. *Review of Educational Research*, 40, 707-720.
- Gall, M. D. (1984). Synthesis of research on teachers' questioning. *Educational Leadership*, 42, 40-47.
-

- Gall, M. D., Ward, B. A., Berliner, D. C., Cahen, L. S., Winne, P. H., Elashoff, J. D. & Stanton, G. C. (1978). Effects of questioning techniques and recitation on student learning. *American Educational Research Journal*, 15, 175-199.
- Gallagher, J. J. (1965). Expressive thought by gifted in children in the classroom. *Elementary English*, 42, 559-568.
- Gardner, H. (1989). *Dem Denken auf der Spur. Der Weg der Kognitionswissenschaft*. Stuttgart: Klett-Cotta.
- Gardner, H. (1993). *Multiple intelligence. The theory in practice*. New York: BasisBooks.
- Garten, H.-K. (1977). *Wechselwirkungen zwischen Schülermerkmalen und Lernbedingungen*. Weinheim: Beltz.
- Gavelek, J. & Raphael, T. E. (1985). Metacognition, instruction, and the role of questioning activities. In D. L. Forrest-Pressley, G. E. MacKinnon & T. G. Walter (Eds.), *Metacognition, cognition, and human performance* (Vol. 2, Instructional practices, pp. 103-136). New York: Academic Press.
- Gerlach, V. & Sullivan, A. (1967). *Constructing statements of outcomes*. Inglewood, CA: Southwest Regional Laboratory for Educational Research and Development.
- Gerstenmaier, J. & Mandl, H. (2000). Wissensanwendung im Handlungskontext: Die Bedeutung intentionaler und funktionaler Perspektiven für den Zusammenhang von Wissen und Handeln. In H. Mandl & J. Gerstenmaier (Hrsg.), *Die Kluft zwischen Wissen und Handeln* (S. 289-311). Göttingen: Hogrefe.
- Gick, M. L. & Holyoak, K. J. (1987). The cognitive basis of knowledge transfer. In S. M. Cormier & J. D. Hagman (Eds.), *Transfer of learning* (pp. 9-47). San Diego: Academic Press.
- Gick, M.L. & Holyoak, K.J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1-38.
- Glaser, R. (1977). *Adaptive education: individual diversity and learning*. New York: Holt, Rinehart & Winston.
-

- Glaser, R. (1984). Education and thinking. The role of knowledge. *American Psychologist*, 39, 93-104.
- Good, T. L., & Brophy, J. E. (1971). Self-fulfilling prophecy. *Today's Education*, 60, 52-53.
- Goodlad, J. I. (1983). A study of schooling: some findings and hypotheses. *Phi Delta Kappan*, 64, 465-470.
- Graesser, A. C. & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31 (1), 104-137.
- Graesser, A. C. (1981). *Prose comprehension beyond the word*. Heidelberg: Springer.
- Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L. & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 21 (4), 347-360.
- Greenfield, P. (1984). A theory of the teacher in the learning activities of everyday life. In B. Rogoff & J. Lave (Eds.), *Everyday cognition* (pp. 117-138). Cambridge, MA: Harvard University Press.
- Greeno, J. G. (1987). Instructional representations based on research about understanding. In A. Schoenfeld (Ed.), *Cognitive science and math instruction* (pp. 61-88). Hillsdale, NJ: Erlbaum.
- Greve, W. & Wentura, D. (1997). *Wissenschaftliche Beobachtung*. Weinheim: Psychologie Verlags Union.
- Gruber, H., Mandl, H. & Renkl, A. (2000). Was lernen wir in Schule und Hochschule: Träges Wissen? In H. Mandl & J. Gerstenmaier (Hrsg.), *Die Kluft zwischen Wissen und Handeln – Empirische und theoretische Lösungsansätze*. (S. 139-156). Göttingen: Hogrefe.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen*. Münster: Waxmann
- Häfele, G. (1995). *Lehrtexte im Selbststudium erarbeiten: Fördern Studienfragen den Wissenserwerb?* Marburg: Phillips-Universität.
-

- Hager, W., Patry, J.-L. & Brezing, H. (2000). *Evaluation psychologischer Interventionsmaßnahmen*. Bern: Huber.
- Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research*, 56 (2), 212-242.
- Hamilton, R. (1989). The role of concept definition, teaching examples and practice on concept learning from prose. *Contemporary Educational Psychology*, 14, 357-365.
- Hancock, G. R. (1994). Cognitive complexity and the comparability of multiple choice and constructed response test formats. *Journal of experimental Education*, 62(2), 143-157.
- Hare, V. C. & Pulliam, C. A. (1978). Teacher questioning: a verification and an extension. *Journal of Reading Behavior*, 12, 69-72.
- Hartley, S. S. (1975). *Meta-analysis of the effects of individually paced instruction in mathematics*. Colorado: University.
- Häußler, P. & Lind, G. (1998). *BLK-Programmförderung "Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts". Erläuterungen zu Modul 1 mit Beispielen für den Physikunterricht. Weiterentwicklung der Aufgabenkultur im mathematisch-naturwissenschaftlichen Unterricht.* Verfügbar unter: <http://www.learnline.de/angebote/sinus/zentral/grundlagen/projektbeschreibungen/uebersicht.html>.
- Hecht, L. W. (1977). *Isolation from learning supports and the processing of group instruction*. Ph.D. dissertation. Chicago: University.
- Heckhausen, H. & Kuhl, J. (1985). From wishes to action: The dead ends and short cuts on the long way to action. *Journal of Personality and Social Psychology*, 54, 5-12.
- Heckhausen, H. (1968). Förderung der Lernmotivation und der intellektuellen Tüchtigkeiten. In H. Roth (Hrsg.), *Begabung und Lernen* (Band 3, S.193-228). Stuttgart: Klett.
- Helmke, A. & Schrader, F.-W. (1988). Successful student practice during seatwork: efficient management and active supervision not enough. *Journal of Educational Research*, 82 (2), 70-75. Helmke, A., Schrader, F.-W. & Weinert, F.E. (1987). Zur Rolle der Übung für den Lernerfolg. *Blätter für Lehrerfortbildung*, 39, 247-252.
-

- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule. Enzyklopädie der Psychologie*. Themenbereich D, Serie I, Bd. 3 (S. 71-176). Göttingen: Hogrefe.
- Helmke, A. (2004). *Unterrichtsqualität erfassen, bewerten, verbessern*. Seelze: Kallmeyer.
- Herber, H.-J. (1979). *Motivationstheorie und pädagogische Praxis*. Stuttgart: Kohlhammer.
- Hill, P. W. & McGaw, B. (1981). Testing the simplex assumption underlying the Bloom's taxonomy. *American Educational Research Journal*, 18, 93-101.
- Hoge, R. D. & Coladarci, T. (1989). Teacher-based judgments of academic achievement: a review of literature. *Review of Educational Research*, 59, 297-313.
- Hornke, L. F. (1977). Antwortabhängige Testverfahren: Ein neuartiger Ansatz psychologischen Testens. *Diagnostica*, 23, 1-14.
- Hosenfeld, I.; Helmke, A. & Schrader, F.-W. (2002). Diagnostische Kompetenz: Unterrichts- und lernrelevante Schülermerkmale und deren Einschätzung durch Lehrkräfte in der Studie SALVE. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen*. *Zeitschrift für Pädagogik*, 45. Beiheft (S. 65-82). Weinheim: Beltz
- Huber, G. (1966). *Lernpsychologische Befunde bei programmierter Unterweisung*. München: Ehrenwirth.
- Hunkins, F. P. (1976). *Involving students in questioning*. Boston: Allyn and Bacon.
- Ingenkamp, K. (1997). *Lehrbuch der Pädagogischen Diagnostik* (4.Aufl.). Weinheim: Beltz.
- Jatzwauk, P. (in Druck). *Unterrichtsstruktur und Aufgabeneinsatz im Biologieunterricht*. Essen: Universität.
- Johnson-Laird, P. N. (1983). *Mental models: towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Johnston, P. H. (1983). *Reading comprehension assessment: a cognitive basis*. Newark, DE: International Reading Association.
-

- Jonassen, D. H., Tessmer, M. & Hannum, W. H. (1999). *Task Analysis Methods for Instructional Design*. Mahwah, NJ: Erlbaum.
- Jorgenson, G. (1977). Relationship of classroom behavior to the accuracy of the match between material difficulty and student ability. *Journal of Educational Psychology*, 69, 24-32.
- Kalyuga, S., Chandler, P., Tuovinen, J. & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, 93, 579-588.
- Keil, F. (1998). Cognitive science and the origins of thought and knowledge. In W. Damon (Series Ed.), D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. I. Theoretical models of human development* (5th ed.) (pp. 341-413). New York: Wiley.
- Keitel, C., Otte, M. & Seeger, F. (1980). *Text-Wissen-Tätigkeit. Das Schulbuch im Mathematikunterricht*. Königstein: Scriptor.
- Keller, F. S. & Sherman, J. G. (1974). *The Keller Plan handbook*. Menlo Park, CA: Benjamin.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, 1988, 95 (2), 163-182.
- Klauer, K. J. (1974). *Methodik der Lehrzieldefinition und Lehrstoffanalyse*. Düsseldorf: Pädagogischer Verlag Schwann.
- Klauer, K. J. (1987). *Kriteriumsorientierte Tests*. Göttingen: Hogrefe.
- Klauer, K. J. (2001). Forschungsmethoden der Pädagogischen Psychologie. In A. Krapp & B. Weidemann (Hrsg.), *Pädagogische Psychologie* (S. 75-97). Weinheim: Beltz.
- Klausmeier, H.J. (1975). IGE: An alternative form of schooling. In H. Talmadge (Ed.), *Systems of individualized education*. (pp. 48-83) Berkeley, CA: McCutchan.
- Klieme, E. (2000). Fachleistungen im voruniversitären Mathematik- und Physikunterricht: Theoretische Grundlagen, Kompetenzen und Unterrichtsschwerpunkt. In J. Baumert, W. Bos & R. Lehmann (Hrsg.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie - Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn, Bd. 2 Mathematische und naturwissenschaftliche Grundbildung am Ende der gymnasialen Oberstufe* (S. 57-128). Opladen: Leske + Buderich.
-

- Kloep, M. & Weimann, F. (1982). Aufgabenschwierigkeit und Mathematikleistungen bei Realschülern. *Psychologie in Erziehung und Unterricht*, 29, 76-80.
- Kloep, M. (1985). Leicht oder mittelschwer? Versuch einer Antwort auf eine Streitfrage in der Pädagogischen Psychologie. *Unterrichtswissenschaft*, 13 (2), 130-139.
- Köller, O., Watermann, R. & Baumert, J. (2001). Skalierung der Leistungstests in PISA. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 517-524). Opladen: Leske+Buderich.
- Kottke, J. L. & Schuster, D. H. (1990). Developing tests for measuring Bloom's learning outcomes. *Psychological Report*, 66, 27-32.
- Kounin, J. S. (1976). *Techniken der Klassenführung*. Stuttgart: Klett.
- Kreitzer, A. E. & Madaus, G. F. (1994). Empirical investigations of the hierarchical structure of the taxonomy. In L. W. Anderson & L. A. Sosniak (Eds.), *Bloom's taxonomy: a forty-year retrospective, ninety-third yearbook of the national society for the study of education* (pp. 64-81). Chicago: University of Chicago Press.
- Kretschmann, R. & Malburg, H. (1976). Generelle und differentielle Effekte dreier verschiedener Formen adaptiven Unterrichts. *Zeitschrift für Pädagogik*, 22 (6), 889-900.
- Kropp, R. P. & Stoker, H. W. (1966). *The construction and validation of tests of the cognitive processes as described in the taxonomy of educational objectives*. ERIC ED 010 044; Tallahassee, FL: Florida State University, Institute of Human Learning and Department of Educational Research and Testing.
- Krumm, V. (1985). Anmerkungen zur Rolle der Aufgaben in der Didaktik. *Unterrichtswissenschaft*, 2, 102-115.
- Kulik, C.-L. & Kulik, J. A. (1991). Effectiveness of computer-based instruction: an updated analysis. *Computers in Human Behavior*, 7, 75-94.
- Kulik, C.-L., Cohen, P. A. & Ebeling, B. J. (1980). Effectiveness of programmed instruction in higher education: a meta-analysis of findings. *Educational Evaluation and Policy Analysis*, 6 (2), 51-64.
-

- Kulik, J. A. & Kulik, C.-L. (1987). Review of recent research literature on computer-based instruction. *Contemporary Educational Research*, 12, 222-230.
- Kulik, J. A., Kulik, C.-L. & Cohen, P. A. (1979). A meta-analysis of outcome studies of Keller's Personalized System of Instruction. *American Psychologist*, 34, 307-318.
- Kunter, M., Dubberke, T., Baumert, J., Blum, W., Brunner, M., Jordan, A., Klusmann, U., Krauss, S., Löwen, K., Neubrand, M. & Tsai, Y.-M. (2006). Mathematikunterricht in den PISA-Klassen 2004: Rahmenbedingungen, Formen und Lehr-Lernprozesse. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003-Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S.161-194). Münster: Waxmann.
- Lave, J. (1992). Word problems: a microcosm of theories of learning. In P. Light & G. Butterworth (Eds.), *Context and cognition: ways of learning and knowledge* (pp.74-92). Hillsdale, NJ: Erlbaum.
- Law, L.-C. (2000). Die Überwindung der Kluft zwischen Wissen und Handeln aus situativer Sicht. In H. Mandl & J. Gerstenmaier (Hrsg.), *Die Kluft zwischen Wissen und Handeln - Empirische und theoretische Lösungsansätze* (S. 253-287). Göttingen: Hogrefe.
- Leutner, D. (1992). *Adaptive Lehrsysteme: Instruktionspsychologische Grundlagen und experimentelle Analysen*. Weinheim: Psychologie Verlags Union.
- Leutner, D. (2004). Instructional design principles for adaptivity in open learning environments. In N. M. Seel & S. Dijkstra (Eds.), *Curriculum, plans and processes of instructional design: international perspectives* (pp. 289-307). Mahwah, NJ: Lawrence Erlbaum.
- Levin, A. (2005). *Lernen durch Fragen*. Münster: Waxmann.
- Levin, T. & Long, R. (1981). *Effective instruction*. Washington DC: Association for Supervision and Curriculum Development.
- Lienert, G.A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6.Aufl.). München: Psychologie Verlags Union.
- Lindfors, J. W. (1980). *Children's language and learning*. Englewood Cliffs, NJ: Prentice-Hall.
-

- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Lundgren, U. P. (1972). *Frame factors and the teaching process. A contribution to curriculum theory and theory on teaching*. Göteborg: University.
- Mandl, H. & Gerstenmaier, J. (Hrsg.) (2000). *Die Kluft zwischen Wissen und Handeln – Empirische und theoretische Lösungsansätze*. Göttingen: Hogrefe.
- Mandler, J. (1998). Representation. In W. Damon (Series Ed.), D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol.2. Cognition, perception, and language* (5th ed.) (pp. 255-308). New York: Wiley.
- Martin, M. O. & Kelly, D. L. (1998). *Third International Mathematics and Science Study (TIMSS). Technical Report (Volume II): Implementation and Analysis. Primary and Middle School Years (Population 1 and Population 2)*. Chestnut Hill, MA: Boston College.
- Marzano, R. J. (1992). *A different kind of classroom*. Alexandria, VA: Association of Supervision and Curriculum Development.
- Mayer, R. E. & Wittrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47-62). New York: Macmillan.
- Mayer, R. E. (1995). Teaching and testing for problem solving. In L. W. Anderson (Ed.), *International encyclopedia of teaching and teacher education*. (2nd ed.) (pp. 4728-4731). Oxford, UK: Pergamon.
- McCarthy, J. & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence, *Machine Intelligence*, 4, 463-502.
- McKeough, A., Lupart, J. & Marini, A. (Eds.). (1995). *Teaching for transfer*. Mahwah, NJ: Erlbaum.
- McKinley, R. L. & Reckase, M. D. (1980). Computer applications to ability testing. *Association for Educational Data Systems Journal*, 13, 193-203.
-

- Messner, R. (1973). Funktionen der Taxonomie für die Planung von Unterricht. In B. S. Bloom (Hrsg.), *Taxonomie von Lernzielen im kognitiven Bereich* (S. 227-251). Weinheim: Beltz.
- Metfessel, N. S., Michael, W. G. & Kirsner, D. A. (1969). Instrumentation of Bloom's and Krathwohl's taxonomies for the writing of educational objectives. *Psychology in the Schools*, 6, 227-231.
- Mietzel, G. (2003). *Psychologie in Unterricht und Erziehung*. Göttingen: Hogrefe.
- Mills, S. R., Rice, C. T., Berliner, D. C. & Rousseau, E. W. (1980). Correspondence between teacher questions and student answers in classroom discourse. *Journal of Experimental Education*, 48, 194-204.
- Mosenthal, P. B. (1998). Defining prose task characteristics for use in computer-adaptive testing and instruction. *American Educational Research Journal*, 35, 269-307.
- Neubrand, J. (2002). *Eine Klassifikation mathematischer Aufgaben zur Analyse von Unterrichtssituationen: Selbstständiges Arbeiten in Schülerarbeitsphasen in den Stunden der TIMSS-Video-Studie*. Hildesheim: Franzbecker.
- Niegemann, H. & Stadler, S. (2001). Hat noch jemand eine Frage? Systematische Unterrichtsbeobachtung zur Häufigkeit und kognitivem Niveau von Fragen im Unterricht. *Unterrichtswissenschaft*, 29 (2), 171-192.
- Niemiec, R. & Walberg, H. J. (1987). Comparative effects of computer-assisted instruction: A synthesis of reviews. *Journal of Educational Computing Research*, 3, 19-37.
- Nix, D. (1990). Should computers know what you can do with them? In D. Nix & R. J. Spiro (Eds.), *Cognition, education, and multimedia. Exploring ideas in high technology* (pp. 143-162). Hillsdale: Erlbaum.
- Nussbaum, A. & Leutner, D. (1986). Entdeckendes Lernen von Aufgabenlösungsregeln unter verschiedenen Anforderungsbedingungen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 18 (2), 153-164.
- Ohlsson, S. & Rees, E. (1991). The function of conceptual understanding in the learning of arithmetic procedures. *Cognition and Instruction*, 8, 103-179.
-

- Park, O. C. & Lee, J. (2003). Adaptive instructional systems. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology*. (2nd ed., pp.651-684). Mahwah, NJ: Lawrence Erlbaum.
- Peterson, P. L. (1979). Direct instruction reconsidered. In P. L. Peterson & H. J. Walberg (Eds.), *Research on teaching. concepts, findings, and implications* (pp. 57-69). Berkeley, CA: McCutchan.
- Phye, G. D. (1997). *Handbook of classroom assessment*. San Diego, CA: Academic Press.
- Prenzel, M., Rost, J., Senkbeil, M., Häußler, P. & Klopp, A. (2001). Naturwissenschaftliche Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.J. Tillmann & M. Weiß (Hrsg.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 191-248). Opladen: Leske + Budrich.
- Quellmalz, E.-S. (1987). Developing reasoning skills. In J. F. Baron, J. Boykoff & R. J. Sternberg (Eds.), *Teaching thinking skills: theory and practice* (pp. 86-105). New York: Freeman.
- Rasch, G. (1980). *Probabilistic modes for some intelligence and attainment tests*. (2.ed.). Chicago: University of Chicago Press. (Originalausgabe 1960)
- Recker, M. M. & Pirolli, P. (1995). Modelling individual differences in students' learning strategies. *Journal of the Learning Sciences*, 4, 1-38.
- Redfield, D. L., & Rousseau, E. W. (1981). Meta-analysis of experimental research on teacher questioning behaviour. *Review of Educational Research*, 51, 237-245.
- Reimann, P. (1997). *Lernprozesse beim Wissenserwerb aus Beispielen*. Bern: Huber.
- Reinmann-Rothmeier, G. & Mandl, H. (1998). Wissensvermittlung: Ansätze zur Förderung des Wissenserwerbs. In F. Klix & H. Spada (Hrsg.), *Wissen, Enzyklopädie der Psychologie*, Serie Kognition, Band 6 (S. 457-500). Göttingen: Hogrefe.
- Reinmann-Rothmeier, G. & Mandl, H. (2001). Unterrichten und Lernumgebungen gestalten. In B. Weidenmann, A. Krapp, M. Hofer, G. L. Huber & H. Mandl (Hrsg.), *Pädagogische Psychologie* (S. 603-648). Weinheim: Beltz.
-

- Renkl, A. & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: a cognitive load perspective. *Educational psychologist*, 38 (1), 15-22.
- Renkl, A. & Helmke, A. (1992). Discriminant effects of performance-oriented and structure-oriented mathematics tasks on achievement growth. *Contemporary Educational Psychology*, 17 (1), 47-55.
- Renkl, A. (1991). *Die Bedeutung der Aufgaben- und Rückmeldungsgestaltung für die Leistungsentwicklung im Fach Mathematik*. Heidelberg: Universität.
- Renkl, A. (1996). Träges Wissen: Wenn Erlerntes nicht genutzt wird. *Psychologische Rundschau*, 47 (2), 78-92.
- Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction*, 12, 529-556.
- Renkl, A., Gruber, H. & Mandl, H. (1996). *Situated learning in instructional settings: From euphoria to feasibility*. Forschungsbericht Nr. 74. München: Ludwig-Maximilians-Universität.
- Renkl, A., Gruner, H., Weber, S., Lerche, T. & Schweizer, K. (2003). Cognitive Load beim Lernen aus Lösungsbeispielen. *Zeitschrift für Pädagogische Psychologie*, 17 (2), 93-101.
- Renkl, A., Schworm, S. & Hilbert, T. S. (2004). Lernen aus Lösungsbeispielen: Eine effektive, aber kaum genutzte Möglichkeit, Unterricht zu gestalten. In: J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule. Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategie der Qualitätsverbesserung* (S. 77-92). Münster: Waxmann.
- Resnick, L. B. & Ford, W. W. (1981). *The psychology of mathematics for instruction*. Hillsdale, NJ: Lawrence Erlbaum.
- Rickards, J. P. & Denner, P. R. (1978). Inserted questions as aids to reading text. *Instructional Science*, 7, 313-346.
- Rickards, J. P. (1979). Adjunct post-questions in text: a critical review of methods and processes. *Review of Educational Research*, 49, 181-196.
-

- Rimmele, R. (2004). *Videograph. Multimedia-Player zur Kodierung von Videos*. Kiel: IPN.
- Rinck, M. (2000). Situationsmodelle und das Verstehen von Erzähltexten: Befunde und Probleme. *Psychologische Rundschau*, 3, 115-122.
- Robitaille, D. F. & McKnight, C. (1993). *Curriculum frameworks for mathematics and science*. Vancouver: Pacific Educational Press.
- Ropo, E. (1990). Teachers' questions: some differences between experienced and novice teachers. In: H. Mandl, E. de Corte, S.N. Bennett & H.F. Friedrich (Eds.), *Learning and Instruction. European Research in an International Context. Volume 2.2: Analysis of Complex Skills and Complex Knowledge Domains* (pp. 114-128). Oxford: Pergamon Press.
- Rosenshine, B. V. (1979). Content, time, and direct instruction. In P. Peterson & H. Walberg (Eds.). *Research on teaching: concepts, findings, and implications* (pp. 28-56). Berkeley, CA: McCutchan.
- Rosenshine, B. V. (1987). Explicit teaching. In D.C. Berliner & B.V. Rosenshine (Eds.). *Talks to teachers. A festschrift for N.L. Gage* (pp. 75-92). New York, NY: McGraw Hill.
- Rosenshine, B.V. (1971). *Teaching behaviours and student achievement*. London: National Foundation for Educational Research.
- Ross, B. H. (1987). This is like that: The use of earlier problems and separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 629-639.
- Rost, J. & Spada, H. (1978). Probabilistische Testtheorie. In K. J. Klauer (Hrsg.), *Handbuch der Pädagogischen Diagnostik* (S. 59-97). Düsseldorf: Schwann.
- Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion*. Bern: Huber.
- Rost, J. Prenzel, M., Carstensen, C. H., Senkbeil, M. & Groß, K. (2004). *Naturwissenschaftliche Bildung in Deutschland. Methoden und Ergebnisse von PISA 2000*. Wiesbaden: VS Verlag für Sozialwissenschaften.
-

- Roth, H. (1963). *Pädagogische Psychologie des Lehrens und Lernens*. (7.Aufl.). Hannover: Schroedel.
- Roth, H. (1966). Lernpsychologie und Unterrichtserfolg. In M. Bönsch (Hrsg.), *Einprägen, Üben und Anwenden im Unterricht* (S. 24-38). München: Ehrenwirth.
- Rothkopf, E. Z. & Bisbicos, E. E. (1967). Selective facilitative effects of interspersed questions on learning from written materials. *Journal of Educational Psychology*, 58, 56-61.
- Rothkopf, E. Z. (1966). Learning from written instructive material: an exploration of the control of inspection behaviour by test-like events. *American Educational Research Journal*, 3, 241-249.
- Rothkopf, E. Z. (1972). Variable adjunct question schedules, interpersonal interaction and incidental learning from written material. *Journal of Educational Psychology*, 63, 87-92.
- Royer, J. M., Ciscero, C. A & Carlo, M. S. (1993). Techniques and procedures for assessing cognitive skills. *Review of Educational Research*, 63, 201-243.
- Ryle, G. (1969). *Der Begriff des Geistes*. Stuttgart: Reclam.
- Salomon, G. & Globerson, T. (1987). Skill may not be enough: the role of mindfulness in learning and transfer. *International Journal of Educational Research*, 6, 623-637.
- Salomon, G. & Perkins, D. N. (1987). Transfer of cognitive skills from programming: When and how? *Journal of Educational Computing Research*, 3, 149-169.
- Salomon, G. (1972). Heuristic models for the generation of aptitude-treatment interaction hypotheses. *Review of Educational Research*, 42, 237-343.
- Samson, G. E., Strykowski, B., Weinstein, T. & Walberg, H. J. (1987). The effects of teacher questioning level on student achievement: a quantitative synthesis. *Journal of Educational Research*, 80, 290-295.
- Sanders, N. (1966). *Classroom questions: what kinds?* New York: Harper and Row.
- Saxe, G.B. (1988). Candy selling and math learning. *Educational Researcher*, 17, 14-21.
-

- Schaich, E. & Hamerle, A. (1984). *Verteilungsfreie statistische Prüfverfahren*. Berlin: Springer.
- Schiefele, H. (1974). *Lernmotivation und Motivlernen*. München: Ehrenwirth.
- Schiever, S. (1991). *Comprehensive approach to teaching thinking*. Needham Heights; MA: Allyn and Bacon.
- Schnotz, W. (1971). Schätzung von Aufgabenschwierigkeiten durch Lehrer. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 3 (2), 106-120.
- Schnotz, W. (1994). *Aufbau von Wissensstrukturen - Untersuchungen zur Kohärenzbildung beim Wissenserwerb mit Texten*. Weinheim: Psychologie Verlags Union.
- Scholz, F. (1980). *Problemlösender Unterricht und Aufgabenstellungen mit einer Klassifikation von Problemlöseaufgaben(stellungen)*. Neue pädagogische Bemühungen, Bd. 82. Essen: Neue Deutsche Schule.
- Schrader, F.-W. & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1, 27-52.
- Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Frankfurt a.M.: Lang.
- Schrader, F.-W. (2001). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (2. Aufl., S. 68-71). Weinheim: Psychologie Verlags Union.
- Schwarzer, R. & Steinhagen, K. (1975). (Hrsg.). *Adaptiver Unterricht. Zur Wechselwirkung von Schülermerkmalen und Unterrichtsmethoden*. München: Kösel.
- Scriven, M. (1972). The concept of comprehension: from semantics to software. In R. O. Freedle & J. B. Carroll (Eds.), *Language comprehension and the acquisition of knowledge* (pp. 31-39). Washington, D.C.: Winston.
- Seel, N. M. (1981). *Lernaufgaben und Lernprozess*. Stuttgart: Kohlhammer.
- Seel, N. M. (2003). *Psychologie des Lernens* (2. Aufl.). München: Reinhardt.
-

- Seeling, D.V. (2005). *Problemlösefähigkeit in generellem und versicherungsspezifischem Kontext-Testentwicklung und Testvalidierung*. Aachen: Shaker.
- Seidel, T. (2003). *Lehr-Lernskripts im Unterricht*. Münster: Waxmann.
- Semb, G. B. & Ellis, J. A. (1994). Knowledge taught in school: what is remembered? *Review of Educational Research*, 64, 253-286.
- Shull, T. J. (1980). Learning theory, instructional theory, and adaptation. In R. E. Snow, P.-A. Federica & W. E. Montague (Eds.), *Aptitude, learning, and instruction. Volume 2: cognitive process analyses of learning and problem solving* (pp. 277-302). Hillsdale, NJ: Lawrence Erlbaum.
- Shulman L. S. (1982). Educational psychology returns to school. In A. G. Kraut (Ed.), *The G. Stanley Hall Lecture Series* (Band 2). Washington, DC: American Psychological Association.
- Shulman L. S. (1986). Paradigms and research programs in the study of teaching: a contemporary view. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.) (pp. 3-36). New York, NY: Macmillan.
- Siegler, R. S. (1989). How domain-general and domain-specific knowledge interact to produce strategy choices. *Merrill-Palme-Quarterly*, 35, 1-26.
- Skinner, B.F. (1958). Teaching machines. *Science*, 128, 969-977.
- Slavin, R. E. (1987). Mastery learning reconsidered. *Review of Educational Research*, 57, 175-213.
- Slavin, R. E. (1994). Quality, appropriateness, incentive, and time: A model of instructional effectiveness. *Journal of Educational Research*, 21, 141-157.
- Slavin, R. E. (1997). *Educational psychology* (5th ed.). Boston: Allyn & Bacon.
- Snow, R. E. & Salomon, G. (1968). Aptitudes and instructional media. *AV Communication Review*, 16 (4), 341-357.
- Snow, R. E. (1977). Research on aptitudes: A process report. In L.S. Shulman (Ed.), *Review on research in education* (pp. 50-105). Itasca, IL: Peacock.
-

- Snow, R. E. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences: advances in theory and research*. (pp. 13-59). New York: Freeman.
- Soar, R. S. & Soar, R. M. (1979). Emotional climate and management. In P. Peterson & H. Walberg (Eds.), *Research on teaching. concepts, findings, and implications* (pp. 97-119). Berkeley, CA: McCutchan.
- Stark, R. (1999). *Lernen mit Lösungsbeispielen. Einfluss unvollständiger Lösungsbeispiele auf Beispielelaboration, Lernerfolg und Motivation*. Göttingen: Hogrefe.
- Stark, R., Gruber, H., Renkl, A. & Mandl, H. (2000). Instruktionale Effekte einer kombinierten Lernmethode- Zählt sich die Kombination von Lösungsbeispielen und Problemlöseaufgaben aus? *Zeitschrift für Pädagogische Psychologie*, 14 (4), 206-218.
- Steiner, G. (1995). Lernverhalten, Lernleistung und Instruktionmethoden. In F.E. Weinert (Hrsg.), *Psychologie des Lernens und der Instruktion* (S. 279-318). Göttingen: Hogrefe.
- Stern, E. (1999). Development of mathematical competencies. In F. E. Weinert & W. Schneider (Eds.), *Individual development from 3 to 12: Findings from the Munich longitudinal study*. New York: Cambridge University Press.
- Sternberg, R. J. (1997). Intelligence and lifelong learning: What's new and how can we use it? *American Psychologist*, 52, 1134-1139.
- Sternberg, R. J. (1998). Principles of teaching for successful intelligence. *Educational Psychologist*, 33, 65-72.
- Stevens, J. (1992). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Erlbaum.
- Sweller, J. (1976). The effect of task complexity and sequence on rule learning and problem solving. *British Journal of Psychology*, 67 (4), 553-558.
- Sweller, J. (1988). Cognitive load during problem solving: effects of learning. *Cognitive Science*, 12, 257-285.
-

- Sweller, J., VanMerriënboer, J. J. G. & Paas, F. G. W. C (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251-296.
- Tiemann, R. (2003). *Aufgaben im naturwissenschaftlichen Unterricht*. Forschungsantrag an die DFG (Nr. TI 336/2-1). Duisburg-Essen: Universität.
- Tobias, S. (1975). Sequenzierung, Materialvertrautheit und Merkmal-Methoden-Interaktionen im programmierten Unterricht. In R. Schwarzer & K. Steinhagen (Hrsg.), *Adaptiver Unterricht* (S. 146-160). München: Tobin, K. & Capie, W. (1982). Relationships between classroom process variables and middle-school science achievement. *Journal of Educational Psychology*, 74, 441-454.
- Treiber, B. (1982). Lehr- und Lernzeiten im Unterricht. In B. Treiber & F. E. Weinert (Hrsg.), *Lehr-Lernforschung* (S.12-36). München: Urban & Schwarzenberg.
- Treiber, B., Petermann, F. (1976). Probleme der Unterrichtsdifferenzierung aus der Sicht des ATI-Forschungsprogramms. Zur Wechselwirkung von Schülermerkmalen und Unterrichtsmethoden. *Zeitschrift für Pädagogik*, 22 (4), 525-546.
- Tudge, J. & Rogoff, B. (1989). Peer influences on cognitive development: Piagetian and Vygotskian perspectives. In M. H. Bornstein & J. S. Bruner (Eds.) *Interaction in Human Development* (pp.17-40). Hillsdale, NJ: Erlbaum.
- Tyler, R. W. (1973). *Curriculum und Unterricht*. Düsseldorf: Schwann.
- Van Dijk, T. A. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Van Gog, T., Paas, F. & Van Merriënboer, J. J. G. (2004). Process-oriented worked examples: Improving transfer performance through enhanced understanding. *Instructional Science*, 32, 83-98.
- Von Glaserfeld, E. (1994). *Radical constructivism in mathematics education*. Dordrecht: Kluwer Academic Publishers.
- Voss, J. F. (1987). Learning and transfer in subject-matter learning: A problem-solving model. *International Journal of Educational Research*, 11, 607-622.
-

- Walter, O., Senkbeil, M., Rost, J., Carstensen, C. H. & Prenzel, M. (2006). Die Entwicklung der naturwissenschaftliche Kompetenz von der neunten zur zehnten Klassenstufe: Deskriptive Befunde. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahrs*. (S.87-118). Münster: Waxmann.
- Wang M. C. & Lindvall, C. M. (1984). Individual differences and school learning environments. *Review of Research in Education*, 11, 161-226.
- Ward, M. & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, 7, 1-39.
- Weinert, F. E. (1997). Notwendige Methodenvielfalt: Unterschiedliche Lernfähigkeit der Schüler erfordern variable Unterrichtsmethoden des Lehrers, *Friedrich Jahresheft (1997), Lernmethoden- Lehrmethoden- Wege zur Selbstständigkeit* (S. 50-52). Velber: Friedrich Verlag.
- Weinert, F. E. (1998). Neue Unterrichtskonzepte zwischen gesellschaftlichen Notwendigkeiten, pädagogischen Visionen und psychologischen Möglichkeiten. In Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und Kunst (Hrsg.), *Wissen und Werte von morgen. Dokumentation zum Bildungskongress* (S. 101-125). München: Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und Kunst.
- Weinert, F. E., Schrader, F.-W. & Helmke, A. (1989). Quality of instruction and achievement outcomes. *International Journal of Educational Research*, 13, 895-914.
- Weinert, F. E., Schrader, F.-W. & Helmke, A. (1990). Educational expertise: closing the gap between educational research and classroom practice. *School Psychology International*, 11, 163-180.
- Weiss, D. J. & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21 (4), 361-375.
- Wellman, H. & Gelman, S. (1998). Knowledge acquisition in foundational domains. In W. Damon (Series Ed.), D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol.2. Cognition, perception, and language*, 5th ed. (pp.523-573). New York: Wiley.
-

- West, R. & Pearson, J. C. (1994). Antecedent and consequent conditions of student questioning: an analysis of classroom discourse across the university. *Communication Education*, 43 (4), 299-311.
- Whitehead, A.N. (1929). *The aims of education*. New York, NY: Macmillan.
- Wilén, W. W. (1987). *Questions, questioning techniques, and effective teaching*. Washington, D.C.: National Education Association.
- Wilén, W. W. (1991). *Questioning skills for teachers* (3th Ed.). Washington, DC: National Education Association.
- Wilén, W., & Clegg, A. (1986). Effective questions and questioning: a research review. *Theory and Research in Social Education*, 14, 153-161.
- Williams, G. & Clarke, D. J. (1997). Mathematical task complexity and task selection, In D. J. Clarke (Ed.), *Mathematics: Imagine the Possibilities* (pp. 406-415). Brunswick, Victoria: Mathematics Association of Victoria.
- Williams, R.G. (1977). A behavioural typology of educational objectives for the cognitive domain. *Educational Technology*, 17 (6), 39-46.
- Willson, I. A. (1973). Changes in mean levels of thinking in grades 1-8 through use of an interaction analysis system based on Bloom's taxonomy. *Journal of Educational Research*, 1973, 66, 423-429.
- Wilson, J.H. (1969). The New Science Teachers are Asking More and Better Questions. *Journal of Research in Science Teaching*, 6, 303- 308.
- Winne, P. H. & Marx, R. W. (1977). Reconceptualising research on teaching. *Journal of Educational Psychology*, 69, 668-678.
- Winne, P. H. (1979). Experiments relating teacher's use of higher cognitive questions to student achievement. *Review of Educational Research*, 49, 13-50.
- Winne, P. H. (1987). Why process-product research cannot explain process-product findings and a proposed remedy: The cognitive mediational paradigm. *Teaching and Teacher Education*, 3, 333-356.
-

-
- Wirtz, M. & Casper, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.
- Wittrock, M. C. (1986). Students' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.) (pp. 297-314). New York: Macmillan.
- Wood, D., Bruner, J. S. & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17, 89-100.
- Wu, M., Adams, R. J. & Wilson, M. R. (1998). *ACER ConQuest. Generalised item response modelling software*. Melbourne: Acer Press.
- Wygotsky, L.S. (1978). *Mind in society: the development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wygotsky, L.S. (1993). *Denken und Sprechen*. Frankfurt a.M.: Fischer.
-

Tabellenverzeichnis

Tabelle 1: Lehrzieltaxonomie nach Anderson et al. (2001).....	51
Tabelle 2: Beispiele einer Lehrzielmatrix nach Klauer (2001).....	62
Tabelle 3: Beschreibung der Untersuchungsstichprobe	70
Tabelle 4: Aufgabenstellung und Aufgabenart am Beispiel einer Unterrichtsstunde zu "Bestandteile und Funktionsweise eines Elektromotors"	84
Tabelle 5: Interrater-Reliabilität der Aufgabenidentifikation	92
Tabelle 6: Verschiedene Interrater-Reliabilitätsmaße zur Aufgabenklassifikation (N = 174 Unterrichtsaufgaben)	93
Tabelle 7: Verteilung der Aufgabenarten der im Unterricht gestellten Aufgaben und der in den Leistungstests wiederholten Unterrichtsaufgaben in Bezug auf die kognitiven Prozesse (für die Gesamtstichprobe von 30 Unterrichtsstunden)	94
Tabelle 8: Einsatz der verschiedenen Aufgabenarten im Physikunterricht	97
Tabelle 9: Schaich-Hamerle-Einzelvergleiche zur Anzahl der Aufgabenarten	99
Tabelle 10: Ränge der verschiedenen Aufgabenarten an Hauptschulen und Gymnasien	99
Tabelle 11: Unterschiede in der Anzahl der verschiedenen Aufgabenarten an Hauptschulen und Gymnasien	100
Tabelle 12: Wilcoxon-Wilcox-Tests zur Anzahl der Aufgabenarten pro Schulform	100
Tabelle 13: Anteil der mündlichen und schriftlichen Aufgaben an Unterrichtsaufgaben	102
Tabelle 14: Häufigkeiten der Aufgabenarten nach Schulform und Aufgabenformat	104
Tabelle 15: Durchschnittliche Lösungshäufigkeit der verschiedenen Aufgabenarten in den Leistungstests (N = 340 Aufgaben).....	107
Tabelle 16: Relative Lösungshäufigkeiten der Testaufgaben in Abhängigkeit von Aufgabenart und Schulform	109

Tabelle 17: Effekte der Schulform und Aufgabenart auf die relative Lösungshäufigkeit....	109
Tabelle 18: Prozentuale Anteile der Aufgabenart und des Aufgabenformats im Unterricht bei der berücksichtigten und der nicht-berücksichtigten Stichprobe.....	112
Tabelle 19: Mittlere Lösungswahrscheinlichkeiten und Lösungshäufigkeiten der verschiedenen Aufgabenarten.....	119
Tabelle 20: Mittlere Schwierigkeitsparameter der Aufgaben in Abhängigkeit von Schulform und Aufgabenart.....	120
Tabelle 21: Effekte der Schulform und Aufgabenart auf den mittleren Schwierigkeitsparameter	121
Tabelle 22: Mittlere Schwierigkeitsparameter der Unterrichtsaufgaben an Gymnasien und Hauptschulen	123
Tabelle 23: Mittlere Fähigkeitsparameter an Gymnasien und Hauptschulen.....	123
Tabelle 24: Effekte der Schulform auf den mittleren Schwierigkeitsparameter.....	124
Tabelle 25: Facetten der Adaption nach Leutner (1992) in Anlehnung an die Modelle von Cronbach (1967) und Glaser (1977)	138
Tabelle 26: Aufgabenauswahl für die Übungsphasen für zwei Schüler der 50%-Gruppe ...	152
Tabelle 27: Personenparameter bei Pretest sowie 1. und 2. Übungsphase	154
Tabelle 28: Lösungshäufigkeiten in den Untersuchungsgruppen zu den Übungszeitpunkten	157
Tabelle 29: Effekte der Gruppenzugehörigkeit und des Übungszeitpunktes auf die Lösungshäufigkeit	157
Tabelle 30: Aggregierte Lösungshäufigkeiten in den Untersuchungsgruppen zu den Übungszeitpunkten.....	158
Tabelle 31: Effekte der Übungszeitpunkte und der Gruppenzugehörigkeit auf die mittlere Lösungshäufigkeit	158
Tabelle 32: Mittlerer IRT-Schwierigkeitsparameter in den drei Untersuchungsgruppen	163

Tabelle 33: Effekt der Untersuchungsgruppe auf den mittleren IRT- Schwierigkeitsparameter in den Übungsphasen	163
Tabelle 34: Korrelationen zwischen den mittleren IRT-Schwierigkeitsparametern und den klassisch berechneten relativen Lösungshäufigkeiten.....	164
Tabelle 35: Personenparameter in den Untersuchungsgruppen bei Pre- und Posttest-Messung	165
Tabelle 36: Effekt des Messzeitpunkts und der Gruppenzugehörigkeit auf den Personenparameter	165
Tabelle 37: Mittlere Posttest-Residuen der Untersuchungsgruppen	169
Tabelle 38: Effekt der Gruppenzugehörigkeit auf die Posttest-Residuen	169

Abbildungsverzeichnis

Abbildung 1: Modell des schulischen Lernens nach Bloom (1976)	42
Abbildung 2: Vom Lehrinhalt zum Lehrziel nach Klauer (1987).....	60
Abbildung 3: Anteil der verschiedenen Aufgabenarten im Unterricht (N = 872 Unterrichtsaufgaben).....	96
Abbildung 4: Mittlere Ränge des Einsatzes der Aufgabenarten an den Schulformen	101
Abbildung 5: Anteile der unterschiedlichen Aufgabenarten bei schriftlichen Unterrichtsaufgaben.....	103
Abbildung 6: Anteile der unterschiedlichen Aufgabenarten bei mündlichen Unterrichtsaufgaben.....	103
Abbildung 7: Mittelwerte der Aufgabenart bei mündlichen und schriftlichen Aufgaben an Gymnasien.....	105
Abbildung 8: Mittelwert der Aufgabenart bei mündlichen und schriftlichen Aufgaben an den Hauptschulen	105
Abbildung 9: Relative Lösungshäufigkeiten der verschiedenen Aufgabenarten für beide Schulformen	108
Abbildung 10: Relative Lösungshäufigkeit der verschiedenen Aufgabenarten für die verschiedenen Stichproben.....	113
Abbildung 11: Skalierung der Aufgaben- und Personenstichprobe	116
Abbildung 12: Mittlere Schwierigkeitsparameter der unterschiedlichen Aufgabenarten an den Schulformen.....	122
Abbildung 13: Zusammenhang zwischen mittlerer Aufgabenschwierigkeit der Unterrichtsaufgaben und mittlerem Personenparameter (N = 20 Klassen)	126
Abbildung 14: Verteilung der mittleren Schwierigkeitsparameter der Unterrichtsaufgaben und der Fähigkeitsparameter	127

Abbildung 15: Versuchsaufbau der Studie 2	149
Abbildung 16: Mittlere relative Lösungshäufigkeiten in den Untersuchungsgruppen zu den Übungszeitpunkten	156
Abbildung 17: Skalierung der Aufgaben- und Personenstichprobe	161
Abbildung 18: Veränderung des Personenparameters zwischen Pretest- und Posttest- Erhebung	166
Abbildung 19: Regressionsgerade der Pretest- auf die Posttestpersonenparameter.....	168
Abbildung 20: Darstellung der Personenparameter beider Messungen mit eingezeichneten Winkelhalbierenden	168

Anhang A: Videoauswertung

Im Folgenden wird lediglich auf die zwei in der vorliegenden Studie mit Hilfe der Videos ausgewerteten Unterrichtsaspekte eingegangen: (1) Aufgabenidentifikation, (2) Aufgabenklassifikation. Für eine genauere Beschreibung des Vorgehens bei der Videoauswertung sei auf Jatzwauk (in Druck) verwiesen.

Aufgabenidentifikation

Um Aufgaben objektiv und reliabel zu identifizieren wurden in der vorliegenden Untersuchung sog. Operatoren als identifizierende Merkmale verwendet. Operatoren sind Handlungsanweisungen bzw. allgemeine Handlungsprogramme (Seel, 2003), die in Form von Verben (in der Imperativform) z.B. „Nenne...“, „Finde heraus...“ oder äquivalenten Formulierungen in Form von Substantiven oder Fragewörtern wie z.B. „Beobachtung...“, „Erklärung...“, „Wie...?“ vorkommen.

Für die Identifikation der Aufgaben wurde das entsprechende Unterrichtsvideo im Programm Videograph (Rimmele, 2004) angeschaut, die eigentliche Kodierung fand in einer zeitgleich geöffneten SPSS-Maske statt.

Zur genauen Identifikation der Aufgabe wurden zunächst folgende Daten erfasst:

- Zeit des Einbringens in den Unterricht
- Schlagwort zur Aufgabe bzw. Aufgabenformulierung
- Aufgabe in Schülerarbeitsphase oder Aufgabe im Unterrichtsgespräch

Im Detail erfolgte die Aufgabenidentifikation wie folgt:

„Zeit des Erscheinens der Lernaufgabe im Video“

Eine Operatoraufgabe wird in der SPSS-Maske zeitlich mit der Anfangszeit des letzten Intervalls der Instruktion erfasst.

Für die Identifikation von Aufgaben im Unterrichtsgespräch wurde per Definition dasjenige 10s-Intervall als letztes Intervall der Instruktion kodiert, in welchem der Lehrer seine Aufgabenstellung beendet hat und für den Schüler die Zeit beginnt, über die Aufgabe nachzudenken.

Für Aufgaben in Schülerarbeitsphasen wurde eine leicht abgewandelte Regelung getroffen. Hier wird das letzte Intervall der Instruktion darüber definiert, dass es das letzte noch vollständig mit Instruktionen des Lehrers gefüllte Intervall ist. Es liegt also vor dem Intervall, in welchem der Lehrer seine Instruktionen endgültig abschließt und die Schüler zu arbeiten beginnen. Diese Regelung wurde aus Gründen der notwendigen Kompatibilität zu der Kodierung der Unterrichtsstruktur nötig, da dort stets diejenigen Aktivitäten kodiert werden, welche in einem Intervall beginnen. Alle Operatoraufgaben, welche innerhalb einer Instruktionsphase gegeben werden, wurden somit mit der gleichen Intervall-Anfangszeit kodiert. Die Anfangszeit dieses Intervalls ist als Zeitpunkt der Erfassung der Operatoraufgaben gewählt worden, da er gut zu erkennen ist und die Schüler sich erst ab diesem Intervall wirklich konzentriert mit den gestellten Aufgaben auseinandersetzen können.

Beispiel: Der Lehrer gibt im Intervall 23:10-23:20 ein Arbeitsblatt an die Schüler, auf dem insgesamt 7 Operatoraufgaben abgedruckt sind. Er erläutert diese Aufgaben bis zum Intervall 25:40-25:50. In diesem Intervall endet seine Instruktion, er sagt „Alles klar? Los geht's!“ und die Schüler beginnen zu arbeiten. Alle 7 Operatoraufgaben erhalten bei der Variable „Zeit“ in der SPSS-Maske den Wert 25:30, denn in diesem Intervall lag das letzte vollständige Intervall der Instruktion.

Die Eingabe erfolgt nach dem Muster h:mm:ss, analog zu der Zeitangabe, wie sie Videograph in SPSS exportiert.

„Schlagwort zur Lernaufgabe“

Bei diesen Variablen wurden der Wortlaut der Lernaufgabe oder mehrerer Schlagwörter zu der entsprechenden Aufgabe kodiert, so dass die Operatoraufgabe bei einem wiederholten Kodierdurchgang - z.B. zur Interrater-Reliabilitätsbestimmung (vgl. Kapitel. 10.1.1) - eindeutig identifiziert werden kann.

Beispiel: Operatoraufgabe: „Erkläre die Funktionsweise des Elektromotors.“
Kodiert bei Schlagwort: „Erkläre Funktionsweise“ oder „Erkläre Elektromotor“

„Art der zugehörigen Lernaufgabe“

Mit dieser Kennzeichnung werden die Lernaufgaben in „Aufgaben in Schülerarbeitsphasen“ und „Aufgaben im Unterrichtsgespräch“ differenziert:

1. Aufgabe in einem Unterrichtsgespräch
2. Aufgabe in einer Schülerarbeitsphase

ad1: Aufgaben im Unterrichtsgespräch

In diese Gruppe gehören sämtliche Lernaufgaben, die vom Lehrer in einem Unterrichtsgespräch gestellt werden. Sie zeichnen sich durch folgende Merkmale aus:

- Sie werden innerhalb eines Unterrichtsgesprächs gestellt und gelöst.
- Sie sind mündlich gestellt und nicht im Wortlaut schriftlich fixiert.
- Sie haben eine sehr kurze Bearbeitungszeit (weniger als eine Minute)
- Die Lösung dieser Aufgaben wird nur von einem Teil der Schüler im Unterrichtsgespräch gegeben.
- Die Schüler erhalten meist ein unmittelbares Feedback auf ihre Wortmeldungen.

ad 2: Aufgaben in Schülerarbeitsphasen

In diese Gruppen gehören die Lernaufgaben im herkömmlichen Sinne. Sie zeichnen sich durch folgende Merkmale aus:

- Sie werden in Schülerarbeitsphasen in Einzel-, Partner- oder Gruppenarbeit bearbeitet.
 - Sie erfahren in irgendeiner Form als Fragestellung oder Lösung eine schriftliche Fixierung. Dies kann als Fragestellung in einem Buch, diktiert in einem Schülerheft, als Arbeitsblatt oder auch nur in Form ihrer Lösung im Schülerheft erfolgen.
 - Die Schüler erhalten stets eine individuelle Bearbeitungszeit, welche per Definition mindestens 1 Minute beträgt.
 - Die Aufgaben sind von allen Schülern, an die sie gerichtet sind, selbstständig oder in einer Gruppe zu bearbeiten.
 - Die Schüler erhalten kein unmittelbares Feedback auf während der Bearbeitung gemachte Fehler oder eingeschlagene Lösungswege.
-

Aufgaben-Klassifikation nach den drei kognitiven Prozessen:

Die identifizierten Aufgaben wurden drei Hauptkategorien kognitiver Prozesse zugeordnet, die die Schwierigkeit einer Aufgabe charakterisieren: 1) Wissen, 2) Verstehen und 3) Anwenden. In dem folgenden Abschnitt werden diese kurz erläutert sowie einige Besonderheiten, die sich in dieser Studie aus der gewählten Kategorisierung ergaben, erwähnt. Für eine genauere Beschreibung der Prozesse sei auf die Kapitel 9.5 verwiesen.

Wissen

Es geht bei diesem Aufgabentyp vor allem um den Abruf von relevantem Wissen aus dem Langzeitgedächtnis, also die Fähigkeit, Fakten, Begriffe usw. zu erinnern, sie aus dem Gedächtnis zu reproduzieren, sie wieder zuerkennen.

Meist lassen sich Wissensfragen durch einen umgrenzten Inhaltsbereich charakterisieren, welche mit Hilfe einer kurzen Antwort des Schülers vollständig bzw. ausreichend zu beantworten ist.

Beobachtungs- und Beschreibungsaufgaben, beispielsweise bei einer Versuchsdurchführung, werden als Wissensaufgabe klassifiziert, z.B. „Was habt ihr beobachtet, als ihr die Spule an die Spannungsquelle angelegt habt?“ oder „Was passiert mit den Elektronen bzw. Löchern in der Sperrschicht?“

Weitere Beispiele für Wissensaufgaben:

- „Ein Newton entspricht welcher Gewichtskraft?“
- „Nenne die Stoffe der 1.Hauptgruppe!“
- „Welche Stoffgruppe riecht bei der Verbrennung so?“
- „Als Hausaufgabe wiederholt ihr das alles noch mal.“
- „Erzähl mal, was wir letzte Stunde zu den Metalloxiden gelernt haben.“

Verstehen

Der kognitive Prozess des Verstehens bzw. Verständnis ist gegeben, wenn Schüler die Bedeutung von Inhalten konstruieren bzw. die Bedeutung von Unterrichtsinhalten verstehen können müssen. Sie müssen das Gelernte bzw. den Unterrichtsstoff in eigenen Worten wiedergeben bzw. zusammenfassen können. Ebenfalls als Verstehensaufgaben werden solche klassifiziert, bei denen eine Erklärung, Begründung oder auch eine begründete Vermutung für ein Phänomen gegeben werden muss. Häufig beginnen solche Aufgaben mit „Warum...“, „Wieso...“ oder ähnliches, denen meist eine längere Antwort des Schülers

folgt. Schüler zeigen Verstehensleistungen, wenn sie Verbindungen zwischen „neu“ erworbenem Wissen und ihrem Vorwissen herstellen können, zudem wird neu hinzukommendes Wissen in existierende Schemata und kognitive Modelle integriert.

Es geht darum, vom Lehrer vermittelte Inhalte aufzunehmen und zu verarbeiten, ohne es notwendigerweise auf andere Bereiche zu übertragen und ohne seine Implikationen zu erkennen - dies zählt bereits zum Prozess der „Anwendung“.

Um sicher zu stellen, dass „Verständnis“ und nicht „Wissen“ überprüft wird, muss für die Beantwortung vom Schüler mehr geleistet werden, als lediglich Fakten und Begriffe zu erinnern. Wenn der Lehrer bei der Aufgabenstellung Sätze, wie „Wie wir bereits durchgenommen, angesprochen haben...“, „Macht es wie bei der Aufgabe...“ benutzt, ist davon auszugehen, dass die präsentierten Informationen nicht neu sind, sondern wiederholt werden und können - je nachdem, was mit den wiederholten Inhalten für die Aufgabenbeantwortung gemacht werden muss - als „Wissen“ klassifiziert werden.

Beispiele für Verstehensaufgaben:

- „Erkläre die Funktionsweise des Elektromotors.“
- „Fassen die wichtigsten Forschungsfragen und –ergebnisse zur Atomphysik zusammen.“
- „Überlege, unter welchen Bedingungen die Spule in einem solchen Aufbau eine volle Umdrehung macht.“
- „Ziehe aus dem Versuch eine Schlussfolgerung.“
- „Macht Vermutungen zum Gewicht

Anwendung

Der Prozess der Anwendung bezieht sich auf das Ausführen und Anwenden von Prozeduren bzw. Methoden in einer gegebenen Situation, um eine Aufgabe zu bearbeiten oder ein Problem zu lösen. Des Weiteren wird diese Kategorie gewählt, wenn der Schüler bereits gelernte Inhalte aus dem Kontext lösen muss, um sie auf eine neue Situation bzw. einen neuen Bereich zu übertragen. Es ist die Fähigkeit gemeint, in konkreten Situationen Abstraktionen, Regeln, Prinzipien, Ideen und Methoden anzuwenden. Soll beispielsweise ein bestimmtes Phänomen oder eine Fragestellung mit Hilfe eines vom Schüler zu planenden Versuchs erklärt werden, wird dies aufgrund der Anwendung bekannter Inhalte als Anwendungsaufgabe klassifiziert.

Eine Besonderheit stellt hierbei die Anwendung von mathematischen und physikalischen Formeln und Gesetzen dar. Müssen diese vom Schüler selbstständig hergeleitet wer-

den, wird dies als Anwendung klassifiziert. Ist jedoch die Formel bekannt und soll lediglich ihr Gebrauch eingeübt werden - z.B. indem die fehlenden Werte in einer Tabelle mit Hilfe der bereits bekannten Formel berechnet werden sollen - wird die Kategorie Wissen klassifiziert. Beispiel einer solchen Anwendungsaufgabe: Eine Formel oder Gesetzmäßigkeit soll aus den Ergebnissen einer Messreihe im Rahmen einer Versuchsdurchführung abgeleitet werden.

Beispiele für Anwendungsaufgaben:

- „Übertragt euer Wissen über Elektromotoren auf die Funktionsweise eines Fahrrad-dynamos.“
- „Was brauchen wir um zu überprüfen, ob die Spannung kleiner wird?“
- „Denkt euch selber einen Versuchsaufbau aus!“

Eine Übersicht über die berücksichtigten kognitiven Prozesse liefert Tabelle 39.

Tabelle 39: Zusammenfassende Darstellung der berücksichtigten kognitiven Prozesse

Niveau	Kognitive Aktivität	Schlüsselwort	Beispielaufgabe
1. Wissen	Erinnern, Abruf, Wiedererkennen von Fakten, Ideen oder Informationen, wie sie gelehrt wurden	Definiere, identifiziere, liste auf, erinnere, nenne, beschreibe, zusammenführen, auswählen,	1. Ein Newton entspricht welcher Gewichtskraft? 2. Nenne die Stoffe der 1.Hauptgruppe! 3. Welche Stoffgruppe riecht bei der Verbrennung so?
2. Verstehen	Verstehen, interpretieren, übertrage eine Idee	Beschreibe, erkläre, vergleiche, fasse zusammen, interpretiere, umformulieren, erweitern, schließen, umschreiben, generalisiere, gebe ein Beispiel, vorhersagen, klassifiziere	1. Ziehe aus dem Versuch eine Schlussfolgerung. 2. Erkläre die Funktionsweise des Elektromotors. 3. Fassen die wichtigsten Forschungsfragen und –ergebnisse zur Atomphysik zusammen.
3. Anwendung	Problemlösen, Lösung finden, Antwort herleiten durch die Anwendung einer Regel oder Gesetzes, Anwendung gelernter Informationen in einer neuen (Lern)situation	Anwenden, Beispiel nennen, nutzen, feststellen, berechne, bediene, entdecke, manipulierte, modifiziere, vorhersagen, zeige, löse, nutze, durchführen	1. Übertragt euer Wissen über Elektromotoren auf die Funktionsweise eines Fahrraddynamos. 2. Was brauchen wir um zu überprüfen, ob die Spannung kleiner wird? 3. Denkt euch selber einen Versuchsaufbau aus!

Besonderheiten

Die gewählte Kategorisierung zur Bestimmung der Aufgabenschwierigkeit ging in der praktischen Umsetzung mit einigen Schwierigkeiten einher:

Wie bereits bei der Beschreibung der drei kognitiven Prozesse verdeutlicht wurde, zeigt auch die Tabelle die Nähe zwischen kognitiven Aktivitäten und Schlüsselwörtern (in Frage- und Aufgabenstellungen). Zum Teil kommen die gleichen Schlüsselwörter in verschiedenen Kategorien vor, so dass eine eindeutige Abgrenzung zwischen den Stufen schwierig ist. Daher wird entsprechend der *code high*-Regel kategorisiert, d.h. es wird jeweils der kognitiv anspruchsvollste Prozess, der zur Lösung der Aufgabe notwendig ist, als determinierend für die Kategorisierung der Aufgabe angesehen.

Eine weitere Schwierigkeit bei der korrekten Einordnung der Lernziele ist die Berücksichtigung des Vorwissens der Schüler. Da jeweils nur eine einzelne Unterrichtsstunde auf Video aufgezeichnet und analysiert wurde, ist nicht bekannt, was in der vorangegangenen Stunde inhaltlich bearbeitet wurde. Daher wurde die Art der Formulierung seitens der Lehrer und auch der Schüler als Hinweis auf das kognitive Niveau betrachtet. Dies wird besonders deutlich bei Testaufgaben, die ein Schüler bereits zuvor im Unterricht bearbeitet hat. In diesem Fall kann eine zuvor kognitiv anspruchsvollere Aufgabe (z.B. Anwendungsaufgabe) zu einer Wissensaufgabe werden, da der Schüler die Aufgabenlösung lediglich aus dem Gedächtnis „erinnern“ muss. Auch kann die identische Aufgabe unterschiedliche kognitive Prozesse in unterschiedlichen Klassen- bzw. Altersstufen von den Schülern verlangen. Eine 100%-ig verlässliche Klassifikation als Wissens- oder Anwendungsaufgabe war daher nicht möglich.

Anhang B: Aufgabeneinsatz im Physikunterricht (Studie 1)

Aufgrund der z.T. beträchtlichen Unterschiede in der Häufigkeit des Aufgabeneinsatzes bzgl. der Aufgabenart (kognitive Prozesse: Wissen, Verstehen und Anwendung) und des Aufgabenformats (mündliche und schriftliche Aufgaben) sind die Ergebnisse der jeweiligen Anteile der verschiedenen Aufgabenarten im Unterricht in der folgenden Tabelle 40 pro Lehrer einzeln aufgeführt.

Tabelle 40: Aufgabeneinsatz im Physikunterricht

		Gesamtanzahl der Aufgaben	Wissensaufgaben	Verstehensaufgaben	Anwendungsaufgaben	Schriftliche Aufgaben	Mündliche Aufgaben
Gymnasium							
Lehrer 1	Anzahl der Aufgaben	16	12	3	1	2	14
	Anteil in %		75.0%	18.8%	6.3%	12.5%	87.5%
Lehrer 2	Anzahl der Aufgaben	21	16	5	0	5	16
	Anteil in %		76.2%	23.8%	0%	23.8%	76.2%
Lehrer 3	Anzahl der Aufgaben	17	4	9	4	3	14
	Anteil in %		23.5%	52.9%	23.5%	17.6%	82.4%
Lehrer 4	Anzahl der Aufgaben	23	13	8	2	9	14
	Anteil in %		56.5%	34.8%	8.7%	39.1%	60.9%

Anhang B: Aufgabeneinsatz im Physikunterricht (Studie 1)

		Gesamtanzahl der Aufgaben	Wissensaufgaben	Verstehensaufgaben	Anwendungsaufgaben	Schriftliche Aufgaben	Mündliche Aufgaben
Lehrer 5	Anzahl der Aufgaben	25	15	7	3	2	23
	Anteil in %		60.2%	29.5%	10.3%	8%	92%
Lehrer 6	Anzahl der Aufgaben	18	16	2	0	3	15
	Anteil in %		88.9%	11.1%	0%	16.7%	83.3%
Lehrer 7	Anzahl der Aufgaben	26	20	4	2	26	0
	Anteil in %		76.9%	15.4%	7.7%	100%	0
Lehrer 8	Anzahl der Aufgaben	27	20	5	2	3	24
	Anteil in %		74.1%	18.5%	7.4%	11.1%	88.9%
Lehrer 9	Anzahl der Aufgaben	34	20	10	4	4	30
	Anteil in %		58.8%	29.4%	11.8%	11.8%	88.2%
Lehrer 10	Anzahl der Aufgaben	20	5	10	5	9	11
	Anteil in %		25%	50%	25%	45.0%	55.0%
Lehrer 11	Anzahl der Aufgaben	28	13	11	4	1	27
	Anteil in %		46.4%	39.3%	14.3%	3.6%	96.4%

Anhang B: Aufgabeneinsatz im Physikunterricht (Studie 1)

		Gesamtanzahl der Aufgaben	Wissensaufgaben	Verstehensaufgaben	Anwendungsaufgaben	Schriftliche Aufgaben	Mündliche Aufgaben
Lehrer 12	Anzahl der Aufgaben	44	30	12	2	3	41
	Anteil in %		68.2%	27.3%	4.5%	6.8%	93.2%
Lehrer 13	Anzahl der Aufgaben	34	21	10	3	3	31
	Anteil in %		61.8%	29.4%	8.8%	8.8%	91.2%
Lehrer 14	Anzahl der Aufgaben	40	26	11	3	40	0
	Anteil in %		65.0%	27.5%	7.5%	100%	0%
Lehrer 15	Anzahl der Aufgaben	27	17	9	1	6	21
	Anteil in %		63%	33.3%	3.7%	22.2%	77.8%
Hauptschule							
Lehrer 16	Anzahl der Aufgaben	34	15	17	2	5	29
	Anteil in %		44.1%	50%	5.9%	14.7%	85.3%
Lehrer 17	Anzahl der Aufgaben	16	7	9	0	1	15
	Anteil in %		43.8%	56.3%	0%	6.2%	93.8%
Lehrer 18	Anzahl der Aufgaben	40	22	15	3	1	39
	Anteil in %						

Anhang B: Aufgabeneinsatz im Physikunterricht (Studie 1)

		Gesamtanzahl der Aufgaben	Wissensaufgaben	Verstehensaufgaben	Anwendungsaufgaben	Schriftliche Aufgaben	Mündliche Aufgaben
Anteil in %			55.0%	37.5%	7.5%	2.5%	97.5%
Lehrer 19	Anzahl der Aufga- ben	39	25	14	1	2	37
	Anteil in %		62.7%	34.5%	2.8%	5.2%	94.8%
Lehrer 20	Anzahl der Aufga- ben	33	20	11	2	5	28
	Anteil in %		60.6%	33.3%	6.1%	15.2%	84.8%
Lehrer 21	Anzahl der Aufga- ben	47	39	8	0	47	0
	Anteil in %		83.0%	17.0%	0%	100%	0%
Lehrer 22	Anzahl der Aufga- ben	18	13	5	0	4	14
	Anteil in %		72.2%	27.8%	0%	22.2%	77.8%
Lehrer 23	Anzahl der Aufga- ben	15	8	5	2	3	12
	Anteil in %		53.3%	33.3%	13.3%	20%	80%
Lehrer 24	Anzahl der Aufga- ben	43	19	22	2	43	0
	Anteil in %		44.2%	51.2%	4.7%	100%	0%
Lehrer 25	Anzahl der Aufga- ben	36	29	7	0	3	33

Anhang B: Aufgabeneinsatz im Physikunterricht (Studie 1)

		Gesamtanzahl der Aufgaben	Wissensaufgaben	Verstehensaufgaben	Anwendungsaufgaben	Schriftliche Aufgaben	Mündliche Aufgaben
Anteil in %			80.6%	19.4%	0%	8.3%	91.7%
Lehrer 26	Anzahl der Aufga- ben	17	13	4	0	1	16
Anteil in %			76.5%	23.5%	0%	5.9%	94.1%
Lehrer 27	Anzahl der Aufga- ben	67	40	26	1	67	0
Anteil in %			59.7%	38.8%	1.5%	100%	0%
Lehrer 28	Anzahl der Aufga- ben	64	50	14	0	3	61
Anteil in %			78.1%	21.9%	0%	4.7%	95.3%
Lehrer 29	Anzahl der Aufga- ben	32	17	15	0	6	26
Anteil in %			53.1%	46.9%	0%	18.8%	81.3%
Lehrer 30	Anzahl der Aufga- ben	35	19	15	1	1	34
Anteil in %			54.3%	42.9%	2.9%	2.9%	97.1%

Anhang C: Skalierung der Aufgaben in Studie 1

Die Beschreibung der Rasch-Skalierung der Aufgaben ist folgendermaßen aufgebaut: In der ersten Spalte ist die Bezeichnung des Items angegeben. In der zweiten Spalte ist die Aufgabenart dargestellt, wobei unterschieden wird zwischen Unterrichtsaufgaben (U) und den drei konstruierten Aufgaben: Wissensaufgaben (W), Verstehensaufgaben (V) und Anwendungsaufgaben (A). Der Itemparameter der Aufgabenschwierigkeit ist in Spalte 3 wiedergegeben und in der Spalte 4 der dazugehörige Standardfehler. Der fünften Spalte ist der Infit-Wert (MNSQ) und der sechsten der T-Wert zu entnehmen. In der siebten Spalte ist die auf Grundlage klassischer Testverfahren berechnete Lösungshäufigkeit (LH) angegeben.

Tabelle 41: Rasch-Skalierung der Aufgaben aus Studie 1

<i>Item</i>	<i>Aufgabenart</i>	<i>Itemparameter</i>	<i>Standardfehler</i>	<i>MNSQ⁸</i>	<i>T-Wert¹</i>	<i>LH</i>
Bo1	U	1.094	0.225	0.88	-0.90	.23
Bo3a	V	0.000	0.218	1.06	0.40	.45
Bo3b	V	-0.577	0.260	1.04	0.30	.59
Bo4b	V	.0531	0.205	1.38	2.50	.33
Bo5a	A	-0.081	0.195	1.18	1.30	.46
Bo7	A	1.412	0.266	0.94	-0.40	.19
Bo10a	V	1.615	0.253	0.76	-1.90	.16
Bo10b	V	1.615	0.253	0.76	-1.90	.16
Bo11	U	-2.357	0.427	0.94	-0.10	.86
Ma1	W	-0.831	0.290	1.07	0.40	.51
Ma2	U	0.241	0.405	0.80	-0.60	.29
Ma10	U	-0.253	0.389	0.72	-0.90	.38
Ma11	U	2.483	0.555	0.62	-2.90	.05
Do5113	A	2.825	0.470	0.92	-0.50	.05
We1	V	-0.816	0.268	1.20	1.00	.59
Moe4	A	-1.674	0.254	0.96	-0.20	.75

⁸ (ungewichteter Wert)

Anhang C: Skalierung der Aufgaben in Studie 1

<i>Item</i>	<i>Aufgabenart</i>	<i>Itemparameter</i>	<i>Standardfehler</i>	<i>MNSQ</i> ⁹	<i>T-Wert</i> ¹	<i>LH</i>
AI4	V	-0.104	0.329	1.11	0.50	.44
AI5b	W	0.851	0.360	0.61	-1.70	.25
AI10b	A	-0.104	0.329	1.20	0.80	.44
Dor1	W	1.321	0.450	1.22	0.90	.20
TIMSS1 (B2) ⁹		-1.448	0.103	0.96	-0.50	.68(.58) ¹⁰
TIMSS2 (K13)		-2.764	0.138	0.76	-4.00	.87(.64)
TIMSS3 (X01)		0.645	0.130	1.00	0.00	.27(.08)
TIMSS4 (B3)		-0.813	0.122	0.97	-0.30	.56(.28)
TIMSS5 (D1)		-0.533	0.097	0.99	-0.10	.49(.42)
TIMSS6 (R1)		-1.393	0.129	1.01	0.20	.68(.21)
HSTIMSS1 (N9)		-2.209	0.187	1.32	2.80	.78(.62)
HSTIMSS2 (B6)		-3.331	0.252	0.82	-1.70	.91(.82)
HSTIMSS3 (K14)		-2.585	0.204	0.90	-0.90	.83(.84)
HSTIMSS5 (D2)		-2.209	0.187	0.95	-0.50	.78(.72)
Bo6	U	0.701	0.368	0.94	-0.10	.29
Vi1	U	1.648	0.431	1.19	0.60	.16
Vi2	U	2.494	0.497	1.09	0.40	.08
Vi3	U	1.347	0.411	0.79	-0.70	.20
Do251a	U	0.017	0.357	0.99	0.10	.38
Do252	U	0.017	0.357	1.03	0.20	.38
Do253	U	1.073	0.406	1.24	0.90	.19
Do254	U	1.713	0.484	0.65	-1.30	.12
Do257b	V	1.713	0.454	0.84	-0.50	.12
Wi6	W	-2.415	0.435	0.90	-0.20	.82
Wi9	A	1.868	0.498	0.99	0.10	.09
Wi5b	U	-2.416	0.435	1.15	0.60	.82
Wi12	U	0.170	0.392	0.91	-0.20	.32

⁹ In Klammern sind die Itembezeichnungen der TIMSS 1995 Erhebung angegeben (vgl. Martin & Kelly, 1998)

¹⁰ (in Klammern LH der TIMSS 1995-Erhebung)

Anhang C: Skalierung der Aufgaben in Studie 1

<i>Item</i>	<i>Aufgabenart</i>	<i>Itemparameter</i>	<i>Standardfehler</i>	<i>MNSQ^β</i>	<i>T-Wert¹</i>	<i>LH</i>
Wi11	V	0.418	0.402	0.80	-0.60	.27
HW7	V	-1.260	0.421	1.05	0.30	.63
HW1	U	-1.260	0.421	1.19	0.60	.63
HW3	U	-0.967	0.415	1.00	0.10	.56
HW6a	V	1.017	0.501	0.93	-0.10	.19
HW8	U	0.202	0.430	1.02	0.20	.31
Dor6	W	-1.629	0.407	1.00	0.10	.70
Dor4a	W	-1.629	0.407	1.25	0.80	.70
Dor5	A	2.594	0.556	1.12	1.80	.05
Dor8	A	0.924	0.439	1.14	0.50	.20
Dor9	U	-0.190	0.393	0.96	0.00	.40
Dor10	V	0.317	0.408	0.98	0.00	.30
Dor11	A	1.309	0.464	1.39	1.20	.15
We2	W	0.480	0.345	1.05	0.30	.37
We3	V	0.004	0.338	0.93	-0.20	.47
We4	V	0.994	0.300	0.88	-0.60	.23
We5	V	-0.151	0.337	0.81	-0.70	.50
We6	V	-0.620	0.342	1.12	0.60	.60
We9	W	1.505	0.331	0.90	-0.50	.16
Wu1	U	-1.126	0.359	1.18	0.70	.67
Wu2	V	0.054	0.349	0.90	-0.30	.41
Wu3	W	-0.443	0.345	1.29	1.10	.52
Wu4	U	2.357	0.493	0.88	-1.30	.07
Wu5	U	1.545	0.425	0.70	-1.20	.15
Wu6	V	2.357	0.493	0.84	-0.55	.07
Wu7a	V	0.405	0.358	0.94	-0.10	.33
Wu10	A	1.545	0.425	1.16	0.60	.15
Ra1	V	-0.615	0.366	0.97	0.00	.54

Anhang C: Skalierung der Aufgaben in Studie 1

<i>Item</i>	<i>Aufgabenart</i>	<i>Itemparameter</i>	<i>Standardfehler</i>	<i>MNSQ^β</i>	<i>T-Wert¹</i>	<i>LH</i>
Ra2	U	0.169	0.373	1.25	0.90	.38
Ra3a	V	-0.228	0.378	0.76	-0.80	.46
Ra6a	V	-0.228	0.378	1.06	0.30	.46
Ra6b	A	1.784	0.460	0.91	-0.20	.13
Ra8a	U	0.379	0.379	0.95	-0.10	.33
Ra8b	U	-0.228	0.366	0.96	-0.10	.46
Ra10	U	-0.614	0.366	0.89	-0.30	.54
Ra11	V	0.842	0.399	0.93	-0.20	.25
Lü1	U	-1.164	0.361	1.10	0.40	.48
Lü2	U	-0.381	0.377	1.18	0.70	.32
Lü3	U	-0.587	0.371	1.32	1.10	.36
Lü1	A	1.494	0.497	0.60	-1.60	.08
Lü6	A	1.010	0.459	0.54	-1.90	.12
Lü9	V	0.643	0.456	0.61	-1.50	.16
Lü10	A	0.340	0.412	0.95	-0.10	.20
Moe1	W	-0.851	0.356	1.05	0.30	.54
Moe3a	W	-2.313	0.410	0.85	-0.50	.81
Moe3b	V	-2.313	0.410	0.85	-0.50	.81
Moe3c	W	0.267	0.374	0.75	-0.90	.31
Moe3d	W	-0.858	0.282	1.06	0.30	.54
Moe3f	V	1.535	0.374	1.10	0.60	.13
Moe6a	W	-0.858	0.282	0.76	-1.20	.54
Do546	U	1.504	0.461	0.84	-0.80	.13
Do543	U	-2.985	0.461	0.97	0.00	.88
Do545	U	1.131	0.435	0.73	-0.90	.17
Do548	V	1.505	0.461	0.79	-0.70	.13
Do549	V	2.782	0.557	0.82	-0.60	.04
Do5410	V	0.306	0.390	0.79	-0.70	.29

Anhang C: Skalierung der Aufgaben in Studie 1

<i>Item</i>	<i>Aufgabenart</i>	<i>Itemparameter</i>	<i>Standardfehler</i>	<i>MNSQ^β</i>	<i>T-Wert¹</i>	<i>LH</i>
Se1	U	-0.616	0.365	0.92	-0.20	.58
Se2	V	0.139	0.364	0.87	-0.40	.42
Se4	W	-0.051	0.362	1.20	0.70	.46
Se7	A	0.536	0.375	0.86	-0.40	.33
Se13	A	0.408	0.307	0.82	-0.60	.25
Se5	W	0.751	0.383	1.31	1.10	.29
Dü2	W	-0.001	0.399	1.02	0.20	.33
Dü3	U	0.869	0.439	0.84	-0.40	.19
Dü4b	V	2.554	0.557	1.15	-0.80	.05
Dü5	A	0.869	0.439	0.84	-0.40	.19
Dü6	V	0.260	0.408	0.99	0.10	.29
Dü9	A	0.546	0.421	0.77	-0.70	.24
Dü10	V	1.255	0.465	1.16	0.60	.14
Dü11	U	1.760	0.501	1.02	0.20	.10
Dü2	U	-0.907	0.400	0.83	-0.50	.52
Dü4a	A	0.862	0.462	1.21	0.70	.19
We4	W	-1.536	0.428	1.55	1.40	.60
We8	W	0.490	0.471	1.87	2.00	.20
We9a	W	-0.918	0.423	1.54	1.40	.47
We9b	A	-1.409	0.292	0.82	-0.80	.50
Dd3	W	0.474	0.402	0.88	-0.30	.27
Dd4	A	1.434	0.461	0.79	-0.70	.14
Dd5a	V	0.225	0.392	0.94	-0.10	.32
Dd5b	A	2.711	0.557	1.41	1.30	.05
Ddo6a	V	-0.007	0.385	0.84	-0.50	.36
Dd6b	A	2.710	0.557	0.86	-0.80	.05
Dd7	W	-4.006	0.555	1.73	2.10	.95
Dd8	A	-1.532	0.392	0.72	-0.90	.68

Anhang C: Skalierung der Aufgaben in Studie 1

<i>Item</i>	<i>Aufgabenart</i>	<i>Itemparameter</i>	<i>Standardfehler</i>	<i>MNSQ³</i>	<i>T-Wert¹</i>	<i>LH</i>
Bo2	W	-2.608	0.278	0.54	-3.40	.81
Bo4a	W	-0.505	0.265	1.22	1.10	.52
Bo8	W	-0.839	0.350	0.91	-0.30	.61
Ma3	V	-2.482	0.438	1.04	0.20	.81
Ma4	V	0.526	0.338	0.87	-1.20	.18
Ma8a	V	-0.008	0.396	0.74	-0.80	.33
Ma8b	A	1.719	0.499	0.78	-0.70	.10
Ma9a	V	2.507	0.557	0.69	-1.20	.05
Do517	A	1.719	0.499	0.87	-0.30	.10
Ma9b	A	2.505	0.557	1.20	0.70	.05
Al1	U	-0.373	0.439	1.20	0.60	.54
Al11	V	1.140	0.475	0.67	-0.80	.23
Al3	U	-2.931	0.559	0.64	-1.10	.92
Al2	W	-1.124	0.339	0.92	-0.20	.66
Al5a	A	0.044	0.331	0.97	-0.10	.41
Al6	W	-0.821	0.331	1.01	0.10	.59
Al7	W	0.675	0.352	0.84	-0.60	.28
Al8	W	-0.246	0.328	0.80	-0.90	.47
Al10a	A	2.586	0.493	0.68	-1.30	.06
HW4	A	0.952	0.468	0.61	-1.20	.19
HW5	V	0.207	0.430	1.07	0.60	.31
HW6b	A	1.467	0.502	1.15	1.00	.13
HW11	W	-0.104	0.420	1.07	0.30	.38
Lü4	W	-1.730	0.366	0.98	0.00	.60
Dd10	W	-3.240	0.499	1.42	1.30	.91
Dd1	W	-2.372	0.436	0.86	-0.80	.82
Dd9	U	1.917	0.498	1.12	0.90	.09
Dd1	U	-0.656	0.376	0.83	-0.50	.50

Anhang C: Skalierung der Aufgaben in Studie 1

<i>Item</i>	<i>Aufgabenart</i>	<i>Itemparameter</i>	<i>Standardfehler</i>	<i>MNSQ^β</i>	<i>T-Wert¹</i>	<i>LH</i>
We2	U	-2.238	0.453	0.89	-0.60	.73
We6	U	-0.613	0.427	0.93	-0.10	.40
We11	U	0.481	0.471	0.82	-0.40	.20
We3	A	0.068	0.449	1.01	1.00	.27
He2	V	0.361	0.410	1.07	0.30	.32
He9	U	-0.639	0.393	1.17	0.60	.53
Dor5110	V	2.152	0.556	0.70	-2.10	.03
Do5118a	V	0.226	0.409	0.95	-0.10	.17
Dor5118b	V	2.150	0.556	0.72	-2.00	.03
Rat4	W	-1.684	0.398	0.89	-0.30	.75
Ra9	W	0.600	0.387	0.96	0.00	.29
Do255	W	-0.678	0.351	1.26	1.00	.54
Do256a	W	-1.837	0.390	1.16	0.70	.77
Moe5	U	-0.238	0.369	1.04	1.20	.42
Ra3b	V	1.106	0.414	0.91	-0.20	.21
Bo9	W	-1.573	0.222	1.02	0.20	.76
Le1	U	0.510	0.271	0.74	-1.40	.36
Do514	U	-1.825	0.342	1.19	1.30	.55
Do515a	U	-4.454	0.494	1.26	1.90	.93
Do5115	U	-2.686	0.365	0.80	-0.70	.72
Dü8	W	-2.899	0.277	0.95	-0.10	.86
Bo3c	V	0.171	0.310	0.96	-0.20	.28
Bo5b	U	0.335	0.341	1.17	1.30	.39
Se3	W	-0.887	0.349	0.84	-0.50	.63
Se9	V	1.413	0.391	0.96	0.00	.17
Se11a	W	-1.092	0.355	1.04	0.20	.67
Se11b	V	2.829	0.476	1.52	1.90	.04
Dd2	V	-1.606	0.369	1.05	0.30	.68

Anhang C: Skalierung der Aufgaben in Studie 1

<i>Item</i>	<i>Aufgaben art</i>	<i>Itempara meter</i>	<i>Standard fehler</i>	<i>MNSQ^β</i>	<i>T-Wert¹</i>	<i>LH</i>
Do251b	V	1.343	0.390	0.83	-0.50	.15
Do512	U	0.775	0.413	1.06	0.30	.10
Do513	U	1.882	0.478	0.80	-1.20	.03
Do519	A	0.776	0.413	0.84	-0.50	.10
Do542	W	-1.203	0.350	1.04	0.20	.58
We8	A	1.156	0.351	1.14	1.00	.23
We10	U	-0.839	0.330	0.80	-0.90	.63
Moe6b	V	-0.725	0.338	0.81	-0.70	.50
Moe6c	W	-0.726	0.338	0.93	-0.20	.50
Wu9b	V	0.540	0.344	0.86	-1.00	.30
Wi3	A	0.593	0.382	1.40	1.30	.23
Wi7	V	0.592	0.382	1.15	0.60	.23
Wi10	V	-2.198	0.390	0.95	-0.10	.77
Ra12	U	-0.872	0.348	0.88	-0.30	.58
HW9	A	-0.198	0.388	1.17	0.60	.38
Lü5	U	0.227	0.378	1.07	1.10	.20
Lü8	W	-2.650	0.373	0.91	-0.20	.76
Lü12	U	0.512	0.485	1.19	1.00	.16

Anhang D: Adaptive Zuordnung der Itemparameter zu den individuellen Fähigkeitsparametern

Da die Adaptation der Aufgabenschwierigkeit in den beiden Übungsphasen für jeden Schüler individuell erfolgte und das zentrale Merkmal der Studie ist, wird dies im Folgenden pro Schüler angegeben. Die zu den einzelnen Items zugehörigen Aufgabenstellungen können auf der beiliegenden CD angeschaut werden. Der Übersichtlichkeit halber sind in der Tabelle lediglich die jeweiligen Itembezeichnungen und die dazugehörigen Itemparameter (IP) sowie die aufgrund der Item- und Personenparameter erwartbaren Lösungshäufigkeiten für beide Übungszeitpunkte angegeben (vgl. Kapitel 14.3).

Tabelle 42: Adaptive Zuordnung der Aufgabenschwierigkeit zu individuellen Fähigkeitsparametern

PP	Gruppe	1. Übung		2. Übung	
		Item	IP	Item	IP
1.334	5	L32	3.856	L323	0.848
		L33	0.562	L324	2.926
		L312	0.796	L326b	2.358
		L52	0.354	L327	0.985
		L56a	0.669	L328	2.064
		L57	1.302	L329	2.402
		L58	2.047	L3212	2.19
		L59	2.046	L522	0.816
		L511	3.239	L524b	1.127
		L82	2.654	L5212	0.776
		L83	0.604	L822	-1.012
		L85	1.041	L824	0.252
		L86c	0.784	L829	-0.345
		L810	0.603	L8211	-1.019
Mittlerer IP			1.468		1.26
LW			46.7%		51.8%
1.334	8	L81	-1.91	L322	-0.385

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		L82	2.654	L3212	2.19
		L83	0.604	L527	-0.507
		L84	-0.99	L821	-1.856
		L85	1.041	L823	-1.857
		L86a	-0.991	L824	0.252
		L86b	-0.788	L825	0.251
		L86c	0.784	L826	0.835
		L87	-0.992	L827	-1.861
		L88	-1.915	L828	-1.862
		L89	-1.915	L829	-0.345
		L810	0.603	L8210	-0.345
		L811	-2.552	L8211	-1.019
				L8212	2.318
		Mittlerer IP	-0.354		-0.299
		LW	84.4%		83.7%
1.700	3	L32	3.856	L323	0.848
		L33	0.562	L324	2.926
		L312	0.796	L326b	2.358
		L52	0.354	L327	0.985
		L56a	0.669	L328	2.064
		L57	1.302	L329	2.402
		L58	2.047	L3212	2.19
		L59	2.046	L522	0.816
		L511	3.239	L523	1.164
		L82	2.654	L524b	1.127
		L83	0.604	L528	2.247
		L85	1.041	L5211	2.247
		L86c	0.784	L8212	2.318

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		L810	0.603		
Mittlerer IP			1.468		1.748
LW			44.2%		48.8%
1.334	5	L32	3.856	L323	0.848
		L33	0.562	L324	2.926
		L312	0.796	L326b	2.358
		L52	0.354	L327	0.985
		L56a	0.669	L328	2.064
		L57	1.302	L329	2.402
		L58	2.047	L3212	2.19
		L59	2.046	L522	0.816
		L511	3.239	L524b	1.127
		L82	2.654	L5212	0.776
		L83	0.604	L822	-1.012
		L85	1.041	L824	0.252
		L86c	0.784	L829	-0.345
		L810	0.603	L8211	-1.019
Mittlerer IP			1.468		1.26
LW			46.7%		51.8%
0.685	8	L33	0.562	L51a	-1.055
		L34	-0.406	L524a	-0.132
		L35	-0.405	L527	-0.507
		L36	-0.803	L821	-1.856
		L37	-0.404	L822	-1.012
		L38	-2.02	L823	-1.857
		L39	0.329	L824	0.252
		L310	-0.39	L825	0.251
		L311	-1.576	L826	0.835

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		L312	0.796	L827	-1.861
		L313	-2.014	L828	-1.862
		L53	-1.537	L829	-0.345
		L54	-0.445	L8210	-0.345
		L56b	-0.694	L8211	-1.019
		Mittlerer IP	-0.643		-0.585
		LW	79.1%		78.1%
0.080	8	L31	-2.028	L51a	-1.055
		L33	0.562	L521b	-1.419
		L38	-2.02	L524a	-0.132
		L39	0.329	L525a	-2.075
		L310	-0.39	L526	-1.417
		L311	-1.576	L527	-0.507
		L313	-2.014	L821	-1.856
		L51	0.354	L822	-1.012
		L53	-1.537	L823	-1.857
		L55	-1.537	L827	-1.861
		L59	2.046	L828	-1.862
		L512	-1.717	L829	-0.345
		L513	-1.719	L8210	-0.345
		L81	-1.91	L8211	-1.019
		Mittlerer IP	-1.087		-1.197
		LW	76.3%		78.2%
1.700	3	L32	3.856	L323	0.848
		L33	0.562	L324	2.926
		L312	0.796	L326b	2.358
		L52	0.354	L327	0.985
		L56a	0.669	L328	2.064

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		L57	1.302	L329	2.402
		L58	2.047	L3212	2.19
		L59	2.046	L522	0.816
		L511	3.239	L523	1.164
		L82	2.654	L524b	1.127
		L83	0.604	L528	2.247
		L85	1.041	L5211	2.247
		L86c	0.784	L8212	2.318
		L810	0.603		
		Mittlerer IP	1.468		1.748
		LW	44.2%		48.8%
2.122	8	L32	3.856	L85	1.041
		L51	0.354	L321	-1.939
		L52	0.354	L322	-0.385
		L53	-1.537	L323	0.848
		L54	-0.445	L324	2.926
		L55	-1.537	L325	-0.968
		L56a	0.669	L326a	-0.968
		L56b	-0.694	L326b	2.358
		L57	1.302	L327	0.985
		L58	2.047	L328	2.064
		L59	2.046	L329	2.402
		L510	0.034	L3210	-0.537
		L511	3.239	L311a	0.741
				L3212	2.19
		Mittlerer IP	0.766		0.657
		LW	79.5%		81.2%
-0.223	3	L37	-0.404	L321	-1.939

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		L39	0.329	L323	0.848
		L51	0.354	L325	-0.968
		L52	0.354	L326a	-0.968
		L53	-1.537	L326b	2.358
		L54	-0.445	L327	0.985
		L55	-1.537	L328	2.064
		L56a	0.669	L329	2.402
		L56b	-0.694	L3210	-0.537
		L57	1.302	L311a	0.741
		L58	2.047	L311b	-0.513
		L59	2.046	L3212	2.19
		L510	0.034	L51a	-1.055
		L511	3.239	L8210	-0.345
		Mittlerer IP	0.386		0.398
		LW	35.2%		24.9%
-0.856	3	L31	-2.028	L322	-0.385
		L32	3.856	L3212	2.19
		L33	0.562	L527	-0.507
		L34	-0.406	L821	-1.856
		L35	-0.405	L823	-1.857
		L36	-0.803	L824	0.252
		L37	-0.404	L825	0.251
		L38	-2.02	L826	0.835
		L39	0.329	L827	-1.861
		L310	-0.39	L828	-1.862
		L311	-1.576	L829	-0.345
		L312	0.796	L8210	-0.345
		L313	-2.014	L8211	-1.019

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		L510	0.034	L8212	2.318
		Mittlerer IP	-0.322		-0.299
		LW	36.1%		36.4%
-0.533	8	L31	-2.028	L321	-1.939
		L311	-1.576	L325	-0.968
		L313	-2.014	L326a	-0.968
		L53	-1.537	L51a	-1.055
		L55	-1.537	L521b	-1.419
		L512	-1.717	L525a	-2.075
		L513	-1.719	L525b	-1.45
		L81	-1.91	L526	-1.417
		L84	-0.99	L529	-1.053
		L86a	-0.991	L5210	-1.417
		L87	-0.992	L821	-1.856
		L88	-1.915	L822	-1.012
		L89	-1.915	L823	-1.857
		L811	-2.552	L827	-1.861
		Mittlerer IP	-1.671		-1.453
		LW	75.7%		71.5%
-0.856	5	L313	-2.014	L51a	-1.055
		L513	-1.719	L521b	-1.419
		L81	-1.91	L524a	-0.132
		L83	0.604	L525a	-2.075
		L84	-0.99	L526	-1.417
		L85	1.041	L527	-0.507
		L86a	-0.991	L821	-1.856
		L86b	-0.788	L822	-1.012
		L86c	0.784	L827	-1.861

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		L87	-0.992	L828	-1.862
		L88	-1.915	L829	-0.345
		L89	-1.915	L8210	-0.345
		L810	0.603	L8211	-1.019
		L811	-2.552		
		Mittlerer IP	-0.911		-1.197
		LW	51.4%		58.4%
-0.856	3	L31	-2.028	L322	-0.385
		L32	3.856	L3212	2.19
		L33	0.562	L527	-0.507
		L34	-0.406	L821	-1.856
		L35	-0.405	L823	-1.857
		L36	-0.803	L824	0.252
		L37	-0.404	L825	0.251
		L38	-2.02	L826	0.835
		L39	0.329	L827	-1.861
		L310	-0.39	L828	-1.862
		L311	-1.576	L829	-0.345
		L312	0.796	L8210	-0.345
		L313	-2.014	L8211	-1.019
		L510	0.034	L8212	2.318
		Mittlerer IP	-0.322		-0.299
		LW	36.1%		36.4%
-0.533	3	L51	0.354	L51a	-1.055
		L52	0.354	L521b	-1.419
		L53	-1.537	L522	0.816
		L54	-0.445	L523	1.164
		L55	-1.537	L524a	-0.132

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		L56a	0.669	L524b	1.127
		L56b	-0.694	L525a	-2.075
		L57	1.302	L525b	-1.45
		L58	2.047	L526	-1.417
		L59	2.046	L527	-0.507
		L510	0.034	L528	2.247
		L511	3.239	L529	-1.053
		L512	-1.717	L5210	-1.417
		L513	-1.719	L5211	2.247
				L5212	0.776
		Mittlerer IP	0.171		-0.141
		LW	33.0%		40.4%
-0.223	3	L37	-0.404	L321	-1.939
		L39	0.329	L323	0.848
		L51	0.354	L325	-0.968
		L52	0.354	L326a	-0.968
		L53	-1.537	L326b	2.358
		L54	-0.445	L327	0.985
		L55	-1.537	L328	2.064
		L56a	0.669	L329	2.402
		L56b	-0.694	L3210	-0.537
		L57	1.302	L311a	0.741
		L58	2.047	L311b	-0.513
		L59	2.046	L3212	2.19
		L510	0.034	L51a	-1.055
		L511	3.239	L8210	-0.345
		Mittlerer IP	.386		.398
		LW	35.2%		34.9%

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
0.080	3	L32	3.856	L321	-1.939
		L35	-0.405	L322	-0.385
		L51	0.354	L323	0.848
		L52	0.354	L324	2.926
		L53	-1.537	L325	-0.968
		L54	-0.445	L326a	-0.968
		L55	-1.537	L326b	2.358
		L56a	0.669	L327	0.985
		L56b	-0.694	L328	2.064
		L57	1.302	L329	2.402
		L58	2.047	L3210	-0.537
		L59	2.046	L311a	0.741
		L510	0.034	L311b	-0.513
		L511	3.239	L3212	2.19
Mittlerer IP			0.663		0.657
LW			35.8%		35.9%
-0.223	5	L81	-1.91	L322	-0.385
		L82	2.654	L3212	2.19
		L83	0.604	L527	-0.507
		L84	-0.99	L821	-1.856
		L85	1.041	L823	-1.857
		L86a	-0.991	L824	0.252
		L86b	-0.788	L825	0.251
		L86c	0.784	L826	0.835
		L87	-0.992	L827	-1.861
		L88	-1.915	L828	-1.862
		L89	-1.915	L829	-0.345
		L810	0.603	L8210	-0.345

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		L811	-2.552	L8211	-1.019
				L8212	2.318
		Mittlerer IP	-0.293		-0.299
		LW	51.7%		51.9%
0.080	5	L81	-1.91	L322	-0.385
		L82	2.654	L3212	2.19
		L83	0.604	L527	-0.507
		L84	-0.99	L821	-1.856
		L85	1.041	L823	-1.857
		L86a	-0.991	L824	0.252
		L86b	-0.788	L825	0.251
		L86c	0.784	L826	0.835
		L87	-0.992	L827	-1.861
		L88	-1.915	L828	-1.862
		L89	-1.915	L829	-0.345
		L810	0.603	L8210	-0.345
		L811	-2.552	L8211	-1.019
				L8212	2.318
		Mittlerer IP	-0.293		-0.299
		LW	59.2%		59.4%
0.685	3	L32	3.856	L323	0.848
		L33	0.562	L324	2.926
		L39	0.329	L326b	2.358
		L312	0.796	L327	0.985
		L56b	-0.694	L328	2.064
		L57	1.302	L329	2.402
		L58	2.047	L3212	2.19
		L59	2.046	L522	0.816

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		L510	0.034	L524b	1.127
		L511	3.239	L5212	0.776
		L82	2.654	L822	-1.012
		L85	1.041	L824	0.252
		L86c	0.784	L829	-0.345
		L810	0.603	L8211	-1.019
		Mittlerer IP	1.329		1.26
		LW	34.4%		36.0%
-0.223	8	L31	-2.028	L51a	-1.055
		L37	-0.404	L521b	-1.419
		L38	-2.02	L524a	-0.132
		L39	0.329	L525a	-2.075
		L310	-0.39	L526	-1.417
		L311	-1.576	L527	-0.507
		L313	-2.014	L821	-1.856
		L53	-1.537	L822	-1.012
		L55	-1.537	L823	-1.857
		L57	1.302	L827	-1.861
		L510	0.034	L828	-1.862
		L512	-1.717	L829	-0.345
		L513	-1.719	L8210	-0.345
		L81	-1.91	L8211	-1.019
		L86b	-0.788		
		Mittlerer IP	-1.141		-1.197
		LW	71.4%		72.9%
0.380	3	L31	-2.028	L323	0.848
		L32	3.856	L324	2.926
		L33	0.562	L326b	2.358

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		L34	-0.406	L327	0.985
		L35	-0.405	L328	2.064
		L39	0.329	L329	2.402
		L312	0.796	L3210	-0.537
		L51	0.354	L311b	-0.513
		L52	0.354	L3212	2.19
		L57	1.302	L522	0.816
		L58	2.047	L524b	1.127
		L59	2.046	L5212	0.776
		L511	3.239	L829	-0.345
		L82	2.654	L8210	-0.345
		Mittlerer IP	1.05		1.054
		LW	33.9%		33.8%
0.685	5	L32	3.856	L326b	2.358
		L311	-1.576	L327	0.985
		L51	0.354	L328	2.064
		L52	0.354	L329	2.402
		L53	-1.537	L3210	-0.537
		L54	-0.445	L311a	0.741
		L55	-1.537	L311b	-0.513
		L56a	0.669	L3212	2.19
		L56b	-0.694	L51a	-1.055
		L57	1.302	L521b	-1.419
		L58	2.047	L525a	-2.075
		L59	2.046	L528	2.247
		L510	0.034	L529	-1.053
		L511	3.239	L5212	0.776
		Mittlerer IP	0.579		0.508

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
LW			52.6%		54.5%
-0.223	8	L31	-2.028	L51a	-1.055
		L37	-0.404	L521b	-1.419
		L38	-2.02	L524a	-0.132
		L39	0.329	L525a	-2.075
		L310	-0.39	L526	-1.417
		L311	-1.576	L527	-0.507
		L313	-2.014	L821	-1.856
		L53	-1.537	L822	-1.012
		L55	-1.537	L823	-1.857
		L57	1.302	L827	-1.861
		L510	0.034	L828	-1.862
		L512	-1.717	L829	-0.345
		L513	-1.719	L8210	-0.345
		L81	-1.91	L8211	-1.019
		L86b	-0.788		
Mittlerer IP			-1.141		-1.197
LW			71.4%		72.9%
0.999	8	L31	-2.028	L322	-0.385
		L32	3.856	L3212	2.19
		L33	0.562	L527	-0.507
		L34	-0.406	L821	-1.856
		L35	-0.405	L823	-1.857
		L36	-0.803	L824	0.252
		L37	-0.404	L825	0.251
		L38	-2.02	L826	0.835
		L39	0.329	L827	-1.861
		L310	-0.39	L828	-1.862

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		L311	-1.576	L829	-0.345
		L312	0.796	L8210	-0.345
		L313	-2.014	L8211	-1.019
				L8212	2.318
		Mittlerer IP	-0.322		-0.299
		LW	79%		78.6%
-0.533	5	L33	0.562	L821	-1.856
		L34	-0.406	L822	-1.012
		L35	-0.405	L823	-1.857
		L36	-0.803	L824	0.252
		L37	-0.404	L825	0.251
		L38	-2.02	L826	0.835
		L39	0.329	L827	-1.861
		L310	-0.39	L828	-1.862
		L311	-1.576	L829	-0.345
		L312	0.796	L8210	-0.345
		L313	-2.014	L8211	-1.019
		L53	-1.537	L8212	2.318
		L56b	-0.694		
		L510	0.034		
		Mittlerer IP	-0.609		-0.542
		LW	51.8%		50.2%
0.381	8	L31	-2.028	L51a	-1.055
		L33	0.562	L521b	-1.419
		L38	-2.02	L524a	-0.132
		L39	0.329	L525a	-2.075
		L310	-0.39	L526	-1.417
		L311	-1.576	L527	-0.507

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		L313	-2.014	L821	-1.856
		L51	0.354	L822	-1.012
		L53	-1.537	L823	-1.857
		L55	-1.537	L827	-1.861
		L510	0.034	L828	-1.862
		L512	-1.717	L829	-0.345
		L513	-1.719	L8210	-0.345
		L81	-1.91	L8211	-1.019
		Mittlerer IP	-1.087		-1.197
		LW	81.3%		82.9%
0.685	8	L33	0.562	L51a	-1.055
		L34	-0.406	L524a	-0.132
		L35	-0.405	L527	-0.507
		L36	-0.803	L821	-1.856
		L37	-0.404	L822	-1.012
		L38	-2.02	L823	-1.857
		L39	0.329	L824	0.252
		L310	-0.39	L825	0.251
		L311	-1.576	L826	0.835
		L312	0.796	L827	-1.861
		L313	-2.014	L828	-1.862
		L53	-1.537	L829	-0.345
		L54	-0.445	L8210	-0.345
		L56b	-0.694	L8211	-1.019
		Mittlerer IP	-0.643		-0.585
		LW	79.1%		78.1%
0.999	3	L32	3.856	L323	0.848
		L33	0.562	L324	2.926

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		L312	0.796	L326b	2.358
		L52	0.354	L327	0.985
		L56a	0.669	L328	2.064
		L57	1.302	L329	2.402
		L58	2.047	L3212	2.19
		L59	2.046	L522	0.816
		L511	3.239	L523	1.164
		L82	2.654	L524b	1.127
		L83	0.604	L528	2.247
		L85	1.041	L5212	0.776
		L86c	0.784	L824	0.252
		L810	0.603	L826	0.835
		Mittlerer IP	1.468		1.495
		LW	38.5%		37.9%
0.593	3	Ma7	2.035	G231	1.351
		Ma9a	2.507	G232	1.357
		Ma9b	2.505	G233	2.043
		Ma9c	2.042	G234	1.335
		Ma11	2.483	G235c	0.495
		Wu7b	1.549	G254	0.027
		Wu11	1.045	G255	0.208
		G132	-1.955	G257	0.539
		G133	0.314	G282	0.532
		G134	-0.138	G284a	0.522
		G135a	1.96	G284b	0.58
		G135b	1.33	G286	0.97
		G135c	1.963	G21c	1.47
		G136	1.342	G2285	1.145

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		G137	-0.124		
Mittlerer IP			1.108		0.989
LW			37.4%		40.2%
1.290	3	Ma8b	1.719	G231	1.351
		Ma7	2.035	G232	1.357
		Ma9a	2.507	G233	2.043
		Ma9b	2.505	G234	1.335
		Ma9c	2.042	G286	0.97
		Ma11	2.483	G21c	1.47
		Wu4	2.357	G2285	1.145
		Wu5	1.545	G153	0.945
		Wu6	2.357	G135c	1.963
		Wu8	-0.399	G136	1.342
		Wu9	1.505	Wu7b	1.549
		Wu10	1.545		
		G134	-0.138		
		G135a	1.96		
Mittlerer IP			1.716		1.547
LW			39.5%		43.7%
-1.396	5	Ma1	-0.831	G235a	-0.341
		Ma3	-2.482	G235b	-0.411
		Wu8	-0.399	G251	-0.41
		Wu12	-2.346	G252	-1.47
		G131	-0.433	G253	-0.326
		G132	-1.955	G256	-1.385
		G134	-0.138	G281	-1.792
		G137	-0.124	G283	-1.401
		G151	-2.026	G21a	-0.137

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		G152	-1.984	G258	-0.796
		G154	-1.15	Wu3	-0.443
		G182	-0.66		
		G183	-2.204		
		G184	-1.132		
		Mittlerer IP	-1.276		-0.9
		LW	47.0%		59%
-0.687	3	Ma1	-0.831	G231	1.351
		Ma5	0.737	G234	1.335
		Ma4	0.526	G235a	-0.341
		Ma8a	-0.008	G235b	-0.411
		Wu2	0.054	G235c	0.495
		Wu3	-0.443	G252	-1.47
		Wu4	2.357	G253	-0.326
		Wu7a	0.405	G256	-1.385
		Wu8	-0.399	G257	0.539
		G133	0.314	G21a	-0.137
		G134	-0.138	G21b	0.454
		G135b	1.33	G21c	1.47
		G137	-0.124	G236	0.292
		G153	0.945	G23	0.292
		Mittlerer IP	0.170		0.154
		LW	29.9%		30.2%
-0.454	3	Ma3	-2.482	G231	1.351
		Ma5	0.737	G234	1.335
		Ma4	0.526	G235a	-0.341
		Ma8a	-0.008	G235b	-0.411
		Ma8b	1.719	G235c	0.495

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		Wu3	-0.443	G252	-1.47
		Wu4	2.357	G253	-0.326
		Wta16	2.357	G256	-1.385
		Wu10	1.545	G257	0.539
		G151	-2.026	G21a	-0.137
		G152	-1.984	G21b	0.454
		G153	0.945	G21c	1.47
		G154	-1.15	G236	0.292
				G23	0.292
		Mittlerer IP	0.161		0.154
		LW	35.1%		35.3%
-0.687	5	Ma1	-0.831	G235a	-0.341
		Ma3	-2.482	G235b	-0.411
		Ma4	0.526	G251	-0.41
		Ma8a	-0.008	G252	-1.47
		Wu3	-0.443	G253	-0.326
		Wu8	-0.399	G256	-1.385
		G132	-1.955	G281	-1.792
		G133	0.314	G283	-1.401
		G134	-0.138	G21a	-0.137
		G151	-2.026	G258	-0.796
		G152	-1.984		
		G153	0.945		
		G183	-2.204		
		G184	-1.132		
		Mittlerer IP	-0.854		-0.91
		LW	54.2%		55.6%
-0.454	8	Ma1	-0.831	G235a	-0.341

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		Ma3	-2.482	G235b	-0.411
		Wu8	-0.399	G251	-0.41
		Wu12	-2.346	G252	-1.47
		G131	-0.433	G253	-0.326
		G132	-1.955	G256	-1.385
		G134	-0.138	G281	-1.792
		G137	-0.124	G283	-1.401
		G151	-2.026	G21a	-0.137
		G152	-1.984	G258	-0.796
		G154	-1.15		
		G182	-0.66		
		G183	-2.204		
		G184	-1.132		
		Mittlerer IP	-1.276		-0.91
		LW	70.0%		61%
-0.214	8	Ma1	-0.831	G235a	-0.341
		Ma3	-2.482	G235b	-0.411
		Wu8	-0.399	G251	-0.41
		Wu12	-2.346	G252	-1.47
		G131	-0.433	G253	-0.326
		G132	-1.955	G256	-1.385
		G134	-0.138	G281	-1.792
		G137	-0.124	G283	-1.401
		G151	-2.026	G21a	-0.137
		G152	-1.984	G258	-0.796
		G154	-1.15		
		G182	-0.66		
		G183	-2.204		

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		G184	-1.132		
		Mittlerer IP	-1.276		-0.91
		LW	74.3%		66.7%
0.035	8	Ma1	-0.831	G235a	-0.341
		Ma3	-2.482	G235b	-0.411
		Wu8	-0.399	G251	-0.41
		Wu12	-2.346	G252	-1.47
		G131	-0.433	G253	-0.326
		G132	-1.955	G256	-1.385
		G134	-0.138	G281	-1.792
		G137	-0.124	G283	-1.401
		G151	-2.026	G21a	-0.137
		G152	-1.984	G258	-0.796
		G154	-1.15		
		G182	-0.66		
		G183	-2.204		
		G184	-1.132		
		Mittlerer IP	-1.276		-0.91
		LW	78.8%		72.0%
0.592	3	Ma7	2.035	Wtal9	1.505
		Ma9a	2.507	Wtal10	1.545
		Ma9b	2.505	G153	0.945
		Ma9c	2.042	G231	1.351
		Ma11	2.483	G232	1.357
		Wu7b	1.549	G233	2.043
		Wu11	1.045	G234	1.335
		G132	-1.955	G235c	0.495
		G133	0.314	G257	0.539

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		G134	-0.138	G282	0.532
		G135a	1.96	G284a	0.522
		G135b	1.33	G284b	0.58
		G135c	1.963	G286	0.97
		G136	1.342	G21c	1.47
		G137	-0.124	G2285	1.145
		Mittlerer IP	1.108		1.371
		LW	37.4%		31.4%
1.290	5	Ma7	2.035	G231	1.351
		Ma9a	2.507	G232	1.357
		Ma9b	2.505	G233	2.043
		Ma9c	2.042	G234	1.335
		Ma11	2.483	G235c	0.495
		Wu7b	1.549	G254	0.027
		Wu11	1.045	G255	0.208
		G132	-1.955	G257	0.539
		G133	0.314	G284a	0.522
		G134	-0.138	G284b	0.58
		G135a	1.96	G286	0.97
		G135b	1.33	G21c	1.47
		G135c	1.963	G2285	1.145
		G136	1.342		
		G137	-0.124		
		Mittlerer IP	1.108		0.989
		LW	54.5%		57.4%
0.035	8	Ma1	-0.831	G235a	-0.341
		Ma3	-2.482	G235b	-0.411
		Wu8	-0.399	G251	-0.41

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		Wu12	-2.346	G252	-1.47
		G131	-0.433	G253	-0.326
		G132	-1.955	G256	-1.385
		G134	-0.138	G281	-1.792
		G137	-0.124	G283	-1.401
		G151	-2.026	G21a	-0.137
		G152	-1.984	G258	-0.796
		G154	-1.15		
		G182	-0.66		
		G183	-2.204		
		G184	-1.132		
		Mittlerer IP	-1.276		-0.91
		LW	78.8%		72%
-1.396	5	Ma1	-0.831	G235a	-0.341
		Ma3	-2.482	G235b	-0.411
		Wu8	-0.399	G251	-0.41
		Wu12	-2.346	G252	-1.47
		G131	-0.433	G253	-0.326
		G132	-1.955	G256	-1.385
		G134	-0.138	G281	-1.792
		G137	-0.124	G283	-1.401
		G151	-2.026	G21a	-0.137
		G152	-1.984	G258	-0.796
		G154	-1.15		
		G182	-0.66		
		G183	-2.204		
		G184	-1.132		
		Mittlerer IP	-1.276		-0.919

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
LW			47.0%		59.0%
-0.214	8	Ma1	-0.831	G235a	-0.341
		Ma3	-2.482	G235b	-0.411
		Wu8	-0.399	G251	-0.41
		Wu12	-2.346	G252	-1.47
		G131	-0.433	G253	-0.326
		G132	-1.955	G256	-1.385
		G134	-0.138	G281	-1.792
		G137	-0.124	G283	-1.401
		G151	-2.026	G21a	-0.137
		G152	-1.984	G258	-0.796
		G154	-1.15		
		G182	-0.66		
		G183	-2.204		
		G184	-1.132		
Mittlerer IP			-1.276		-0.919
LW			74.3%		66.7%
-0.687	5	Wu3	-0.443	G131	-0.433
		Wu7a	0.405	G235a	-0.341
		Wu8	-0.399	G235b	-0.411
		G133	0.314	G251	-0.41
		G137	-0.124	G252	-1.47
		G151	-2.026	G253	-0.326
		G152	-1.984	G254	0.027
		G153	0.945	G255	0.208
		G154	-1.15	G256	-1.385
		G181	0.453	G257	0.539
		G182	-0.66	G281	-1.792

Anhang D: Adaptive Zuordnung der Itemparameter zu den Fähigkeitsparametern

<i>PP</i>	<i>Gruppe</i>	<i>1. Übung</i>		<i>2. Übung</i>	
		<i>Item</i>	<i>IP</i>	<i>Item</i>	<i>IP</i>
		G183	-2.204	G283	-1.401
		G184	-1.132	G21a	-0.137
				G258	-0.796
	Mittlerer IP		-0.572		-0.581
	LW		47.1%		47.4%

Anmerkungen:

PP= Personen- bzw. Fähigkeitsparameter

IP= Itemparameter (Schwierigkeit)

LW= Lösungswahrscheinlichkeit

Anhang E: Skalierung der Aufgaben in Studie 2

Die Aufgabenbeschreibung ist folgendermaßen aufgebaut: neben der Itembezeichnung (Spalte 1) wird der Testzeitpunkt (Pretest, Übungsphase 1, Übungsphase 2, Posttest) in Spalte 2 angegeben. Anschließend sind in Spalte 3 der Itemparameter der Aufgabenschwierigkeit mit dem zugehörigen Standardfehler in Spalte 4 dargestellt. Der fünften Spalte ist der Infit-Wert (MNSQ) und der sechsten Spalte der T-Wert zu entnehmen. Zudem ist abschließend in Spalte 7 die auf der Grundlage der klassischen Testverfahren ermittelte Lösungshäufigkeit abgebildet.

Tabelle 43: Rasch-Skalierung der Aufgaben in Studie 2

<i>Item</i>	<i>T</i>	<i>Itemparameter</i>	<i>Standardfehler</i>	<i>MNSQ¹¹</i>	<i>T-Wert¹</i>	<i>LH</i>
Bo1	Pretest	1.094*	0.225	0.88	-0.90	.23
Bo2	Pretest	-2.608*	0.278	0.54	-3.40	.92
Do511	Pretest	2.825*	0.410	0.68	-2.30	.25
Le1	Pretest	0.510*	0.272	0.74	-1.40	.32
Bo4a	Pretest	-0.505*	0.265	1.22	1.10	.78
Bo4b	Pretest	0.531*	0.205	1.38	1.30	.89
Bo7	Pretest	1.412*	0.266	0.94	-0.40	.67
Bo9	Pretest	-1.573*	0.222	1.02	0.20	.96
Bo5a	Pretest	-0.081*	0.195	1.18	1.30	.50
L2	Pretest	1.271	0.290	1.07	0.40	.04
Bo10a	Pretest	1.615*	0.253	0.76	-1.90	.61
Bo10b	Pretest	1.615*	0.253	0.76	-1.90	.29
TIMSSB2	in allen 4 Tests eingesetzt	-1.448*	0.103	0.96	-0.50	.74
TIMSSB3	in allen 4 Tests eingesetzt	-0.813*	0.122	0.97	-0.30	.44
TIMSSX01	in allen 4 Tests eingesetzt	0.645*	0.130	1.00	0.00	.73
TIMSSD1	in allen 4 Tests eingesetzt	-0.533*	0.097	0.99	-0.10	.36

¹¹ (ungewichteter Wert)

Anhang E: Skalierung der Aufgaben in Studie 2

<i>Item</i>	<i>T</i>	<i>Itemparameter</i>	<i>Standardfehler</i>	<i>MNSQ¹¹</i>	<i>T- Wert¹</i>	<i>LH</i>
TIMSSH05	in allen 4 Tests eingesetzt	-0.333	0.329	1.11	0.50	.87
TIMSSJ05	in allen 4 Tests eingesetzt	-0.883	0.360	0.61	-1.70	.67
L31	Übungsphase 1	-2.028	0.278	0.84	-1.10	.60
L32	Übungsphase 1	3.856	0.410	0.68	-2.30	.00
L33	Übungsphase 1	0.562	0.568	0.96	0.10	.52
L34	Übungsphase 1	-0.406	0.615	0.84	-0.30	.56
L35	Übungsphase 1	-0.405	0.615	0.88	-0.20	.56
L36	Übungsphase 1	-0.803	0.436	0.86	-0.80	.32
L37	Übungsphase 1	-0.404	0.265	1.22	1.10	.48
L38	Übungsphase 1	-2.020	0.453	0.89	-0.60	.32
L39	Übungsphase 1	0.329	0.430	1.07	0.60	.42
L310	Übungsphase 1	-0.390	0.702	0.69	-0.50	.42
L311	Übungsphase 1	-1.576	0.366	0.98	0.00	.36
L312	Übungsphase 1	0.796	0.547	1.11	0.40	.32
L313	Übungsphase 1	-2.014	0.277	0.95	-0.10	.38
L51	Übungsphase 1	0.354	0.675	1.04	0.30	.67
L52	Übungsphase 1	0.354	0.675	1.13	0.40	.57
L53	Übungsphase 1	-1.537	0.342	1.19	1.30	.67
L54	Übungsphase 1	-0.445	0.351	1.26	1.00	1.00
L55	Übungsphase 1	-1.537	0.799	0.89	-1.20	.67
L56a	Übungsphase 1	0.669	0.638	0.94	-0.70	.67
L56b	Übungsphase 1	-0.694	0.376	0.83	-0.50	.51
L57	Übungsphase 1	1.302	0.653	1.05	0.30	.43
L58	Übungsphase 1	2.047	0.707	1.15	0.50	.29
L59	Übungsphase 1	2.046	0.707	0.96	0.10	.40
L510	Übungsphase 1	0.449	1.010	1.00	-1.00	.67
L511	Übungsphase 1	3.239	0.476	1.52	1.90	.29

Anhang E: Skalierung der Aufgaben in Studie 2

<i>Item</i>	<i>T</i>	<i>Itemparameter</i>	<i>Standardfehler</i>	<i>MNSQ¹¹</i>	<i>T- Wert¹</i>	<i>LH</i>
L512	Übungsphase 1	-1.717	0.355	1.04	0.20	.51
L513	Übungsphase 1	-1.719	0.357	1.02	0.20	.67
L81	Übungsphase 1	-1.910	0.770	1.15	-0.20	.80
L82	Übungsphase 1	2.654	0.390	0.95	-0.10	.40
L83	Übungsphase 1	0.604	0.567	1.29	0.80	.60
L84	Übungsphase 1	-0.990	0.331	1.01	0.10	.90
L85	Übungsphase 1	1.041	0.566	0.72	-0.60	.50
L86a	Übungsphase 1	-0.991	0.336	0.99	0.20	.90
L86b	Übungsphase 1	-0.788	0.689	0.62	-0.80	.88
L86c	Übungsphase 1	0.784	0.587	0.95	0.00	.68
L87	Übungsphase 1	-0.992	0.282	0.76	-1.20	.80
L88	Übungsphase 1	-1.915	0.410	0.85	-0.50	1.00
L89	Übungsphase 1	-1.915	0.410	0.85	-0.50	1.00
L810	Übungsphase 1	0.603	0.567	0.83	-0.30	.78
L811	Übungsphase 1	-2.552	0.731	0.68	-2.20	.90
L321	Übungsphase 2	-1.939	0.366	0.98	0.00	.50
L322	Übungsphase 2	-0.385	0.594	0.81	-0.30	.44
L323	Übungsphase 2	0.848	0.645	0.87	-0.10	.33
L324	Übungsphase 2	2.926	0.771	0.74	1.80	.00
L325	Übungsphase 2	-0.968	0.355	1.04	0.20	.33
L326a	Übungsphase 2	-0.968	0.355	1.04	0.20	.40
L326b	Übungsphase 2	2.358	0.788	1.44	1.60	.33
L327	Übungsphase 2	0.985	0.579	0.84	-0.30	.40
L328	Übungsphase 2	2.064	0.659	1.47	1.10	.25
L329	Übungsphase 2	2.402	0.788	1.49	1.10	.00
L3210	Übungsphase 2	-0.537	0.393	1.17	0.60	.25
L311a	Übungsphase 2	0.741	0.551	1.47	1.20	.44
L311b	Übungsphase 2	-0.513	0.583	1.25	0.70	.59

Anhang E: Skalierung der Aufgaben in Studie 2

<i>Item</i>	<i>T</i>	<i>Itemparameter</i>	<i>Standardfehler</i>	<i>MNSQ¹¹</i>	<i>T- Wert¹</i>	<i>LH</i>
L3212	Übungsphase 2	2.190	0.654	1.44	1.10	.33
L51a	Übungsphase 2	-1.055	0.516	0.90	-0.20	.80
L521b	Übungsphase 2	-1.419	0.544	0.99	0.10	.80
L522	Übungsphase 2	0.816	0.599	0.83	-0.30	.30
L523	Übungsphase 2	1.164	0.660	0.78	-0.40	.30
L524a	Übungsphase 2	-0.132	0.485	1.15	0.50	.73
L524b	Übungsphase 2	1.127	0.535	0.81	-0.40	.44
L525a	Übungsphase 2	-2.075	0.675	0.91	-0.10	.80
L525b	Übungsphase 2	-1.450	0.615	0.81	-0.30	.73
L526	Übungsphase 2	-1.417	0.544	0.73	-0.70	.80
L527	Übungsphase 2	-0.507	0.563	1.25	0.70	.60
L528	Übungsphase 2	2.247	0.556	0.72	-2.00	.11
L529	Übungsphase 2	-1.053	0.517	0.90	-0.20	.73
L5210	Übungsphase 2	-1.417	0.391	0.96	0.00	.73
L5211	Übungsphase 2	2.247	0.556	0.72	-2.00	.11
L5212	Übungsphase 2	0.776	0.510	1.00	0.10	.56
L821	Übungsphase 2	-1.856	0.342	1.19	1.30	1.00
L822	Übungsphase 2	-1.012	0.679	0.65	-0.70	.80
L823	Übungsphase 2	-1.857	0.342	1.19	1.30	1.00
L824	Übungsphase 2	0.252	0.626	0.97	0.10	.80
L825	Übungsphase 2	0.251	0.626	0.97	-0.10	.80
L826	Übungsphase 2	0.835	0.635	0.98	0.10	.71
L827	Übungsphase 2	-1.861	0.342	1.24	1.50	1.00
L828	Übungsphase 2	-1.862	0.342	1.24	1.50	1.00
L829	Übungsphase 2	-0.345	0.369	1.04	1.20	.71
Le8210	Übungsphase 2	-0.345	0.369	1.04	1.20	.83
L8211	Übungsphase 2	-1.019	0.678	0.99	0.10	.86
L8212	Übungsphase 2	2.318	0.545	0.67	-0.60	.56

Anhang E: Skalierung der Aufgaben in Studie 2

<i>Item</i>	<i>T</i>	<i>Itemparameter</i>	<i>Standardfehler</i>	<i>MNSQ¹¹</i>	<i>T- Wert¹</i>	<i>LH</i>
We7	Pretest	-1.041*	0.365	0.92	-0.20	.67
We10	Pretest	-0.839*	0.330	0.78	-0.90	.21
Moe6a	Pretest	-0.858*	0.282	0.76	-1.20	.69
Moe6b	Pretest	-0.725*	0.338	0.79	-0.70	.64
Moe6c	Pretest	-0.726*	0.338	0.93	-0.20	.67
Ma1	Übungsphase 1	-0.831*	0.290	1.07	0.40	.67
Ma3	Übungsphase 1	-2.482*	0.438	1.04	0.20	.16
Ma5	Übungsphase 1	0.737*	0.439	1.89	2.40	.88
Ma4	Übungsphase 1	0.526*	0.338	0.77	-1.20	.75
Ma8a	Übungsphase 1	-0.008*	0.396	0.74	-0.80	.83
Ma8b	Übungsphase 1	1.719*	0.499	0.78	-0.70	.63
Ma7	Übungsphase 1	2.035*	0.421	0.77	-0.70	.66
Ma9a	Übungsphase 1	2.507*	0.557	0.59	-1.50	.49
Ma9b	Übungsphase 1	2.505*	0.557	1.20	0.70	.00
Ma9c	Übungsphase 1	2.042*	0.386	0.64	-1.30	.33
Ma11	Übungsphase 1	2.483*	0.555	0.33	-2.90	.00
Wu2	Übungsphase 1	0.054*	0.349	0.90	-0.30	.75
Wu3	Übungsphase 1	-0.443*	0.345	1.29	1.10	.63
Wu4	Übungsphase 1	2.357*	0.493	0.48	-2.30	.63
Wu5	Übungsphase 1	1.545*	0.425	0.70	-1.20	.67
Wu16	Übungsphase 1	2.357*	0.493	0.84	-0.50	.71
Wu7a	Übungsphase 1	0.405*	0.358	0.94	-0.10	.33
Wu7b	Übungsphase 1	1.549*	0.392	0.94	-0.10	.79
Wu8	Übungsphase 1	-0.399*	0.557	1.41	1.30	1.00
Wu9	Übungsphase 1	1.505*	0.385	0.84	-0.50	.53
Wu10	Übungsphase 1	1.545*	0.425	1.16	0.60	.25
Wu11	Übungsphase 1	1.045*	0.555	1.73	2.10	.33
Wu12	Übungsphase 1	-2.346*	0.392	0.72	-0.90	.67

Anhang E: Skalierung der Aufgaben in Studie 2

<i>Item</i>	<i>T</i>	<i>Itemparameter</i>	<i>Standardfehler</i>	<i>MNSQ¹¹</i>	<i>T- Wert¹</i>	<i>LH</i>
G1	Pretest	-0.052	0.418	0.87	-0.30	.78
G2a	Pretest	1.995	0.454	1.04	0.20	.63
G2b	Pretest	0.578	0.415	0.87	-0.30	.91
G4b	Pretest	3.226	0.520	1.20	1.00	.23
G5	Pretest	0.268	0.414	0.73	-0.70	.77
G6a	Pretest	-1.657	0.478	0.87	-1.10	.91
G7	Pretest	0.267	0.415	0.81	-0.50	.91
G8	Pretest	0.266	0.415	0.81	-0.70	.91
G131	Übungsphase 1	-0.433	0.097	0.99	-0.10	.66
G132	Übungsphase 1	-1.955	0.254	0.96	-0.20	.33
G133	Übungsphase 1	0.314	0.757	1.12	0.70	.49
G134	Übungsphase 1	-0.138	0.337	0.81	-0.70	.33
G135a	Übungsphase 1	1.960	0.454	1.14	-1.20	.00
G135b	Übungsphase 1	1.330	0.501	0.93	-0.10	.49
G135c	Übungsphase 1	1.963	0.478	1.18	-1.30	.16
G136	Übungsphase 1	1.342	0.501	1.03	-0.50	.16
G137	Übungsphase 1	-0.124	0.354	0.85	-0.90	.49
G151	Übungsphase 1	-2.026	0.902	1.41	2.00	.73
G152	Übungsphase 1	-1.984	0.905	1.10	0.40	.69
G153	Übungsphase 1	0.945	0.626	0.81	-0.40	.21
G154	Übungsphase 1	-1.150	0.644	0.86	-0.20	.51
G181	Übungsphase 1	0.453	0.345	1.05	0.30	.67
G182	Übungsphase 1	-0.660	0.859	0.90	-0.20	.62
G183	Übungsphase 1	-2.204	0.913	1.49	1.90	1.00
G184	Übungsphase 1	-1.132	0.359	1.18	0.70	.75
G231	Übungsphase 2	1.351	0.705	1.18	1.10	1.00
G232	Übungsphase 2	1.357	0.706	1.19	1.10	.12
G233	Übungsphase 2	2.043	0.678	1.38	1.20	.24

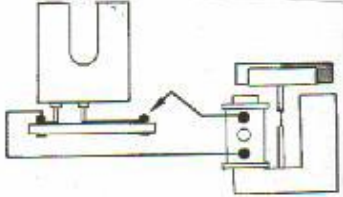
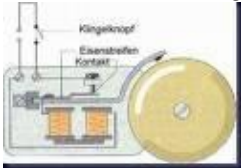
Anhang E: Skalierung der Aufgaben in Studie 2


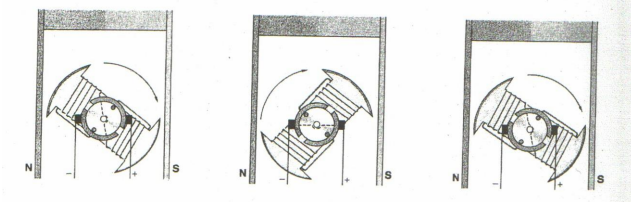
<i>Item</i>	<i>T</i>	<i>Itemparameter</i>	<i>Standardfehler</i>	<i>MNSQ¹¹</i>	<i>T- Wert¹</i>	<i>LH</i>
G234	Übungsphase 2	1.335	0.536	1.12	-0.50	.33
G235a	Übungsphase 2	-0.341	0.377	1.18	0.70	.33
G235b	Übungsphase 2	-0.411	0.428	1.13	0.40	.33
G235c	Übungsphase 2	0.495	0.359	0.96	-0.90	.00
G251	Übungsphase 2	-0.410	0.628	1.13	0.70	.42
G252	Übungsphase 2	-1.470	0.674	1.39	0.80	1.00
G253	Übungsphase 2	-0.326	0.676	0.92	0.10	.42
G254	Übungsphase 2	0.027	0.555	0.91	0.00	.56
G255	Übungsphase 2	0.208	0.439	0.81	-0.80	.14
G256	Übungsphase 2	-1.385	0.693	0.86	0.10	.56
G257	Übungsphase 2	0.539	0.681	1.25	0.60	.56
G281	Übungsphase 2	-1.792	0.679	0.89	0.10	1.00
G282	Übungsphase 2	0.532	0.681	1.24	0.60	.72
G283	Übungsphase 2	-1.401	0.694	0.87	0.10	1.00
G285	Übungsphase 2	0.069	0.694	1.23	0.50	.86
G284a	Übungsphase 2	0.522	0.680	1.23	0.50	.66
G284b	Übungsphase 2	0.580	0.492	1.15	0.50	.56
G286	Übungsphase 2	0.970	0.507	0.86	-0.20	.62
G21a	Übungsphase 2	-0.137	0.484	1.06	0.30	.56
G21b	Übungsphase 2	0.454	0.601	0.78	-0.20	.72
G21c	Übungsphase 2	1.470	0.670	1.26	1.50	.86
G236	Übungsphase 2	0.292	0.412	0.95	-0.10	.14
G258	Übungsphase 2	-0.796	0.196	1.08	0.40	.22
G2285	Übungsphase 2	1.145	0.211	1.17	1.30	.86
G23	Übungsphase 2	0.292	0.205	0.98	0.00	.31
L41	Posttest	-1.285	0.360	0.89	-0.30	.74
L42	Posttest	2.139	0.440	1.15	1.50	.24
L43	Posttest	-2.623	0.431	0.94	-1.20	.88

Anhang E: Skalierung der Aufgaben in Studie 2

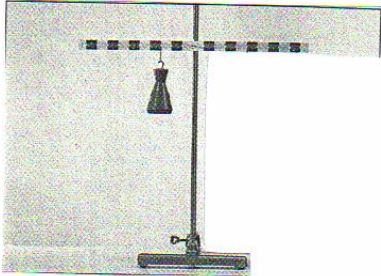
<i>Item</i>	<i>T</i>	<i>Itemparameter</i>	<i>Standardfehler</i>	<i>MNSQ¹¹</i>	<i>T- Wert¹</i>	<i>LH</i>
L44a	Posttest	-0.721	0.343	0.86	-0.70	.78
L44b	Posttest	0.940	0.370	0.81	-0.60	.31
L45	Posttest	-0.372	0.439	1.20	0.60	.63
L46	Posttest	-0.372	0.439	1.20	0.60	.48
L47a	Posttest	-3.076	0.461	1.19	1.40	.88
L47b	Posttest	-1.493	0.367	1.07	0.30	.95
L48	Posttest	-3.076	0.461	1.24	-2.10	.88
L49	Posttest	1.446	0.397	0.85	-0.50	.23
L410	Posttest	-1.492	0.366	0.85	-0.50	.69
L411	Posttest	-1.719	0.377	0.93	-0.20	.86
L412a	Posttest	-2.265	0.408	1.16	0.70	.83
L412b	Posttest	-0.196	0.340	1.02	1.30	.79
L413	Posttest	-1.085	0.352	1.03	-1.00	.76
GP1a	Posttest	-4.655	0.568	1.37	-0.90	.11
GP1b	Posttest	0.506	0.437	1.00	0.10	.89
GP2	Posttest	-1.729	0.444	1.18	0.60	.78
GP3a	Posttest	-0.248	0.420	1.07	0.30	.44
GP3b	Posttest	0.961	0.455	0.98	-0.80	.11
GP4	Posttest	-1.317	0.429	1.04	1.20	.78
GP5	Posttest	-0.250	0.419	1.12	0.90	.66
GP6a	Posttest	-0.597	0.417	1.00	0.10	.66
GP6b	Posttest	0.109	0.425	0.91	-1.00	.33
GP7	Posttest	-0.251	0.419	1.20	-0.70	.44

Anhang F. Aufgaben in Studie 1

<i>Item-Code</i>	<i>Aufgabenstellung</i>
1 Bo1	Welche Funktion hat der Kommutator (Stromwender) bei dem Elektromotor?
Bo3a	Das nebenstehende Bild zeigt, wie ein einfacher „Elektromotor“ gebaut werden kann. Als Rotor wird ein Stabmagnet verwendet, als Stator eine Spule mit Eisenkern. Wie kann man den Magnet zum Drehen bringen? 
2	
3 Bo3b	Welche Nachteile besitzt dieser „Motor“?
4 Bo3c	Wie kann man diese Nachteile beheben?
5 Bo4b	Erkläre die Funktionsweise eines Elektromotors.
Bo5a	Eine elektrische Klingel hat den in der Abbildung dargestellten Aufbau. 
6	Beschreibe den Aufbau und erkläre die Wirkungsweise einer solchen Klingel! Nutze dafür dein Wissen über Elektromotoren.

<i>Item-Code</i>	<i>Aufgabenstellung</i>
Bo7	In manchen Motoren findet man einen Anker mit drei Spulen. Verfolge den Stromkreis und erläutere, wie sich dieser Anker bewegt. Welchen Vorteil hat er gegenüber einem Anker mit nur zwei gegenüberliegenden Spulen?
	
7	
Bo10a	Die unten stehenden Bilder zeigen einen Elektromotor mit verschiedenen Stellungen des Rotors. Erkläre, warum sich der Rotor jeweils weiterdreht. Kennzeichne die Pole des Elektromagneten in allen drei Bildern.
	
8	
9 Bo10b	Wenn der Rotor waagrecht steht, fließt kein Strom durch ihn. Welches Problem tritt dadurch auf?
Bo11	Vervollständige bitte den Text mit Hilfe der folgenden Begriffe:
	<i>wechselnde, Kommutator, Stator, Stromanschlüsse, kreisende, Rotor, Spule, Polwender, Wechselwirkung, umgepolt</i>
	Ein Gleichstrommotor besteht aus einem _____, der durch einen U- förmigen Permanentmagneten gebildet wird, einem _____, der aus einem Elektromagnet und einem _____ besteht. Durch Umpolung der _____ kann die Drehrichtung des Rotors umgekehrt werden. Die Arbeitsweise beruht auf der _____ zweier Magnetsysteme. Das feststehende Magnetsystem, der Stator, ist ein u- förmig gestalteter Magnet. Das sich drehende Magnetsystem, der Rotor, wird durch eine _____ mit Eisenkern auf einer Achse gebildet. Durch einen mit auf der Achse sitzenden Stromwender (_____ genannt) wird dieses Magnetsystem entsprechend der Drehzahl regelmäßig _____, so dass zwischen den beiden Magnetsystemen ständig _____, anziehende und abstoßende Kraft wirksam wird, die den Rotor in _____ Bewegung versetzt.
10	

Item-Code	Aufgabenstellung
Ma1	Vervollständige den folgenden Satz zu Halbleitern mit den Begriffen:
11	<p>Störstellenleitung, Elemente, n- Halbleitern, Isolatoren, Außenelektronen, Halbleiters, Löchern, Leitfähigkeit, Dotieren</p> <p>Halbleiter sind Stoffe, deren elektrische _____ zwischen der von Leitern und _____ liegt. Die Leitfähigkeit eines _____ kann gezielt erhöht werden, wenn man Atome anderer _____ (Fremdatome) einbringt, die mehr oder weniger _____ haben als die Halbleiter. Diesen Vorgang nennt man _____. Beim Dotieren entstehen Störstellen mit freien Elektronen oder _____. Die darauf basierende Leitung wird _____ genannt. Je nach Dotierung unterscheidet man zwischen _____ und p- Halbleitern.</p>
12 Ma2	Wie schaltet man ein Strommessgerät bzw. ein Spannungsmessgerät?
Ma10	Fertige eine Versuchsbeschreibung mit Schaltbild, Durchführung, Beobachtung und Schlussfolgerung des Versuchs von letzter Woche an.
13	Zur Erinnerung: Folgende Instrumente wurden bei dem Versuch eingesetzt: Diode, Widerstand, Spannungsquelle und ein Strommessgerät.
Ma11	Drei Silicium-Dioden sind in Reihe geschaltet. Welche Spannung ist ungefähr nötig, damit es zu einem Strom kommt? Erkläre dabei auch die Schleusenspannung.
14	
Do5113	Ein stromdurchflossener Leiter wirkt wie ein Magnet. Besonders stark ist die magnetische Wirkung, wenn der Leiter als Spule aufgewickelt ist und einen Eisenkern enthält. Schaltet man den Strom ab, hört die magnetische Wirkung auf. Große Elektromagnete werden genutzt, um schwere Lasten zu heben und zu transportieren.
15	Wie muss solch ein Lasthebemagnet aufgebaut sein? Wie kann man seine Wirkungsweise erklären?
We1	Widerstände können schnell mit speziellen Messgeräten gemessen werden. Stehen diese nicht zur Verfügung, kann man sich anders helfen. Nenne eine Möglichkeit.
16	
Moe4	In der Abbildung sind zwei Glühlampen sind an eine elektrische Quelle angeschlossen. Was passiert, wenn jeweils eine der beiden Lampen aus der Fassung geschraubt wird? Begründe deine Antwort.
17	

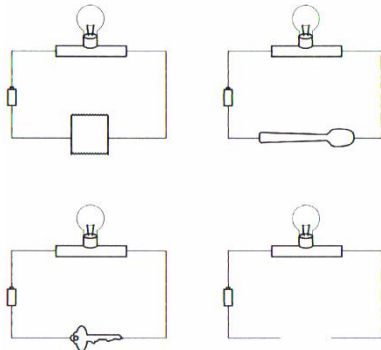
<i>Item-Code</i>	<i>Aufgabenstellung</i>
A14	Das Gewicht in der Abbildung hat eine Gewichtskraft von 0,2 N. Zeichne in die Abbildung, wo ein weiteres Gewicht an den Hebel gehangen werden muss und welche Gewichtskraft dieses haben muss, damit der Hebel im Gleichgewicht ist.
	18
	19 A15b
20 A110b	Wie groß ist das Produkt, wenn die Platte nur horizontal bewegt wird?
Dor1	Was versteht man unter der Gewichtskraft eines Körpers? Was versteht man unter der Masse eines Körpers?
21	Weshalb ist die Gewichtskraft eines Körpers nicht überall gleich?
TIMSS1 (B2)	Der größte Teil der chemischen Energie, die bei der Benzinverbrennung in einem Automotor freigesetzt wird, wird nicht für den Antrieb des Autos verwendet, sondern umgewandelt in: A. Elektrizität B. Wärme C. Magnetismus D. Schall
22	

Item-Code

Aufgabenstellung

TIMSS2
(K13)

Die folgenden Abbildungen zeigen jeweils eine Taschenlampenbatterie und eine Glühlampe, die über Drähte mit verschiedenen Stoffen verbunden sind. Welche Glühlampen leuchten auf?



- A. Nur 1
- B. Nur 2 und 3
- C. Nur 1 und 3
- D. Nur 1, 3 und 4
- E. Nur 1, 2 und 3.

23

TIMSS3 (X01) Elektrische Energie wird eingesetzt, um eine Lampe leuchten zu lassen. Wie kann die Menge an eingesetzter Energie mit der Menge an produzierter Lichtenergie verglichen werden?

- a) Die Menge an elektrischer Energie ist:
- 1. größer, als die Menge an produzierter Lichtenergie
 - 2. geringer, als die Menge an produzierter Lichtenergie
 - 3. genauso groß, wie die Menge an produzierter Lichtenergie

b) Bitte begründe deine Antwort:

24

Item-Code

Aufgabenstellung

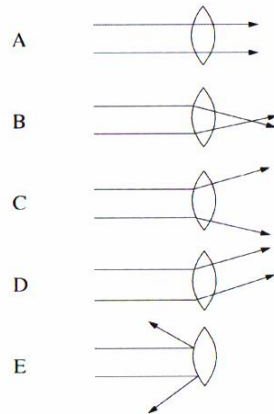
TIMSS4 (B3) Welcher Gegenstand aus der Tabelle hat die größte Dichte?

Gegenstand	Masse des Gegenstands	Volumen des Gegenstands
W	11,0 Gramm	24 Kubikzentimeter
X	11,0 Gramm	12 Kubikzentimeter
Y	5,5 Gramm	4 Kubikzentimeter
Z	5,5 Gramm	11 Kubikzentimeter

- A. W
- B. X
- C. Y
- D. Z

25

TIMSS5 (D1) Welche Zeichnung gibt am besten wieder, was geschieht, wenn Licht durch ein Vergrößerungsglas fällt?

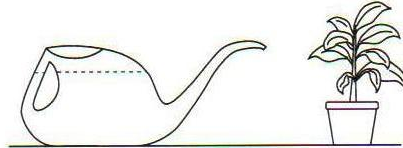


26

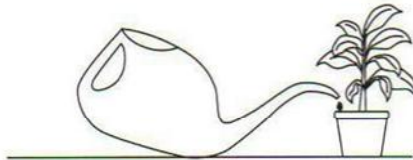
Item-Code

Aufgabenstellung

TIMSS6 (R1) Eine Gieskanne ist fast voll mit Wasser, wie dargestellt.

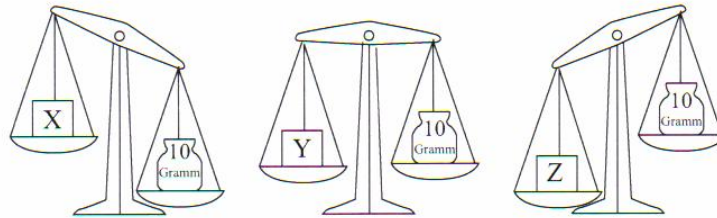


Die Gieskanne wird leicht gekippt, sodass das Wasser gerade beginnt herauszutropfen.









27 Zeichne eine Linie, die den jetzigen Wasserstand in der Gieskanne anzeigt.

TIMSS1 (N9) Welche der Schachteln X, Y und Z hat die GERINGSTE Masse?



- A. Z
 - B. X
 - C. Y
 - D. Alle drei Schachteln haben die gleiche Masse.
-

28

<i>Item-Code</i>	<i>Aufgabenstellung</i>
TIMSS2 (B6)	Die Wände eines Gebäudes sollen so gestrichen werden, dass sie möglichst viel Licht reflektieren. In welcher Farbe sollten sie gestrichen werden? A. weiß B. rot C. schwarz D. rosa
29	
HSTIMSS3 (K14)	Ein Schüler hat je 100 ml Wasser in jeden dieser offenen Behälter gefüllt und lässt sie einen Tag in der Sonne stehen. Aus welchem Behälter verdunstet vermutlich am meisten Wasser?
	<p>A. </p> <p>B. </p> <p>C. </p> <p>D. </p>
30	
TIMSS5 (D2)	Jeder der drei abgebildeten Magneten ist in den Stoff unter ihm eingetaucht worden. Welcher Stoff könnte Kaffee sein?
	<p></p> <p></p>
31	

Item-Code	Aufgabenstellung
Bo6 32	Was passiert, wenn du bei einem Elektromotor an die senkrecht zur Achse gesteckte Spule mit Eisenkern eine Gleichspannung von 4 Volt anschließt? Wie verhält sich die Spule? Und wie verhält sich die Spule, wenn du die Stromzuleitung an der Gleichspannungsquelle umpolst?
33 Vi1	Wie wird der Wirkungsgrad definiert?
34 Vi2	Wie ist elektrische Leistung definiert? Gib Formelzeichen und Einheit an.
35 Vi3	Warum braucht die Spule eines Elektromagneten manchmal „Starthilfe“? Warum muss die Spule manchmal angedreht werden?
Do251a	<p>Hier ist noch mal das Diagramm von eurem Arbeitsblatt abgebildet. Zur Erinnerung: Es geht um den Anlasser beim Auto. In dem Diagramm ist auf der horizontalen Achse der Strom I aufgetragen. Die Kennlinien für Drehzahl (n), Drehmoment (M) und die Klemmenspannung (U) der Batterie sind ebenfalls eingezeichnet. Die Leistungskennlinie (P) wurde aus den Werten von n und M berechnet. Beantworte folgende Fragen bitte zu dem Diagramm: Beschreibe jeweils die Kennlinien von Drehmoment (m) und Spannung (U) unabhängig voneinander durch eine „Je... desto...“- Beziehung.</p>
36	
37 Do251b	Erkläre, warum es Sinn macht, dass die Linien jeweils so verlaufen.
38 Do252	Überlege, wie es experimentell überprüfbar ist, ob eine Batterie Spannung verliert, wenn sie in Betrieb war?
39 Do253	Stelle folgende Geradengleichung: $f(x) = mx + b$ auf, die den Verlauf der Batteriespannung beschreibt.
40 Do254	Warum kann das Ohm'sche Gesetz beim Elektromotor nicht angewendet werden? Was ist hier anders?

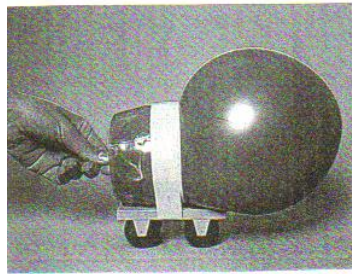
<i>Item-Code</i>	<i>Aufgabenstellung</i>
Wi6 41	Welche der folgenden Aussagen ist richtig? A. Die Gleitreibung ist immer größer als die Haftreibung. B. Die Haftreibung ist immer größer als die Gleitreibung. C. Die Gleitreibung ist gegeben, wenn ein Gegenstand nicht bewegt wird. D. Die Haftreibung ist gegeben, wenn ein Gegenstand bewegt wird.
Wi9 42	Eine Holzkiste mit der Anpresskraft $F_G = 1850 \text{ N}$ wird über einen Steinfußboden geschoben. Zum Anschieben ist eine Kraft von 1295 N notwendig. Um die Kiste in Bewegung zu halten, reicht die Kraft von 555 N . Berechne die Haft- und Gleitreibungszahl.
43 Wi5b	Wie viel N entsprechen 100 gr ?
Wi2	Vervollständige den Lückentext mit folgenden Begriffen: Gewichtskraft, Haftreibung, glatt, Kraft, Reibung, Kontakt, bewegt, Gleitreibung, Unebenheiten, Bewegung Überall dort, wo Gegenstände gezogen, geschoben und überhaupt _____ werden, ist _____ im Spiel. Dabei kann die Oberfläche dieser Körper noch so _____ aussehen, unter dem Mikroskop betrachtet weisen sie stets Unebenheiten auf. Beim Ziehen oder Schieben über eine feste Unterlage verhaken sich diese _____ ineinander und es entsteht Reibung, eine _____, die erst einmal überwunden werden muss, wenn _____ zustande kommen soll. Wenn die _____ eines Körpers dabei besonders groß ist, erhöht sich die Reibung. Die Größe der Reibung hängt auch von der Beschaffenheit der Oberflächen ab, die miteinander _____ haben. Am Anfang ist die Reibung immer am größten. Dies nennt man _____. Sobald sich ein Gegenstand bewegt, wird die Reibung kleiner. Das nennt man _____.
Wi11 45	Welcher Zusammenhang besteht bei folgenden Tätigkeiten zur Reibungskraft? a) Einen Teig ausrollen. b) Gläser spülen. c) Fußboden polieren.
HW7 46	Ein Schüler sitzt auf einem Stuhl mit Rollen und hält ein Paket in den Händen. Was passiert, wenn sie das Paket nach vorne wirft? A. Das Paket übt Druck auf den Schüler aus, aber der Schüler übt wiederum keinen Druck auf das Paket aus. B. Der Schüler übt Druck auf das Paket aus und auch das Paket übt Druck auf den Schüler aus, wodurch er mit dem Stuhl nach hinten rollt. C. Nur der Schüler übt Druck auf das Paket aus, das Paket übt keinen Druck auf den Schüler aus. D. Schüler und Paket üben keinen Druck aufeinander aus. Der Schüler rollt zurück, weil er sich bewegt.
47 HW1	Erkläre, wie du eine Rakete in kurzer Zeit auf hohe Geschwindigkeit und Höhe beschleunigen kannst.
48 HW3	Beschreibe ein dir bekanntes Beispiel bei dem das Rückstoßprinzip wirkt.
HW6a 49	Zwei Schüler stehen sich auf Skates gegenüber. Robert stößt sich von Katharina ab. Beschreibe die Bewegungen der beiden Schüler nach dem Abstoßen! Begründe deine Aussage.

Item-Code

Aufgabenstellung

HW8

Beschreibe jeweils die Versuche in den beiden Abbildungen und erkläre, was ihnen gemeinsam ist.



50

Dor6

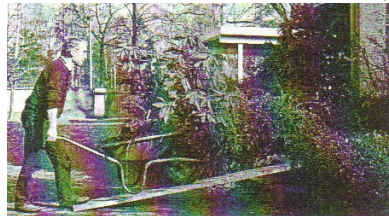
Welche der folgenden Aussagen ist richtig?
A. Die Gewichtskraft hängt von dem Ort ab, an dem sich der Körper befindet.
B. Die Gewichtskraft ist an jedem Ort gleich.
C. Masse und Gewichtskraft sind identisch.

51

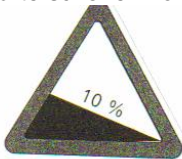
D. Die Anziehungskraft der Erde und aller anderen Planeten auf Gegenstände ist gleich.

Dor4a

Was musst du in dem Bild ändern, um mit weniger Kraft die Karre samt Inhalt schieben zu können?



52

Item-Code	Aufgabenstellung																		
Dor5 53	<p>Bei einer Radtour treffen Anna und Max auf das nebenstehende Verkehrsschild (Neigung: 10%). Beide sind sich einig, dass sie nun ihre Kräfte schonen können. Was bedeutet es aber genau?</p> 																		
Dor8 54	<p>Finde heraus, welche der folgenden schiefen Ebenen gleich steil sind. Die erste Zahl gib die zu überwindende Höhe an, die zweite den zurückzulegenden Weg: 20/40, 40/80, 25/75, 40/ 120, 45/90 und 30/180.</p>																		
Dor9 55	<p>Wie lang muss ein Brett mindestens sein, damit ein Junge eine 50 kg schwere Schubkarre über eine 30 cm hohe Stufe schieben kann, wenn seine Muskelkraft in Wegrichtung maximal 300 N beträgt? Welche Auswirkung hätte eine Verdoppelung der Brettlänge auf die dann aufzubringende Kraft?</p>																		
Dor10 56	<p>Schrauben sind nichts anderes als „aufgewickelte“ schiefe Ebenen. Wie wirkt sich die Anzahl der Windungen auf die Kraft aus, die beim Schrauben aufgebracht werden muss?</p>																		
Dor11	<p>Vervollständige den Text mit folgenden Begriffen:</p> <p>Verringerung, gleich, Weg, Verhältnis, Arbeit, Höhe, Kraft, länger, Proportionalität, kleiner, Verdoppelung, Hälfte.</p> <p>Kraft kann man „sparen“, indem man einen längeren _____ wählt, die _____, die dabei verrichtet werden muss, bleibt allerdings jedes Mal _____. Das _____ von Höhe zu Weg gibt an, wie viel _____ zur Überwindung des Höhenunterschieds notwendig ist. Im gleichen Verhältnis wie die Kraft _____ wird, muss der Weg _____ werden, um eine bestimmte _____ zu erreichen. Es herrscht eine _____ zwischen Kraft und Weg, eine _____ der Kraft bedeutet eine _____ des Weges um die _____.</p>																		
We2 58	<p>Berechne die fehlenden Größen in der Tabelle:</p> <table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <tbody> <tr> <td>U</td> <td>30 V</td> <td>200 V</td> <td>...</td> <td>0,5 V</td> <td>...</td> </tr> <tr> <td>I</td> <td>...</td> <td>...</td> <td>5 A</td> <td>3 mA</td> <td>2 mA</td> </tr> <tr> <td>R</td> <td>210 Ω</td> <td>1,5 Ω</td> <td>10 Ω</td> <td>...</td> <td>5 Ω</td> </tr> </tbody> </table>	U	30 V	200 V	...	0,5 V	...	I	5 A	3 mA	2 mA	R	210 Ω	1,5 Ω	10 Ω	...	5 Ω
U	30 V	200 V	...	0,5 V	...														
I	5 A	3 mA	2 mA														
R	210 Ω	1,5 Ω	10 Ω	...	5 Ω														
59 We3	<p>Wie kannst du erreichen, dass eine Lampe in einem Stromkreis mal schwach, mal sehr hell leuchtet?</p>																		

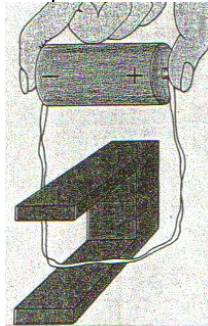
Item-Code	Aufgabenstellung
We4 60	Es gibt elektrische Lichterketten mit 16 oder 10 in Reihe geschalteten „Kerzen“. Für eine 16er-Kette muss eine Kerze nachgekauft werden. Du hast die Wahl zwischen Kerzen mit der Aufschrift „14V“ und „23V“. Für welche entscheidest du dich? Begründe deine Antwort. Was geschieht, wenn du aus Versehen eine 14-V-Kerze in eine Lichterkette schraubst, die für 23-V-Kerzen vorgesehen ist? Was passiert im umgekehrten Fall?
We5 61	Welche der folgenden Aussagen zur Veränderung der Anzahl der Verbraucher in einem Stromkreis und dem Widerstand ist richtig? A. Je größer die Zahl der gleichen Verbraucher, desto größer wird der Widerstand der Gesamtverbraucher. B. Je kleiner die Zahl der gleichen Verbraucher, desto kleiner wird der Widerstand der Gesamtverbraucher. C. Je größer die Zahl der gleichen Verbraucher, desto kleiner wird der Widerstand der Gesamtverbrauch. D. Der Widerstand ändert sich nicht durch die Anzahl der gleichen Verbraucher.
We6 62	Erläutere am Beispiel der Fahrradbeleuchtung mit einem Dynamo den Zusammenhang zwischen der elektrischen Spannung und der Stromstärke im Stromkreis.
We9 63	Beschreibe den Unterschied zwischen Reihen- und Parallelschaltung und nenne die Gleichungen für Stromstärke und Spannung bei den beiden Schaltungarten.
Wu1	Vervollständige mit folgenden Begriffe den Satz: Stärke, Leiter, Kraft, Richtung, Leiterstück, Stromstärke, Magnetfeld, Stromrichtung
64	Jeder stromdurchflossene _____ ist von einem _____ umgeben. Die _____ des Magnetfeldes hängt von der _____ ab. Die _____ des Magnetfeldes hängt von der _____ ab. Die _____, die auf die Leiterschleife wirkt, ist umso größer, je länger das im Magnetfeld befindliche _____ ist.
65 Wu2	Erkläre die Drei-Finger-Regel und ihren Zweck.

Item-Code

Aufgabenstellung

Wu3

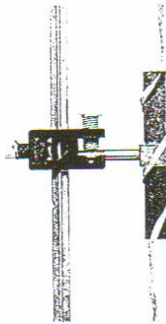
Was passiert, wenn der Lamettafaden wie im Folgenden mit der Batterie verbunden wird? Warum, und in welche Richtung bewegt sich der Lamettafaden? Begründe deine Antwort. Was geschieht, wenn die Stromrichtung verändert wird oder wenn man die Lage von Nord- und Südpol vertauscht?



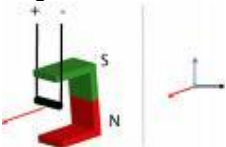
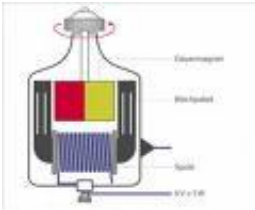
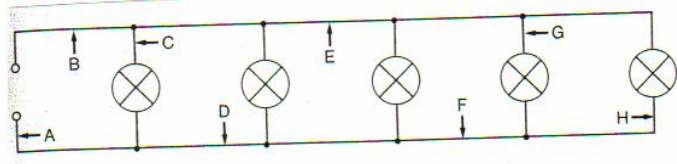
66

Wu4

Was passiert, wenn der Leiter mit Strom durchflossen wird? Wie kann man erreichen, dass sich der Leiterstreifen in unterschiedliche Richtungen bewegt?



67

Item-Code	Aufgabenstellung
Wu5	Zwischen die Pole des Hufeisenmagneten hängt senkrecht zu den Feldlinien eine „Leiterschaukel“, durch die ein starker Gleichstrom geschickt wird. In welche Richtung bewegt sich die Leiterschaukel, wenn Strom fließt? Begründe deine Vermutung.
68	
Wu6 69	Zeichne die Möglichkeiten, die für Feld- und Stromrichtung der stromdurchflossenen Leiterschaukel im Feld des Hufeisenmagneten bestehen können. Bestimme mit der Drei-Finger-Regel jeweils die Krafrichtung
Wu7a 70	Beschreibe die Energieumwandlungen bei einem a) Elektromotor b) Wechselstromgenerator.
Wu10	Erkläre mit Hilfe der Abbildung die Funktionsweise eines Fahrraddynamos. Wie wird hier die Energie erzeugt?
71	
72 Ra1	Wie lässt sich bei einer Schaltung mehrerer Bauteile prüfen, ob eine Reihen- oder eine Parallelschaltung vorliegt?
73 Ra2	Im Haushalt werden elektrische Geräte parallel geschaltet. Begründe, weshalb diese Art der Schaltung gewählt wird.
74 Ra3a	Welche Aufgabe hat die Sicherung in einem Stromkreis?
Ra6a	In der Schaltskizze sind verschiedene Stellen durch Buchstaben gekennzeichnet. An welcher Stelle könnte ein Strommesser eingebaut werden, der den Gesamtstrom anzeigt? Begründe deine Wahl.
75	

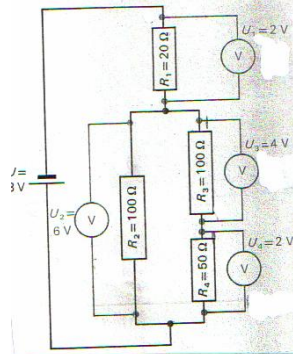
<i>Item-Code</i>	<i>Aufgabenstellung</i>
76 Ra6b	Wo würdest du einen Schalter einsetzen, mit dem zwei der fünf Lampen gesondert geschaltet werden können?
Ra8a 77	An ein Netzgerät werden nacheinander mehrere Lampen parallel angeschlossen. Was wird der Strommesser anzeigen, wenn immer mehr Lampen hinzugeschaltet werden?
Ra8b 78	Was passiert, wenn zusätzlich ein Heizdraht (z.B. eines Toasters) und ein weiterer Eisendraht als „Sicherung“ parallel angeschlossen werden?
79 Ra10	Warum nennt man eine Schmelzsicherung so und was schmilzt in ihr?
Ra11 80	Eine Sicherung brennt bei einem Kurzschluss oder einer Überlastung durch. Dabei fließt der Strom beim Kurzschluss durch gar keine Geräte und bei Überlastung durch zu viele und jedes Mal springt die Sicherung an. Was ist die gemeinsame Ursache?
81 Lü1	Wie kannst du ein elektromagnetisches Feld herstellen?
82 Lü2	Ein Magnetfeld ist unsichtbar. Wie kannst du trotzdem feststellen, dass es vorhanden ist?
83 Lü3	Ist ein Magnetfeld immer gleich stark oder gibt es unterschiedlich starke Magnetfelder? Begründe deine Antwort.
Lü11 84	Zwei Stabmagnete werden zusammengehalten: einmal mit gleichnamigen Polen und einmal mit ungleichnamigen Polen nebeneinander. Dann werden jeweils Nägel damit hochgehoben. In welchem Fall ergibt sich eine große, in welchem eine kleine magnetische Kraft? Erkläre, warum dies so ist.
85 Lü6	Beschreibe den Aufbau und erkläre die Wirkungsweise einer Klingel.
Lü9 86	Ein Magnet ist nicht gekennzeichnet. Du weißt daher nicht, wo sich der Nordpol und der Südpol befinden. Wie kannst du herausfinden, welches der Nordpol ist?
Lü10 87	Ein Magnet zieht eine Büroklammer an und hält sie fest. Die Büroklammer zieht eine zweite Büroklammer an und hält diese fest. Wie ist das zu erklären?
Moe1	Vervollständige bitte folgenden Satz zur Maschenregel mit Hilfe der Begriffe: Null, minus, Summe, positive, negativ, plus.
88	Die _____ der Spannungen in einer Masche ist _____. Dabei werden alle Spannungen die von _____ nach _____ durchlaufen werden, _____ gerechnet, die anderen _____.

Item-Code

Aufgabenstellung

Moe3a

Beantworte mit Hilfe der Abbildung folgende Fragen:
 Trage jeweils die Plus- und Minusseite für jeden Widerstand in die Skizze ein



89

90 Moe3b

Berechne jeweils die fehlenden Stromstärken.

91 Moe3c

Was bezeichnet man im Zusammenhang mit einem Stromnetz als Masche?

92 Moe3d

Wie viele Maschen hat das Netz?

93 Moe3f

Welche Zusammenhänge gelten für die Stromstärke in dem Netz?

Moe6a

Zwei Widerstände mit $50\ \Omega$ und mit $200\ \Omega$ sind in einer elektrischen Quelle in Reihe geschaltet. Am Widerstand mit $50\ \Omega$ liegt eine Spannung von $12\ \text{V}$ an. Wie groß ist die am anderen Widerstand anliegende Spannung?

94

95 Do546

Welche Wirkung hat eine an die Spule angeschlossene Gleichspannung?

96 Do543

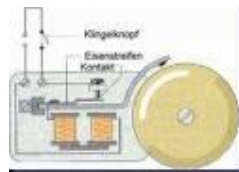
Mit welchem Gerät kann man eine Spannung z.B. von $230\ \text{V}$ auf $10,5\ \text{V}$ verringern?

97 Do545

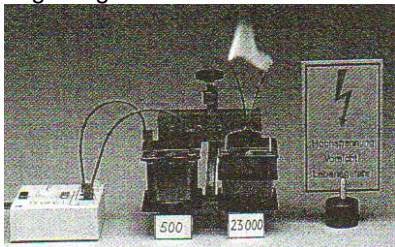
Warum entsteht in der zweiten Spule eines Transformators eine Wechselspannung?

Do548

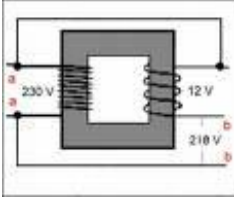
Eine elektrische Klingel hat den in der Abbildung dargestellten Aufbau. Beschreibe den Aufbau und erkläre die Wirkungsweise einer solchen Klingel! Nutze dafür dein Wissen über Transformatoren.

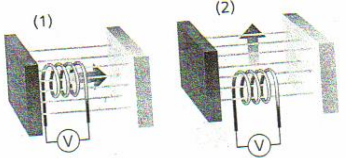
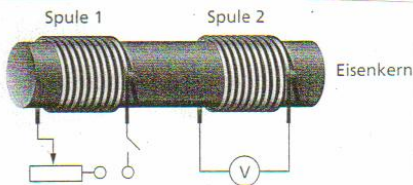
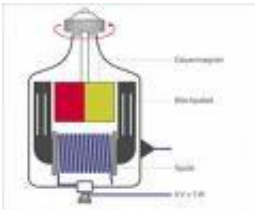


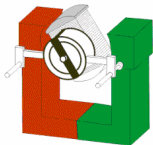
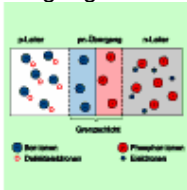
98

Item-Code	Aufgabenstellung
99 Do549	Was ist das Besondere an den Spulen eines Transformators?
Do5410	Fertige für den in der Abbildung dargestellten Versuch eine Versuchsbeschreibung mit Skizze, Durchführung, Beobachtung und Schlussfolgerung an.
100	
101 Se1	Wie lautet die Formel zur Berechnung der Dichte eines Stoffes?
102	Se2 Du hast bei einem Versuch zunächst Aluminium erhitzt und dieses dann in Wasser gegeben, dabei hast du einen Unterschied zwischen berechneter und gemessener Mischungstemperatur beobachtet. Wie lässt sich dieser Unterschied erklären?
103 Se4	Wie lautet die Formel zur Berechnung der Mischungstemperatur zweier unterschiedlich temperierter Wassermengen?
104	Se7 Ist das Wasser in einem unbeheizten Freibad erst mal warm, so kühlt es nur sehr langsam wieder ab. Ein oder zwei kühle Tage haben kaum Auswirkungen auf die Wassertemperatur. Erkläre.
Se13	Die spezifische Wärmekapazität von trockenem Sand beträgt $c = \frac{0,84 \text{ kJ}}{\text{kg} * \text{K}}$ und ist wesentlich kleiner als Wasser.
105	Begründe damit die hohen Temperaturschwankungen in Wüstengebieten im Unterschied zu Gebieten am Meer.
Se5	Bringt man zwei Körper unterschiedlicher Temperatur in Berührung, so stellt sich eine gemeinsame Temperatur ein, die zwischen den beiden ursprünglichen Temperaturen liegt. Welche der folgenden Aussagen ist richtig? A. Die Abnahme der Temperatur des wärmeren Körpers ist gleich dem Zuwachs der Temperatur des kälteren Körpers. B. Die Abnahme der Temperatur des wärmeren Körpers ist größer als der Zuwachs der Temperatur des kälteren Körpers. C. Der Zuwachs der Temperatur des kälteren Körpers ist größer als die Temperaturabnahme des wärmeren Körpers. D. Es hängt von der Höhe der Temperatur des wärmeren Körpers ab, ob deren Abnahme größer oder kleiner ist, als der Temperaturzuwachs des kälteren Körpers.
106	
107 Dü2	Erkläre die Begriffe Energieerzeugung und Energieverbrauch.
108 Dü3	Nenne einige dir bekannte Energieformen mit jeweils einem Beispiel.

Item-Code	Aufgabenstellung
109 Dü4b	Bitte begründe deine Antwort in a.
110 Dü5	Stell dir vor, ein Fahrzeug fährt auf einer geneigten Fahrbahn abwärts und prallt gegen eine Feder. Warum ist es nach dem Energieerhaltungssatz unmöglich, dass das Fahrzeug seine ursprüngliche Höhe wieder erreicht? Welche Energieform bleibt am Schluss übrig?
111 Dü6	Erkläre den Unterschied zwischen Wärmeenergie und Strahlungsenergie.
112 Dü9	Solarzellen wandeln Licht direkt in elektrischen Strom um. Da wäre es doch eine gute Idee, den Strom zu nutzen, um damit eine Lampe zum Leuchten zu bringen. Das Licht der Lampe fällt dann auf die Solarzelle, die damit wieder Strom erzeugen kann. Wird das funktionieren? Begründe deine Antwort.
113 Dü10	Erläutere, warum auch Pflanzen Energiewandler sind.
114 Dü11	Was passiert bei den folgenden Beispielen mit der Energie? Trage die beteiligten Energieformen wie im Beispiele ein: Beispiel: Chemische Energie → Auto, das fährt → Bewegungsenergie, Wärmeenergie, elektrische Energie. → Holz im Kamin → → elektrische Glühlampe → → Sonne → → Fahrraddynamo → → Benzinmotor eines Rasenmähers → → Windmühle → → Gasherd →
Dü12	Vervollständige den Lückentext mit den folgenden Begriffen: Erscheinungsart, verringert, Joule, andere, übertragen, Energie, hergestellt, Ursprung, Arbeit, Energieformen Physikalisch betrachtet ist _____ die Fähigkeit, _____ zu verrichten. Die Maßeinheit der Energie ist das _____. Je nach ihrer _____ unterscheidet man bestimmte _____, z.B. mechanische Energie, elektrische Energie, Wärmeenergie usw. Energie kann nicht aus dem Nichts gewonnen werden, sie kann auch nicht einfach _____ werden. Energie kann bei einer Umwandlung nicht vermehrt oder _____ werden. Energie braucht als _____ immer eine _____ Energieform. Energie kann von einem Körper auf einen anderen _____ werden.
115	

<i>Item-Code</i>	<i>Aufgabenstellung</i>
Dü4a	<p>Elektrische Energie wird eingesetzt, um eine Lampe leuchten zu lassen. Wie kann die Menge an eingesetzter Energie mit der Menge an produzierter Lichtenergie verglichen werden? Kreuze die richtige Antwort an. Die Menge der elektrischen Energie ist... A: größer als die Menge der produzierter Lichtenergie. B: geringer als die Menge der produzierten Lichtenergie. C: genauso groß wie die Menge der produzierter Lichtenergie. D: Es hängt von der Menge der elektrischen Energie ab, wie groß die Menge der produzierten Lichtenergie ist.</p>
116	
We4	<p>Welcher der folgenden Stoffe ist radioaktiv? A. Blei B. Thorium C. Zink D. Titan</p>
117	
We8	<p>Welcher der folgenden radioaktiven Stoffe wird häufig mit der Nahrung aufgenommen? A. Radon B. Polonium C. Uran D. Cäsium</p>
118	
119 We9a	<p>Welche Gefahren sind für den Menschen mit Radioaktivität verbunden?</p>
120 We9b	<p>Warum müssen die Keimzellen des Menschen besonders vor radioaktiver Strahlung geschützt werden?</p>
121 Dd3	<p>Weshalb wird die induzierte Spannung höher, wenn man Spulen mit Eisenkern verwendet?</p>
Dd4	<p>Ein Transformator dient der Umwandlung von elektrischen Spannungen und Stromstärken. Dabei wird die elektromagnetische Induktion genutzt. Ein Transformator besteht aus zwei Spulen, die sich auf einem geschlossenen Eisenkern befinden. Erkläre mit Hilfe deines Wissens zur elektromagnetischen Induktion und der Abbildung die Funktionsweise eines Transformators.</p>
122	

Item-Code	Aufgabenstellung
Dd5a	<p>In einem Magnetfeld wird eine Spule bewegt. Gib an, ob in Fall (1) oder in Fall (2) in der Spule eine Spannung induziert worden ist. Begründe deine Antwort.</p>
123	
124 Dd5b	<p>Nenne eine weitere Variante, in der keine Spannung induziert worden ist.</p>
Dd6a	<p>In der Abbildung ist eine Experimentieranordnung dargestellt, mit der in der feststehenden Spule 2 eine Spannung induziert werden soll. Beschreibe eine (experimentelle) Möglichkeit, wie mit dieser Anordnung in der Spule 2 eine Spannung induziert werden kann.</p>
125	
126 Dd6b	<p>Durch welche experimentellen Maßnahmen kann die in der Spule 2 induzierte Spannung vergrößert werden? Begründe.</p>
Dd7	<p>Welche der folgenden Aussagen zur elektromagnetischen Induktion ist richtig!</p> <p>A: Ein elektrischer Strom wird erzeugt, wenn ein Magnet relativ zur Spule bewegt wird. B: Elektrischer Strom wird erzeugt, wenn Magnet und Spule in Ruhe zueinander sind. C: Elektrischer Strom kann nicht erzeugt werden, wenn statt des Magneten (nur) die Spule bewegt wird. D: Elektrischer Strom entsteht unabhängig von Position des Magneten.</p>
127	
Dd8	<p>Erkläre mit Hilfe deines Wissens über die elektromagnetische Induktion und der Abbildung die Funktionsweise eines Fahrraddynamos.</p>
128	

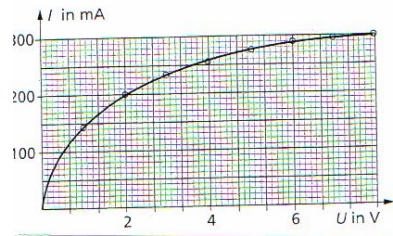
<i>Item-Code</i>	<i>Aufgabenstellung</i>
Bo2	Welche der folgenden Aussagen ist richtig? A. Jedes Magnetfeld hat einen Plus- und einen Minuspol. B. Der Nordpol wird in Abbildungen immer grün und der Südpol immer rot dargestellt. C. Jedes Magnetfeld hat einen Nord- und einen Südpol. D. Ungleiche Pole stoßen sich gegenseitig ab, gleiche Pole ziehen sich an.
129	
Bo4a	Beschrifte in der Abbildung die wichtigsten Bestandteile des Elektromotors mit Hilfe der folgenden Begriffe: Kommutator, Anker mit Drehspule, Feldmagnet, Schleifkontakt
130	
Bo8	Zeichne jeweils einen Stabmagneten, einen Hufeisenmagneten und einen Elektromagneten mit dem dazugehörigen Magnetfeld.
131	
132 Ma3	Warum ist ein dotierter Halbleiter nach außen elektrisch neutral?
Ma4	Erkläre mit Hilfe der Abbildung den Bau einer Halbleiterdiode und den Leitungsvorgang, der in ihr stattfindet. Beschreibe dabei auch die Vorgänge in der Grenzschicht: einmal, wenn die Diode in Sperrrichtung und einmal, wenn sie in Durchlassrichtung geschaltet ist.
133	

Item-Code

Aufgabenstellung

Ma8a

Für ein Bauelement wurde die nachfolgend dargestellte I-U-Kennlinie aufgenommen. Interpretiere das Diagramm.



134

135 Ma8b

Um was für ein Bauelement könnte es sich handeln? Begründe deine Vermutung.

Ma9a

Ein elektrisches Fieberthermometer wird mit einer kleinen Knopfzelle betrieben, die für eine konstante Spannung sorgt. Dabei dient ein Thermistor (temperaturabhängiger Halbleiterwiderstand) als Messfühler. Entwirf einen Schaltplan für ein solches Fieberthermometer.

136

Do517

Erkläre die Funktionsweise eines Kompasses. Weshalb besteht das Gehäuse eines Kompasses aus Kunststoff oder Messing und nicht aus Eisen? Wie kommt es, dass ein Kompass die Orientierung erleichtert?

137

138 Ma9b

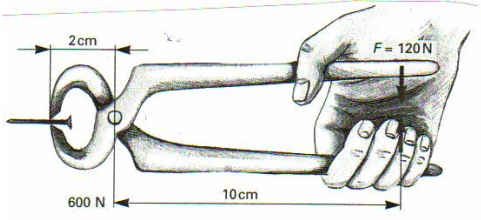
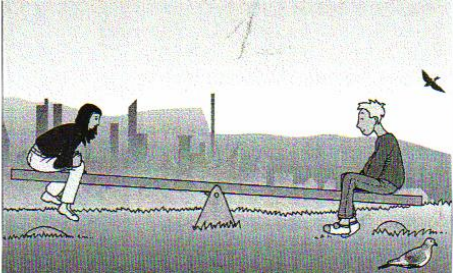
Was für eine Art Thermistor (Kaltleiter oder Heißleiter) kann für die Spitze des Thermometers verwendet werden? Begründe deine Antwort.

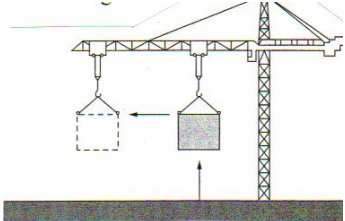
AI1

Berechne die fehlenden Werte:

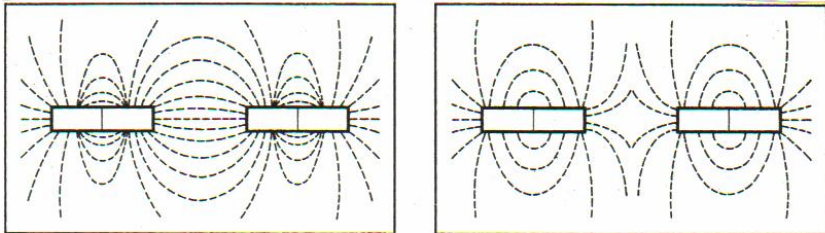
	a	b	c
F1 in N	420	320	45
S1 in cm	150		6
F1 * S1			
F2 in N		400	54
S2 in cm	315	50	
F2 * S2			

139

Item-Code	Aufgabenstellung
AI11	Bestimme jeweils das Produkt aus Kraft und Strecke in der Abbildung.
140	
141 AI3	Wofür steht die Abkürzung s in der Physik?
142 AI2	Welche Gewichtskraft hat eine Masse von 100 gr?
AI5a	Anna und Jörg wollen eine Wippe ins Gleichgewicht bringen, Anna ist jedoch leichter als Jörg. Wie kann die Wippe trotzdem ins Gleichgewicht gebracht werden?
143	
AI6	<p>Welche der folgenden Aussagen ist richtig?</p> <p>A. Damit Gleichgewicht zustande kommen kann, muss das Gewicht auf beiden Seiten eines Hebels gleich groß sein.</p> <p>B. Damit Gleichgewicht zustande kommen kann, muss die Summe aus Gewicht und Entfernung zum Mittelpunkt auf beiden Seiten gleich sein.</p> <p>C. Damit Gleichgewicht zustande kommen kann, muss das Produkt aus Gewicht und Entfernung zum Mittelpunkt auf beiden Seiten gleich sein.</p>
144	D. Damit ein Gleichgewicht zustande kommen kann, muss die Entfernung zum Mittelpunkt auf beiden Seiten eines Hebels gleich sein.
AI7	<p>An einem zweiseitigen Hebel, der um eine Achse drehbar ist, hängen folgende Gewichte:</p> <p>Links von der Drehachse: 20 N im Abstand von 30 cm; 25 N im Abstand von 20 cm und 15 N im Abstand von 10 cm.</p> <p>Rechts von der Drehachse: 18 N im Abstand von 40 cm; 12 N im Abstand von 35 cm und X (unbekannte Größe) N im Abstand von 11 cm.</p>
145	Wie groß muss X sein, damit der Hebel im Gleichgewicht ist?

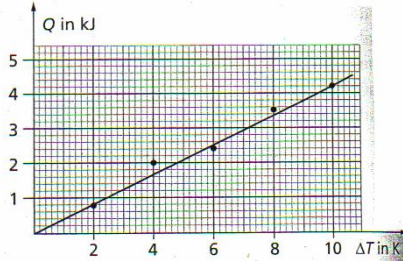
Item-Code	Aufgabenstellung
A18 146	An einem Hebel hängt links von der Drehachse in 20 cm Abstand ein Körper mit 0,75 N Gewichtskraft. Wie groß muss der Abstand eines Körpers auf der anderen Seite der Drehachse mit 0,6 N Gewichtskraft sein, damit der Hebel im Gleichgewicht ist?
A110a	Ein Kran hat eine Betonplatte (Masse $m = 500$ kg) um 5 m hochgehoben. Dann wird die Betonplatte in horizontaler Richtung um 4 m bewegt. Wie groß ist das Produkt aus Kraft und Strecke?
147	
148 HW14	Beim schnellen Anfahren eines Autos wird man in den Sitz gedrückt. Erkläre, wie es dazu kommt.
149 HW5	Stell dir vor, du bläst einen Luftballon auf und lässt ihn dann los. Erkläre, warum und wie er sich bewegt.
150 HW6b	Erkläre, was passiert und warum, wenn beide gleichzeitig das Seil in ihre Richtung ziehen.
HW11	<p>Wenn ein Auto auf einem Kiesweg abfährt, wird der Kies nach hinten weggeschleudert. Den Druck, den die Räder auf die Straße ausüben, ist also nach hinten gerichtet. Warum fährt das Auto dann nach vorne?</p> <p>A. Beim Anfahren übt das Auto einen Druck auf den Kiesweg aus und es kommt zu einem Rückstoß, wodurch der Kiesel nach hinten weggeschleudert wird.</p> <p>B. Der Kiesweg übt Druck auf das Auto aus, aber das Auto übt keinen Druck auf den Weg aus, es kommt zu keinem Rückstoß.</p> <p>C. Nur das Auto übt Druck auf den Weg aus, wodurch das Auto nach vorne fährt.</p>
151	D. Das Auto übt einen größeren Druck auf den Kiesweg aus, als der Weg auf das Auto.
Lü4	<p>Welche der folgenden Aussagen ist richtig?</p> <p>A. Jedes Magnetfeld hat einen Plus- und einen Minuspol.</p> <p>B. Es gibt auch Magneten mit nur einem Pol.</p> <p>C. Jedes Magnetfeld hat einen Nord- und einen Südpol.</p>
152	D. Ungleiche Pole stoßen sich gegenseitig ab, gleiche Pole ziehen sich an.
Dd10	<p>Welche der folgenden Aussagen zur elektromagnetischen Induktion ist richtig?</p> <p>A. Der in einer Spule induzierte Strom ist umso größer, je kleiner die Windungszahl n der Spule ist.</p> <p>B. Der in einer Spule induzierte Strom ist umso größer, je größer die Windungszahl n der Spule ist.</p> <p>C. Der in einer Spule induzierte Strom ist unabhängig von der Anzahl der Windungen der Spule.</p>
153	D. Der in einer Spule induzierte Strom ist umso kleiner, je größer die Windungszahl n der Spule ist.

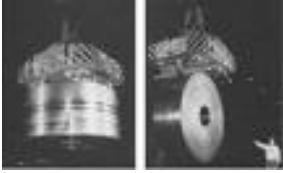
Item-Code	Aufgabenstellung
Dd1 154	Erkläre, wie es möglich ist, mit Magnetismus elektrischen Strom zu erzeugen und beschreibe kurz ein Experiment zur Überprüfung dieses Vorganges.
Dd9 155	Was passiert mit der Anzeige des Amperemeters, wenn man mit dem Nordpol eines Magneten in die Spule reingeht und den Magneten am Südpol des Magneten wieder herauszieht?
Dd11	<p>Bitte vervollständige den Text mit folgenden Begriffen:</p> <p>elektrischer, Stromstärke, Ausrichtung, Magnet, Ruhe, Windungszahl, größer, Spule, Bewegungsrichtung, Relativgeschwindigkeit</p> <p>Ein _____ Strom wird erzeugt, wenn ein _____ relativ zur _____ bewegt wird. Sind Spule und Magnet zueinander in _____, so gibt es keinen elektrischen Strom. Je _____ die _____ zwischen Magnet und Spule ist, desto größer ist die elektrische _____.</p> <p>Wie stark dabei das Amperemeter ausschlägt, hängt von</p> <ol style="list-style-type: none"> 1. der _____ der Pole des Magneten 2. der _____ des Magneten zur Spule und 3. der _____ der Spule ab.
156	
157 We2	Welche Arten der radioaktiven Strahlung gibt es? Worin unterscheiden sich diese?
158 We6	Was geschieht bei der Kernspaltung?
We11	<p>Vervollständige den Lückentext mit folgenden Begriffen:</p> <p>geladenen, Elektronen, α- Strahlung, Isotope, γ- Strahlung, Neutronen, Becquerel, Abschirmbarkeit, Protonen, β- Strahlung</p> <p>Atomkerne bestehen aus _____ und _____, die Hülle des Atoms wird von _____ gebildet. Atome eines Elementes, die sich nur in der Anzahl der Neutronen unterscheiden, heißen _____. Es gibt verschiedene radioaktive Strahlungsarten: _____ besteht aus zwei Neutronen und zwei Protonen, _____ besteht aus Elektronen. Diese Strahlungsarten bestehen aus _____ Teilchen.</p> <p>Die _____ ist eine Energiestrahlung und wird z. B. bei Röntgenaufnahmen eingesetzt. Die Strahlungsarten unterscheiden sich in ihrer Reichweite und _____. Als Aktivität eines radioaktiven Stoffes wird die Anzahl der Kernumwandlungen in einer Sekunde bezeichnet. Die Einheit der Aktivität ist _____.</p>
159	
160 We3	Warum kann es keine Nahrung geben, die keine radioaktiven Stoffe enthält?
161 He2	Wie kannst du erklären, dass die Knoten eines Mobilés im Gleichgewicht sind, obwohl unterschiedliche Gewichte daran hängen?

Item-Code	Aufgabenstellung
He9	Vervollständige den Text mit folgenden Begriffen:
	Newton, Gegenkraft, groß, Kraft, Masse, ortsunabhängig, Körper, gerichtet, Ort
162	Kräfte können _____ bewegen oder verformen. Zu jeder Kraft gib es eine _____. Diese Kräfte, die zwei Körper gegenseitig aufeinander ausüben, sind gleich _____, aber entgegengesetzt _____. Mit Kraftmessern kann die _____ gemessen werden. Die Einheit der Kraft ist _____ (N). Die Größe der Gewichtskraft hängt nicht nur von der _____ des Gegenstandes ab, sondern auch von dem _____ an dem sich der Gegenstand befindet. Die Masse eines Gegenstandes ist im Gegensatz dazu an jedem Ort gleich, sie ist also _____.
163 Do5110	Aus welchen Stoffen müssen Wandtafeln bestehen, damit man Gegenstände mithilfe von Magneten daran befestigen kann?
Do5118a	In den Abbildungen sind schematisch für unterschiedliche Polanordnungen zweier Stabeisenmagnete die Feldlinienbilder dargestellt. Welche Pole stehen sich jeweils gegenüber?
	
164	Abb.1: Abb.2:
Do5118b	Welche Wirkung wird jeweils zwischen den Stabeisenmagneten auftreten? (siehe Abbildung Do5118a)
165	Abb.1: Abb.2:
Ra4	Welche der folgenden Aussagen zur Sicherung ist richtig? A. Sicherungen dürfen in einem Stromkreis nie an der „schwächsten Stelle“ eingebaut werden. B. Die Aufgabe einer Sicherung ist es, den Stromkreis zu unterbrechen, bevor die Leitungen so heiß werden, dass es zu Schäden kommen kann. C. Die Sicherung sorgt dafür, dass der Strom auch dann weiter fließen kann, wenn viele Geräte parallel angeschlossen sind. D. Die Sicherung verhindert einen Kurzschluss.
166	

<i>Item-Code</i>	<i>Aufgabenstellung</i>
Ra9 167	Welche der folgenden Aussagen zur Parallelschaltung ist richtig? A. Die Gesamtstromstärke entspricht der Teilstromstärke. B. Die Gesamtstromstärke ist gleich dem Produkt der Teilstromstärken. C. Die Gesamtstromstärke ist gleich der Summe der Teilstromstärken. D. Die Gesamtstromstärke ist unabhängig von der Teilstromstärke.
Do255 168	Welche der folgenden Aussagen trifft bei dem Elektromotor zu: A. Bei geringer Spannung verändert sich die Stromstärke nur gering. B. Bei hoher Spannung verändert sich die Stromstärke nur gering. C. Sowohl bei geringer, wie bei hoher Spannung verändert sich die Stromstärke stark. D. Bei hoher Spannung verändert sich die Stromstärke stark.
Do256a 169	Welche der folgenden Aussagen trifft auf die Drehzahl bei einem Elektromotor zu? A. Je höher die Stromstärke, desto höher die Drehzahl. B. Je niedriger die Stromstärke, desto höher die Drehzahl. C. Je niedriger die Stromstärke, desto niedriger die Drehzahl. D. Bei mittlerer Stromstärke ist die Drehzahl am höchsten.
170 Do257b	Erkläre deine Antwort mit Hilfe eines Beispiels.
171 Moe5	Erkläre kurz das Ohm'sche Gesetz.
Ra3b 172	Welche der folgenden Aussagen zur Parallelschaltung ist richtig? A. Die Gesamtstromstärke entspricht der Teilstromstärke. B. Die Gesamtstromstärke ist gleich dem Produkt der Teilstromstärken. C. Die Gesamtstromstärke ist gleich der Summe der Teilstromstärken. D. Die Gesamtstromstärke ist unabhängig von der Teilstromstärke.
Bo9 173	Welche der folgenden Aussagen ist im Zusammenhang mit einem Elektromotor richtig? A. Sowohl die gleichnamigen und die ungleichnamigen Pole von Rotor und Stator ziehen sich gegenseitig an. B. Die gleichnamigen Pole von Rotor und Stator ziehen sich an, die ungleichnamigen stoßen sich ab. C. Sowohl die gleichnamigen, wie die ungleichnamigen Pole von Rotor und Stator stoßen sich gegenseitig ab. D. Die gleichnamigen Pole von Rotor und Stator stoßen sich gegenseitig ab, die ungleichnamigen ziehen sich an.
Le1 174	Wie kannst du mit Hilfe eines Dauermagneten und eines Elektromagneten elektrische Energie in Bewegungsenergie umwandeln? Verdeutliche deine Idee mit einer Skizze.
175 Do514	Was sind ferromagnetische Stoffe? Gib Beispiele.

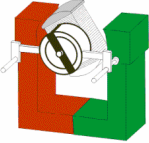
<i>Item-Code</i>	<i>Aufgabenstellung</i>
Do515a 176	Vervollständige die folgenden Sätze: Jeder Magnet hat einen _____ Pol und einen _____ Pol. Bringt man zwei Magneten zueinander, so stellt man fest: Gleichnamige Pole _____, ungleichnamige Pole _____. Die beiden bekanntesten Magneten sind der _____ und der _____. Auch die Erde besitzt ein Magnetfeld. Der magnetische Südpol befindet sich am _____, wohingegen der magnetische Nordpol am _____ liegt. Wie kann man die letzte Aussage überprüfen?
177 Do5115	Was passiert, wenn man einen Magneten teilt?
Dü8 178	Kreuze an, welche der folgenden Aussagen zur Energie falsch ist. A. Man kann verschiedene Energieformen unterscheiden. B. Energie kann aus dem Nichts gewonnen werden, indem sie einfach hergestellt wird. C. Energie kann von einer Form in eine andere Energieform umgewandelt werden. D. Energie kann bei der Umwandlung nicht verringert oder vermehrt werden.
179 Bo5b	Welche anderen Geräte kennst du, die von Elektromotoren betrieben werden?
Se3 180	Welches Material gleicher Masse erreicht beim Erhitzen am schnellsten denselben Temperaturanstieg? A. Aluminium B. Glas C. Eisen D. Holz
Se9 181	Drei gleich große Zylinder aus Aluminium, Eisen und Messing hängen an Fäden in siedendem Wasser. Die auf 100°C erwärmten Metallzylinder werden dann auf eine Paraffinplatte aufgesetzt. Das Paraffin schmilzt, die Zylinder sinken ein. Welcher sinkt am tiefsten ein? Begründe deine Antwort.


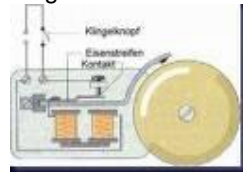
Item-Code	Aufgabenstellung
Se11a	In einem Becherglas wurden 200g Flüssigkeit erwärmt. Das Diagramm zeigt den experimentell ermittelten Zusammenhang. Interpretiere das Diagramm!
182	
183 Se11b	Berechne für die Flüssigkeit die spezifische Wärmekapazität. Entnimm die notwendigen Daten dem Diagramm.
184 Dd2	Nenne mehrere Möglichkeiten, wodurch die Spannung, die in einer Spule induziert wird, vergrößert werden kann.
185 Do512	Wie ordnen sich die Feldlinien bei einem Hufeisenmagnet an?
186 Do513	Erkläre was homogene und inhomogene Magneten sind? Gib jeweils ein Beispiel.
Do519 187	Was passiert, wenn du an einen Stabeisenmagnet immer einen Nagel unter den anderen hängst, bis kein weiterer mehr hängen bleibt und du dann vorsichtig den obersten Nagel vom Magnet löst?
188 Do542	Was ist der Unterschied zwischen Wechsel- und Gleichspannung?
We8 189	Erläutere am Beispiel der Fahrradbeleuchtung mit einem Dynamo den Zusammenhang zwischen der elektrischen Spannung und der Stromstärke im Stromkreis.
We10 190	Wie bezeichnet man die Beziehung zwischen der Veränderung der Stromstärke, und des Widerstandes in einem Stromkreis, wenn der Widerstand um 2/3 verringert wird? A. proportional B. linear C. konstant D. antiproportional
191 Moe6b	Welche Spannung hat die Quelle?
192 Moe6c	Wie groß ist die Stromstärke, die durch die Widerstände fließt?
Wu9b 193	Die Drehung einer „Spule“ mit einer einzigen Windung lässt sich mit Hilfe der Lorentzkraft erklären. Damit die „Spule“ rotiert, muss die Stromrichtung umgekehrt werden, wenn die „Spule“ senkrecht steht. Begründe, warum dies so ist!


Item-Code	Aufgabenstellung
Wi3 194	Erkläre den unterschiedlichen Kraftaufwand beim Ziehen eines Schlittens über Schnee und über Asphalt mit den Begriffen Haftreibung und Gleitreibung.
195 Wi7	Im Schwimmbad rutschst du auf den nassen Fliesen sehr leicht aus. Warum ist das so?
Wi10	<p>Welche der folgenden Aussagen ist richtig?</p> <p>A. Je größer die Reibung zwischen Gegenstand und Boden ist, desto leichter ist es den Gegenstand zu bewegen.</p> <p>B. Je geringer die Reibung zwischen Gegenstand und Boden ist, desto leichter ist es den Gegenstand zu bewegen.</p> <p>C. Es hängt nur vom Gewicht des Gegenstandes ab, wie leicht er zu bewegen ist. Die Reibung spielt keine Rolle.</p> <p>D. Je geringer die Reibung zwischen Gegenstand und Boden ist, desto schwerer ist es den Gegenstand zu bewegen.</p>
196	
Ra12	<p>Vervollständige den Lückentext mit folgenden Begriffen:</p> <p>eigenen, einzigen, Reihenschaltung, direkt, Sicherung, Schaltungen, Teilströme, gleiche, Überlastung, Parallelschaltung</p> <p>Geräte kann man an eine Stromquelle in verschiedenen _____ anschließen.</p> <p>Bei der _____ werden alle Lampen hintereinander in einen Stromkreis eingesetzt. Diese hintereinander geschalteten Lampen und die Stromquelle bilden einen _____ Stromkreis. An jeder Stelle des Stromkreises misst man daher die _____ Stromstärke.</p> <p>Bei der _____ werden die Lampen parallel an eine Stromquelle angeschlossen. Jede Lampe ist _____ mit der Stromquelle verbunden und bildet mit ihr zusammen einen _____ Stromkreis. Die Ströme, die in den einzelnen Stromkreisen fließen, heißen _____ und deren Summe ergibt die Gesamtstromstärke.</p> <p>Wenn zu viele Geräte gleichzeitig parallel in einem Stromkreis geschaltet werden, kann es zu _____ des Stromkreises kommen. Um die dadurch entstehende Brandgefahr zu mindern, muss der Stromkreis durch eine _____ rechtzeitig unterbrochen werden.</p>
197	
198 HW9	Ein Gartenschlauch mit offenem Ende liegt ausgerollt auf der Wiese. Was geschieht, wenn der Wasserzulauf aufgedreht wird?
Lü5	Erkläre den Aufbau und die Funktionsweise des abgebildeten Lasthebemagneten.
199	

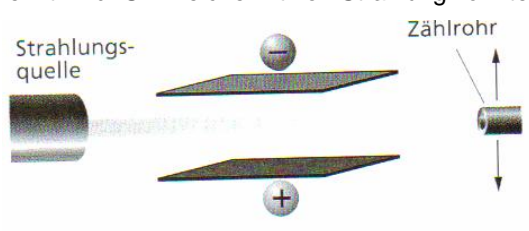
Item-Code	Aufgabenstellung
Lü8	Welche der folgenden Aussagen ist richtig? A. Je größer die Windungszahl, desto schwächer ist das Elektromagnetfeld. B. Je kleiner die Windungszahl, desto stärker ist das Elektromagnetfeld. C. Je größer die Windungszahl, desto stärker ist das Elektromagnetfeld. D. Die Stärke des Elektromagnetfeldes ist unabhängig von der Windungszahl.
200	
Lü12	Vervollständige den Text mit folgenden Begriffen: <i>magnetisch, ziehen sich an, Hufeisenmagnet, Süd, geographischer Nordpol, stoßen sich ab, Magnetfeld, Nord, geographischer Südpol, Magneten, Stabmagnet</i> Jeder Magnet hat einen _____ Pol und einen _____ Pol. Bringt man zwei Magneten zueinander, so stellt man fest: Gleichnamige Pole _____, ungleichnamige Pole _____. Die beiden bekanntesten Magneten sind der _____ und der _____. Auch die Erde besitzt ein Magnetfeld. Der magnetische Südpol befindet sich am _____, wohingegen der magnetische Nordpol am _____ liegt. Wickelt man einen Leiter zu einer Spule und lässt einen Strom fließen, so wird es _____. Die Lage der Pole hängt vom _____ ab. Das Magnetfeld der Spule kann durch einen Eisenkern im Inneren verstärkt werden. Das Magnetfeld der Spule kann man mit dem eines _____ vergleichen.
201	

Anhang G: Aufgaben in Studie 2

	<i>Item-Code</i>	<i>Aufgabenstellung</i>
1	Bo1	Welche Funktion hat der Kommutator (Stromwender) bei dem Elektromotor?
2	Bo2	Welche der folgenden Aussagen ist richtig? A. Jedes Magnetfeld hat einen Plus- und einen Minuspol. B. Der Nordpol wird in Abbildungen immer grün und der Südpol immer rot dargestellt. C. Jedes Magnetfeld hat einen Nord- und einen Südpol. D. Ungleiche Pole stoßen sich gegenseitig ab, gleiche Pole ziehen sich an.
3	Do5113	Ein stromdurchflossener Leiter wirkt wie ein Magnet. Besonders stark ist die magnetische Wirkung, wenn der Leiter als Spule aufgewickelt ist und einen Eisenkern enthält. Schaltet man den Strom ab, hört die magnetische Wirkung auf. Große Elektromagnete werden genutzt, um schwere Lasten zu heben und zu transportieren. Wie muss solch ein Lasthebemagnet aufgebaut sein? Wie kann man seine Wirkungsweise erklären?
4	L1	Wie kannst du mit Hilfe eines Dauermagneten und eines Elektromagneten elektrische Energie in Bewegungsenergie umwandeln? Verdeutliche deine Idee mit einer Skizze.
5	Bo4a	Beschrifte in der Abbildung die wichtigsten Bestandteile des Elektromotors mit Hilfe der folgenden Begriffe: Kommutator, Anker mit Drehspule, Feldmagnet, Schleifkontakt 
6	Bo4b	Erkläre die Funktionsweise eines Elektromotors.

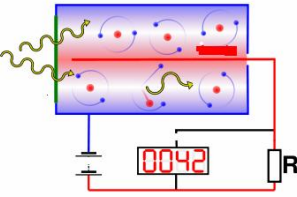
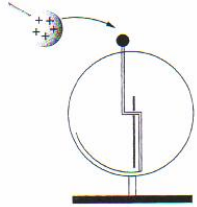
	Item-Code	Aufgabenstellung
7	Bo7	<p>In manchen Motoren findet man einen Anker mit drei Spulen. Verfolge den Stromkreis und erläutere, wie sich dieser Anker bewegt. Welchen Vorteil hat er gegenüber einem Anker mit nur zwei gegenüberliegenden Spulen?</p> 
8	Bo9	<p>Welche der folgenden Aussagen ist im Zusammenhang mit einem Elektromotor richtig?</p> <p>A. Sowohl die gleichnamigen und die ungleichnamigen Pole von Rotor und Stator ziehen sich gegenseitig an.</p> <p>B. Die gleichnamigen Pole von Rotor und Stator ziehen sich an, die ungleichnamigen stoßen sich ab.</p> <p>C. Sowohl die gleichnamigen, wie die ungleichnamigen Pole von Rotor und Stator stoßen sich gegenseitig ab.</p> <p>D. Die gleichnamigen Pole von Rotor und Stator stoßen sich gegenseitig ab, die ungleichnamigen ziehen sich an.</p>
9	Bo5a	<p>Eine elektrische Klingel hat den in der Abbildung dargestellten Aufbau. Beschreibe den Aufbau und erkläre die Wirkungsweise einer solchen Klingel! Nutze dafür dein Wissen über Elektromotoren.</p> 
10	L2	<p>Durch welche Maßnahmen lässt sich die von einem Generator erzeugte Spannung steigern?</p>

	Item-Code	Aufgabenstellung
11	Bo10a	<p>Die unten stehenden Bilder zeigen einen Elektromotor mit verschiedenen Stellungen des Rotors. Erkläre, warum sich der Rotor jeweils weiterdreht. Kennzeichne die Pole des Elektromagneten in allen drei Bildern.</p> 
12	Bo10b	Wenn der Rotor waagrecht steht, fließt kein Strom durch ihn. Welches Problem tritt dadurch auf?
13	L31	Erkläre den Aufbau eines Atoms.
14	L32	<p>Was geschieht, wenn ein Atom ein zusätzliches Elektron in seine Hülle aufnimmt?</p> <p>A. Das Atom gibt dafür ein Proton ab. B. Das Atom wird zum positiven Ion. C. Das Atom gibt dafür ein Neutron ab. D. Das Atom wird zum negativen Ion.</p>
15	L33	Welche Informationen über den Aufbau der Atome geben die Massen- und Kernladungszahl?
16	L34	Beschreibe den Aufbau und die Funktionsweise eines Geiger-Müller-Zählrohrs.
17	L35	Welche Experimente geben Ausschluss über den Aufbau von Atomen? Beschreibe ein dir bekanntes Experiment.
18	L36	Beschreibe die wesentlichen Ergebnisse des Rutherford'schen Streuversuchs und erläutere, welche Schlussfolgerungen Rutherford für den Atomaufbau daraus ziehen konnte.
19	L37	Wie hängen die Reichweite und die Durchdringungsfähigkeit radioaktiver Strahlung zusammen?

	Item-Code	Aufgabenstellung
20	L38	<p>Was geschieht, wenn aus der Hülle eines Atoms ein Elektron entfernt wird?</p> <p>A. Das Atom wird zum positiven Ion. B. Das Atom zerfällt. C. Das Atom wird zum negativen Ion. D. Das Atom bleibt neutral.</p>
21	L39	<p>Radioaktive Strahlung wird zwischen zwei elektrisch geladenen Platten hindurchgelenkt. Mit einem Zählrohr stellt man fest, dass sie abgelenkt wird. Um welche Art von Strahlung könnte es sich hierbei handeln? Begründe deine Antwort bitte.</p> 
22	L310	Wie können auch solche Atome ionisiert werden, die z.B. von den α -Teilchen nicht getroffen werden?
23	L311	Warum musste Rutherford seinen Streuversuch in einer abgedichteten, luftleer gepumpten Apparatur durchführen?
24	L312	<p>Wodurch unterscheiden sich die Atomkerne verschiedener chemischer Elemente voneinander?</p> <p>A. Durch die Elektronenzahl B. Durch die Massenzahl C. Durch die Kernladungszahl D. Durch die Neutronenzahl.</p>
25	L313	Welche Arten der radioaktiven Strahlung gibt es? Worin unterscheiden sich diese?
26	L51	Beschreibe den Aufbau eines Heliumatoms, eines Kohlenstoffatoms und eines Eisenatoms. Worin unterscheiden sich diese Atome?

	Item-Code	Aufgabenstellung
27	L52	Durch die Massenzahl wird angegeben, wie viele A: Neutronen B. Protonen C. Neutronen und Protonen D. Elektronen ein Atomkern enthält.
28	L53	Wie kann man nachweisen, dass radioaktive Strahlung Luft ionisiert?
29	L54	Beschreibe den Aufbau und die Funktionsweise einer Nebelkammer.
30	L55	Erkläre mit Hilfe einer Skizze den Aufbau und die Durchführung des Rutherford'schen Streuversuchs.
31	L56a	Wie lässt sich die Reichweite einer radioaktiven Strahlung nachweisen?
32	L56b	Wie hängen Reichweite und Durchdringungsfähigkeit der Strahlung zusammen?
33	L57	Warum unterscheidet sich die Masse eines Stickstoffatoms nur unwesentlich von der Masse eines Stickstoffkerns?
34	L58	Wie viele Wasserstoffatome würden aneinander gereiht eine Strecke von etwa 1mm ergeben? Etwa A: 1.000 Wasserstoffatome. B. 10.000 Wasserstoffatome. C. 10.000.000 Wasserstoffatome. D. 100.000.000 Wasserstoffatome.
35	L59	Woran lagern sich die Flüssigkeitströpfchen bei einer Nebelspur in einer Nebelkammer bevorzugt ab?
36	L510	Aus wie vielen Protonen, Neutronen und Elektronen besteht ein neutrales Aluminiumatom, dessen Kern durch das Symbol ${}_{13}^{27}\text{Al}$ gekennzeichnet ist? Wie lauten die entsprechenden Zahlen für ${}_{83}^{209}\text{Bi}$ und ${}_{92}^{235}\text{U}$?

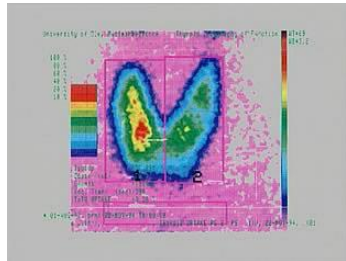
	Item-Code	Aufgabenstellung
37	L511	Die Kernladungszahl gibt die A. Anzahl der Neutronen B. Anzahl der Protonen C. Anzahl der Elektronen D. Anzahl der Protonen und Neutronen an.
38	L512	Welche Arten der radioaktiven Strahlung gibt es?
39	L513	Erkläre den Begriff der Ionisation.
40	L81	Woraus setzt sich ein Atom zusammen?
41	L82	Wie nennt man die kleinsten Teilchen eines chemischen Elements? A. Moleküle B. Protonen C. Atome D: Elektronen
42	L83	Woraus besteht die Nebelspur in einer Nebelkammer?
43	L84	Beschreibe mit Hilfe der Abbildung den Rutherford'schen Streuversuch. 

	<i>Item-Code</i>	<i>Aufgabenstellung</i>
44	L85	<p>Beschrifte die Abbildung des Geiger-Müller-Zählrohrs mit den angegebenen Begriffen: Elektronen, isolierte Durchführung, Glimmerfenster, Draht, ionisierte Gasteilchen, Gas, radioaktive Strahlung, Metallgehäuse</p> 
45	L86a	<p>Eine positiv geladene Kugel wird einem Elektroskop genähert. Was kann man beobachten, wenn man die Kugel dem Elektroskop nähert bzw. davon entfernt?</p> 
46	L86b	Erkläre die Beobachtung.
47	L86c	Was kann man beobachten, wenn die Kugel negativ geladen ist?
48	L87	<p>Beschreibe ein einfaches Atommodell: Ein Atom besteht aus _____. Um den Kern bewegen sich _____. Das Volumen des Kerns ist sehr viel _____.</p>

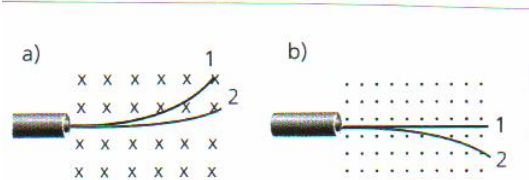
	Item-Code	Aufgabenstellung
49	L88	<p>Woraus besteht der Kern eines Atoms?</p> <p>A. Elektronen und Neutronen. B. Protonen und Elektronen. C. Protonen und Neutronen. D. Ausschließlich aus Protonen.</p>
50	L89	<p>Vervollständige den Lückentext mit folgenden Begriffen:</p> <p>geladenen, Elektronen, α-Strahlung, Isotope, γ-Strahlung, Neutronen, Becquerel, Abschirmbarkeit, Protonen, β-Strahlung</p> <p>Atomkerne bestehen aus _____ und _____, die Hülle des Atoms wird von _____ gebildet. Atome eines Elements, die sich nur in der Anzahl der Neutronen unterscheiden, heißen _____.</p> <p>Es gibt verschiedene radioaktive Strahlungsarten: _____ besteht aus zwei Neutronen und zwei Protonen, _____ besteht aus Elektronen. Diese Strahlungsarten bestehen aus _____ Teilchen. Die _____ ist eine Energiestrahlung und wird z. B. bei Röntgenaufnahmen eingesetzt. Die Strahlungsarten unterscheiden sich in ihrer Reichweite und _____.</p> <p>Als Aktivität eines radioaktiven Stoffes wird die Anzahl der Kernumwandlungen in einer Sekunde bezeichnet. Die Einheit der Aktivität ist _____.</p>

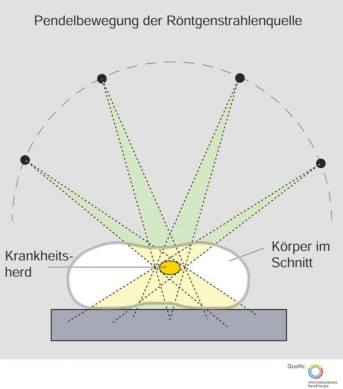
	Item-Code	Aufgabenstellung																																			
51	L810	<p>Aus der Symbolschreibweise für Atomkerne kann man für jeden einzelnen Kern die Protonen- und Neutronenzahl sowie die Massenzahl und Ordnungszahl entnehmen. Vervollständige die Übersicht:</p> <table border="1"> <thead> <tr> <th>Kern</th> <th>Protonenzahl</th> <th>Neutronenzahl</th> <th>Massenzahl</th> <th>Ordnungszahl</th> </tr> </thead> <tbody> <tr> <td>${}^1_1\text{H}$</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>${}^4_2\text{He}$</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>${}^{12}_6\text{C}$</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>${}^{23}_{11}\text{Na}$</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>${}^{35}_{17}\text{Cl}$</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>${}^{56}_{26}\text{Fe}$</td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Kern	Protonenzahl	Neutronenzahl	Massenzahl	Ordnungszahl	${}^1_1\text{H}$					${}^4_2\text{He}$					${}^{12}_6\text{C}$					${}^{23}_{11}\text{Na}$					${}^{35}_{17}\text{Cl}$					${}^{56}_{26}\text{Fe}$				
Kern	Protonenzahl	Neutronenzahl	Massenzahl	Ordnungszahl																																	
${}^1_1\text{H}$																																					
${}^4_2\text{He}$																																					
${}^{12}_6\text{C}$																																					
${}^{23}_{11}\text{Na}$																																					
${}^{35}_{17}\text{Cl}$																																					
${}^{56}_{26}\text{Fe}$																																					
52	L811	<p>Welche Teilchen bilden die Hülle des Atoms?</p> <p>A. Neutronen. B. Protonen. C. Elektronen. D. Protonen und Neutronen.</p>																																			
53	L321	<p>Welche Eigenschaften und Wirkungen der radioaktiven Strahlung werden in der Medizin genutzt? Welche Chancen, aber auch welche Risiken hat diese Nutzung?</p>																																			
54	L322	<p>Die Intensität der kosmischen Strahlung ist von der Höhe über dem Meeresspiegel abhängig. Sie ...</p> <p>A. steigt mit zunehmender Höhe. B. sinkt mit zunehmender Höhe. C. steigt und sinkt abwechselnd alle 1000 m. D. steigt bis 1000m und bleibt dann gleich.</p>																																			

	Item-Code	Aufgabenstellung												
55	L323	Wie kann man künstliche Radionuklide herstellen? Wodurch unterscheiden sich diese von natürlichen Radionukliden?												
56	L324	Was versteht man unter der effektiven Dosis bzw. der Äquivalentdosis?												
57	L325	Wie kannst du feststellen, ob die Leuchtziffern einer Uhr radioaktiv strahlen. Beschreibe zwei Möglichkeiten.												
58	L326a	<p>Bei einem Barium-137-Präparat wurde mit einem Zählrohr in 30 s eine Zählrate von 2070 Impulsen gemessen. Nach jeweils zwei Minuten wurden die Messungen wiederholt. Dabei erhielt man folgende Ergebnisse:</p> <table border="1" data-bbox="793 553 1608 643"> <tbody> <tr> <td>Zeit in min.</td> <td>2</td> <td>4</td> <td>6</td> <td>8</td> <td>10</td> </tr> <tr> <td>Zählrate in Impulsen</td> <td>1100</td> <td>650</td> <td>416</td> <td>241</td> <td>179</td> </tr> </tbody> </table> <p>Stelle die Messwerte graphisch dar.</p>	Zeit in min.	2	4	6	8	10	Zählrate in Impulsen	1100	650	416	241	179
Zeit in min.	2	4	6	8	10									
Zählrate in Impulsen	1100	650	416	241	179									
59	L326b	Ermittle aus dem Diagramm die Halbwertszeit von Barium-137.												
60	L327	<p>Die Organe des Menschen werden durch die natürliche Strahlung unterschiedlich stark belastet. Die stärkste Belastung ergibt sich für</p> <p>A. die Keimdrüsen. B. die Knochen. C. die Lunge. D. die Haut.</p>												
61	L328	Welchen Grund könnte es haben, dass Bäume an verkehrsreichen Straßen nach der C-14-Methode ein Alter von etwa 500 Jahren haben?												
62	L329	<p>Die zivilisatorisch bedingte Strahlenexposition wird im Wesentlichen durch</p> <p>A. das Fernsehen B. die Kernkraftwerke C. die medizinische Strahlenanwendungen D. den Flugverkehr verursacht.</p>												
63	L3210	Warum müssen die Keimzellen des Menschen besonders vor radioaktiver Strahlung geschützt werden?												

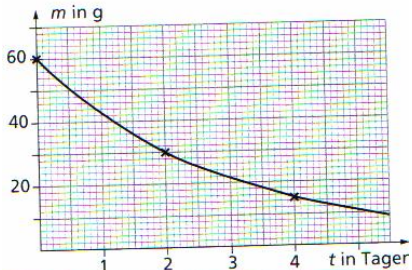
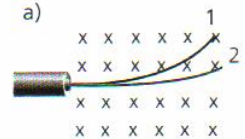
	Item-Code	Aufgabenstellung
64	L311a	<p>Die Abbildung zeigt das Ergebnis einer Schilddrüsen- Szintigrafie. Erkläre anhand der Abbildung die Wirkungsweise der Szintigrafie und interpretiere das Ergebnis.</p> 
65	L311b	Warum verwendet man bei der Szintigrafie γ -Strahler und nicht β - oder α -Strahler?
66	L3212	Was geschieht beim α -Zerfall bzw. β -Zerfall? Wie kommt es zur γ -Strahlung? Gib jeweils die zugehörigen Reaktionsgleichungen des Zerfalls an.
67	L51a	In einem Klassenraum wird an einem Tag eine Nullrate von 20 Impulsen/min gemessen. An einem anderen Tag in einer anderen Schule stellt man mit demselben Zählgerät eine Nullrate von 18 Impulsen/min fest. Was bedeutet die Nullrate?
68	L521b	Welche Gründe sind für die unterschiedlichen Nullraten denkbar?
69	L522	<p>Was wird durch die Energiedosis angegeben?</p> <p>A. Absorbierte Strahlungsenergie. B. Absorbierte Strahlungsenergie pro Zeit. C. Absorbierte Strahlungsenergie pro Masse. D. Absorbierte Strahlungsenergie pro Masse multipliziert mit dem Wichtungsfaktor.</p>
70	L523	Radioaktive Stoffe werden in der Medizin in verschiedenen Anwendungen eingesetzt. Eine davon ist die Szintigrafie. Erkläre ihre Wirkungsweise.

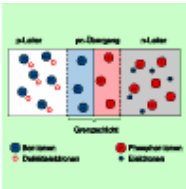
	Item-Code	Aufgabenstellung												
71	L524a	<p>In einem Versuch wird experimentell untersucht, wie die Strahlung eines radioaktiven Präparats durch Aluminiumfolie absorbiert wird. Dazu wird zunächst der Nulleffekt mit 20 Impulsen je Minute ermittelt. Anschließend werden die Messungen mit dem radioaktiven Präparat und Aluminiumfolie zwischen Präparat und Zählrohr durchgeführt. Dabei erhält man folgendes Ergebnis:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>Dicke der Aluminiumfolie in mm</td> <td>0</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> </tr> <tr> <td>Zählrate in Impulsen je Minute</td> <td>650</td> <td>305</td> <td>210</td> <td>185</td> <td>170</td> </tr> </table> <p>Erhöht man die Schichtdicke des Aluminiums weiter, so verändert sich die Zählrate nicht. Stelle den Zusammenhang zwischen Zählrate und Dicke der Aluminiumschicht in einem Diagramm dar. Interpretiere das Ergebnis.</p>	Dicke der Aluminiumfolie in mm	0	1	2	3	4	Zählrate in Impulsen je Minute	650	305	210	185	170
Dicke der Aluminiumfolie in mm	0	1	2	3	4									
Zählrate in Impulsen je Minute	650	305	210	185	170									
72	L524b	Wie ist es zu erklären, dass sich die Zählrate bei Erhöhung der Schichtdicke nicht verändert?												
73	L525a	Nenne einige Strahlenbelastungen, denen der Mensch ständig ausgesetzt ist.												
74	L525b	Welche Maßnahmen zum Strahlenschutz kennst du?												
75	L526	Erläutere die Unterteilung von Strahlenschäden in „somatische“ und „genetische“ Schäden sowie in Früh- und Spätschäden. Nenne jeweils einige Beispiele.												
76	L527	<p>Wodurch wird die terrestrische Strahlung verursacht?</p> <p>A. Durch besonders hartes Gestein. B. Durch die Erdmaterie. C. Durch Lava von Vulkanausbrüchen. D. Durch Radionuklide im Wasser.</p>												
77	L528	In vielen Gräbern sind Mumien zu finden. Bei einer dieser Mumien wurde festgestellt, dass der C-14-Anteil nur noch 25 % des heutigen Anteils beträgt. Auf welches Alter der Mumie kann man daraus schließen?												
78	L529	<p>Die terrestrische Strahlung</p> <p>A. ist überall auf der Erde gleich. B. tritt nur im Gebirge oberhalb 1000 m auf. C. schwankt in Abhängigkeit vom geologischen Untergrund. D. tritt nur im Wasser unterhalb von 1000 m auf.</p>												

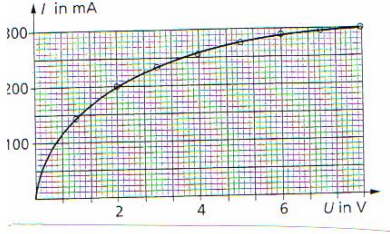
	Item-Code	Aufgabenstellung
79	L5210	Warum kann es keine Nahrung geben, die keine radioaktiven Stoffe enthält?
80	L5211	Warum ist Blei besonders gut geeignet, um Röntgenstrahlung abzuschirmen? Begründe deine Antwort.
81	L5212	<p>Radioaktive Strahlung wird jeweils durch ein Magnetfeld gelenkt (s. Skizzen a und b). Das Magnetfeld zeigt in die Blattebene hinein (a) bzw. hinaus (b). Um welche Art von Strahlung könnte es sich handeln? Begründe deine Antwort.</p> 
82	L821	Von welchen Faktoren hängt die schädigende Wirkung der radioaktiven Strahlung ab?
83	L822	Nenne einige Strahlenschutzmaßnahmen und begründe deren Funktion.
84	L823	<p>Wie nennt man die Kerne radioaktiver Atome?</p> <p>A. Radionuklide B. Isotope C. Nuklide D. Protonen</p>

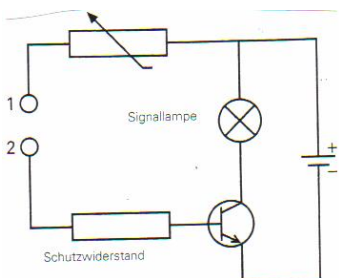
	Item-Code	Aufgabenstellung
85	L824	<p>Erkläre mit Hilfe der Abbildung das Prinzip der Röntgenbestrahlung und gehe dabei auf die Pendelbewegung ein. Welchen Vorteil hat diese Pendelbewegung?</p>  <p>Das Diagramm zeigt die Pendelbewegung einer Röntgenstrahlenquelle um einen Körper im Schnitt. Die Quelle bewegt sich in einem Halbkreis um den Körper. Strahlen gehen von der Quelle zum Körper und zurück zur Quelle. Beschriftungen: 'Pendelbewegung der Röntgenstrahlenquelle', 'Krankheitsherd', 'Körper im Schnitt'.</p>
86	L825	Nenne dir bekannte natürliche Strahlungsquellen.
87	L826	Der Nulleffekt ist praktisch überall messbar. Was versteht man darunter? Warum kann er an verschiedenen Orten unterschiedlich sein?

	<i>Item-Code</i>	<i>Aufgabenstellung</i>																						
88	L827	<p>Zwischen einen γ-Strahler und ein Zählrohr wurden unterschiedlich dicke Kupferplatten gestellt. Stelle die Impulsrate in Abhängigkeit der Schichtdicke graphisch dar und interpretiere das Diagramm.</p> <table border="1" data-bbox="961 381 1438 868"> <thead> <tr> <th>Schichtdicke d in mm</th> <th>Impulsrate 1/min</th> </tr> </thead> <tbody> <tr><td>0</td><td>869</td></tr> <tr><td>4</td><td>779</td></tr> <tr><td>8</td><td>639</td></tr> <tr><td>12</td><td>565</td></tr> <tr><td>16</td><td>477</td></tr> <tr><td>20</td><td>385</td></tr> <tr><td>24</td><td>372</td></tr> <tr><td>28</td><td>364</td></tr> <tr><td>32</td><td>318</td></tr> <tr><td>36</td><td>268</td></tr> </tbody> </table>	Schichtdicke d in mm	Impulsrate 1/min	0	869	4	779	8	639	12	565	16	477	20	385	24	372	28	364	32	318	36	268
Schichtdicke d in mm	Impulsrate 1/min																							
0	869																							
4	779																							
8	639																							
12	565																							
16	477																							
20	385																							
24	372																							
28	364																							
32	318																							
36	268																							
89	L828	<p>Der menschliche Körper besitzt eine Eigenstrahlung. Sie tritt auf, weil</p> <p>A. der Mensch durch die kosmische Strahlung dauernd bestrahlt wird.</p> <p>B. natürliche radioaktive Elemente mit der Nahrung und der Atmung in den Körper aufgenommen werden.</p> <p>C. die Haut des Menschen durch radioaktiven Staub verunreinigt ist.</p> <p>D. der Mensch durch die Kernkraftwerke dauernd radioaktiver Strahlung ausgesetzt ist.</p>																						
90	L829	<p>Zu Beginn eines Versuchs sind 12.000.000 Atome radioaktiv. Wie viele Atome sind nach Ablauf zweier Halbwertszeiten insgesamt verfallen?</p> <p>A. 6.000.000 Atome.</p> <p>B. 3.000.000 Atome.</p> <p>C. 9.000.000 Atome.</p> <p>D. 12.000.000 Atome.</p>																						
91	L8210	<p>Nenne einige radioaktiven Stoffe, die der Mensch mit der Nahrung aufnimmt und erkläre, wie diese in die Nahrung gelangen.</p>																						

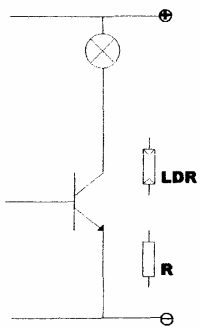
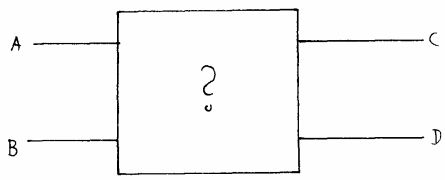
	Item-Code	Aufgabenstellung
92	L8211	<p>Für ein Radionuklid gilt die im Diagramm dargestellte Zerfallskurve. Wie groß ist die Halbwertszeit? Erkläre dabei (auch) den Begriff Halbwertszeit.</p> 
93	L8212	<p>Die Abbildung zeigt radioaktive Strahlung im Magnetfeld. Um welche Strahlungsarten handelt es sich dabei? Erkläre die Richtung der Ablenkung der verschiedenen Strahlungen und ihre Ladungen.</p> 
94	We7	<p>Welche der folgenden Aussagen zum Zusammenhang zwischen der Anzahl der Glühbirnen und dem Widerstand in einem Stromkreis ist richtig?</p> <p>A. Je größer die Zahl der Glühbirnen ist, desto größer wird der Gesamtwiderstand. B. Je kleiner die Zahl der Glühbirnen ist, desto größer wird der Gesamtwiderstand. C. Je größer die Zahl der Glühbirnen ist, desto kleiner wird der Gesamtwiderstand. D. Der Gesamtwiderstand verändert sich nicht durch die Anzahl der Glühbirnen.</p>

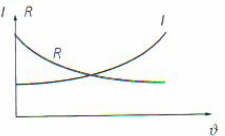
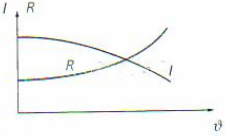
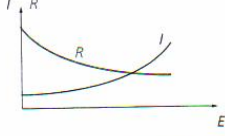
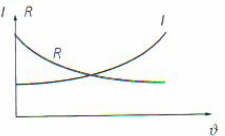
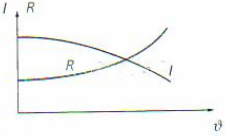
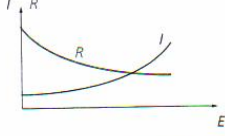
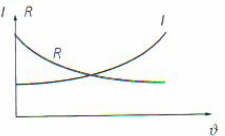
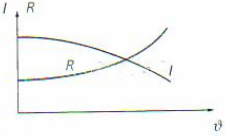
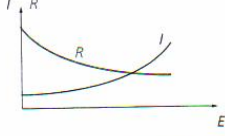
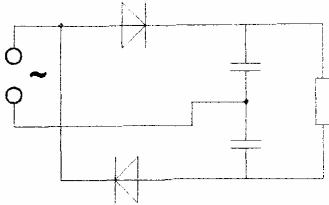
	Item-Code	Aufgabenstellung
95	We10	Wie bezeichnet man die Beziehung zwischen der Veränderung der Stromstärke und des Widerstandes in einem Stromkreis, wenn der Widerstand um $\frac{2}{3}$ verringert wird? A. Proportional B. Linear C. Konstant D. Antiproportional
96	Mor6a	Am Widerstand mit 50Ω liegt eine Spannung von 12 V an. Wie groß ist die am anderen Widerstand anliegende Spannung?
97	Moe6b	Welche Spannung hat die Quelle? Begründe deine Antwort.
98	Moe6c	Wie groß ist die Stromstärke, die durch die Widerstände fließt?
99	Ma1	Vervollständige den Lückentext mit folgenden Begriffen: Pluspol, Kollektor, Anschlüsse, Dioden, Basis, Transistor, Elektronenstrom, Emitter, Durchlass Ein _____ hat drei Anschlüsse. Sie heißen Emitter, Basis und _____. Ein Transistor besteht eigentlich aus zwei _____. Beim Transistor muss die Teildiode Emitter- Basis in _____ - Richtung geschaltet werden, dies ist gegeben, wenn an der _____ der Pluspol der Batterie anliegt. Die Prüfdiode leuchtet dann auf. Entsprechend lässt sich auch ein _____ zwischen Basis und Kollektor nachweisen, wenn die Basis wieder am _____ der Batterie liegt. Zwischen _____ und Kollektor leuchtet die Prüfdiode nicht, gleichgültig wie die beiden _____ gepolt sind.
100	Ma3	Warum ist ein dotierter Halbleiter nach außen hin elektrisch neutral?
101	Ma4	Erkläre mit Hilfe der Abbildung den Bau einer Halbleiterdiode und den Leitungsvorgang, der in ihr stattfindet. Beschreibe dabei auch die Vorgänge in der Grenzschicht, einmal wenn die Diode in Sperrrichtung und einmal in Durchlassrichtung geschaltet ist. 
102	Ma5	Wie entstehen p-Leiter und n-Leiter?

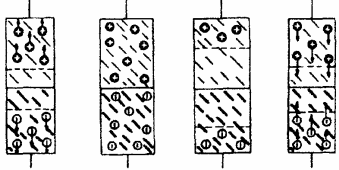
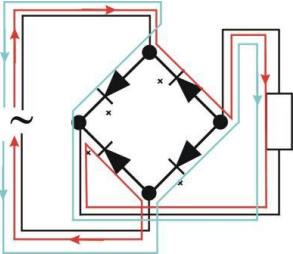
	Item-Code	Aufgabenstellung
103	Ma7	Einer von zwei äußerlich nicht zu unterscheidenden Festkörperwiderständen wurde aus einem Halbleitermaterial, der andere aus einem metallischen Material hergestellt. Wie kannst du ermitteln, welcher der beiden Widerstände aus dem Halbleitermaterial gefertigt wurde?
104	Ma8a	Für ein elektrisches Bauelement wurde die nachfolgend dargestellte I-U-Kennlinie aufgenommen. Interpretiere das Diagramm. 
105	Ma8b	Um was für ein Bauelement könnte es sich handeln? Begründe deine Vermutung.
106	Ma9a	Ein elektrisches Fieberthermometer wird mit einer kleinen Knopfzelle betrieben, die für eine konstante Spannung sorgt. Dabei dient ein Thermistor (temperaturabhängiger Halbleiterwiderstand) als Messfühler. Entwirf den Schaltplan für ein das Fieberthermometer.
107	Ma9b	Was für eine Art Thermistor kann für die Spitze des Thermometers verwendet werden? Begründe deine Antwort.
108	Ma9c	Erkläre die Wirkungsweise dieses Fieberthermometers.
109	Ma11	Drei Silicium-Dioden sind in Reihe geschaltet. Welche Spannung ist ungefähr nötig, damit Strom fließen kann? Erkläre dabei auch die Schleusenspannung.
110	Wu2	Wie heißen die Anschlüsse des Transistors?
111	Wu3	Wo müssen beim Transistor Plus- und Minuspol des Stroms angeschlossen werden, damit Strom fließen kann?

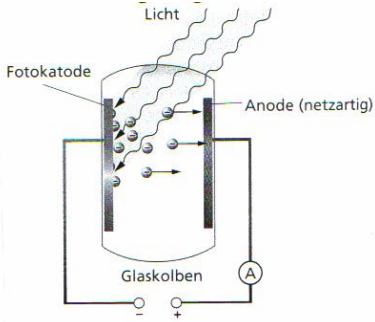
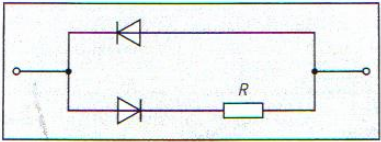
	Item-Code	Aufgabenstellung
112	Wu4	<p>Die Schaltskizze zeigt die Transistorgrundschaltung, zwischen den beiden Anschlussstellen 1 und 2 wurde ein Sensor eingebaut. Beschreibe die Schaltskizze und erkläre die Funktionsweise des Transistors.</p> 
113	Wu5	<p>Welche der folgenden Aussagen ist richtig?</p> <p>A. Ob Strom durch einen Transistor fließt oder nicht, ist unabhängig vom Widerstand.</p> <p>B. Ist der Widerstand eines Sensors klein, schaltet der Transistor durch.</p> <p>C. Ist der Widerstand eines Sensors klein, sperrt der Transistor und es fließt kein Strom mehr.</p> <p>D. Ist der Widerstand eines Sensors groß, schaltet der Transistor durch und es fließt Strom.</p>
114	Wu6	<p>Was bedeutet es, wenn ein Transistor „sperrt“ bzw. „leitet“?</p>
115	Wu7a	<p>Was passiert mit der Helligkeit einer Leuchtdiode, wenn der Widerstand in einer Transistorschaltung verändert wird?</p>
116	Wu7b	<p>Was passiert mit der Diode, wenn mehrere Widerstände hintereinander geschaltet werden?</p>
117	Wu8	<p>Nenne Geräte, bei denen Leuchtdioden oder Transistoren eingebaut sind.</p>

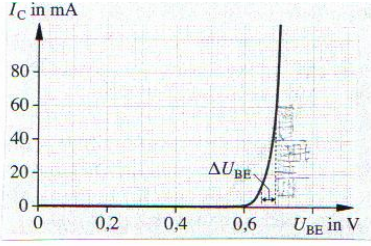
	Item-Code	Aufgabenstellung
118	Wu9	<p>Welche der folgenden Aussagen ist richtig? Am größten Widerstand in einer Spannungsteilerschaltung eines Transistors misst man....</p> <p>A. die kleinste Teilspannung. B. die größte Teilspannung. C. dieselbe Teilspannung, wie bei den anderen Widerständen. D. eine Teilspannung, deren Höhe unabhängig vom Widerstand ist.</p>
119	Wu10	Welches sind die zwei möglichen Aufgaben eines Transistors?
120	Wu11	Welche Vorteile hat ein Transistor als Schalter? Und welche Nachteile hat er?
121	Wu12	<p>Vervollständige den folgenden Satz zu Halbleitern mit den folgenden Begriffen: Störstellenleitung, Elemente, n- Halbleitern, Isolatoren, Außenelektronen, Halbleiters, Löchern, Leitfähigkeit, Dotieren</p> <p>Halbleiter sind Stoffe, deren elektrische _____ zwischen der von Leitern und _____ liegt. Die Leitfähigkeit eines _____ kann gezielt erhöht werden, wenn man Atome anderer _____ (Fremdatome) einbringt, die mehr oder weniger _____ haben als die Halbleiter. Diesen Vorgang nennt man _____. Beim Dotieren entstehen Störstellen mit freien Elektronen oder _____. Die darauf basierende Leitung wird _____ genannt. Je nach Dotierung unterscheidet man zwischen _____ und p-Halbleitern.</p>
122	G1	Zwischen den beiden Platten eines Plattenkondensators liegt eine Spannung von 22,5 V an und eine Ladung von $6 \cdot 10^{-8}$ Q. Berechne die Kapazität C des Kondensators.
123	G2a	Bei Blitzgeräten von Fotoapparaten werden Kondensatoren als Energiespeicher genutzt. Wie groß ist die im Kondensator gespeicherte und beim Auslösen des Blitzes frei werdende elektrische Ladung, wenn der eingebaute Kondensator eine Kapazität von 100 µF hat und die Ladespannung 6 V beträgt?
124	G2b	Warum kann ein zweiter Blitz erst nach einer bestimmten Zeit ausgelöst werden?
125	G4b	Bestimme den elektrischen Widerstand des Bauelements für 15 V und 30 V.
126	G5	Was versteht man unter einem Spannungsteiler? Erkläre deine Antwort.
127	G6a	Zwei Widerstände mit 50 Ω und mit 200 Ω sind mit einer elektrischen Quelle in Reihe geschaltet. Zeichne den dazugehörigen Schaltplan.
128	G7	Erkläre den Aufbau und die Funktionsweise eines Kondensators.

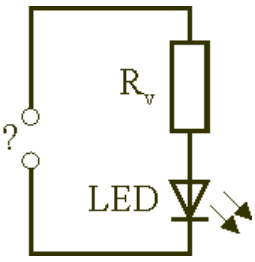
	Item-Code	Aufgabenstellung
129	G8	Vergleiche das unterschiedliche Verhalten eines Kondensators in einem Gleichstrom- und in einem Wechselstromkreis.
130	G131	Erkläre mit Hilfe einer Skizze den Aufbau einer Halbleiterdiode und beschreibe den Leitungsvorgang. Erkläre die Vorgänge in der Grenzschicht, einmal wenn die Diode in Sperrrichtung und einmal, wenn die Diode in Durchlassrichtung geschaltet ist.
131	G132	Erkläre die Funktionsweise eines Transistors.
132	G133	Ergänze die nebenstehende Schaltung jeweils so, dass einmal eine Dunkel- und einmal eine Hellschaltung entstehen. 
133	G134	In einer Black-Box ist eine Schaltung versteckt, die aus einer Wechselspannung an A und B eine Gleichspannung an C und D machen soll; an C soll der Pluspol liegen. Nenne drei Lösungen. 

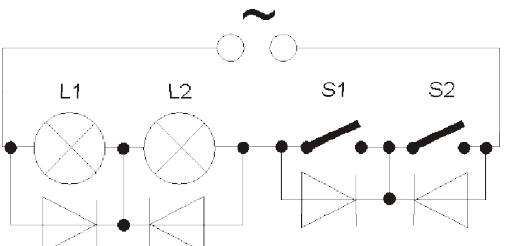
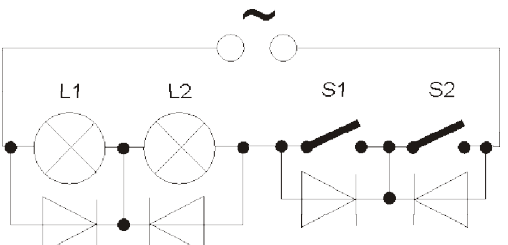
	Item-Code	Aufgabenstellung									
134	G135a	<p>Erkläre mit Hilfe der Abbildungen die Unterschiede zwischen Fotowiderständen, Heiß- und Kaltleitern.</p> <table border="1" data-bbox="415 349 1171 581"> <thead> <tr> <th colspan="2" data-bbox="415 349 919 386">Thermistoren</th> <th data-bbox="919 349 1171 386">Fotowiderstände (LDR)</th> </tr> <tr> <th data-bbox="415 386 667 423">Heißleiter (NTC-Widerstand)</th> <th data-bbox="667 386 919 423">Kaltleiter (PTC-Widerstand)</th> <th data-bbox="919 386 1171 423"></th> </tr> </thead> <tbody> <tr> <td data-bbox="415 423 667 581">  </td> <td data-bbox="667 423 919 581">  </td> <td data-bbox="919 423 1171 581">  </td> </tr> </tbody> </table>	Thermistoren		Fotowiderstände (LDR)	Heißleiter (NTC-Widerstand)	Kaltleiter (PTC-Widerstand)				
Thermistoren		Fotowiderstände (LDR)									
Heißleiter (NTC-Widerstand)	Kaltleiter (PTC-Widerstand)										
											
135	G135b	Beschreibe dabei jeweils den Zusammenhang zwischen Temperatur und Widerstand.									
136	G135c	Überlege dir außerdem, für welche Anwendungen die jeweiligen Leiter sinnvoll genutzt werden können.									
137	G136	Welche Vorteile haben das Dotieren von Halbleitern und die darauf basierende Störstellenleitung gegenüber der Eigenleitung?									
138	G137	Begründe, warum Heißleiter stets mit einem Vorwiderstand in Reihe geschaltet werden sollen.									
139	G151	<p>Beschreibe die Funktion der nebenstehenden Schaltung. Begründe, warum am Ausgang das Doppelte der Eingangsspannung zur Verfügung steht.</p> 									
140	G152	Was versteht man unter Dotierung? Erläutere die dadurch hervorgerufene Veränderung der Leitfähigkeit.									
141	G153	Erkläre den Begriff der Eigenleitung.									
142	G154	Wie verändert sich der Widerstand eines Heißleiters bei der Erhöhung der Temperatur. Wie ist diese Veränderung zu erklären?									

	Item-Code	Aufgabenstellung
143	G181	Was versteht man unter einem Halbleiter? Beschreibe den Leitungsvorgang modellhaft und erkläre die Unterschiede zwischen metallischen Leitern, Halbleitern und Nichtleitern (Isolatoren).
144	G182	Beschreibe das Verhalten einer Halbleiterdiode an Hand der Schemazeichnungen. Bringe sie dazu in die richtige Reihenfolge und ergänze sie. 
145	G183	Begründe, warum die Basis eines Transistors extrem dünn sein muss.
146	G184	Erkläre die Begriffe n- und p-Halbleiter und ihre Funktionsweise.
147	G231	Luft ist normalerweise ein elektrischer Isolator. Bei einem Gewitter gibt es jedoch elektrische Leitungsvorgänge in der Luft. Erkläre das Zustandekommen eines elektrischen Leitungsvorgangs bei einem Gewitter.
148	G232	Beschreibe die Gleichrichter-Wirkung der abgebildeten Schaltung. Deute das abgebildete Oszillogramm. Welchen Vorteil hat diese Gleichrichterschaltung gegenüber der Gleichrichterschaltung einer Diode? 
149	G233	Wie lässt sich mit einem Widerstandsmessgerät der Basisanschluss eines Transistors finden?
150	G234	Begründe, warum der Widerstand dotierter Halbleiter deutlich kleiner ist als die der reinen Halbleiter.

	Item-Code	Aufgabenstellung
151	G235a	<p>In Fotozellen wird die Fotoemission zur Erzeugung eines elektrischen Leitungsvorgangs genutzt. Beschreibe den Aufbau und erkläre die Wirkungsweise einer Fotozelle.</p> 
152	G235b	Wie verändert sich die Stromstärke bei intensiverer Beleuchtung? Begründe deine Antwort.
153	G235c	Wozu kann eine Fotozelle eingesetzt werden?
154	G251	Erläutere, wie der Transistoreffekt zustande kommt.
155	G252	Erkläre, warum es im n-dotierten Halbleiter praktisch keine Löcher und im p-dotierten keine Leitungselektronen gibt.
156	G253	In welcher Weise lassen sich Leuchtdioden als Polprüfer einsetzen?
157	G254	<p>In einer „Black-Box“ mit zwei Anschlüssen befindet sich ein unbekannter Widerstand R und zwei gleiche Dioden. Wie kannst du auf experimentelle Weise die Größe des Widerstandes R bestimmen?</p> 

	Item-Code	Aufgabenstellung
158	G255	<p>Bestimme mithilfe der Steuerkennlinie des Transistors die Zunahme des Kollektorstroms, wenn die Basisspannung von 0,6V auf 0,7V anwächst.</p> 
159	G256	Erkläre die Unterschiede zwischen NTC, PTC und LDR.
160	G257	<p>Bei einer Überwachungsanlage soll eine Kontrolllampe aufleuchten, sobald in dem zu überwachenden Raum die Temperatur unzulässig ansteigt. Entwickle eine sinnvolle Schaltung bei der ein Heißleiter als Temperaturfühler eingesetzt wird. Gebe auch eine kurze Erläuterung der Funktion der Schaltung. (Solltest Du keine Idee haben, mit den folgenden Materialien ist z.B. eine Lösung möglich: Heißleiter, Stromquellen, Kontrolllampe, Relais und Schaltmaterial.)</p>
161	G281	Wie wird eine Halbleiterdiode in Durchlass-, wie in Sperrrichtung geschaltet?
162	G282	Was sind Thermistoren?
163	G283	Welche Elemente eignen sich zum Dotieren von Silicium oder Germanium?
164	G285	Erkläre die Unterschiede zwischen NTC (Heißleiter), PTC (Kaltleiter) und LDR (Fotowiderstand). Gehe dabei auch auf die unterschiedlichen Beziehungen zum Widerstand ein.
165	G284a	<p>Eine Halbleiterdiode ist mit einer Glühlampe in Reihe an eine Gleichspannungsquelle in Durchlassrichtung geschaltet, die Glühlampe leuchtet. Jetzt wird die Gleichspannungsquelle durch eine Wechselspannungsquelle gleicher Spannung ersetzt. Was ist an der Glühlampe zu erkennen?</p> <p>A. Die Glühlampe leuchtet nicht. B. Die Glühlampe leuchtet, aber dunkler als an der Gleichspannungsquelle. C. Die Glühlampe leuchtet genau so hell wie an der Gleichspannungsquelle. D. Die Glühlampe leuchtet, aber heller als an der Gleichspannungsquelle.</p>

	Item-Code	Aufgabenstellung
166	G284b	Bitte begründe deine Antwortauswahl.
167	G286	Was versteht man unter der Rekombination in einem Halbleiterkristall?
168	G21a	Nenne zwei Einsatzmöglichkeiten für eine Leuchtdiode.
169	G21b	Nenne zwei Vorteile einer Leuchtdiode gegenüber einer Glühlampe.
170	G21c	<p>Wie muss die untenstehend skizzierte Spannungsquelle gepolt sein, damit die LED leuchtet?</p> 
171	G236	Berechne den für die Leuchtdiode nötigen Vorwiderstand, wenn für die Leuchtdiode gilt: Schwellenspannung $U_S = 1,6V$; maximaler Strom $I_{max} = 12mA$. Die Batteriespannung soll $U_{batt} = 9,0V$ sein.
172	G258	<p>Berechne den für die Leuchtdiode nötigen Vorwiderstand, wenn für die Leuchtdiode gilt: Schwellenspannung $U_S = 1,6V$; maximaler Strom $I_{max} = 12mA$. Die Batteriespannung soll $U_{batt} = 9,0V$ sein.</p> $R_y = \frac{U_{batt} - U_S}{I_{max}}$ <p>Zur Erinnerung die Formel zur Berechnung des Widerstandes:</p>






	Item-Code	Aufgabenstellung
173	G2285	<p>In einer Reihenschaltung aus zwei Glühlampen, zwei Schaltern und einer 6V-Wechselspannungsquelle sind die Lampen und Schalter wie in der Abbildung zu sehen mit Dioden überbrückt.</p>  <p>Was zu beobachten, wenn der Schalter S1 geschlossen wird?</p> <p>A. Keine Lampe leuchtet. B. Nur Lampe L1 leuchtet. C. Nur Lampe2 leuchtet. D. Beide Lampen leuchten.</p> <p>Begründe deine Antwortauswahl.</p>
174	G23	<p>In einer Reihenschaltung aus zwei Glühlampen, zwei Schaltern und einer 6V-Wechselspannungsquelle sind die Lampen und Schalter wie in der Abbildung zu sehen mit Dioden überbrückt.</p>  <p>Was ist zu beobachten, wenn der Schalter S1 geschlossen wird? Erkläre deine Beobachtung.</p>
175	L41	<p>Stelle in einer Übersicht die wichtigsten Eigenschaften und Nachweismöglichkeiten von α-, β- und γ-Strahlung zusammen.</p>

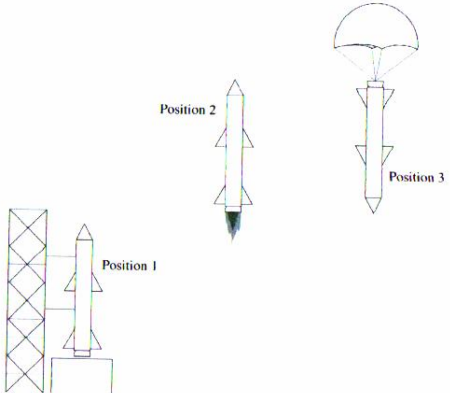
Anhang G: Aufgaben in Studie 2

	Item-Code	Aufgabenstellung
176	L42	Das radioaktive Kohlenstoffisotop $^{14}_6\text{C}$ wandelt sich in Stickstoff (N) um und sendet eine β -Strahlung aus. Stelle die Gleichung für den Zerfall auf.
177	L43	Erkläre den Ablauf der Kernspaltung. Was ist der Unterschied zwischen gesteuert und ungesteuert in diesem Zusammenhang?
178	L44a	Bei einem Präparat registriert man mit einem Zählrohr 306 Impulse in 3,0 Minuten. Ohne Präparat misst man in 10 Minuten 120 Impulse (Nulleffekt). Worauf ist der Nulleffekt zurückzuführen?
179	L44b	Welche Zählrate (Impulse/Minute) stellt man für das Präparat nach Abzug des Nulleffekts fest?
180	L45	Erkläre, was ein Moderator ist.
181	L46	Bei der Kernspaltung entstehen zwei Trümmerkerne und zwei bis drei Neutronen. Welche Strahlung wird zusätzlich ausgesandt? A. Alphastrahlung B. Betastrahlung C. Gammastrahlung D. Neutronenstrahlung
182	L47a	Erkläre den Explosionsvorgang bei einer Atombombe. Gehe in diesem Zusammenhang auf die Begriffe Kettenreaktion, Druckwelle, Pilzwolke und Fallout ein.
183	L47b	Beschreibe die Auswirkungen einer Atombombenzündung.
184	L48	Wo wird der radioaktive Müll gelagert? Welche Gefahren gehen von diesem radioaktiven Müll aus?
185	L49	Stelle die Reaktionsgleichung der Kettreaktion von $^{235}_{92}\text{U}$ auf.
186	L410	Durch die Notkühlsysteme eines Reaktors soll sichergestellt werden, dass A. keine zusätzliche Strahlung in die Umgebung eines Kernkraftwerks gelangt. B. die Neutronen durch Wasser weiter moderiert werden. C. die Brennelemente auch bei einem großen Störfall mit Kühlmittel versorgt werden. D. die Wasservorräte innerhalb und außerhalb des Sicherheitsbehälters in ausreichender Menge vorhanden sind.

	Item-Code	Aufgabenstellung
187	L411	Welche Vorteile und welche Nachteile haben Kernkraftwerke gegenüber anderen Kraftwerksarten? Welche Risiken sind mit ihrem Betrieb verbunden?
188	L412a	Erkläre den Aufbau eines (einfachen) Atommodells.
189	L412b	Welche Experimente geben Auskunft über den Aufbau von Atomen?
190	L413	Erkläre die Unterschiede zwischen der Kernspaltung und der Kernfusion.
191	GP1a	Ein Transistor kann als elektronischer Schalter genutzt werden. Erläutere, wie ein Transistor als Schalter funktioniert. Nenne dabei seine drei Anschlüsse.
192	GP1b	Welche Vor- und Nachteile hat ein Transistor als Schalter gegenüber einem elektromagnetischen Relais?
193	GP2	Fertige eine Übersicht über wichtige Halbleiterbauelemente und deren Anwendungen an. Denke dabei auch an Geräte, die im Haushalt und im Freizeitbereich genutzt werden.
194	GP3a	Netzgeräte sind Stromquellen, die zum Betrieb an der Steckdose angeschlossen werden. Einfache Netzgeräte enthalten als Bauteile einen Transformator, einen Brückengleichrichter und gegebenenfalls eine Sicherung. Am Ausgang des Netzgerätes, d.h. zwischen den Polen, liegt eine (pulsierende) Gleichspannung an. Zeichne das Schaltbild zu diesem Netzgerät.
195	GP3b	Erläutere die Wirkungsweise der einzelnen Bauteile.
196	GP4	Erkläre die Unterschiede zwischen Leiter, Halbleiter und Nichtleiter/Isolatoren.
197	GP5	Nenne ein Verfahren, mit dem Halbleiter dotiert werden können.
198	GP6a	Welcher Halbleiter liegt vor, wenn Silizium mit fünfwertigem Arsen dotiert wird? Skizziere eine flächenhafte schematische Darstellung eines solchen Halbleiters.
199	GP6b	Welcher Halbleiter liegt vor, wenn Silizium mit dreiwertigem Indium dotiert wird? Skizziere eine flächenhafte schematische Darstellung eines solchen Halbleiters.
200	GP7	Beschreibe ein Experiment, mit dessen Hilfe du unterscheiden kannst, ob ein Stoff ein metallischer Leiter oder ein Halbleiter ist.

	Item-Code	Aufgabenstellung															
201	TIMSSB2	<p>Der größte Teil der chemischen Energie, die bei der Benzinverbrennung in einem Automotor freigesetzt wird, wird nicht für den Antrieb des Autos verwendet, sondern umgewandelt in:</p> <p>A. Elektrizität B. Wärme C. Magnetismus D. Schall</p>															
202	TIMSSB3	<p>Welcher Gegenstand aus der Tabelle hat die größte Dichte?</p> <table border="1"><thead><tr><th>Gegenstand</th><th>Masse des Gegenstands</th><th>Volumen des Gegenstands</th></tr></thead><tbody><tr><td>W</td><td>11,0 Gramm</td><td>24 Kubikzentimeter</td></tr><tr><td>X</td><td>11,0 Gramm</td><td>12 Kubikzentimeter</td></tr><tr><td>Y</td><td>5,5 Gramm</td><td>4 Kubikzentimeter</td></tr><tr><td>Z</td><td>5,5 Gramm</td><td>11 Kubikzentimeter</td></tr></tbody></table> <p>A. W B. X C. Y D. Z</p>	Gegenstand	Masse des Gegenstands	Volumen des Gegenstands	W	11,0 Gramm	24 Kubikzentimeter	X	11,0 Gramm	12 Kubikzentimeter	Y	5,5 Gramm	4 Kubikzentimeter	Z	5,5 Gramm	11 Kubikzentimeter
Gegenstand	Masse des Gegenstands	Volumen des Gegenstands															
W	11,0 Gramm	24 Kubikzentimeter															
X	11,0 Gramm	12 Kubikzentimeter															
Y	5,5 Gramm	4 Kubikzentimeter															
Z	5,5 Gramm	11 Kubikzentimeter															

	<i>Item-Code</i>	<i>Aufgabenstellung</i>
203	TIMSSX01	<p>Elektrische Energie wird eingesetzt, um eine Lampe leuchten zu lassen. Wie kann die Menge an eingesetzter Energie mit der Menge an produzierter Lichtenergie verglichen werden?</p> <p>a) Die Menge an elektrischer Energie ist:</p> <ol style="list-style-type: none"> 1. größer, als die Menge an produzierter Lichtenergie 2. geringer, als die Menge an produzierter Lichtenergie 3. genauso groß, wie die Menge an produzierter Lichtenergie <p>b) Bitte begründe deine Antwort:</p> <hr/>
204	TIMSSD1	<p>Welche Zeichnung gibt am besten wieder, was geschieht, wenn Licht durch ein Vergrößerungsglas fällt?</p> <div style="display: flex; flex-direction: column; align-items: flex-start;"> <div style="margin-bottom: 10px;"> <p>A </p> </div> <div style="margin-bottom: 10px;"> <p>B </p> </div> <div style="margin-bottom: 10px;"> <p>C </p> </div> <div style="margin-bottom: 10px;"> <p>D </p> </div> <div style="margin-bottom: 10px;"> <p>E </p> </div> </div>

	Item-Code	Aufgabenstellung
205	TIMSSH05	<p>Menschen erhalten Energie durch die Nahrung, die sie essen. Woher stammt die Energie, welche in der Nahrung gespeichert wird?</p> <p>A. Dünger B. Sonne C. Vitamine D. Boden</p>
206	TIMSSJ05	<p>Die Abbildung zeigt eine Rakete, die zunächst von der Erde abgeschossen wird und dann zurückkehrt.</p>  <p>In welcher der drei Positionen wirkt die Schwerkraft auf die Rakete?</p> <p>A. Nur in Position 3 B. In den Positionen 1 und 2 C. In den Positionen 2 und 3 D. In den Positionen 1, 2 und 3.</p>